Maximizing Fairness for Resource Allocation in Heterogeneous 5G Networks

Ajay Pratap, Rajiv Misra, *Senior Member, IEEE*, and Sajal K. Das, *Fellow, IEEE*

Abstract—In this article, we first formulate the joint resource allocation, interference minimization, user-level, and cell-level fairness for maximum resource reuse in 5G heterogeneous small cell networks as an NP-hard problem. We then propose three algorithms—centralized, distributed, and randomized distributed algorithms—to efficiently solve the formulated resource allocation problem while minimizing interference, maximizing fairness, and resource reuse. Through extensive real data analysis and network simulations, we show that our proposed solutions outperform state-of-the-art schemes, namely interfering model (INT) and distributed random access (DRA), for both low and high-density 5G networks.

Index Terms—5G cellular networks, randomized and distributed algorithms, heterogeneous small cell networks

1 Introduction

The 5G wireless technology is expected to have a heterogeneous multi-tier deployment of a large number of femtocells, picocells, and microcells along with macrocell structured cellular networks. The *Small Cell Access Points* (SAPs) yield efficient spectrum reuse and reduced traffic load from *Macro Base Station* (MBS). The deployment of SAPs such as *Femto Access Point* (FAP), *Pico Access Point* (PAP), *Micro Access Point* (MAP) is an economically better solution than using more MBSs. The SAPs are of different capacity and coverage areas (ranging from a few hundred meters to few kilometers) but smaller than MBS. The MAP is an early form of small cells that complements MBS over a remote location like a hotspot, while the PAP and FAP are derived from the MAP concept. The coverage area of PAP 30 percent smaller than a MAP, and the PAP improves the capacity and coverage area over a hotspot region. To fulfill continuous high data demand for indoor and in-vehicle, the Third Generation Partnership Project (3GPP) came up with FAP to further reduce the transmit power. The FAP is cost-effective and low powered due to which this technique benefits both the operators and users. The method to deploy FAP not only manages the traffic but also improves the capacity and coverage area for indoor wireless users.

A *Small Cell User Equipment* (SUE) can connect to a SAP in three access control modes, such as open access, closed access, and hybrid access [1], [2]. In the open access mode, whenever a user of the same operator comes in a region of a small cell, it gets connected to it. In closed access, only the subscribed users of the small cell get services. This improves the *quality of service* (QoS) for high paying subscribers in terms of resource sharing. In the hybrid access mode, a limited resource of SAPs is available for all users while the rest are dedicated to the subscribed users. In this paper, we model the hybrid access mode where the high paying subscribers have higher priority; and lower-paying subscribers can avail resources only if they are available.

The deployment of SAPs faces many challenges due to different coverage areas, non-uniform nature, and random deployment scenarios, most notably, resource allocation, interference management, resource reuse, user-level, and cell level fairness. A FAP may get interference from PAP, MAP, or MBS, and vice versa. Therefore, an important question is how to allocate resources to the SAPs and respective users to meet the desired performance criteria without affecting others. Shared and split spectrum allocation schemes are used to allocate the limited available resources in small cell networks [3]. In the first scheme, the SAPs including the MBS can use the same frequency bands; whereas in the second scheme, the SAP uses different frequency bands as the MBS. The shared scheme results in more dynamic resource allocation, although the interference among different SAPs may result in a performance degradation. Similar to the conference version of this work [4], we have considered here a split spectrum allocation for macro-cell users. In contrast, the SUEs can reuse the resources following interference, user-level, and small cell level fairness constraints.

Due to the random deployment of SAPs, the coverage area of neighboring SAPs gets interfered, resulting in the deterioration of 5G cellular network's efficiency. As the total available resources is limited compared to the user demands, it is constitutive to make some juxtaposition among the subscribers and the respective serving SAPs before assigning them actual resources. Thus, a priority model is important, which is based on the users' subscription plan [5], [6]. To this end, we propose new resource

- *A. Pratap and S.K. Das are with the Department of Computer Science, Missouri University of Science and Technology, Rolla, MO 65409 USA. E-mail: {ajaypratapf, sdas}@mst.edu.*
- *R. Misra is with the Department of Computer Science and Engineering, Indian Institute of Technology Patna, Patna Bihar 801106, India. E-mail: rajivm@iitp.ac.in.*

Manuscript received 19 Jan. 2019; revised 11 Oct. 2019; accepted 13 Oct. 2019. Date of publication 23 Oct. 2019; date of current version 7 Jan. 2021. (Corresponding author: Ajay Pratap.)
Digital Object Identifier no. 10.1109/TMC.2019.2948877

allocation methods that simultaneously maximize the reusability of resources, minimize interference among cells, and maintain the user-level and cell level fairness. To achieve these objectives, we first convert the heterogeneous 5G network topology into a *Link Conflict Graph* (LCG), $G(V, E)$, and then present our proposed heuristics.

1.1 Our Contributions

The contributions of this paper are summarized below:

1) The resource allocation problem in 5G networks that maximizes fairness and minimizes interference for maximum resource reuse is shown as NP-hard.
2) The formulated problem is solved by proposing efficient heuristic algorithms based on centralized, distributed, and randomized distributed models.
3) The worst-case time complexity of the centralized algorithm is derived as $O(\mathbb{N}^2 + N\mathbb{N}\Delta)$, where $\mathbb{N} = |V|$ is the number of nodes, N is the total number of *Physical Resource Blocks* (PRBs) and Δ is the maximum degree of the link conflict graph, LCG.
4) The worst-case time and message complexity of the proposed distributed algorithm is shown as $O(\Delta)$.
5) With the help of real data and simulation experiments, the proposed algorithms are validated to demonstrate that they outperform state-of-the-art methods in the literature.

The rest of the paper is organized as follows. Section 2 summarizes the limitations of existing works. Section 3 introduces the system model and problem formulation. Section 4 proposes a centralized algorithm to solve the resource allocation problem, while Sections 5 and 6 describe a distributed algorithm and a randomized distributed heuristic, respectively. Section 7 presents simulation results. Section 8 offers conclusions and future research directions.

2 Related Works

In this section, we review the existing literature on resource allocation for centralized and distributed schemes with fairness, interference, and power management criteria.

2.1 Centralized Method

We summarize the works dealing with no fairness followed by those that consider user-level fairness, cell-level fairness, and power management in the network.

2.1.1 No Fairness

In [7], a centralized algorithm is proposed for interference avoidance in *Orthogonal Frequency Division Multiple Access* (OFDMA) femtocell network by assuming a central entity, called *Femto Management System* (FMS), to monitor the allocation procedure. However, this model can not handle non-uniform heterogeneous SAPs serving different priority users. In [8], a centrally controlled resource partitioning method was developed, based on graph coloring, to balance between interference protection and spatial frequency reuse in randomly deployed femtocells. However, this approach handles only a single user association per femtocell. In [9], the spectrum reuse ratio in 5G networks is shaped as a matrix graph-based centralized multi-coloring model to minimize the interference and maximize the reuse ratio in a network, without discussing the distribution of resources at cell level for different priority users. In [10], the authors proposed a fractal geometry-based resource allocation technique for 5G networks. This approach matches the fractal geometry pattern of small cells based on their respective sizes. However, the consideration of small cells as a circumscribed circle of a triangle is not well motivated for real small cell networks. In [11], the authors discussed resource sharing and hand-off mitigation in *macro* and *micro* cell networks. But this work is not applicable to a priority-based model where some users have higher priority than others due to limited availability of resources. Moreover, this work does not mention the distribution of users versus available resources at respective micro and macro cells.

2.1.2 User-Level Fairness

In [12], a dynamic graph partitioning approach is proposed to optimize resource allocation problem. However, this approach cannot handle the distribution of different priority users in the system. In [13], a graph-based resource management algorithm is proposed, which first generates a chordal graph from the LCG, and then performs priority differentiation-based admission control sub-algorithm, followed by the final resource assignment with maximum effect rank allocation sub-algorithm. This approach considered only two priority levels. In [14], the authors formulated the resource allocation as a max-min optimization problem. Further, an interfering (INT) model is proposed based on the partitioned greedy graph coloring method. In this approach, in each iteration, a maximal independent set (MIS) popped out from the LCG gets assigned with a color (tile), and these steps continue until each node gets a color. The remaining colors are re-allocated to each node based on the max-min fairness criteria. However, the computational complexity of the INT algorithm is a significant shortcoming for real-world applications in ultra-dense small cell networks.

