Near-Optimal Resource Allocation Algorithms for 5G+ Cellular Networks

Huda Y. Alsheyab *IEEE Student Member*, Salimur Choudhury, *IEEE Member*, Ebrahim Bedeer, *IEEE Member* and Salama Ikki, *IEEE Senior Member*

***Abstract*—Fifth generation and beyond (5G+) systems will support novel cases, and hence, require new network architecture. In this work, network flying platforms (NFPs) as aerial hubs are considered in future 5G+ networks to provide fronthaul connectivity to small cells (SCs). We aim to find the optimal association between the NFPs and SCs to maximize the total sum rate subject to quality-of-service (QoS), bandwidth, and supported number of links constraints. The formulated optimization problem is an integer linear program and the optimal association between the NFPs and SCs is found using numerical solvers at the expense of high computational complexity. We propose two algorithms (centralized and distributed) to reach a sub-optimal association at reduced complexity. Simulation results show that the performance of the proposed algorithms approaches the counterpart of its optimal solution and outperforms the state-of-the-art techniques from the literature.**

***Index Terms*—5G+, integer linear program, network flying platforms (NFPs), small cells (SCs), unmanned aerial vehicles (UAVs)**

I. INTRODUCTION

In recent years, wireless communication needs have witnessed a continuous growth, which demands more complex infrastructure to cope with. Fifth generation and beyond (5G+) systems are expected to adopt several changes in their network architecture, construction, and deployment to be compatible with the latest technologies and end users' needs. 5G+ systems should support a 10 to 100 times greater number of connected devices and typical user data rate [1], [2].

Some of the key technologies that can be used in 5G+ wireless systems to satisfy the desired performance are massive MIMO, Device to Device communication, spectrum sharing with cognitive radio, ultra dense networks, multi-radio access technology, full duplex communication, millimeter wave communication, energy harvesting communication and Cloud Technologies [3], [4]. Fiber has been used for fronthaul links; however, it has some issues, such as high cost and the need to minimize time-to-market [5], [6]. On the other hand, free-space optical (FSO)/microwave links are cost-effective, easy-to-deploy and carry traffic for SCs from the core network.

H. Y. Alsheyab, S. Choudhury, and S. S. Ikki are with Lakehead University, Thunder bay, On, Canada (e-mail:{halsheya, salimur.choudhury, sikki}@lakeheadu.ca).
E. Bedeer is with Ulster University, Newtownabbey, UK (e-mail: e.bedeer.mohamed@ulster.ac.uk).
This work was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

However, FSO/microwave systems lead to less coverage due to short-range communication, and are affected by any obstacles or animals in the environment, which hinders the transmission. Furthermore, FSO is affected by weather (rain, snow, and fog). In contrast, NFPs are cost-effective and scalable. NFPs are capable of hovering at an altitude ranging from a few hundred meters and up to 20 kms to mitigate unfavorable weather conditions [7], which cannot be achieved using a fixed FSO/microwave.

Moreover, the regulators of drone operations can dictate when to operate the NFPs depending on the weather conditions. An example of practical applications for NFPs that are capable of operating at different weather conditions is weather resistant drones equipped with a built in WiFi chip to support all-time connectivity even in raining, snowing, lightning, and windy weather [8]. Additionally, waterproof drones in [8] can assist operation in raining weather. Furthermore, artificial intelligence can be used to predict weather conditions and improve the operation of NFPs [9].

As can be seen from the previous discussion, the effect of the bad weather conditions on the NFPs can be reduced or at least controlled which is not the case in the FSO/ microwave links. Another advantage of NFPs is that NFPs can be considered an affordable solution to extend coverage in rural areas that do not have the required infrastructure [10]. NFPs also assist in maintaining the communications in case of failure of the existing infrastructure that may happen in case of disasters such as: earthquakes, tsunami, flooding, and land sliding [11]–[13]. It is expected that NFPs can overcome the shortcomings of the fiber and FSO in 5G+ systems [7].

Network flying platforms (NFPs) such as unmanned aerial vehicles (UAVs), drones, and unmanned balloons systems have attracted industry and academic attention in the last couple of years [14], [15] as candidate solutions to support coverage of dense networks. The integration of NFPs with 5G+ capabilities will allow much greater connectivity, lower latency, and quick transfer of high-precision data. This aggregation of 5G+ networks and NFPs is powerful, giving way to many new capabilities and improvements in wireless applications. In comparing with the static ground macro base station NFP is more scalable. the integration of NFPs with wireless and mobile networks is expected to bring very high spectral efficiency and solves many communication challenges. The rapid and dynamic deployment of NFPs and their reliable line-of-sight (LoS) communication links are the main advantages of NFP-based communications.

Naqvi *et al.* [16] presented a case study of combining UAVs in a wireless network, with both high- and low-power BSs. The objective was to investigate if UAV deployment could fulfill higher data rate requirements in a 5G network utilizing mmWave technology while ensuring an acceptable level of the power consumed. They concluded that using UAVs with the conventional cellular network improved the energy efficiency while sustaining the QoS requirements.

Ahmadi *et al.* [14] has presented a novel layered architecture of the future cellular networks. NFPs of various types flying in low / medium / high layers are considered to provide additional capacity and expand the coverage. In [17], NFPs are employed as fronthaul hubs to connect SCs with the core network and a greedy algorithm was designed to solve the association problem of NFPs and SCs. Utilizing drone small cells (DSCs) as aerial base stations to support cellular networks was proposed in [18], where the optimal height of DSCs to minimize the required transmit power for covering a target area was found. A drone-BSs deployment plan was provided in [19] to serve the users based on their traffic requirements while minimizing the number of required drones. In particular, the number of drone-BSs and their 3D placement was estimated while satisfying the coverage and capacity constraints of the system.

Few algorithms (centralized and distributed) have been proposed to connect air drones and balloons with traditional small cells of the cellular network while maximizing the system capacity. In this work, a heterogeneous network that consists of SCs, NFPs, and the ground core network is studied as well as; the association problem between NFPs with SCs to maximize the network total sum rate is also studied. Two algorithms (centralized and distributed) are proposed, where each NFP is associated with one or more SCs. We ensure that necessary quality-of-service (QoS) requirements are met.

Practical constraints were included in our optimization problem, such as considering interference between NFPs and SCs, the maximum number of links and the maximum bandwidth that the NFP can support. NFPs act as a hub to provide fronthaul connectivity between the SCs and the core network. Hence, the association problem of SCs and NFPs is an important problem to enhance the performance of the system. These algorithms enhance the running time speed and perform a greedy search with higher data rate demands.

The main contributions of this work can be summarized as follows.

- We design a centralized resource allocation based on a weighted bipartite matching algorithm (Hungarian algorithm) to find the best association between NFPs and SCs that maximizes the system's total sum rate subject to QoS constraints (the SINR). This guarantees the performance while satisfying all constraints and notably outperforming other existing algorithms.
- We design a distributed algorithm based on a stable marriage matching scheme to maximize the total sum rate, while requiring only local information of NFPs and SCs, subject to QoS constraints. The main advantage of this distributed algorithm is the reduction of the necessary feedback overhead, hence reducing the system complexity.
- Furthermore, we compare our work with what is studied in [20]. Indeed, our solution is different from [20] and this difference can be explained as follows. The SCs in [20] send association requests to the NFP with the highest SINR. Each NFP selects the SCs depending on its available bandwidth and number of links. The distributed algorithm in [20] does not re-associate the rejected SCs with other NFPs. To overcome this drawback, in our proposed distributed algorithm we used the staple marriage matching algorithm which guarantees that each SC is matched with one NFP (either real or dummy).
- Finally, we provide extensive simulation results to assess the performance of the proposed algorithms in realistic conditions.

The rest of this paper is organized as follows. Related works are presented in Section II. A description of the system model and the problem formulation is provided in Section III. We discuss existing algorithm in IV. Section V discusses the proposed algorithms. In Section VI, the performance of the proposed algorithms is presented. Finally, our conclusion is presented in Section VII.

II. Related Work

5G and 5G+ aim for systems with higher capacity, increased data rate, reduced latency and cost. Moreover these systems are expected to be energy efficient and capable of handling massive device connectivity [21].

Yu *et al.* [22], proposed a new paradigm of the 5G-enabled vehicular network, to provide efficient and elastic services for mobile applications which require vast bandwidth resource and high computing capability. They discussed the cloudlet resource management approach which includes resource allocation and sharing. Furthermore, they employed a matrix game to solve the problem and use Karush-Kuhn-Tucker conditions to work out the explicit solutions of global optimization.

Lotfi *et al.* [23], studied a cell planning problem in 5G to determine the number and the location of base stations (BSs) with fiber backhaul and BSs with wireless self-backhauled. They proposed an algorithm to minimize deployment costs to meet coverage and capacity constraints with the minimum number of BSs. Furthermore, they applied a meta-heuristic algorithm to solve the proposed cell planning problem. They further developed an efficient meta-heuristic algorithm.

Shah *et al.* [20] employed NFP at different altitudes, as aerial backhaul hubs. The association problem of NFP-hubs and small-cells was solved while taking into consideration backhaul links and NFP related constraints. These include the maximum number of supported links and bandwidth. They presented a distributed solution, which performed a greedy search to maximize the total sum rate of the overall network, where they depended on the preferred SINR as the first step to associate the SCs and NFPs. A drawback of this approach is that if more than one SC try to associate with a single NFP, and that NFP cannot associate with all of the SCs; then the unassigned SC will not try to associate with another NFP.

Shah *et al.* [17], solved the association problem of NFPs and SCs for the network. The objective was to serve the maximum number of SCs without any consideration (like the total sum rate). Their work presented two simple greedy algorithms, centralized and distributed algorithms. Thus, centralization was used when the system needed to decrease power consumption and distribution was used when it needed to reduce latency.