2.1.3 Cell-Level Fairness

The work in [15] achieves inter-cell fairness using the cooperative algorithm, which is an enhanced, modified iterative water-filling (EMIWF) approach. However, this approach is pertinent only for femtocell networks. In [16], the authors studied a sub-channel allocation problem in OFDMA-based networks. To solve the allocation problem, a graph coloring based model is proposed. But this method does not consider the priority levels of different users. In [17], the authors modeled the resource allocation problem with cell level fairness as an optimization problem and proposed a centralized sub-optimal algorithm to maximize resource utilization and equity in the long term evolution (LTE) network.

2.1.4 Interference Mitigation and Power Management

In [18] the authors have proposed an energy-efficient resource allocation framework for *macro* and *micro* cells by adjusting their power levels. This approach does not address the re-usability of limited available resources for different priority users. In [19], stochastic geometry theory is used to

determine the optimal macro and micro BS density for energy-efficient network with QoS constraints. The lower and upper bounds of interference are also estimated in the network, and gamma distribution is applied to the micro and macro cell network model. However this work does not talk about any fairness constraint. In [20], the power consumption is estimated based on the coverage area of micro and macro cells, without the distribution of radio resources at different cells. Furthermore, this work does not apply to different priority users. In [21], the authors proposed an interference management mechanism by controlling the transmit power of femtocell users at MBS in the cross-tier network model. This approach also does not consider fairness and different priority users in the system.

2.2 Distributed Method

Similar to the centralized approach, the distributed methods are also categorized according to fairness, interference, and power management criteria.

2.2.1 No Fairness

In [22], authors have proposed a decentralized Frequency-ALOHA spectrum allocation approach in two-tier networks. Each femtocell accesses a random subset of reserved resources to avoid interference. However, due to pseudo-random nature, this approach cannot guarantee QoS in a realistic scenario. In [23], an interference coordination scheme was proposed for interference mitigation in the network based on the information gained in regular interval of time from a central coordinator. However, this scheme does not handle different priority levels of users. In [24], a distributed algorithm was presented for interference management in LTE-A environments, namely, Autonomous Component Carrier Selection. A femtocell selects a secondary set of resources from neighboring femtocells if its QoS demand is not fulfilled, without deteriorating the QoS of neighboring one's. However, this scheme mainly relies on measurement reports due to high correlation with environment sensing. In [25], the authors formulated the resource allocation as a Min-Max optimization problem and further proposed a clustering-based approach to develop a hybrid centralized-decentralized method for resource allocation. This scheme does not consider different priority levels of users.

2.2.2 User-Level Fairness

In [26], a graph-based multi-cell scheduling framework is proposed based on dynamic clustering combined with channel-aware resource allocation for providing QoS based services. In [27], a graph-based distributed scheme is proposed to manage the reuse radio of resources among femtocells, achieving a good trade-off between the system throughput and user's fairness. However, [26], [27] considered a uniform distribution of resources at each small cell, which is not a prudent model concerning the non-uniform load of traffic at different cells. In [28], a clustering-based mechanism is introduced to solve the resource allocation problem under two priority levels of users. This scheme did not acknowledge the dynamic model, where a user may come or leave the network. In [29], we proposed a distributed resource allocation provisioning in the networks, in which our main aim was to maximize the reuse ratio. This method is not applicable to different priority subscribers. In [30], we proposed a random graph-based model for resource allocation in femtocellular networks. However, it is not applicable to heterogeneous non-uniform small-cell structured model enabled with different priory subscribers.

2.2.3 Cell-Level Fairness

In [31], a distributed random access (DRA) scheme is proposed for medium-sized networks. Each femtocell locally executes DRA to reserve a set of resources using a random hashing function. Each femtocell also divides the total available resources into blocks in proportion to the interfering neighbors. The resource allocation is done at the cell level regardless of different user priorities. In [32] is proposed an adaptive frequency reuse method based on a semi-static inter-cell interference coordination scheme that sorts out the inter-cell and intra-cell interference in the network considering cell-level fairness. However, this scheme does not cover multi-tier priority levels of users in the system.

2.2.4 Interference Mitigation and Power Management

In [33], the user's uplink transmission power and rate are determined in a semi-joint manner, a distributed algorithm is proposed to maximize the utility of user. This method always gives higher priority to macrocell users compared to femtocell users, thus restricting the variability in hierarchical priority levels. In [34], the authors formulated a joint utility-based customized price and power control in multi-service networks as a two-variable optimization problem and designed a non-cooperative distributed game. This method does not consider heterogeneous radio-enabled SAPs with multi-tier user-level priority.

2.3 Shortcomings of Existing Methods

The above methods are not prudent for 5G due to their convoluted computation and rigid structure of femtocells. Interference minimization, fairness, and resource allocation for maximum spectrum reuse become even more challenging due to the nonuniform capacity, cell radius, and users' association with small cells. To achieve these objectives, we design the concept of fairness at the user and cell levels, minimizing interference while maximizing spectrum reuse in the network. Unlike existing approaches, we deal with simultaneously the heterogeneous network, limited available resources at an individual small cell, and fairness issues at user and cell levels. We exploit the heterogeneity of users who get services from different cells by distinguishing them based on their subscription plans of various services. To the best of our knowledge, there is no prior work that addressed the re-usability of PRBs, fairness at the user and cell levels, interference management in the heterogeneous small cells, such as micro, pico, and femto at the same time. Thus, considering the above constraints, objectives, and hardness of the problem, the main focus of this work is to design a deterministic offline scheme to solve resource allocation problem with lower computational complexity.

We propose centralized, distributed, and randomized distributed resource allocation schemes to achieve

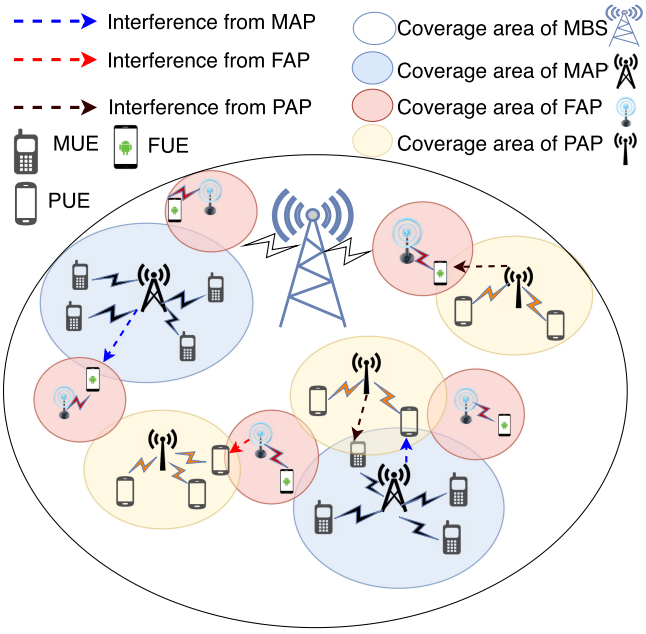

Fig. 1. Heterogeneous 5G networks.

maximum efficiency of resources in heterogeneous 5G networks. For the proposed centralized algorithm, we need to congregate information of the whole system at a particular central entity. Most of the existing works considered the availability of FMS, but that can only handle the case of femtocells; it is not suitable if the picocell and microcell also exist together. Thus, there is a need for a central entity to monitor all heterogeneous small cells; for this purpose, Cloud Random Access network (Cloud-RAN) is a well-accepted candidate [35]. A central entity is a Cloud computing center with a pool of available digital processing units, such as baseband units (BBUs), which can be used to serve the computation tasks that arrive at the cloud center. The X2 interface can enable the neighbor's information in order to execute the proposed distributed algorithms [36].

3 System Model

We consider an OFDMA multi-tier SAP-based cellular 5G network model (see Fig. 1). A physical resource block (PRB) is the smallest unit of resource that can be assigned to a user. It corresponds to 0.5 ms time and 180 kHz frequency band. A resource frame has 20 time slots derived from a frame length of 10 ms [37]. We assume that macrocell users are assigned with orthogonal resources, whereas SUEs share the remaining PRBs [7], [22], [31]. Let the number of MAP, PAP, and FAP be M, P, and F, respectively, and SUEs associated with each SAP be U_M, U_P, and U_F, respectively. Thus, the total number of SUEs in the system can be written as $U = U_M + U_P + U_F$. To fulfill the traffic demands, we have considered that SAPs can reuse the total available N PRBs given that they adhere to the interference constraints. We further assume that a user $u \in U$ demands one PRB, i.e., $d_u = 1$ at a time. Users associated with MAP, PAP, and FAP are denoted as MUE, PUE, and FUE, respectively. Fig. 1 explains our system model where non-uniform SAPs are deployed in association with a different number of SUEs and interference. We assume that MAP, PAP, and FAP work at fixed power levels P_M, P_P, and P_F, respectively.