Mozaffari *et al.* [24] proposed an approach for deploying UAVs to provide wireless service to ground users while minimizing the overall UAV transmit power needed to satisfy the users data rate. Hence, they tried to derive the optimal coverage area and locations of UAVs that minimize the required transmit power. Furthermore, Mlika *et al.* [25], studied the user association problem under QoS requirements in a heterogeneous and small cells network (HetSNet). This piece of work proposed a completely distributed algorithm which assumes no coordination between the base stations. Their ultimate goal was to design a fully distributed algorithm to maximizes the number of associated users.

In [26], they proposed low-complexity Subgrouping algorithm, to effectively perform subgroup formation in Satellite-Long Term Evolution systems. The goals of their algorithm were achieving high scalability; the computational cost didn't depend on the available resources, and reaching close to optimum performance through a novel resource allocation strategy.

AlQerm *et al.* [27] studied the power allocation problem for the downlink transmission in a spectrum sharing multi-tier 5G environment. They proposed an online learning based approach to assigning transmission power to reduce the overall power consumption while maintaining QoS. Also, the scheme employed an approximation mechanism for the Q-value, which reduced the state/action space and accelerate the speed of convergence.

Abdelhadi *et al.* [28] introduced an approach for optimal resource allocation from multiple carriers for users in fourth generation long term evolution (4G-LTE) system. They used logarithmic and sigmoidal-like utility functions to represent the user applications running on different user equipment (UE)s. To solve the problem, they implemented a distributed rate allocation algorithm that approximated the optimal rate and optimal price. All UEs requested for resources from the nearby Evolved NodeBs (eNodeBs). eNodeB set a price for resource based on the sent UE requests. In case UE had more than one nearby eNodeB, it chose the one with the lowest price and started requesting resources. If the allocated rate was not enough or the price of the resources increased, the UE began to allocate the rest of the required resources from another eNodeB. This was done iteratively until the optimal rates were allocated in the 4G-LTE mobile network.

Abdelhadi *et al.* [29] discussed another resource allocation optimization problem in 4G-LTE with users running multiple applications. Where the objective was to allocate the resources optimally with a utility proportional fairness policy, they solved the problem by implementing a two-stage algorithm. First-stage allocated the rates to the UEs. Each UE started with sending an initial request $w_i(n)$ to the eNodeB. The eNodeB calculated the difference between the received request $w_i(n)$ and the previously received request $w_i(n1)$, if the difference was less than a pre-specified threshold the algorithm exited. If the value was greater than the threshold, eNodeB calculated the shadow price and sent it to all the UEs. Each UE received the shadow price to solve the rate. The rate was used to calculate the new request $w_i(n)$. This process was repeated until $|w_i(n)w_i(n1)|$ was less than the threshold. In the second stage of resource allocation, the rates to the j_{th} application in i_{th} UE ri, j were allocated internally in the UE. The UE used the allocated rate in the first-stage and solved the problem.

Zulhasnine *et al.* [30] proposed a greedy heuristic algorithm that could mitigate interference to the primary cellular network using channel gain information. They provided two phases downlink (DL) and uplink (UL) algorithm to maximize the sum rate of the primary cellular UEs and secondary D2D UEs while maintaining a minimum SINR. Both in the UL and DL phases, the algorithm sorted the cellular user list C in descending order based on the channel quality indicator (CQI). After that, the algorithm repeated the following until there was no more D2D or cellular users in the list: first, the algorithm selected the resource blocks (RBs) with c_{th} largest value and found the D2D transmitter d for which the channel gain was minimum. Then, the algorithm found the SINR of the cellular UE and D2D communication, respectively. If the SINR guaranteed the target SINR, then share all RBs of the UE c with D2D connection d.

Liang *et al.* [31] defined a utility function to refer to a user's benefit from all possible resource allocation amounts. They denoted the utility functions as $U(x)$ where x was the amount of resource and $U(x)$ was the utility obtained from that allocation. They proposed a Maximum Segmental Slope (MSS) Resource Allocation Algorithm to maximize the utility derived by user's flow for a bandwidth allocation while the total bandwidth did not exceed the total link capacity. MSS started by initializing the flow bandwidth to zero; then it initialized the remaining bandwidth to channel capacity. After that, the algorithm kept repeating the following until the remained bandwidth equalled zero: first, for each flow, it considered a line on the utility function graph and determined the point to which a line would connect to the highest slope on the Utility Function graph. Second, the algorithm selected the flow that had the most lucrative incremental allocation. If the remaining bandwidth was less than the purchasing bandwidth, the algorithm repeated the second point. Otherwise, the algorithm would update the flow and remaining bandwidth.

Padaganur *et al.* [32] Proposed an optimal resource and power allocation using a Feed Forward Neural Network. They defined the communication radius of the covered area by the evolved node base station (eNB) nodes as Femtocells to improve the effective throughput of the system. In their algorithm, each Femtocell operated different subcarriers and had different allocated power. A mobile device accessed the service through Femtocells. The user could be assigned to different subcarriers based on his demand. The appropriate cells selected based on the quality of transmission. When the base station found that SNR was low for a mobiles transmission specified to a cell, it assigned the mobile to another cell.

Wen *et al.* [33] proposed joint mode selection and resource allocation (JMSRA) schemes to maximize the throughput of the hybrid network. JMSRA had two scheduling schemes one for dedicated users, and the other one underlay users. The DMSRA scheme was designed for dedicated users, including cellular users and dedicated D2D users, while the UMSRA scheme was for underlay users, who shared resources with other dedicated or underlay users.

They divided users into two sets, i.e., S_d for dedicated users and S_u for underlay users. Cellular and D2D users with poor direct-link gain were added to S_d while other D2D users were classified into S_u. After the initialization phase in DMSRA algorithm, the scheduler assigned RBs in sequential order. The scheduler assigned feasible RBs to the user with the highest priority, which is calculated by Proportional Fair (PF) function $Q_d(i) = Ar_i/R_i$ where A was service modified factors (if A was not available at the beginning, priority was randomized). The scheduling process was terminated when all the RBs were allocated, or users with requests were all scheduled. In case the number of RBs was not enough, the scheduler added the remaining users into S_u.

On the other hand, After the initialization phase in UMSRA, each user in S_u was measured and scheduled in descending order of $Q_u(i) = delay^i PL_{u,ii}/PL_b$. It would give higher priority to users with poor satisfaction and channel state since the scheduler could record average rate and delay. If there was no dedicated user on the feasible RB group, the scheduler decided the best link mode for UE_i, which was measured by the received SINR of the cellular and direct links. Otherwise UE_i would choose underlay D2D mode and limited the transmit power to avoid interference to the cellular link. If there was no feasible RB group meeting the throughput requirement, UE_i added to S_d and had a high fairness priority among dedicated users.

Alamouti *et al.* [34] proposed an energy-efficient resource allocation scheme for D2D communications as an underlay of LTE-A (4G) cellular network. They considered a scenario where at most one CU and one D2D pair shared the same uplink channel. They solved the optimization problem by first utilizing the scheme proposed in [35] for admissibility and determining candidate CU reused partners. They found the minimum transmit power levels for CU i as a candidate reused partner and for D2D pair j that satisfy their required SINRs from [35].

After that, they used the Hungarian algorithm to solve this problem. However, the Hungarian algorithm needed the bipartite graph to be symmetric. To satisfy this, they added CS virtual vertices to the set of admissible D2D pairs S, where C was the union of all candidate reused partner sets for all D2D pairs. If the vertex i was not connected to the vertex j, they connected them with a large-valued weighted edge. After that, they used the Hungarian algorithm to solve the minimum weighted matching problem on the transformed bipartite graph.

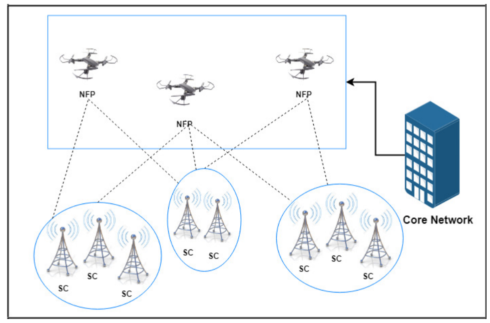

Fig. 1. Graphical illustration of NFPs and SCs in 5G+.

As shown, there are a lot of works that tried to solve different resource allocation problems in 4G-LTE and 5G. However, they had a different system model and problem formulation than here, and their solutions may not suit our problem. Most of the work related to NFPs use them to enhance the network coverage and other arguments. To the best of the authors' knowledge, [20] is the only work which designed the association problem of NFP and SCs and they provided a distributed greedy solution of the optimization problem.

III. SYSTEM MODEL AND PROBLEM FORMULATION

A. System model

As shown in Fig. 1, a heterogeneous network is investigated, which includes three categories of wireless nodes: ground SCs, NFP-hubs, and the ground core network. The SCs carry the traffic directly between the end users and the core network using fronthaul links; while the NFPs act as a hub to provide additional fronthaul connectivity between SCs and the core network. The system model in [17] and [20] is considerably similar to ours, but it is worth mentioning that our proposed algorithms are divergent. In particular, [20] proposed a distributed algorithm to solve the association problem between NFPs and SCs; however, in our work we proposed a more efficient distributed algorithm and a centralized algorithm to address the same association problem.

On the other hand, the problem formulation in [17] maximizes the number of associated SCs and NFPs which is different than our problem formulation that targets maximizing the total sum rate. NFPs can be considered an affordable solution to extend coverage in rural areas that do not have the required infrastructure [10]. NFPs also assist in maintaining the communications in case of failure of the existing infrastructure that may happen in case of disasters such as: earthquakes, tsunami, flooding, and land sliding [11]–[13]. Hence, the association between the SCs and NFPs is crucial to improve the overall network performance.

It is assumed that the NFPs are placed at a pre-defined height h_D (i.e., LAP, MAP, and HAP) according to safety and security policies [20] and uniformly distributed on a horizontal plane parallel to the ground; while, SCs are uniformly distributed on the ground level. This work considers a system with I NFPs and J SCs pairs $(I << J)$ where the set of NFPs is represented as $F = \{f_1, f_2, ..., f_I\}$ and the set of SCs is represented as $S = \{s_1, s_2, ..., s_J\}$. For the remainder of the paper, NFPs are denoted as f_i, $1 \leq i \leq I$ and SCs are denoted as s_j, $1 \leq j \leq J$. To facilitate the implementation of the centralized algorithm, NFPs can share the control information (required bandwidth, required data rate, and SINR) with the core network for association purposes. Based on the control information, the system can specify the association between SCs and NFPs. However, this does not include the data information.