The users are categorized based on the QoS criteria, such as their service plans subscribed with the operators. Inspired by the works [5], [6], we classify different SUEs based on their subscription plans with network operator. The QoS model of SUEs is broken down into four levels, such as premium, gold, silver, and bronze. These weights are assigned by the network operator based on the subscription plan of SUEs. Our model considers the weight of the premium, gold, silver, and bronze subscribers as 4, 3, 2, and 1, respectively. These QoS levels ensure that an SUE with more expensive (or higher) subscription plan gets a higher priority. For example, a premium subscriber is given the topmost priority or weight compared to the subscribers in lower plans. In our model, *fairness* means how fairly we can allocate resource to the higher priority SUEs over lower priority ones. As each SAP has limited available PRBs to serve the users, there is a need to contemplate cell level fairness while allocating resources to the users. The cell-level fairness is addressed by the capacity of cells, as illustrated in the following section.

3.1 Problem Formulation

Our objective is to maximize the resource reuse while considering user-level fairness, interference, and demand of each SAP. However, the joint consideration of user-level fairness and cell's PRB demands has a trade off with cell level fairness for the subscribers and availability of resources. The matrix of PRB allocation is given by:

$$T_{(U_M+U_P+U_F)\times N} = \begin{pmatrix} A^M_{U_M\times N} \\ B^P_{U_P\times N} \\ C^F_{U_F\times N} \end{pmatrix}, \tag{1}$$

where, $A^M_{U_M\times N} = [\alpha^m_{i,n}]$, $B^P_{U_P\times N} = [\beta^p_{j,n}]$ and $C^F_{U_F\times N} = [\gamma^f_{k,n}]$. The respective values are as follows:

$$\alpha^m_{i,n} = \begin{cases} 1, & PRB\ n\ is\ allocated\ to\ user\ i \\ & \quad associated\ with\ Microcell\ m \\ 0, & Otherwise, \end{cases} \tag{2}$$

$$\beta^p_{j,n} = \begin{cases} 1, & PRB\ n\ is\ allocated\ to\ user\ j \\ & \quad associated\ with\ Picocell\ p \\ 0, & Otherwise, \end{cases} \tag{3}$$

$$\gamma^f_{k,n} = \begin{cases} 1, & PRB\ n\ is\ allocated\ to\ user\ k \\ & \quad associated\ with\ Femtocell\ f \\ 0, & Otherwise. \end{cases} \tag{4}$$

To achieve optimal PRB assignment, we formulate the problem as PRB reuse ratio $\bar{f}$ as follows:

$$\mathbf{P1}: \max_{\alpha,\beta,\gamma} \bar{f} = \frac{1}{\sum_{n=1}^{N} f(n)} \left(\sum_{n=1}^{N} \left(\sum_{m=1}^{M} \sum_{i=1}^{U_M} w^m_{i,n} \alpha^m_{i,n} + \sum_{p=1}^{P} \sum_{j=1}^{U_P} w^p_{j,n} \beta^p_{j,n} + \sum_{f=1}^{F} \sum_{k=1}^{U_F} w^f_{k,n} \gamma^f_{k,n} \right) \right), \tag{5}$$

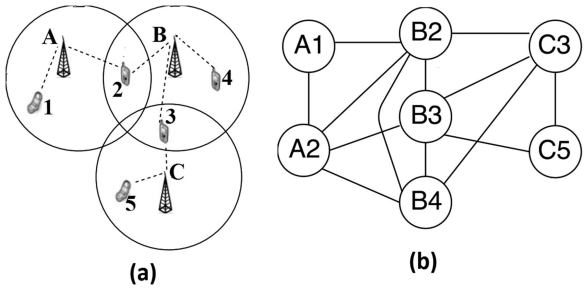

Fig. 2. (a) Heterogeneous topology and (b) LCG corresponding (a).

Subject to the constraints:

$$\alpha^m_{i,n}, \beta^p_{j,n}, \gamma^f_{k,n} \in \{0,1\}, \tag{6}$$

$$\sum_{m=1}^{M} \alpha^m_{o,n} + \sum_{p=1}^{P} \beta^p_{o,n} + \sum_{f=1}^{F} \gamma^f_{o,n} \leq 1, \forall z : o \in I_z, z \in U, \tag{7}$$

$$\sum_{n=1}^{N} \sum_{i=1}^{U_M} \alpha^m_{i,n} \leq D_m, \tag{8}$$

$$\sum_{n=1}^{N} \sum_{j=1}^{U_P} \beta^p_{j,n} \leq D_p, \tag{9}$$

$$\sum_{n=1}^{N} \sum_{k=1}^{U_F} \gamma^f_{k,n} \leq D_f, \tag{10}$$

$$f(n) = \begin{cases} 1, & \left(\sum_{m=1}^{M} \sum_{k=1}^{U_M} w^m_{k,n} \alpha^m_{k,n} + \right. \\ & \left. \sum_{p=1}^{P} \sum_{j=1}^{U_P} w^p_{j,n} \beta^p_{j,n} + \sum_{f=1}^{F} \sum_{i=1}^{U_F} w^f_{i,n} \gamma^f_{i,n} \right) \geq 1 \\ 0, & Otherwise. \end{cases} \tag{11}$$

The goal of the objective function in Eq. (5) is to maximize the reuse ratio of PRBs keeping in mind fairness, interference, and demand of resources at each SAP. Here $w^m_{i,n}$, $w^p_{j,n}$, and $w^f_{k,n}$ represent the weight of users i, j and k, respectively associated with the MAP m, PAP p and FAP k at a PRB n. However, the weight denotes the importance of an SUE given an individual's subscription plan. The network constraints are defined in Eqs. (6), (7), (8), (9), and (10), where Eq. (6) is a binary variable of Eqs. (2), (3), and (4). In Eq. (7), I_z is the list of SUEs interfering with SAPs. This constraint shows that no two interfering SUEs can use the same PRB. Constraints (8)-(10) confer that the maximum number of PRBs that can be allocated to the SUEs within a small cell is limited to the respective small cell's PRB demand. In other words, the total number of users that can get services from an SAP is bounded by the maximum capacity of the respective SAP. Eq. (11) defines a function of PRB allocation. If a PRB is allocated to any user in the heterogeneous 5G cellular network, then the function value is one, otherwise it is zero.

The graph κ-coloring problem can be mapped to the formulated problem, where the objective is to maximize reusability of colors with additional set of constraints for weight and availability of resources at each SAP. The next subsection will prove NP-hardness of the problem.

TABLE 1
Vertex κ-Coloring and Resource Allocation

Vertex coloring	Resource allocation
Set of κ colors	Set of N PRBs
Set of V vertices	Set of $\mathbb{N}$ communication links
Set of E edges	Interference constraint
Graph G	Heterogeneous topology

3.2 Link Conflict Graph Construction

We construct the LCG, denoted as $G(V, E)$ from the heterogeneous topology. Each vertex $v \in V$ represents the communication link between user $u \in U$ and its corresponding SAP, $a \in \{M, P, F\}$, i.e., $v = (u, a)$ and $\mathbb{N} = |V|$. The priority, P_v, of a vertex v in LCG is denoted as the surrogate of user's weight. For a user, there may be more than one link corresponding to various FAPs, PAPs, and MAPs. All such communication links that cannot be activated simultaneously are connected by an edge $e \in E$ in the LCG. Two links may not be activated simultaneously due to interference or *conflicting user* occupancy (see Eq. (7)). This scenario may occur due to inter-cell or intra-cell interference. Fig. 2a shows a heterogeneous topology and Fig. 2b represents the corresponding LCG. In Fig. 2a, users 2 and 3 are conflicting because both of them lie in the overlapping area of more than one small Cell. Although the LCG construction is outside the scope of this paper, it can be done using the technique in [14] which assume binary interference model such that if two links interfere with each other, then there is an edge between them; otherwise, there will be no edge. Some prior works on LCG construction used bandwidth testing-based model [38], [39], which is adopted in recent small cell networks [14], [40]. A bandwidth test framework can also be used to construct LCG.

Theorem 1. *The resource allocation problem **P1** is NP-hard.*

Proof. We map the graph κ-coloring problem to the formulated problem **P1**, where κ is the total number of colors. A special case, where $w^m_{i,n}$=$w^p_{j,n}$=$w^f_{k,n}$=1, implies all users are given the same priority. For simplicity, we remove the constraints in Eqs. (8)-(10). As the number of communication links is larger than the available PRBs (i.e., $\mathbb{N} > N$), more than one communication link is assigned the same PRB. However, since the assignment of the same PRB to any two neighboring vertices creates interference in the network, we map the vertex κ-coloring to our optimal PRB assignment problem as follows:

- Map the set of κ-colors to N PRBs.
- Map the set of V vertices to $\mathbb{N}$ communication links.
- Map the set of E edges to interference-free property.
- Map the LGG to the entire heterogeneous topology.