In this work, it has also been assumed that each NFP has a different number of links to support the communication with SCs, i.e., each NFP can serve one or more SCs depending on the number of links L_i and maximum bandwidth it can support B_i.

Since NFPs are spread in a horizontal parallel plane to the SCs at a height h_d from ground level, we use the air-to-ground path loss channel model; this is in contrary to the conventional terrestrial communications that use log-distance path loss model. Thus, the wireless link between NFPs and SCs is mainly vertical. Hence, in the following subsection, the air-to-ground (ATG) path loss model is discussed.

B. Air-to-Ground Path Loss Model

The same Air-to-Ground (ATG) path loss (PL) model is used as in [20], which is widely adopted in the NFP literature [7]. The adopted ATG model considers two propagation groups: *i*) line-of-sight (LoS) receivers where the SCs are placed in LoS / near-LoS to the NFPs and *ii*) Non-line-of-sight (NLoS) receivers where SCs depend on reflections and refractions for coverage. One factor that plays an important role in determining the PL in the ATG model is the probability of LoS. This depends on the surrounding environment (urban, rural, etc.) and the orientation of NFPs and SCs. Hence, the probability of LoS is formulated as in [20]

$$P(\text{LoS}) = \frac{1}{1 + \alpha \exp[-\beta(\frac{180}{\pi}\theta - \alpha)]}, \tag{1}$$

where α and β are constants depending on the environment (rural, urban, etc) and $\theta = \arctan(\frac{h_D}{s})$ is the angle between the SC and the NFP, where $s = \sqrt{(x - x_D)^2 + (y - y_D)^2}$ is the horizontal distance between the SC and the NFP. The locations of the SCs and NFPs in the Cartesian coordinate are given as (x, y) and (x_D, y_D, h_D), respectively. The average PL is given as

$$PL(d)|_{\text{dB}} = 10 \log \left(\frac{4\pi f_c d}{c}\right)^{\gamma} + \eta_{\text{LoS}} P(\text{LoS}) + \eta_{\text{NLoS}} P(\text{NLoS}), \tag{2}$$

where $PL(d)|_{\text{dB}}$ represents the free space path loss in dB, f_c is the carrier frequency, c is the speed of light, γ is the PL exponent, and $d = \sqrt{{h_D}^2 + s^2}$ is the distance between the NFP and SC. η_{LoS} and η_{NLoS} represent the additional losses of the LoS and NLoS links, and $P(\text{NLoS}) = 1 - P(\text{LoS})$.

C. Problem formulation

This work aims to find the optimal association between SCs and NFPs to maximize the total sum rate subject to constraints on the QoS and the maximum number of links and bandwidth supported by each NFP. This work denotes the requested data of the s_j associated with the f_i by $r_{i,j}$, where Shannon capacity formula is used to compute $r_{i,j}$, and we denote the association between the s_j and the f_i by $A_{i,j}$ that can be formally defined as

$$A_{i,j} = \begin{cases} 1, & \text{if } f_i \text{ is connected with } s_j, \\ 0, & \text{Otherwise.} \end{cases}$$

Thus, the data rate supported by the f_i is $\sum_{j=1}^{J} r_{i,j}.A_{i,j}$ and the total sum rate over all NFPs is $\sum_{i=1}^{I} \sum_{j=1}^{J} r_{i,j}.A_{i,j}$. The SINR between the s_j and the f_i is defined as

$$\text{SINR}_{i,j} = \frac{P_{i,j} PL(d_{i,j})}{\sum_{k=1,k\neq i}^{I} P_{k,j} PL(d_{k,j}) + \sigma_i^2}, \tag{3}$$

where $P_{k,j}$ is the transmit power from f_k to s_j and σ_i^2 represents the received noise power at the f_i and $d_{i,j}$ is the distance between f_i and s_j. The Interference depend on path loss in equation (2). The QoS constraint should guarantee that the $\text{SINR}_{i,j}$ is greater than a minimum required value $\text{SINR}_{\min}$. Hence, the QoS constraint can be expressed as

$$A_{i,j}.\text{SINR}_{\min} \leq \text{SINR}_{i,j}, \quad \forall\, i, j. \tag{4}$$

It is denoted that the requested bandwidth of the s_j associated with the f_i by $b_{i,j}$ and the maximum available bandwidth of the f_i is B_i. Thus, the bandwidth constraint can be formulated as

$$\sum_{j=1}^{J} A_{i,j}.b_{i,j} \leqslant B_i \ , \quad \forall\, i. \tag{5}$$

The maximum number of links that the f_i can support is declared as L_i. Hence, the number of NFP links constraint can be expressed as

$$\sum_{j=1}^{J} A_{i,j} \leqslant L_i \ , \quad \forall\, i. \tag{6}$$

The maximum number of links that the s_j can support is one link. Hence, the number of SC links constraint can be formulated as

$$\sum_{i=1}^{I} A_{i,j} \leqslant 1 \ , \quad \forall\, j. \tag{7}$$

Taking into consideration all constraints mentioned previously, for a specific time when NFPs and SCs have fixed positions, this work seeks to find the association between the SCs and the NFPs in order to maximize total sum rate of the system. The SC association with NFP problem can be formulated as

$$\max_{A_{i,j}} \quad \sum_{i=1}^{I}\sum_{j=1}^{J} r_{i,j}.A_{i,j} \tag{8a}$$

$$\text{s. t.} \tag{8b}$$

$$A_{i,j}.\text{SINR}_{\text{min}} \leq \text{SINR}_{i,j}, \quad \forall\, i, j. \tag{8c}$$

$$\sum_{j=1}^{J} A_{i,j}.b_{i,j} \leqslant B_i\ , \quad \forall\, i. \tag{8d}$$

$$\sum_{j=1}^{J} A_{i,j} \leqslant L_i\ , \quad \forall\, i. \tag{8e}$$

$$\sum_{i=1}^{I} A_{i,j} \leqslant 1\ , \quad \forall\, j. \tag{8f}$$

$$A_{i,j} \in \{0,1\}\ , \quad \forall\, i, j. \tag{8g}$$

This is an integer linear program that can be solved numerically to get the optimal solution. However, in worst case the system will try to connect each NFP with all SCs to find the best association and this can take exponential time [36]. Therefore, this is an NP-hard problem as explained in [37], which proves that the provided association problem is equivalent to generalized assignment problem (GAP). Using this relation with the GAP, they show the NP-hard complexity of the association problem. (the problem can be reduced to a maximum knapsack problem [38]). In the coming section, two polynomial-time algorithms to obtain sub-optimal solutions are proposed.

IV. Existing Algorithm

A Distributed Maximal Demand Minimum Servers algorithm $(DM)^2S$ is proposed in [20] to provide an efficient solution of the optimization problem (8). They have proposed a greedy method to solve the association problem. The major steps of the algorithm are:

1. Each SC sends a message to the NFP with the maximum SINR. Basically, each SC wants to connect that NFP which will give it the best SINR value.
2. Once an NFP receives messages (an NFP can get messages from more than one SC), it selects that SC to be connected with which will maximize the total sum rate along with satisfying maximum NFP bandwidth and links constraints.

Lemma 1 In the worst case, the performance of $(DM)^2S$ is unbounded.

Proof Consider a scenario where there is J SCs $(s_1, s_2, ...s_J)$ and 2 NFPs (f_1, f_2). Now, if SINR of each SCs and f_1 is the same say, $\text{SINR}_{1,j}$, for all j. And for all j, $\text{SINR}_{2,j}$ is same too. Now, if $\text{SINR}_{2,j} = \text{SINR}_{1,j} - \epsilon$ where ϵ is a very small constant, then all SCs will send message to f_1 as shown in Fig. 2.

Assume that f_1 has only one link available and f_2 has $(J-1)$ available links. After sending messages from all SCs, f_1 eventually will only consider one small cell. Now, if all SCs

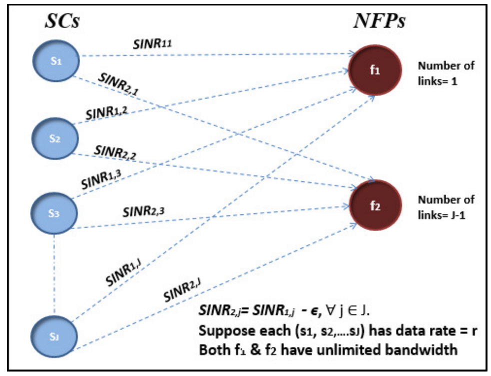

Fig. 2. Example of the SINR and data rate between SCs and NFPs

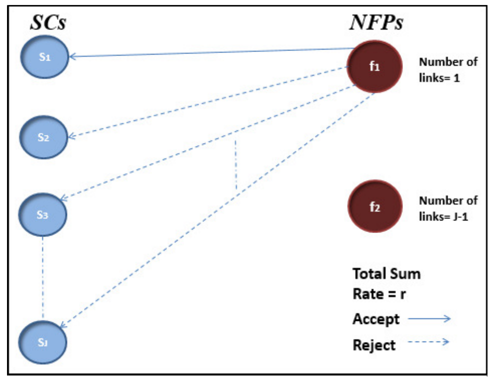

Fig. 3. SCs ask NFPs to associate using $(DM)^2S$ algorithm

provide the same data rate (say r for every connection) and f_1 picks one of them (say s_1). Hence, at the end of the algorithm, only one SC will be connected as shown in Fig. 3. No small cell will be connected with f_2 since no message was sent to f_2.

The total sum rate of this solution will be r. On the other hand, as one can see in Fig. 4, in an optimal solution, total $J * r$ can be obtained by connecting one small cell to f_1 and the rest with f_2. Hence, the performance ratio of $(DM)^2S$ is $\frac{r}{J*r}$ which $\frac{1}{J}$ which is unbounded (will be increase with the increase of small cells).