Table 1 shows the mapping of the resource allocation problem to the vertex κ-coloring which is NP-complete [41], implying the formulated problem **P1** is an NP-hard. □

Next we aim to solve the resource allocation problem in heterogeneous 5G networks in computationally feasible time and message complexity. Given different constraints

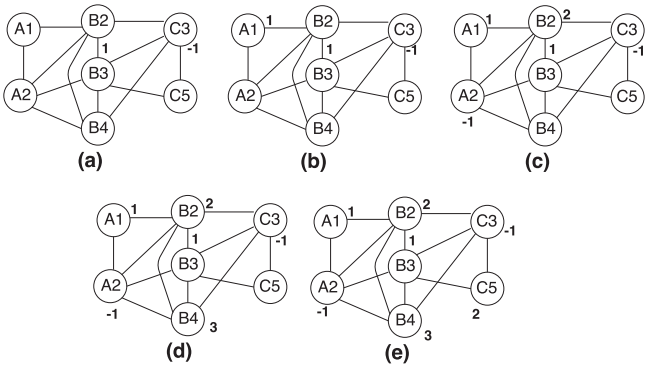

Fig. 3. Illustrative example of Centralized Algorithm. (a) Iteration 1. (b) Iteration 2. (c) Iteration 3. (d) Iteration 4. (e) Iteration 5.

such as interference, varying resource demands, limited available PRBs, and a large number of users within densely deployed nonuniform cells, we propose here deterministic offline algorithms to satisfy these constraints yet coming up with a computationally feasible and near-optimal solution.

4 A Centralized Algorithm

A vertex $v \in V$ corresponding to a link (u, a) is assigned a priority P_v based on user's weight as given in the Eq. (5). Let $R(a_i)$ be the maximum available PRBs at $a_i \in \{M, P, F\}$, which specifies the maximum orthogonal PRBs that can be assigned to the respective SUEs.

Definition 1 (One-hop view [42]). *Let i be a node and τ_i be the set of IDs of one-hop neighbors of i. We say the pair (i, τ_i) the one-hop view of node i.*

The input to the centralized algorithm is LCG $G(V, E)$, total number of PRBs N, available PRBs at each SAP $R(a_i)$ (= demand D_{a_i}), users associated with each SAP $U(a_i)$, $v.assign = 0$ for each node and priority P_v (= weight), randomly generated number Rb_v (called random-backoff). The output of the algorithm is PRB allocated nodes. Algorithm 1 is a two-phase approach. In Phase 1, the highest priority unallocated node is selected (line 2). If two nodes have the same priority, then node with smaller randomly generated number (Rb) is the winner. (Here we have considered an ideal model where nodes do not try to manipulate the random numbers to win a higher rank.) If the selected node contains the conflicting user u_c, it means other unassigned PRB nodes contain the same user u_c. In case of conflicting users, two or more nodes present the same user associated with different SAPs. Thus, it is necessary to assign PRB to one of them. Moreover, a node with the highest value of $R(a_i) - |U(a_i)|$ goes for Phase 2 while rest of the nodes containing the conflicting user u_c are assigned with -1 (line 3). In other words, the same user is assigned with valid PRB by other higher priority SAP (cell-level fairness). In Phase 2, if the total available PRBs at the respective SAPs is more than zero (line 5), then the smallest unassigned PRB within one-hop view is assigned to the node and the respective SAPs update the total available PRBs (lines 6-9). The process terminates when each node is assigned with some PRB, or the total available PRBs at each SAP gets exhausted.

Algorithm 1. Centralized Resource Allocation Algorithm

Input: LCG $G(V, E)$; Set of PRBs N; $R(a_i)$, $U(a_i)$, $\forall a_i \in \{M, P, F\}$; $v.assign = 0$, Rb_v and P_v $\forall v \in V$
Output: PRB allocated nodes
Repeat
1: **Phase 1: Deciding Priority**
2: Select the highest priority (P_v) node $v \in V$ such that $v.assign = 0$. If two nodes have the same highest priority, then a node with smaller random number Rb_v is considered as the highest priority node.
3: If the selected highest priority node contains conflicting user u_c, then among all nodes containing user u_c, the one with maximum value of ($R(a_i)$-$|U(a_i)|$) only retains and the rest are assign with -1 /* Here $U(a_i)$ is the unallocated users associated with SAP a_i*/
4: **Phase 2: PRB Assignment**
5: If $R(a_v) > 0$ then /* Available resources corresponding to the highest priority node v */
6: for $n = 1$ to N
7: If ($\tau_v.assign \neq n$) /* Where τ_v is one-hop view of v */
8: $v.assign = n$
9: $R(a_v) = R(a_v) - 1$
Until $\forall v \in V$ such that $v.assign \neq 0$ or $R(a_v) = 0$ /* PRBs can not be allocated any further. */

4.1 Illustrating the Centralized Algorithm

Fig. 3, shows an illustrative example by executing various steps of Algorithm 1. Recall the LCG in Fig. 2b. Let the non-increasing priority of SUEs be $u3 > u1 > u2 > u4 > u5$. (Two users may have at the same priority level based on their subscription plan with the operator.) Accordingly, we write the priority of nodes in LCG as $B3 = C3 > A1 > B2 = A2 > B4 > C5$. Nodes $B3$ and $C3$ hold the same priority because they represent the same user $u3$; likewise for nodes $B2$ and $A2$. Let the total available resources at SAPs A, B and C be $R(A) = 3$, $R(B) = 5$ and $R(C) = 2$, respectively. Based on Phase 1 of Algorithm 1, the central entity can select any of the highest priority nodes, say $B3$ or $C3$. Based on line 3 of Phase 1, node $B3$ moves to Phase 2 and $C3.assign = -1$ (cell-level fairness). Based on lines 5-9 in Phase 2, $B3.assign = 1$ and $R(B) = 4$. Thus, in iteration 1 as illustrated in Fig. 3a, $C3.assign = -1$, $B3.assign = 1$, $R(A) = 3$, $R(B) = 4$ and $R(C) = 2$. Note that no update took place in the available PRBs at SAPs A and C because these SAPs did not spend any of its PRBs. In the 2nd iteration (see Fig. 3b), node $A1$ is selected in Phase 1 and assigned with smallest available PRB within one-hop view in Phase 2, i.e., $A1.assign = 1$, $R(A) = 2$, $R(B) = 4$ and $R(C) = 2$. In the 3rd iteration as shown in Fig. 3c, nodes $B2$ and $A2$ are selected, $A2.assign = -1$ and $B2.assign = 2$, $R(A) = 2$, $R(B) = 3$ and $R(C) = 2$ based on the Phase 1 and Phase 2 of Algorithm 1 as described in iteration 1. In the 4th iteration (Fig. 3d), node $B4$ is selected and $B4.assign = 3$, $R(A) = 2$, $R(B) = 2$ and $R(C) = 2$. Similarly, in the 5th iteration (Fig. 3e) $C5.assign = 2$, $R(A) = 2$, $R(B) = 2$ and $R(C) = 1$. The final PRB assignment satisfies the set of constraints given in the formulated problem P1.

4.2 Analysis of Centralized Algorithm

In this subsection, we analyze the time complexity and correctness of the proposed Algorithm 1.

Theorem 2. *The time complexity of the proposed centralized resource allocation algorithm is* $O(\mathbb{N}^2 + N\mathbb{N}\Delta)$.

Proof. Lines 1-9 of the centralized Algorithm 1 is executed for each vertex in the LCG. The time complexity of lines 2 and 3 are $O(\mathbb{N})$ and $O(\Delta)$ respectively, where Δ is the degree of LCG. Line 5 takes constant time while lines 6-9 requires $O(N\Delta)$ time. Phase 1 and Phase 2 are executed for each node, thus the centralized Algorithm 1 requires $O(\mathbb{N})(\mathbb{N} + \Delta + N\Delta) = O(\mathbb{N}^2 + N\mathbb{N}\Delta)$ time. □

Theorem 3. *The proposed centralized resource allocation algorithm guarantees valid PRB assignment in the network.*

Proof. Execution of the centralized Algorithm 1 ensures a valid PRB assignment $\forall\, v \in V$ in $G(V, E)$. Before selecting a PRB, a node scrutinizes the total available PRBs at each SAP and respective user's priority. A node v is only allowed to select $n \in N$ if $R(a_v) > 0$, and its priority is higher or equal to any other unassigned node. Thus, this statement corroborates the total available PRBs at the respective SAP and user-level fairness. However, the conflicting users are served by the SAP that has higher remaining resources, thereby substantiating cell-level fairness. A node v always gets assigned with the following criteria: $v.assign = min\ n \notin \{\tau_v.assign\}$, i.e., re-use the smallest PRB that is not assigned to any other node within a one-hop view. This statement ensures interference-free re-usability of resources. Thus, Algorithm 1 demonstrates a valid PRB assignment in the network, and this assignment validates user-level fairness, cell-level fairness and interference-free re-usability of resources in a heterogeneous 5G network. □

In the following, we show that the proposed centralized algorithm satisfies demand of each SAP a_i with the objective given in Eq. (5). The algorithm only fails when the available PRBs are exhausted, but some SAPs with positive demand remains un-allocated. Based on this algorithm, we can define the re-usability of PRB n_i as follows:

$$L_i = \begin{cases} \frac{\Lambda}{N} + 1, & if\ i \leq \Lambda \% N \\ \frac{\Lambda}{N}, & if\ i > \Lambda \% N, \end{cases} \tag{12}$$

where $\Lambda = D_m + D_p + D_f$ is the total demand of PRBs in the 5G network. To show that the proposed algorithm satisfies the requirement of Λ, we prove the following lemma. Let D_a^i and Λ^i represent the demand of SAP a and total demand before the ith PRB has been assigned, respectively.

Lemma 1. *The PRB n_i is assigned to L_i SAPs with positive demand in the ith iteration. Moreover, for every SAP a, it holds $D_a^i \leq N - i + 1$ and $L_i \leq \Lambda^i (N - i + 1)^{-1}$.*

Proof. We prove this lemma by induction over i, where $i = 1$ is the base case. As per the constraint, the demand for any SAP should not exceed the total available PRBs. Thus, we can write $D_a^1 \leq N$. This proves the first condition of the statement, and also satisfies the base case. Summing up, the SAP results in $\Lambda = \Lambda^1 \leq (M + P + F)N$. Dividing the above outcome with total PRBs yields $L_1 = \Lambda/N = \Lambda^1/N \leq (M + P + F)$, thus proving the second condition of the lemma. Since $L_1 \leq (M + P + F)$, some L_1 SAP with positive demand will be assigned with n_1. This proves the base case. Following the inductive steps, next we assume the lemma is true for i and prove it for $i + 1$. We prove the following condition first.

$$L_{i+1} \leq \frac{\Lambda^{i+1}}{N - i}. \tag{13}$$

By the inductive hypothesis:

$$L_i (N - i + 1) \leq \Lambda^i. \tag{14}$$

We can write $L_i = \Lambda^i - \Lambda^{i+1}$. Substituting L_i in equation (14) and rearranging the inequalities, we obtain:

$$L_{i+1} \leq L_i \leq \frac{\Lambda^i}{N - i + 1} \leq \frac{\Lambda^{i+1}}{N - i}. \tag{15}$$

Hence, condition (13) proved.

Let us contradict that after $(i+1)$th steps, $D_a^{i+1} > N - i$ for some SAP a. Using inductive hypothesis, we conclude that for each SAP a, $D_a^i \leq N - i + 1$. As, $D_a^{i+1} \leq D_a^i \leq N - i + 1$, we should have $D_a^{i+1} = D_a^i = N - i + 1$ and SAP a is not assigned with PRB i. Thus, at least $L_i + 1$ SAPs demand $N - i + 1$ PRBs before n_i has assigned. This signifies $\Lambda^i \geqslant (L_i + 1)(N - i + 1)$. Thus,

$$\Lambda^i = \sum_{j=i+1}^{N} L_j \leqslant (N - i + 1) L_i. \tag{16}$$

The above Equation (16) is contradicted. Thus, $D_a^{i+1} \leq N - i$ for SAP a. A relation can be written as $L_{i+1} \leq (\Lambda^{i+1})/(N - i) \leq (\Lambda^{i+1})/(D_a^{i+1})$. The number of SAPs with positive demand is at least $\Lambda^{i+1}/D_{a*}^{i+1}$, where $a*$ is the SAP with the maximum demand of PRBs. This proves the lemma. □

The above centralized algorithm depends on a central entity, which makes the system less adaptive for any local decision in the network. If any change ensues, the whole system needs to be straightened out. To fend off such convoluted scenarios, we propose a distributed approach where the resource allocation is done locally, without sole dependency over any central entity.

5 Distributed Algorithm

The distributed algorithm differs from the centralized one in the following sense. In the distributed approach, the resource allocation problem is solved locally because multiple nodes can decide on their resources at the same time. As the interference constraint depends on the one-hop view, each node makes its decision over a given PRB depending on the nodes within one-hop view.

The input to Algorithm 2 is a one-hop view of each node, total available PRBs N, total available resources to each node $R(a_i)$, the number of users associated to each SAP $U(a_i)$, priority P_v and $v.assign = 0$ for each node. The output of the distributed algorithm is PRB allocated nodes. The algorithm is described in four phases as follows.

Phase 1: Each unassigned node v for which total available PRBs at respective SAP is not zero (i.e., $R(a_v) > 0$), sends PRB_Request $< v, R(a_v), U(a_v), TS, P_v, v.assign, Rb_v >$ to all

nodes within one-hop view τ_v and starts a timer t_1. Here, TS is the time stamp of PRB_Request message and Rb_v is the random number generated by node. A node v retries after t_1 time if it did not receive PRB_Reply message from all nodes within one-hop view and $R(a_v) > 0$.

Algorithm 2. Distributed Resource Allocation Algorithm

Input: One-hop view of nodes, Set of PRBs N, $R(a_i)$, $U(a_i)$, $\forall a_i \in \{M, P, F\}$, $v.assign = 0$ and P_v; $\forall v \in V$,
Output: PRB allocated nodes
Repeat
1: **Phase 1: Sending PRB Request**
2: For $\forall v \in V$ such that $v.assign = 0$; generate a random number Rb_v
3: If $R(a_v) > 0$
4: Send PRB_Request $< v, R(a_v), U(a_v), P_v, TS, v.assign, Rb_v >$ message to all one-hop view nodes τ_v and start timer t_1
Until $v.assign \neq 0$ or $R(a_v) = 0$
5: **Phase 2: Deciding Priority**
6: Node x receives message PRB_Request $< v, R(a_v), U(a_v), P_v, TS, v.assign, Rb_v >$ from node v
7: If $(P_v > P_x) \parallel x.assign \neq 0$
8: Send PRB_Reply $< x, R(a_x), U(a_x), P_v, TS, x.assign >$ message to node v
9: Else if x contains the same user as v /*conflicting user*/
10: If $(R(a_x) - |U(a_x)|) \leq (R(a_v) - |U(a_v)|)$
11: $x.assign = -1$
12: Send PRB_Reply $< x, R(a_x), U(a_x), P_x, TS, x.assign >$ message to node v
13: Else if $Rb_x > Rb_v$, node x sends PRB_Reply $< x, R(a_x), U(a_x), P_x, TS, x.assign >$ message to node v /* if priorities of nodes v and x are the same and both contain different users */
14: **Phase 3: PRB Assignment**
15: Upon receiving PRB_Reply message $< x, R(a_x), U(a_x), P_x, TS, x.assign >$
16: If node v receives PRB_Reply message from all one-hop view nodes τ_v
17: $v.assign = min\ n \notin \{\tau_v.assign\}$
18: $R(a_v) = R(a_v) - 1$
19: Send PRB_Update $< v, R(a_v), U(a_v), P_v, TS, v.assign >$
20: **Phase 4: Update**
21: Upon receiving PRB_Update message $< v, R(a_v), U(a_v), P_v, TS, v.assign >$
22: Update locally maintained PRB assignment of nodes in the one-hop view of LCG and update the timer $t1$ based on time stamp of message.

Phase 2: This phase describes the steps when a node x receives PRB_Request from a v. If $x.assign \neq 0$ and x's priority is less than that of node v, then node x sends PRB_Reply $< x, R(a_x), U(a_x), P_v, TS, x.assign >$ message to v following lines 6-8. If x contains the same user as node v (i.e., conflicting user), then a comparison takes place between their respective SAPs. As node x and node v both contain the same user associated with different SAPs, only one SAP is allowed to assign PRB to this user. Specifically, the SAP having higher value of $R(a) - |U(a)|$ assigns a PRB to the user ensuring cell-level fairness. If $(R(a_x) - |U(a_x)|) \leq (R(a_v) - |U(a_v)|)$, then $x.assign = -1$. In other words, no PRB is assigned to this node, which sends PRB_Reply $< x, R(a_x), U(a_x), P_x, TS, x.assign >$ message to node v as shown in lines 9-12. If nodes v and x have the same priority but contain different users, then a comparison takes place between the corresponding random numbers. If $Rb_x > Rb_v$, then x sends PRB_Reply to v (line 13).