V. Proposed Algorithms

In this section, two efficient algorithms are proposed to solve the SCs and NFPs association problem in (8) in polynomial time complexity. The first algorithm works in a centralized manner; while the second one provides a distribution solution.

A. Proposed Hungarian Based Centralized Algorithm (HBCA)

The centralized solutions are designed to move all processing work to a central location in support of multiple remote radio heads. The central location could store both the communication and the user account information, as well as all the necessary information from the SCs and NFPs. The proposed Hungarian based centralized algorithm (HBCA) maximizes the total sum rate after receiving all necessary

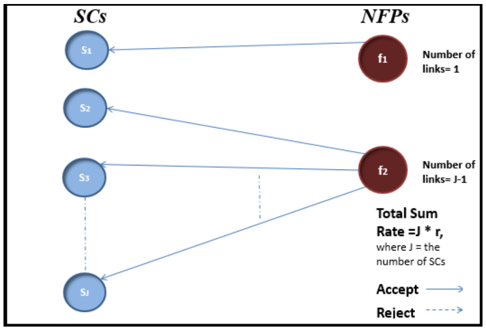

Fig. 4. Optimal solution

`HBCA - part 1`

1: **Input:** $(S, F, \text{SINR}, \text{SCLink}, \text{SINR}_\text{m}, L, B, b, r)$
2: Let $W_{J\times J}$ be a new Matrix ▷ $W_{i,j}$ is the weight of the edge between f_i and s_j
3: **while** there is free $f_i \in F$ **do**
4: **for** each $f_i \in F$ **do**
5: **for** each $s_j \in S$ **do**
6: **if** $SCLink_j > 0$ and $L_i > 0$ **then**
7: **if** $B_i - b_{i,j} \geq 0$ and $\text{SINR}_{i,j} \geq \text{SINR}_m$ **then**
8: $W_{i,j} = r_{i,j}$
9: **end if**
10: **end if**
11: **end for**
12: **end for**
13: **for** each $J - I$ dummy NFP pair **do**
14: $W_{i,j} = 0$
15: **end for**
16: Let $H_{J\times J}$ be a new matrix ▷ a boolean matrix, a true value in i, j index depicts, f_i is assigned to s_j $H = \text{HUNGARIAN}(W)$ ▷ Hungarian Algorithm is the bipartite matching algorithm which will return a boolean matrix
17: **for** each $f_i \in F$ **do**
18: **for** each $s_j \in S$ **do**
19: **if** $H_{i,j} = 0$ **then**
20: Associate s_j with f_i
21: $L_i = L_i - 1$
22: $B_i = B_i - bi, j$
23: $SCLink_j = SCLink_j - 1;$
24: **end if**
25: **end for**
26: **end for**
27: **end while**

information about both SCs and NFPs, such as $r_{i,j}$, $b_{i,j}$, $\text{SINR}_{i,j}$, SINR_min , L_i, and B_i.

The main idea of the proposed HBCA is based on extending the Hungarian algorithm to handle the unbalanced association problem between SCs and NFPs, where the number of SCs is much larger than the number of NFPs. The Hungarian algorithm gives the optimal solution in case of one-to-one matching (which is not the case in our SCs-NFPs association problem). The first part of the proposed HBCA, i.e., `HBCA - part 1`, can be briefly explained as follows.

`HBCA - part 2`

Input: (S, F)
Improve=true
while Improve **do**
Improve=false
for each pair $(f_i, f_j) \in F$ **do** ▷ where $i \neq j$ ▷ $s_i, s_j \in S$ are assigned with f_i and f_j respectively
Find Total Sum rate : SumRate
Swap association (f_i, s_j) , (f_j, s_i)
Find Total Sum rate: SumSwap
if $SumSwap > SumRate$ && all constraints still satisfied **then**
Swap association (f_i, s_j) , (f_j, s_i)
Improve=true
end if
end for
end while

`HBCA - part 3`

Input: (S, F)
Improve=true
while Improve **do**
for each pair $(f_i, f_j) \in F$ **do** ▷ where $i \neq j$ ▷ $s_j \in S$ are assigned with f_j
Improve=false
if f_i is free **then**
Find Total Sum rate : SumRate
Unassigned s_j from f_j and assigned it to f_i
Find Total Sum rate: SumSwap
if $SumSwap > SumRate$ && all constraints still satisfied **then**
Unassigned s_j from f_j and assigned it to f_i
Improve=true
end if
end if
end for
end while

- The algorithm starts by checking if there is a free NFP (has remaining links and enough bandwidth), then fills the W matrix of the SCs and NFPs with $r_{i,j}$ or zero based on the constraints. (lines 4-12).
- Since the Hungarian algorithm accepts only a squared matrix and the number of SCs is much larger than the number of NFPs, a number of dummy NFPs is added that represents the difference between the number of SCs and NFPs. (lines 13-15).
- The Hungarian algorithm returns the association between SCs and NFPs, and then, the available bandwidth and number of remaining links for non-dummy NFPs is updated accordingly. (lines 16-26).
- This process will repeat until all SCs are assigned or there are no free NFPs.

The simulation results presented in Section VI show a gap between the performance of the proposed `HBCA - part 1` and the optimal solution. In the following, the performance of `HBCA - part 1` is improved by introducing `HBCA - part 2` to reduce the gap to the optimal solution. The basic idea of `HBCA - part 2` can be explained as follows. `HBCA - part 2` checks if swapping the already existing association between SCs and NFPs obtained from `HBCA - part 1` can further lead to a higher total sum rate. If yes, `HBCA - part 2` swaps the association; otherwise, the association kept as in `HBCA - part 1`. It is worth mentioning that if algorithm `HBCA - part 2` swapped the association then the new matrix will be used to find the total sum rate.

The performance of the HBCA can be further improved as shown in `HBCA - part 3`. The basic idea of `HBCA - part 3` is to check if any NFP has free links and if so, then dissociating one SC and associating it with the free NFP can lead to higher total sum rate.

Fig. 5 shows an example that helps illustrate the `HBCA` algorithm more clearly. We have two NFPs f_1 and f_2, where f_1 has one link and f_2 has 2 links. We also have three SCs and each SC requires a specific data rate from each NFP as shown in Fig. 5. Please note that we suppose all links between SCs and NFPs satisfy the minimum SINR and each NFP has enough bandwidth. Since the Hungarian algorithm accept only a square matrix, the `HBCA` starts by adding a dummy NFPs to create and fill a $3X3$ W matrix with the requested $r_{i,j}$ if all constraints are satisfied and with zero otherwise.

After that, `HBCA` sends the matrix to the Hungarian algorithm to get the optimal one-to-one match between the SCs and the NFPs as shown in Fig. 6. Each SC that has a real association drops from the W matrix. This process repeats until all the NFPs links are used, all the NFPs bandwidth are used or all SCs are associated. Fig. 7 shows the initial association between the SCs and the NFPs at the end of `HBCA - part 1`. Fig. 8 shows the association between the SCs and the NFPs at the end of `HBCA- part 2`, where some swapping between the associated SCs and NFPs has been done to enhance the total sum rate and the new associations should satisfy all constraints.

There is a swap here, where the association between (s_1 associates instead with f_1) and (s_2 associates with f_2) to be (s_1 associates with f_2) and (s_2 associates with f_1) enhances the total sum rate. Finally if there is available any free links `HBCA- part 3` checks if dropping some SCs association and associate it with the free NFPs can improve the total sum rate or not, However in this example there are no free links; therefore we give another example to explain `HBCA- part 3` as shown in Fig. 9. We have two NFPs f_1 and f_2, where f_1 has one link and f_2 has 2 links, we also have two SCs and each SC requires a specific data rate from each NFP as shown in Fig. 9.

Please note that we assume all links between SCs and NFPs satisfy the minimum SINR and each NFP has enough bandwidth. Fig. 10 shows the initial association between the SCs and NFPs after `HBCA- part 1`, `HBCA- part 2` does not change the association in this case. On the other hand, `HBCA- part 3` disassociates s_1 from f_1 and associates it with f_2, since f_2 satisfy all constraints and has a free link. This process enhances the total sum rate as shown in Fig. 11.

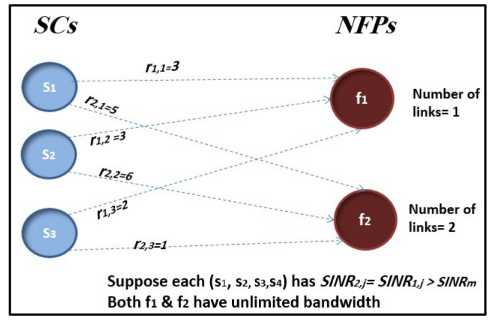

Fig. 5. SCs ask NFPs to associate

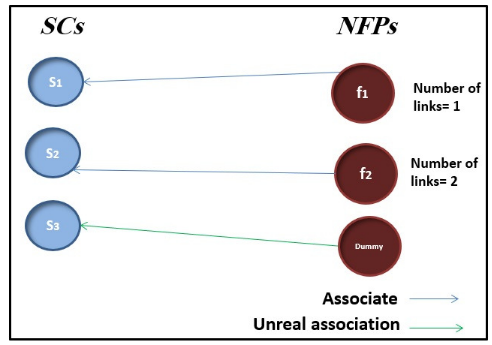

Fig. 6. Hungarian Algorithm returned matching

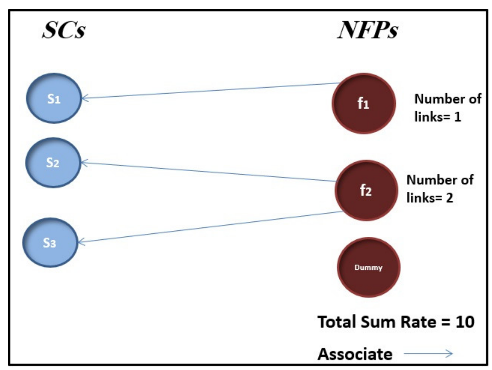

Fig. 7. SCs and NFPs Association at the end of `HBCA-part 1`

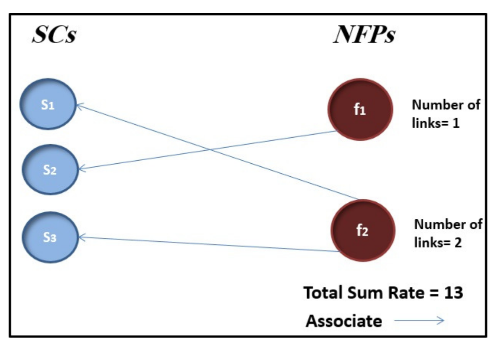

Fig. 8. SCs and NFPs Association at the end of `HBCA`

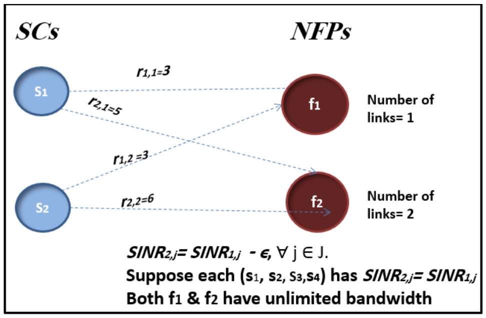

Fig. 9. SCs ask NFPs to associate

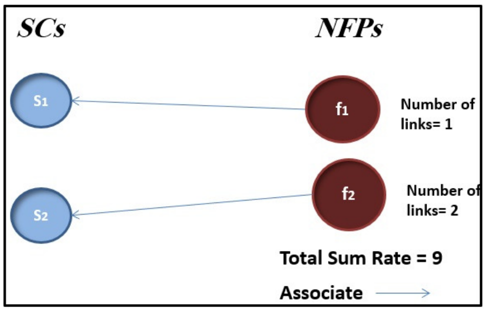

Fig. 10. SCs and NFPs Association at the end of `HBCA-part 1`

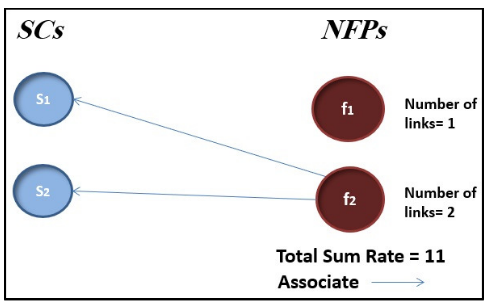

Fig. 11. SCs and NFPs Association at the end of `HBCA`

B. Stable Marriage Based Distributed Algorithm (SMBDA)

The stable marriage algorithm is the problem of finding a stable matching between two equally sized sets of elements given an ordering of preferences for each element [39]. A matching is a mapping from the elements of one set to the elements of the other set. Matching is not stable if there is an element A of the first matched set that prefers some given element B of the second matched set over the element to which A is already matched, or if B prefers A over the element to which B is already matched. In the end, each element in both lists will have a matched element from the other list.

In this section, a stable marriage based distributed algorithm (SMBDA) is proposed to efficiently solve the SCs and NFPs association problem in low complexity. In the distributed algorithm, each SC and NFP stores only local information. This means that each SC and NFP is responsible for its association, where SC sends the request to associate with NFP or vice-versa.

In a `preferable list` algorithm, each SC and NFP

`preferable list`

Input: (S, F, r)
Each free $f_i \in F$ will broadcast B_i to all $s_j \in S$
Each s_j calculate $\text{SINR}_{i,j}$ and $b_{i,j}$
Each s_j will broadcast it's local information to all $f_i \in F$
Each s_j will fill it's $PrefSC$ with the indices of the preferred NFPs based on $r_{i,j}$ from max to min
Each f_i will fill it's $PrefNFP$ with the indices of the preferred SCs based on $r_{i,j}$ from max to min

are filled a preferable list, based on the maximum data rate between SC and NFP, starting from the most preferable down to least preferred. As previously mentioned, the number of SCs is much larger than the number of NFPs and considering the fact that the size of the first list and second list in the stable marriage algorithm should be the same, we added dummy NFPs. The preferable list will help each SC to connect with the most appropriate NFP which is done in the `SMBDA` algorithm.

SMBDA Algorithm starts at the beginning with all SCs and NFPs are free, (line 2). Each NFP and SC broadcasts its local information; after that the NFP f_i first selects the preferable SC s_j from it's $PrefNFP$ and if the constraints are satisfied, f_i sends a association request to that s_j, (lines 3-7). There are three cases:

- Case 1: SC s_j is free and it sends accept and call `Accept`$(s_j, f_i, L_i, B_i, b_{i,j})$, (lines $8-9$).
- Case 2: SC s_j is engaged (no final association) to NFP f_k and SC s_j prefers NFP f_i more, then SC s_j sends disassociate message to NFP f_k and sends accept message to NFP f_i and call `Accept`$(s_j, f_i, L_i, B_i, b_{i,j})$, (lines $10-12$).
- Case 3: SC s_j is engaged (no final association) to NFP f_k and SC s_j prefers NFP f_k more, then SC s_j sends Reject message to NFP f_i. after that, NFP f_i selects the next SC in its $PrefNFP$ list, (lines $13-15$).

After that, the SC married (final associated) with NFP, that provide it a stable matching.

As can be seen in `Accept` algorithm, when s_j calls `Accept`$(s_j, f_i, L_i, B_i, b_{i,j})$, then SC s_j and NFP f_i local information will be updated as following:

- Associate SC s_j with NFP f_i
- Decrease B_i by $b_{(i,j)}$
- Decrease L_i by 1
- if L_i less than or equal to zero, set f_i to be not free.

On the other hand, in the `Disassociate` algorithm, when s_j calls `Disassociate` $(s_j, f_i, L_i, B_i, b_{i,j})$, then SC s_j, NFP f_i local information will be updated as followed:

- Disassociate SC s_j with NFP f_i
- Increase B_i by $b_{(i,j)}$
- Increase L_i by 1
- Set f_i to be free.

Using `SMBDA` we found the best association for the fig. 5 example. In this example there are three SCs and two NFPs with three Free links. Fig. 12 shows each SC and NFP with

```
SMBDA
```

Input: $(S, F, \text{PrefNFP}, \text{PrefSC}, L, B, b)$
At the beginning All $f_i \in F$ and $s_j \in S$ are free
Each free f_i broadcasts B_i to all s_j
Each s_j calculates $\text{SINR}_{i,j}$ and $b_{i,j}$
Each s_j broadcasts it's local information to all f_i
$s_j = PrefNFP_i$
If all constraints are satisfied f_i Sends Request to associate with s_j
Case 1: s_j **is free**
Calls `Accept`$(s_j, f_i, L_i, B_i, b_{i,j})$
Case 2: s_j **is engaged to** f_k **and** f_i **is more preferable, as in** $PrefSC_j$**, for** s_j **than** f_k
Calls `Accept`$(s_j, f_i, L_i, B_i, b_{i,j})$
Calls `Disassociate`$(s_j, f_k, L_k, B_k, b_{k,j})$
Case 3: s_i **prefers** f_k **more than** f_i **as in** $PrefSC_j$ **list**
s_i Rejects to associate with f_i
Go to the next preferred SC in $PrefSC_j$ preferable list

```
Accept
```

Input: $(s_j, f_i, L_i, B_i, b_{i,j})$
s_j connect with f_i
$L_i = L_i - 1$
$B_i = B_i - b_{i,j}$
if $L_i == 0$ **then**
set f_i is not free
end if

its preferable lists. For example, as can be seen from Fig. 12, s_1 prefers f_2 then f_1; however, f_1 prefers s_1 first then s_2 and finally s_3. `SMBDA` starts when f_1 sends request to s_1, which accepts to engage (not fixed associates) with f_1.

After that, as shown in Fig. 13, f_2 sends request to s_2 which accepts to engage to f_2. Since f_2 has 2 links and enough bandwidth f_2 sends request to the second SC in its preferable list which is s_1. Therefore, s_1 disengages from f_1 and engages to f_2 as shown in Fig. 14. Again f_1 becomes free. Hence, f_1 sends request to s_2 which is reject, since s_2 is engaged to f_2 and it prefers f_2 more than f_1. Hence, f_1 sends request to s_3 which accepts to engage with f_1 as shown in Fig. 15. At the end there are no free links, and all SCs and NFPs have stable match. Fig. 16 shows the final association between the SCs and the NFPs.

C. Complexity analysis

In this subsection, the worst case complexity analysis of the proposed algorithms is provided. We find that the time complexity of `HBCA - part 1` in the worst case is $O(IJ^3)$. This can be explained as follows:

- The while loop in line 3 requires a complexity of $O(I)$
- The for loops in lines 4 and 5 require a complexity of $O(I)$ and $O(J)$, respectively.
- The for loop in line 13 requires a complexity of $O(J - I)$, and the Hungarian algorithm in line 16 requires a complexity of $O(J^3)$.
- The for loops in lines 17 and 18 require a complexity of $O(I)$ and $O(J)$, respectively.

```
Disassociate
```

Input: $(s_j, f_i, L_i, B_i, b_{i,j})$
s_j disconnect from f_i
B_i=B_i + $b_{i,j}$
set f_i is free
$L_i = L_i + 1$

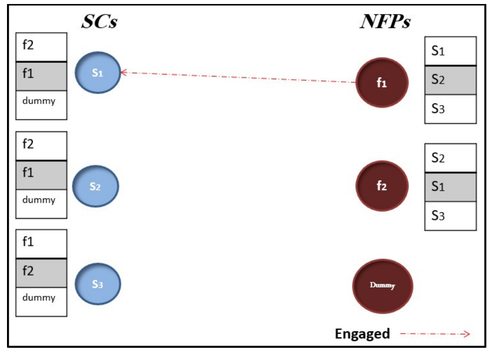

Fig. 12. NFP f_1 is engages to s_1

Hence the worst case complexity of `HBCA - part 1` is $O(I)\big(O(IJ) + O(J - I) + O(J^3) + O(IJ)\big) = O(IJ^3)$.