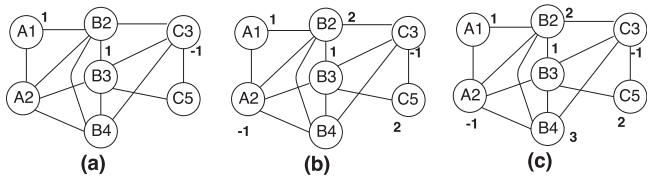

Fig. 4. Illustrative example of distributed algorithm. (a) End of round-1. (b) End of round-2. (c) End of round-3.

Phase 3: This phase describes the steps when v receives PRB_Reply from all of its one-hop view nodes. Node v selects the smallest unassigned PRB in τ_v, and the respective SAPs update the available PRBs. Node v also sends PRB_Update $< v, R(a_v), U(a_v), P_v, TS, v.assign >$ message. If v did not receive PRB_Reply within t_1 time from all nodes within one-hop view, then it re-executes Phase 1.

Phase 4: This phase illustrates the actions of nodes upon receiving the PRB_Update message. When a node gets PRB_Update message from its one-hop view nodes, it refurbishes the total available PRB set as well as the waiting timer t_1 based on TS of the message. The total number of unassigned users are updated at the respective SAPs.

The process terminates when either each node is assigned with some PRB or the total number of available PRBs at each SAP gets exhausted.

5.1 Illustrating the Distributed Algorithm

Fig. 4 illustrates the execute of steps of the distributed Algorithm 2, by considering the same LCG in Fig. 2b. Priorities and resources available to each SAP are considered the same as in Fig. 3. Each node sends PRB_Request to nodes within one-hop view and starts a timer t_1 as shown in Phase 1 of Algorithm 2. As node $B3$ has the highest priority within its one-hop view, it receives PRB_Reply from all nodes within one-hop view, and for node $C3$ having the same user as node $B3$, we have $C3.assign = -1$ as shown in Phase 2. In the same round, node $A1$ receives PRB_Reply from all others in one-hop view because $A1$ is the highest priority node among all within its one-hop view (i.e., $A1 > B2 = C2$). Thus, booth $A1$ and $B3$ move to Phase 3 and select the smallest unassigned PRB, update resources to the respective SAPs and send PRB_Update message to the nodes within one-hop view. In other words, $B3.assign = 1$, $A1.assign = 1$ and $R(A) = 2$, $R(B) = 4$ as per Phase 3 (Fig. 4a). Upon receiving PRB_Update message, each one-hop view node updates its timer $t1$ and gets to know the total available resources at the respective SAP. In the second round, as shown in Fig. 4b, all unassigned nodes repeat the same steps as described above in the first round. At the end of this round, $B3.assign = 1$, $A1.assign = 1$, $B2.assign = 2$, $C5.assign = 2$, $C3.assign = -1$, $A2.assign = -1$, $R(A) = 2$, $R(B) = 3$, $R(C) = 1$. Likewise, in the third round, node $B4$ sends PRB_Request and receives PRB_Reply from all of its one hop-view nodes and selects the minimum unassigned PRB within one-hop view, i.e., $B4.assign = 3$. Fig. 4c

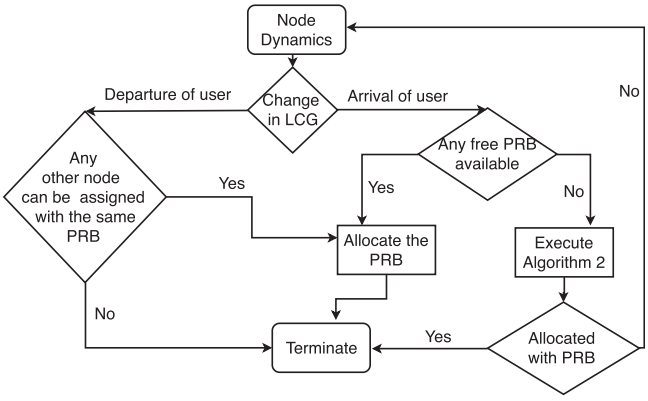

Fig. 5. Dynamic resource allocation algorithm.

shows the final assigned PRBs to each node and this allocation meets the set of constraints in Eqs. (5), (6), (7), (8), (9), and (10).

5.2 Dynamic Resource Allocation

Let us propose a localized implementation of the PRB assignment in a dynamic network model, in which an SUE may arrive/depart or switch-on/off. This implies the LCG needs to be dynamically adapted. Fig. 5 shows a dynamic PRB allocation scheme. To avoid the overhead of LCG reconstruction, we update it locally in the coverage area. For example, if a new user is coming/turning-on or moving away/turning-off in the coverage area of any SAP a_q, then we define one-hop view of the local LCG, $G_l^q \subseteq G$. Accordingly the LCG is updated locally without affecting the whole network topology.

When a user arrives in the coverage area of a_q, a new partition shows up on the subgraph G_l^q. The node corresponding to the newly arrived user sends PRB_Request to those within one-hop view (see Phase 1 of Algorithm 2). Upon receiving PRB_Request, the one-hop view nodes send PRB_Reply message containing the assigned PRBs to the respective nodes (Phase 2). Next, upon receiving this Reply message from the one-hop view nodes, the newly appeared node selects the smallest available PRB and sends the PRB_Update message (Phases 3-4 of Algorithm 2). However, if no PRB is available within a one-hop view, then Algorithm 2 will be executed to reassign PRBs to all the nodes in the LCG. Finally, if a node does not get a valid PRB, it waits for PRB availability at the occorrrence of any change in the network topology.

If a user departs from an SAP a_q, an update occurs in LCG locally, such as the node representing this user and its incident edges are removed from the LCG. If any node is waiting for PRB within a one-hop view, then it can be assigned with the same PRB as the departing user given none of the other nodes within one-hop view is assigned the same PRB; otherwise, the process terminates. This procedure continues locally with the arrival and departure of any user in the network.

5.3 Analysis of Distributed Algorithm

Each node i in the network executes Algorithm 2 for rounds of time period T_i. A *round* is defined as the execution of steps 1-22 of the algorithm. The value of T_i is set to be $3\mathbb{D}_i$, where $\mathbb{D}_i$ is the one-way message delay for node i. Nodes are not required to be synchronized in each round. The value of T_i should be sufficiently large to send PRB_Request, receive PRB_Reply from its one-hop view nodes, select minimum available PRB, and send PRB_Update message. Let $\mathbb{D}_{max}$ and $\mathbb{D}_{min}$ be the maximum and minimum one-hop message delays. Let $\delta = \lceil T_{max}/T_{min} \rceil$ and $T_{max} = 3\mathbb{D}_{max}$. Ignoring the time for executing the internal steps of checking priorities, selecting the smallest unassigned PRB and generating a random number, each node sends PRB_Request message at least once and at most δ times during T_{max} time period.

Theorem 4. *Both the time complexity and message complexity of the proposed distributed resource allocation algorithm are $O(\Delta)$.*

Proof. *(Time Complexity)* The maximum number of rounds that a node can wait before deciding its PRB depends on the maximum number of nodes within a one-hop view. Since a node can take up to T_{max} time to decide on its PRB, the maximum number of rounds is O(Δ), implying the maximum time to execute the algorithm is $O(T_{max}\Delta)$. As T_{max} is a constant, the time complexity of the distributed algorithm is $O(\Delta)$.

(Message Complexity). In one round, a contender node can try its PRB_Request message acknowledged by all nodes within one-hop view for δ times. In each try, the node sends $O(1)$ messages; hence $O(\delta)$ messages in one round. For a total of $O(\Delta)$ rounds on an average, each node thus sends $O(\delta\Delta)$ messages. Assuming the message delay δ is bounded by a constant, the message complexity is $O(\Delta)$. □

Theorem 5. *The distributed resource allocation algorithm results in a valid PRB assignment.*

Proof. To prove this theorem, it is enough to convince that no two nodes within a one-hop view select the same PRB, i.e., the assignment follows fairness and re-usability of resources in the network. Interference and re-usability can be ensured as follows: 1) a node must receive PRB_Reply from all nodes within a one-hop view, and 2) any two nodes within a one-hop view must share a common edge. The PRB_Reply contains the information of assigned PRBs to the respective nodes. This property ensures that when a node decides on a PRB, it always picks the minimum indexed PRB not seized by any one-hop view node, and no other node within one-hop can pick the same PRB. This validates the interference-free re-usability of resources. However, as the highest priority node gets PRB_Reply from all others within one-hop view, the user-level fairness is ensured. As conflicting users receive a valid PRB from the SAP with higher remaining resources, the cell-level fairness is also ensured. □

Let there be a complete graph topology where each node is adjacent to all other nodes. In this scenario, the distributed algorithm takes almost $\mathbb{N}$ rounds to terminate. This is because each node is adjacent to all other nodes with the same priority. Hence, in each round, only one node gets the privilege to select a valid PRB. Considering this scenario,

we propose below a distributed randomized heuristic to solve the worst-case condition of the PRB assignment.

6 Randomized Distributed Resource Allocation (RDRA)

In this approach, each node estimates a rank $\wp$ of other nodes within one-hop view including itself, based on priority levels. In case two nodes (say, x, v) contain the same user, the ranking is done based on the total available PRBs and the number of users connected with the respective SAP. That is, if $(R(a_x) - |U(a_x)|) \leq (R(a_v) - |U(a_v)|)$, then $\wp_x = 0$ and $\wp_v > 0$. If two nodes contain different users but show the same priority level, then the node with smaller random number (i.e., Rb) gets higher rank. A node Q of degree $Z-1$ consists of Z nodes within its one-hop view. The highest priority node grabs rank 1, the next highest priority node gets rank 2, and so on. Let rank of Q be $\wp$ and the smallest unassigned PRB be n such that $1 \leq n \leq N$. Node Q selects -1, if $\wp_Q = 0$; else if $\wp_Q = 1$, node Q selects PRB n; otherwise Q selects PRB $n + \wp_Q - 1$ with probability $p > 0$. If the selected PRB endures interference (constraint in Eq. 7) from any one-hop view node, then Q re-selects a new PRB $n' + c$ with probability p_c which is computed according to a truncated geometric distribution such that $\sum_{c=1}^{N} p_c = 1$. Here, n' is the highest PRB index assigned to any node in the one-hop view of Q and $c \in N$. Let the node Q and its one-hop view nodes consist of Z nodes. Then $p_c = p\pounds^{|c-\wp|}$, where $\pounds < 1$, if $|c - \wp| \leq Z - 1$; otherwise, $p_c = 0$. Thus:

$$\sum_{c=1}^{N} p_c = \sum_{c=1}^{N} p\pounds^{|c-\wp|}. \tag{17}$$

The above equation can be written as follows:

$$\begin{aligned}\sum_{c=1}^{N} p_c &= \sum_{c=1}^{\wp-1} p\pounds^{c-\wp} + p\pounds^{0} + \sum_{c=\wp+1}^{N} p\pounds^{\wp-c},\\ &= p(\pounds^{\wp-1} + \pounds^{\wp-2} + \cdots + \pounds^{1} + \pounds^{0} + \pounds^{1} + \pounds^{2} + \cdots + \pounds^{N-\wp}),\\ &= p\left(\pounds\frac{1-\pounds^{\wp-2+1}}{1-\pounds} + \frac{1-\pounds^{N-\wp+1}}{1-\pounds}\right)\\ &= p\left(\frac{\pounds - \pounds^{\wp} + 1 - \pounds^{N-\wp+1}}{1-\pounds}\right) = 1.\end{aligned} \tag{18}$$

As p is the probability, its value must lie between zero and one. To validate, note that the value of £ lies between 0 and 1. Thus, from Eq. (18), we can say that $p > 0$. We can also write $\pounds^{\wp-1} + \pounds^{N-\wp} < 2$ based on the £ value. Solving this inequality, we get $0 < p < 1$ which satisfies the probability constraint.

6.1 Illustrating the RDRA Algorithm

Fig. 6a shows the assigned rank to each node based on the user priorities and available PRBs at each SAP. The priorities and the number of available PRBs at each SAP are assumed the same as in the example Fig. 3. Each node keeps track of the highest-ranked node within one-hop view. Lower priority nodes that contain conflicting users, are ranked as zero (i.e., $\wp_{C3} = \wp_{A2} = 0$). The node with the highest priority in one-hop view is ranked as 1 (i.e., $\wp_{B3} = \wp_{A1} = 1$), the next highest priorty node as rank 2, and so on, as illustrated in Fig. 6a. Now nodes with rank zero pick out -1; nodes with rank 1 pick out PRB $n = 1$, and nodes with $\wp > 1$ select PRB $n + \wp - 1$ with probability p as shown in Fig 6b. For instance, node $B2$ selects PRB $n + \wp - 1 = 1 + 2 - 1 = 2$ with probability p. The PRB assigned to each node satisfies the constraints in the problem definition. If any node does not satisfy a constraint after the PRB assignment, then that node will re-select PRB $n' + c$ with probability p_c as described in the RDRA heuristic.

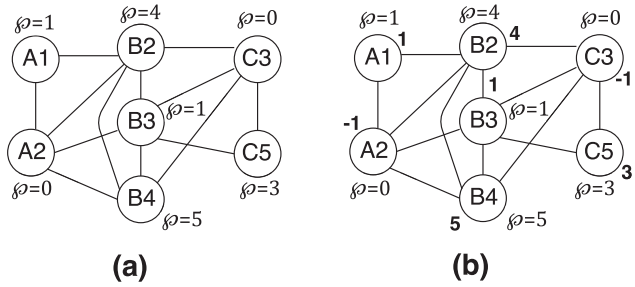

Fig. 6. Illustration of RDRA. (a) Rank assignment. (b) PRB assignment.

6.2 Analysis of RDRA Algorithm

If the maximum and minimum message delays (i.e., $\mathbb{D}_{max}$ and $\mathbb{D}_{min}$ as discussed in Section 5.3) are bounded by some constant, then we can prove the following.

Theorem 6. *The expected time complexity of RDRA is $O(\Delta)$.*

Proof. Let $\mathbb{C}(i,k)$ be a set of conflicting nodes of a node i within its one-hop view, who are yet to be assigned with a valid PRB in the kth round. Let $S(i,k)$ stand for the event that node i selects a valid PRB in the kth round. Thus, the selected PRB of node i is valid if no other node within one-hop view is assigned the same PRB. Since the maximum number of contender nodes could be Δ in the kth round, the probability $P_b(i\ get,\ k)$ that a node i achieves a valid PRB in the kth round is bounded as follows:

$$P_b(i\ get,\ k) \geq P_b(S(i\ get,\ 1)). \tag{19}$$

The selection of PRB at node i depends on the number of contending nodes. Thus,

$$\begin{aligned}P_b(i\ get,\ k) &\geq P_b(S(i,1)) \prod_{j\in\mathbb{C}(i)} (1 - P_b(S(j,k))),\\ &\geq \frac{1}{(\Delta+1)} \prod_{j\in\mathbb{C}(i)} \left(1 - \frac{1}{|\mathbb{C}(i)|+1}\right),\\ &= \frac{1}{(\Delta+1)} \left(1 - \frac{1}{|\mathbb{C}(i)|+1}\right)^{\mathbb{C}(i)},\\ &> \frac{1}{(\Delta+1)} \left(\frac{1}{\sqrt{e}}\right).\end{aligned} \tag{20}$$

The above equation is valid because: (a) Δ is the maximum contender set size of any node in the network, and (b) the competitor node j selects a PRB which is inverse of the maximum neighborhood size of its competitor set involving node i, and (c) the inequality $(1 - \frac{1}{|\mathbb{C}(i)|+1})^{\mathbb{C}(i)} > \frac{1}{\sqrt{e}}$ holds.