For `HBCA - part 2` time complexity can be clarified as follows:

- The while loop in line 3 requires a complexity of $O(W)$ where $W = \sum_{i=1}^{I}\sum_{j=1}^{J} r_{i,j}$.
- The for loop in line 5 requires a complexity of $O(IJ)$.
- The SumRate and SumSwap in lines 6 and 8 both require a complexity of $O(IJ)$.

Hence, `HBCA - part 2` time complexity in worst case is $O(W)(O(IJ)O(IJ)) = O(WI^2J^2)$. Similarly the worst case computational complexity of `HBCA - part 3` is $O(WI^2J^2)$. Finally, the overall time complexity for the HBCA algorithm will be $O(IJ^3) + O(WI^2J^2) + O(WI^2J^2) = O(WI^2J^2)$.

Regarding the time complexity of `SMBDA`, it is found that in the worst case each NFP at maximum sends requests to all SCs. Therefore, the time complexity in the worst case occurs when all NFPs send to all SCs, is $O(IJ)$. Furthermore, in the worst case, each SCs rejects all NFPs. Hence, the time complexity in the worst case occurs when all SCs reject all NFPs, is $O(IJ)$. Subsequently, the worst time complexity of `SMBDA` is $O(IJ)$. This explains that `SMBDA` termination is assured, as each NFP at maximum sends request to all SCs depend on each NFP $PrefNFP$ list, and each SC at maximum sends reject or accept to all NFPs depend on each SC $PrefSC$ list.

Table I summaries the worst case time complexity of the proposed algorithms with the $(DM)^2S$ algorithm. It can be noticed that the proposed algorithms are slightly more computationally expensive than the $(DM)^2S$ algorithm in the worst case. However, the proposed algorithms are computationally acceptable and are practically applicable.

Additionally, the message complexity of the `SMBDA` distributed algorithm is found. The NFP message complexity in

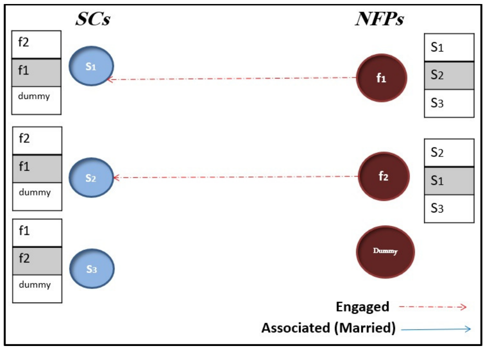

Fig. 13. NFP f_2 engages to s_2

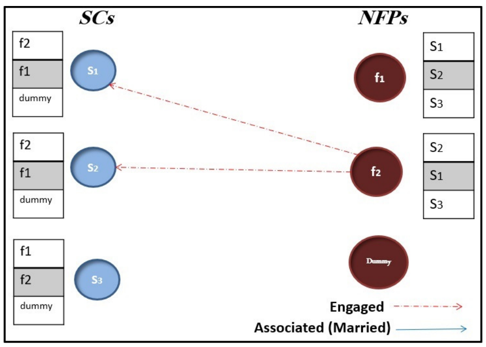

Fig. 14. NFP f_2 engages to s_1

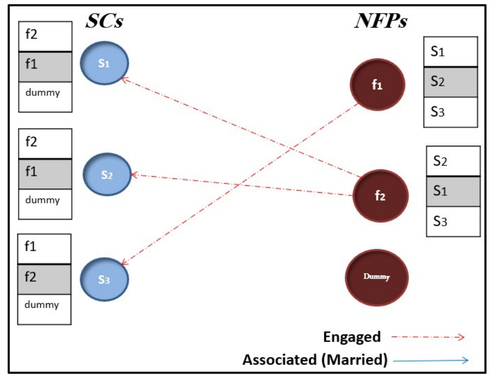

Fig. 15. NFP f_1 engages to s_3

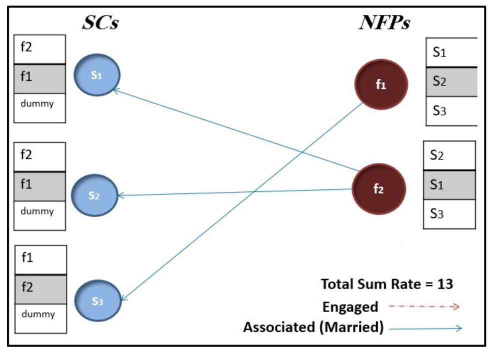

Fig. 16. NFP f_1 fix associates to s_3 and f_2 fix associates to both s_1 and s_2

worst case happens when the NFP sends a request message to each single SC. Therefore, the NFP message complexity is $O(IJ)$ in worst case. The SC message complexity in the worst case happens when the SC sends accept or reject message to each single f. Therefore, the SC message complexity is $O(IJ)$ in the worst case. Hence, the SMBDA message complexity is $O(IJ) + O(IJ) = O(IJ)$.

Table II compares the worst case time complexity of the proposed SMBDA algorithm and the $(DM)^2S$ algorithm. It can be noticed that the proposed SMBDA algorithm message complexity is the same as its counterpart of the $(DM)^2S$ algorithm.

VI. Performance Evaluation

We use the Gurobi optimization tool [40] to find the the integer linear program (ILP) solution for problem (8), which takes an exponential time. In this section, the performance of the proposed HBCA and SMBDA algorithms are investigated and the results with their counterparts obtained from the optimal solution of ILP of the problem in (8) are compared with the proposed distributed algorithm $(DM)^2S$ in [20].

A 5G+ system is considered, where the SCs and NFPs are uniformly distributed within a 4 km by 4 km area. The data rates used in [20] is considered, then the bandwidth $b_{i,j}$ and $\mathrm{SINR}_{i,j}$ are calculated. Without loss of generality, we assume that all NFPs have the same hight, $h_{d_i} = h_d = 300$ m $\forall i$, and all NFPs have the same bandwidth $B_i = B = 250$ MHz $\forall i$. Following [20], the rest of parameters are defined in Table III.

Fig. 17 shows the total sum rate of the proposed HBCA, SMBDA, ILP, $(DM)^2S$ versus the number of SCs at 30 NFPs. As can be seen, the HBCA and the SMBDA performances approach that of the optimal results of the ILP. One can also see from Fig. 17 that the proposed SMBDA outperforms the $(DM)^2S$. As previously discussed in the $(DM)^2S$ algorithm, the SC sends a request to associate with the NFP of the highest SINR, and if rejected then that SC will not associate with another NFP. Even more, as shown in Fig. 17, the total sum rate in both proposed HBCA and SMBDA along with the ILP increases when the number of SCs increases, on the other hand, the total sum rate of $(DM)^2S$ has a little increment.

Fig. 18 shows the total sum rate of the proposed HBCA, SMBDA, ILP, $(DM)^2S$ versus the number of NFP at 100 SCs. Again as shown in Fig. 18, the HBCA and the SMBDA performances approach that of the optimal results of the ILP. Further, Fig. 18 shows that the proposed SMBDA outperforms the $(DM)^2S$ as discussed in Fig. 17. Additionally, as shown in Fig. 18, the total sum rate of the proposed HBCA and SMBDA along with the ILP at the beginning increases with the increase of NFP. After that, the total sum rate saturates or slightly increases. This can be explained as when the number of NFPs reaches 40, almost all SCs are associated which causes the total sum rate to saturate. However, the total sum rate of $(DM)^2S$ almost unchanged.

Fig. 19 shows the total sum rate of the proposed HBCA, SMBDA, ILP, $(DM)^2S$ versus the number of SCs at 50 NFPs. As can be seen, the HBCA and the SMBDA performances

TABLE I
COMPUTATIONAL TIME COMPLEXITY OF THE PROPOSED ALGORITHMS.

Algorithm	Time Complexity Order
HBCA	$O(WI^2J^2)$
SMBDA	$O(IJ)$
$(DM)^2S$	$O(IJ)$

TABLE II
COMPUTATIONAL MESSAGE COMPLEXITY OF THE ALGORITHMS.

Algorithm	Message Complexity Order
SMBDA	$O(IJ)$
$(DM)^2S$	$O(IJ)$

approach that of the optimal results of the ILP. However the ILP takes around 46882 second running time, which is a very long time comparing to HBCA which takes around 138 second and the SMBDA which takes around 48 seconds. where is the used computer processor is Intel core i7-8750h 2.20 GHz and the ram is 16 GB. As shown in Fig. 19, the proposed SMBDA outperforms the $(DM)^2S$. Even more, the total sum rate in both proposed HBCA and SMBDA along with the ILP highly increases when the number of SCs increase in contradiction with $(DM)^2S$ where the total sum rate is slightly increased.

Fig. 20 shows the total sum rate of the proposed HBCA, SMBDA, ILP, $(DM)^2S$ versus the number of NFP at 200 SCs. As shown in Fig. 20, the HBCA and the SMBDA performances approach that of the optimal results of the ILP. Fig. 20 shows that the proposed SMBDA outperforms the $(DM)^2S$.

As one can see from figure (21), as the number of SCs increases the total number of associated SCs decreases. However, the number of SCs associate with NFPs in our algorithms are very close to the number of SCs associate with NFPs in the optimal case. Moreover, the number of associate SCs in our algorithms outperform the number of associate SCs in case of $(DM)^2S$.

As one can see from the previous examples both HBCA and SMBDA outperform $(DM)^2S$. As mentioned before $(DM)^2S$ is a distributed algorithm, therefore, it has only a local information and this explain why HBCA outperform $(DM)^2S$, since HBCA algorithm is a centralized algorithm. As discussed in IV, in the first step of $(DM)^2S$ algorithm each SC sends a message to the NFP with the maximum SINR. Basically, each SC wants to connect NFP with the best SINR link. However, based on other constraints such as the number of links or the bandwidth each NFP can support, the NFP could send rejections to some SCs and these SCs will not attempt to associate with another NFP.

The SMBDA is also a distributed algorithm; however, the SMBDA tries to find the best association for each SC as explained in Section V-B taking into consideration all constraints. The SMBDA tries to find the best association by exploiting the idea of the stable marriage algorithm, which finds the stable match between two lists. This explains why the SMBDA outperforms the $(DM)^2S$ algorithm.

TABLE III
SIMULATION PARAMETERS

Parameter	Value	Parameter	Value
α	9.61	β	0.16
η_{Los}	1 dB	η_{NLos}	20 dB
f_c	2 GHz	P_t	5 Watts
σ_j	1 dB	h_d	300 m
S	$(50-100)$ $(60-200)$	F	$(20-50)$ $(50-100)$
SINR_{min}	-5 dB	L_i	$2-5$

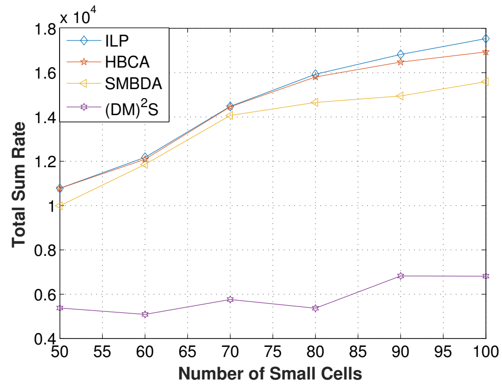

Fig. 17. Total Sum Rate versus the number of SCs for the proposed HBCA, SMBDA, ILP, and $(DM)^2S$ at 30 NFPs.

Fig. 22 shows the total sum rate versus the number of SCs for the proposed HBCA, SMBDA, ILP, and $(DM)^2S$ at 30 NFPs. In this scenario, the $\text{SINR}_{i,j}$ is given random values between -10 and 0 dB to put the proposed HBCA and SMBDA algorithm under a critical limitation. As it can be seen, in Fig. 22, the HBCA, and the SMBDA performances approach that of the optimal results of the ILP. Moreover, one can see from Fig. 22 that the proposed SMBDA outperforms the $(DM)^2S$, which becomes clear as previously discussed. Fig. 23 shows the total number of associated SCs versus the number of SCs for the proposed HBCA, SMBDA, ILP, and $(DM)^2S$ at 30 NFPs. As can be seen the total number of associated SCs of the proposed HBCA and SMBDA algorithms approximated the total number of associated SCs of the ILP. However, the number of associated SC of $(DM)^2S$ is less than SMBDA and HBCA. We find different results with different numbers of SCs and NFPs other than the previous one. We find that regardless the number of SCs or NFPs, the two proposed algorithms approach their counterparts obtained from the optimal solution of the integer linear program (ILP) of the problem in (8) and outperform the proposed distributed algorithm $(DM)^2S$ in [20].

VII. CONCLUSION

In this work, the association problem of the NFPs with SCs of future cellular network is studied to maximize the system sum rate while taking into consideration each NFP bandwidth, the number of supported links, and minimum required SINR. We proposed a centralized (HBCA) and a distributed (SMBDA) algorithm to find a sub-optimal association between the SCs and NFPs, at reduced computational complexity. The

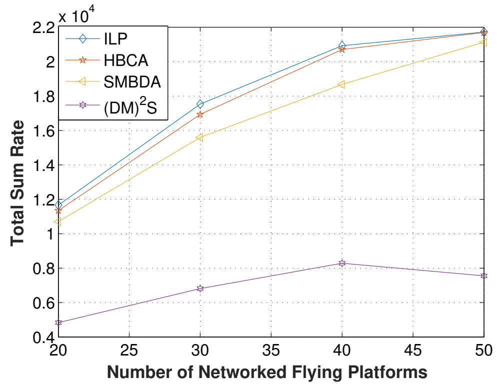

Fig. 18. Total Sum Rate versus the number of NFPs for the proposed HBCA, SMBDA, ILP, and $(DM)^2S$ at 100 SCs.

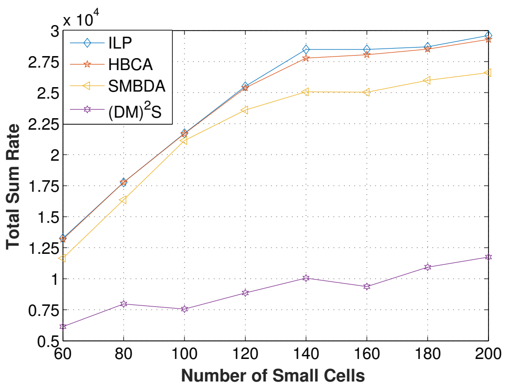

Fig. 19. Total Sum Rate versus the number of NFPs for the proposed HBCA, SMBDA, ILP, and $(DM)^2S$ at 50 NFPs.

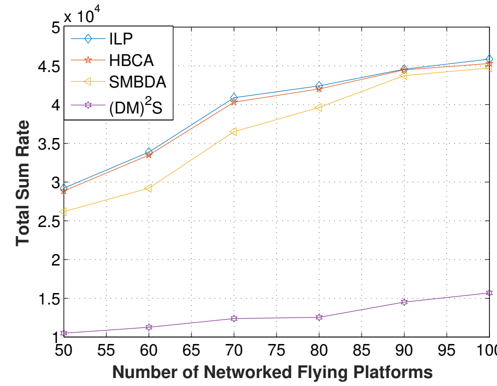

Fig. 20. Total Sum Rate versus the number of NFPs for the proposed HBCA, SMBDA, ILP, and $(DM)^2S$ at 200 SCs.

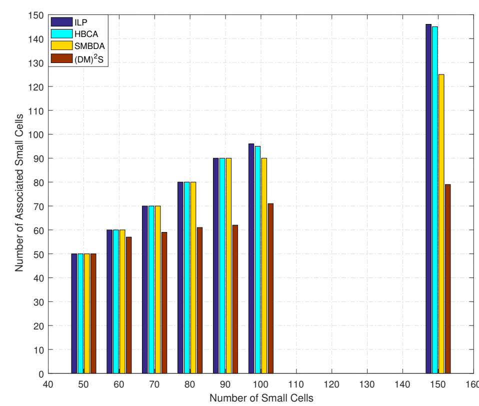

Fig. 21. Total Number of Associated SC versus the number of SCs at 50 NFPs.

numerical evaluation of the considered case study has shown that the performance of the proposed algorithms outperform the performance of the existing algorithm in terms of the number of connected SCs and the total sum rate. In future works, it is of interest to investigate other resource allocation problems such as minimizing the total interference while satisfying a target sum rate.

REFERENCES

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?, *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 1065 - 1082, Jun. 2014.

[2] A. Pizzinat, P. Chanclou, F. Saliou, and T. Diallo, "Things you should know about fronthaul," *IEEE J. Lightw. Technol.*, vol. 33, no. 5, pp. 1077-1083, Mar. 2015.

[3] A. GUPTA, and R. KUMAR JHA, "A Survey of 5G Network: Architecture and Emerging Technologies," *IEEE Access*, vol. 3, pp 1206-1232, Jul. 2015.

[4] W. Vincent WS, R. Schober, D. Wing Kwan Ng, and L. Wang, "Key technologies for 5G wireless systems", *Cambridge university press*, 2017.

[5] "Microwave towards 2020," *Ericsson*, Stockholm, Sweden, White Paper, Sep. 2015, accessed on May 18, 2017.

[6] "Buisness case elements for small cell virtualization," *Real Wireless Ltd.*, Pulborough, U.K., White Paper SCF 158, Jun.2015, accessed on May 18, 2017.

[7] M. Alzenad, M. Z. Shakir, H. Yanikomeroglu, and M. S. Alouini, "FSO-based vertical backhaul/fronthaul framework for 5G+ wireless networks.," *IEEE Commun. Mag*, vol. 56, pp. 218 - 224, Jan. 2018.

[8] https://www.thatsclever.co.uk/drones-vs-bad-weather/

[9] https://www.aerisweather.com/blog/2018/04/30/impact-of-weather-within-uas/

[10] M. Mozaffari, W. Saad, M Bennis, Y. Nam, and M. Debbah "A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems," *arXiv preprint arXiv:1803.00680,* Mar. 2018.

[11] M. Erdelj and E. Natalizio. "UAV-Assisted Disaster Management: Applications and Open Issues," in *Computing, Networking and Communications (ICNC)*, Kauai, HI, USA, Feb. 2016, pp. 1-5.

[12] A. Restas, "Drone applications for supporting disaster management," *World Journal of Engineering and Technology*, pp. 316-321 , Oct. 2015.

[13] T. Joseph Tanzi, M. Chandra, J. Isnard, D. Camara, O. Sbastien, and F. Harivelo, "Towards "drone-borne" disaster management: future application scenarios," In *Proc. XXIII ISPRS Congress, Commission VIII* , Prague, Czech Republic, Jul. 2016, pp. 316-321.

[14] H. Ahmadi, K. Katzis, M. Z. Shakir, "A Novel airbone self-organising architecture for 5G+ networks," in *Proc. IEEE Vehicular Technology Conference (VTC)*, Toronto, ON, Canada, Sep. 2017, pp. 1-5.

[15] S. Meng, X. Su, Z. Wen, X. Dai, Y. Zhou, and W. Yang, "Robust drones formation control in 5G wireless sensor network using mmWave," in *Wireless Communications and Mobile Computing,* vol. 2018, pp.1-7, May 2018.

[16] S. A. R. Naqvi, S. A. Hassan, H. Pervaiz, and Q. Ni, "Drone-Aided Communication as a Key Enabler for 5G and Resilient Public Safety Networks," *IEEE Commun. Mag*, vol 56, pp. 36-42, Jan. 2018.

[17] S. A. Shah, T. Khattab, M. Z. Shakir, and M. O. Hasna, "Association of networked flying platforms with small cells for network centric 5G+ C-RAN," in *Proc. IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Montreal, Canada, Oct. 2017, pp. 1-7.

[18] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Drone small cells in the clouds: design, deployment and performance analysis," in *Proc.*

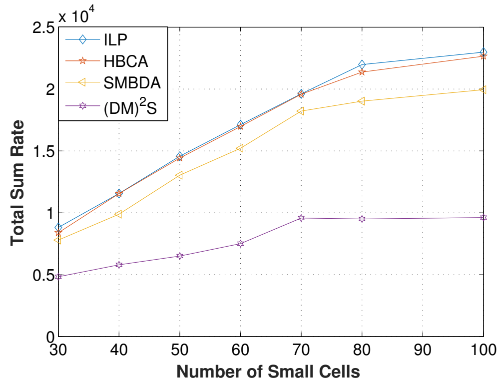

Fig. 22. Total Sum Rate versus the number of SCs for the proposed HBCA, SMBDA, ILP, and $(DM)^2S$ at 30 NFPs.

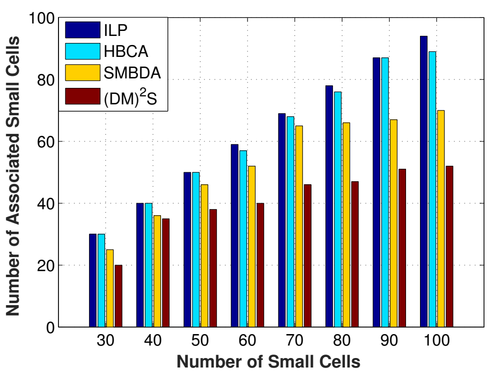

Fig. 23. Total Number of Associated SC versus the number of SCs for the proposed HBCA, SMBDA, ILP, and $(DM)^2S$ at 30 NFPs.

IEEE Global Communications Conference (GLOBECOM), San Diego, CA, USA, Dec. 2015, pp. 1-6.

[19] E. Kalantari, H. Yanikomeroglu, and A. Yongacoglu, "On the number and 3D placement of drone base stations in wireless cellular networks," in *Proc. IEEE Vehicular Technology Conference (VTC)*, Montreal, QC, Canada, Sep. 2016, pp. 1-6.

[20] S. A. W. Shah, T. Khattab , M. Z. Shakir and M. O. Hasna, "A distributed approach for networked flying platform association with small cells in 5G+ network," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Singapore, Dec. 2017, pp. 1-7.

[21] S. Onoe, "Evolution of 5G mobile technology toward 2020 and beyond," in *Proc. IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, CA, USA, Jan. 2016, pp. 23-28.

[22] R. Yu, J. Ding, X. Huang, M. Zhou, S. Gjessing, Y. Zhang, "Optimal resource sharing in 5G-enabled vehicular networks: A matrix game approach," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 7844-7856, Mar. 2016.

[23] A. Lotfi, H. Beyranvand, J. A. Salehi, and M. Maier, "Ultra-Dense 5G Small Cell Deployment for Fiber and Wireless Backhaul-Aware Infrastructures," *IEEE Transactions on Vehicular Technology*, DOI 10.1109/TVT.2018.2875114.

[24] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Optimal transport theory for power-efficient deployment of unmanned aerial vehicles," in *Proc. IEEE International Conference on Communications (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1-6.

[25] Z. Mlika, E. Driouch, W. Ajib, and H. Elbiaze, "A completely distributed algorithm for user association in HetNets," in *Proc. IEEE International Conference on Communications (ICC)*, London, UK, Jun. 2015, pp. 2172-2177.

[26] G. Araniti, M. Condoluci, A. Orsino, A. Iera, and A. Molinaro, "Effective Resource Allocation in 5G-Satellite Networks," in *Proc. IEEE International Conference on Communications (ICC)*, London, UK, Jun. 2015, pp. 844-849.

[27] I. AlQerm, B. Shihada, "Energy Efficient Power Allocation in Multi-tier 5G Networks Using Enhanced Online Learning", *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 11086 - 11097, Jul. 2017.

[28] A. Abdelhadi, and C. Clancy"An Optimal Resource Allocation with Joint Carrier Aggregation in 4G-LTE," in *Proc. International Conference on Computing, Networking and Communications (ICNC),* Garden Grove, CA, USA, Mar. 2015.

[29] A. Abdelhadi, C. Clancy, and J. Mitola, "A Resource Allocation Algorithm for Users with Multiple Applications in 4G-LTE," in*Proc. ACM workshop on Cognitive radio architectures for broadband,* Miami, Florida, USA, Oct. 2013, pp. 13-20.

[30] M. Zulhasnine, C. Huang and A. Srinivasan, "Efficient Resource Allocation for Device-to-Device Communication Underlaying LTE Network," in *Proc. IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications,* Niagara Falls, ON, Canada, Nov. 2010, pp. 368-375.

[31] X. LIANG, M. W. Kim, S. Uno and S. MIURA, "Adaptive Resource Allocation Method for Mobile Network," in *Proc. 13th IEEE International Conference,* Seoul, South Korea, Apr. 2011, pp. 350-355.

[32] S. K. Padaganur and J. D. Mallapur, "A Neural Network Based Resource Allocation Scheme for 4G LTE Heterogeneous Network," *Proc. 2nd International Conference for Convergence in Technology,* Mumbai, India, Dec. 2017, pp. 250-254.

[33] S. WEN, X. ZHU, X. ZHANG, and D. Yang, "QoS-Aware Mode Selection and Resource Allocation Scheme for Device-to-Device (D2D) Communication in cellular networks," *Proc. IEEE International Conference on Communications Workshops,* Budapest, Hungary, Oct. 2013, pp. 101 - 105.

[34] S. Alamouti and A. Sharafat, "Resource Allocation for Energy-Efficient Device-to-Device Communication in 4G Networks," in *Proc. International Symposium on Telecommunications,* Tehran, Iran, Jan. 2015, pp. 1058 - 1063.

[35] D. Feng, L. Lu, Y. Yuan-Wu, G. Y. Li, G. Feng, and S. Li, "Devicetodevice communications underlaying cellular networks," *IEEE Transactions on Communications*, vol. 61, no. 8, pp. 3541-3551, Aug. 2013.

[36] M. Garey and D. Johnson, "Computers and Intractability; A Guide to the Theory of NP-Completeness," *W. H. Freeman and Co.*, 1990.

[37] S. Shah, T. Khattab, M. Shakir, M. Khafagy, M. Hasna. "Small Cell Association with Networked Flying Platforms: Novel Algorithms and Performance Bounds," *arXiv preprint arXiv:1802.01117*, 2018 Feb 4.

[38] E. Balas, and E. Zemel, "An algorithm for large zero-one knapsack problems," *Operations Research*, vol. 28, pp. 1130-1154, Oct. 1980.

[39] D. Gusfield, and R. W. Irving, "The Stable Marriage Problem: Structure and Algorithms," *Cambridge, MA: MIT Press*, 1989.

[40] http://www.gurobi.com/products/gurobi-optimizer

Huda AlSheyab is a second-year Doctoral student of Electrical and Computer Engineering at Lakehead University, Thunder Bay, ON. Her research interests include Wireless Communications including heterogeneous small cell networks, and 5G (5G+) technologies, such networked-flying platforms. She received her MS in Computer Engineering from Jordan University of Science and Technology, Irbid, Jordan in 2016. She received her Bachelors degree in Computer Engineering from Jordan University of Science and Technology, Irbid, Jordan in 2012.

Salimur Choudhury received the Ph.D. degree in computing from Queens University, Kingston, ON, Canada, in 2012. He is currently an Assistant Professor with the Department of Computer Science, Lakehead University, Thunder Bay, ON. His research interests include designing algorithms for wireless communication, optimization, cellular automata, approximation algorithms, and so on. He was the Technical Program Chair for the SGIoT 2017, SGIoT 2018 and iThings 2018 Conferences. He has been serving as an Editor for the Parallel Processing Letters and the International Journal of Computers and Applications.

Ebrahim Bedeer received the B.Sc. (Hons.) and M.Sc. degrees from Tanta University, Tanta, Egypt and the Ph.D. degree from Memorial University, St. Johns, NL, Canada, all in Electrical Engineering. He was a postdoctoral fellow at University of British Columbia (UBC), BC, Canada and Carleton University, Ottawa, ON, Canada. Currently, he is an Assistant Professor (Lecturer UK) at Ulster University, Jordanstown, United Kingdom (UK). His general research interests are in the area of wireless communications and signal processing, with a current focus of resource management in wireless networks and applications of optimization is signal processing.

Dr. Bedeer is an Editor of the IEEE COMMUNICATIONS LETTERS and IEEE ACCESS. He has served on the Technical Program Committees of numerous major international communication conferences, such as IEEE GLOBECOM, IEEE ICC, and IEEE VTC. Dr. Bedeer received numerous awards including the Exemplary Reviewer of IEEE COMMUNICATIONS LETTERS and IEEE WIRELESS COMMUNICATIONS LETTERS.

Salama S. Ikki received his Ph.D. degree in Electrical Engineering from Memorial University, St. Johns, NL, Canada, in 2009. From February 2009 to February 2010, he was a Postdoctoral Researcher at the University of Waterloo, ON, Canada. From February 2010 to December 2012, he was a Research Assistant with INRS, University of Quebec, Montreal, QC, Canada. He is currently an Associate Professor of wireless communications with Lakehead University, Thunder Bay, ON, Canada. He is the author of 100 journal and conference papers and has more than 4000 citations and an *H*-index of 30. His research interests include cooperative networks, multiple-input-multiple-output, spatial modulation, and wireless sensor networks. Dr. Ikki served on the Editorial Board of *IEEE Communications Letters* and Institution of Engineering and Technology Communications. He Also served as a Technical Program Committee member for various conferences, including *IEEE International Conference on Communications, IEEE Global Communications Conference, IEEE Wireless Communications and Networking Conference, IEEE Spring/Fall Vehicular Technology Conference*, and *IEEE International Symposium on Personal, Indoor, and Mobile Communications*. Dr Ikki received Best Paper Award for his paper published in the EURASIP Journal on Advanced Signal Processing. He also received an *IEEE Communications Letters, IEEE Wireless Communications Letters, IEEE Transactions on Vehicular Technology* and *IEEE Transactions on Communications* exemplary reviewer certificates for 2012, 2012 and 2014, respectively.