The above bound represents a lower bound probability for a node to get a valid PRB, and it does not depend on i and k. Let X be the number of rounds before a node

finds a valid PRB. Then this relation can be written as a geometric distribution:

$$P_b(X=k) = P_{lower}(1-P_{lower})^{k-1}, \tag{21}$$

where $P_{lower} = \frac{1}{(\Delta+1)}(\frac{1}{\sqrt{e}})$. Thus, the upper bound of the expected number of rounds required by a node to select a valid PRBs is given by:

$$E(M) = (\Delta+1)\sqrt{e}. \tag{22}$$

Hence, the expected time complexity of RDRA is $O(\Delta)$. □

To maximize the re-usability of PRBs, observe that nodes of individual PRB form an independent set. Thus finding a minimum PRB assignment can equivalently be formulated as finding the minimum number of independent sets such that each node is enclosed in one set. Following this observation, we map the PRB allocation as an antisymmetric relation [43] on the ID set [$\mathbb{N}$]. Let the probability based heuristic assigns a PRB from N to one-hop view (v,τ). For each PRB $n \in N$, there is a relation $\rhd_n$ s.t., $\forall\ v,x \in [\mathbb{N}]$ $v \not\rhd_n x \vee x \not\rhd_n v$. A PRB n can be assigned to one-hop view (v,τ) if and only if $\forall x \in \tau : v \rhd_n x$. Let $\psi_n(v) := \{x \in [\mathbb{N}] : v \not\rhd_n x\}$ be the set of nodes that must be nonadjacent to a node v and assigned with PRB n. We select v_r uniformly at random from the set [$\mathbb{N}$]. The remaining $\mathbb{N}-1$ are independently included to a set τ_r with probability p. For a PRB n, let η_n be the event that $\tau_r \cap \psi_n(v_r) \neq \phi$. In other words, ψ_n is an event that PRB n can not be assigned to a randomly chosen (v_r,τ_r).

Lemma 2. *The probability that any randomly selected one-hop view cannot be assigned a PRB $n \in N$ is bounded by $p(\bigcap_{n\in N}\eta_n) \geqslant \prod_{i=1}^{N}(1-1/e^{|\psi_n(v_r)|p})$.*

Proof. We can write the following equation for PRB $n \in N$:

$$p(\eta_n) = p(\tau_r \cap \psi_n(v_r) = \phi) = (1-p)^{|\psi_n(v_r)|}. \tag{23}$$

As $p \in [0,1]$, we can write:

$$(1-p)^{|\psi_n(v_r)|} \leq \frac{1}{e^{p|\psi_n(v_r)|}}. \tag{24}$$

Based on Eqs. (23)-(24), we havd:

$$p(\eta_n) \leq \frac{1}{e^{|\psi_n(v_r)|p}}. \tag{25}$$

Now, we prove that the probability that all events ψ_n occur is lower bounded by the probability that would come up from the independent events as follows:

$$p\left(\bigcap_{n\in N}\eta_n\right) = \prod_{i=1}^{N} p\left(|\psi_{n_i}|\bigcap_{j=1}^{i-1}\psi_{n_j}\right). \tag{26}$$

Now, events ψ_n are positively correlated and nodes are independently added to τ_r. Thus, we say:

$$\prod_{i=1}^{N} p\left(|\psi_{n_i}|\bigcap_{j=1}^{i-1}\psi_{n_j}\right) \geq \prod_{i=1}^{N} p(\eta_{n_i}). \tag{27}$$

From Eqs. (25), (26), and (27) we can write:

$$p\left(\bigcap_{n\in N}\eta_n\right) \geqslant \prod_{i=1}^{N}\left(1-\frac{1}{e^{|\psi_n(v_r)|p}}\right). \tag{28}$$

□

Similar to the above analysis, we can derive the probability for a SUE to get an interference-free PRB assuming that the same PRB has not been assigned to any two nodes within a one-hop view. Let $\mathbb{P}_n$ be the probability of selecting an interference-free PRB. Then we have the following lemma.

Lemma 3. *The probability ($\mathbb{P}_n$) that each node can be assigned with interference-free PRB using the RDRA algorithm should satisfy $\mathbb{P}_n = \sum_{i=1}^{N}(-1)^{i-1}\binom{N}{i}(1-i\mathbb{P}_n/N)^{\Delta}$.*

Proof. Based on RDRA, the probability that a node is assigned with a PRB is given by $\mathbb{P}_n = \sum_{i=1}^{N}\mathbb{P}_{n_i}$, where $\mathbb{P}_{n_i}$ is the probability that a node can be assigned with PRB n_i. Let $\bar{n_i}$ denotes that a node does not get a PRB n_i. Thus, if a node is assigned with a PRB n_i, then none of its one-hop view nodes gets the PRB n_i. This fact can be written as $\mathbb{P}(\bigcup_{i=1}^{N}\bar{n_i})$. Thus,

$$\mathbb{P}_n = \sum_{i=1}^{N}\mathbb{P}_{n_i} = \mathbb{P}\left(\bigcup_{i=1}^{N}\bar{n_i}\right). \tag{29}$$

Furthermore, we can write the following equation based on the principle of inclusion-exclusion.

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{N}\bar{n_i}\right) &= \sum_{i=1}^{N}\mathbb{P}(\bar{n_i}) - \sum_{1\leq i<j\leq N}\mathbb{P}(\bar{n_i}\cap\bar{n_j}) \\ &+ \sum_{1\leq i<j<k\leq N}\mathbb{P}(\bar{n_i}\cap\bar{n_j}\cap\bar{n_k}) - \cdots + (-1)^{N-1}\mathbb{P}\left(\bigcap_{i=1}^{N}\bar{n_i}\right) \\ &= \sum_{i=1}^{N}(1-\mathbb{P}_{n_i})^{\Delta} - \sum_{1\leq i<j\leq N}(1-\mathbb{P}_{n_i}-\mathbb{P}_{n_j})^{\Delta} \\ &+ \sum_{1\leq i<j<k\leq N}(1-\mathbb{P}_{n_i}-\mathbb{P}_{n_j}-\mathbb{P}_{n_k})^{\Delta} \\ &- \cdots + (-1)^{N-1}(1-\mathbb{P}_{n_1}-\mathbb{P}_{n_2}-\cdots-\mathbb{P}_{n_N})^{\Delta}.\end{aligned} \tag{30}$$

Due to the symmetry in each PRB, we can write:

$$\mathbb{P}_{n_1} = \mathbb{P}_{n_2} = \cdots = \mathbb{P}_{n_N} = \frac{\mathbb{P}_n}{N}. \tag{31}$$

Thus, based on Eqs. (30) and (31) we get:

$$\mathbb{P}_n = \sum_{i=1}^{N}(-1)^{i-1}\binom{N}{i}\left(1-i\frac{\mathbb{P}_n}{N}\right)^{\Delta}. \tag{32}$$

Hence the proof of the lemma. □

7 Experimental Results

In this section, we present the performance of the proposed algorithms through simulation experiments. The simulation environment is modeling using JAVA language following the 3GPP standard as shown in Table 2 [44], [45]. The transmit power of MBS, MAP, PAP, and FAP are set to 40W, 5W, 1W, and 0.1W, respectively. The carrier frequency, system bandwidth, the modulation scheme, and the number of PRBs are

TABLE 2
Simulation Parameters

Parameters	Details
System bandwidth	20 MHz
Carrier frequency	2 GHz
Transmit power of Macro base station	40 W
Transmit power of Micro access point	5 W
Transmit power of Pico access point	1 W
Transmit power of Femto access point	0.1 W
Weights as priority levels	1,2,3,4
Number of users per access point	{1,2,..100}
Number of PRBs	100
Small cells layout	circle
Transmission Time Interval	1 ms
Radio frame length in physical channel	10 ms
Subcarrier Spacing	15 kHz

considered as 2 GHz, 20 MHz, 64 QAM, and 100, respectively, as per the 3 GPP standard. The Transmission Time Interval (TTI) and the radio frame length in the physical channel are considered as 1 ms and 10 ms, respectively as per the 3GPP standard [44], [45]. Our results are obtained on an Intel Core 2 Duo 3.0 GHz with 8 GB RAM under Windows 7 Enterprise. The number of users in each SAP is varied for different outcomes. A square region of 1000m x 1000m is used for the simulation. The macro base station (MBS) is deployed in the center while the FAP, PAP, and MAP are randomly deployed underlying the macrocell region. The radius of FAP, PAP, and MAP are considered as 15m, 30m, and 50m, respectively. We used four priority levels of nodes, $P_1 > P_2 > P_3 > P_4$ based on the user's subscription plan, such as premium, gold, silver, and bronze respectively. We have compared our proposed algorithms with two competing works in the literature [31] and [14], based on the objectives of re-usability of resources and fairness constraints, as given by Eqs. (5), (6), (7), (8), (9), and (10).

7.1 Real Data Analysis

We have evaluated the performance of our proposed algorithm using real data set of the *Wireless Topology Discovery* project [46]. This data set contains information of approximately 275 personal digital assistant (PDA) users for a period between September 22, 2002 and December 8, 2002. All devices are connected with WiFi access point (AP). A PDA may detect more than one APs due to the short distance among them. Similar to [47], [48], where small cells are considered equivalent to WiFi, we also simulated a WiFi is equivalent to a SAP. This approximation is reasonable for SAPs because the radius of WiFi APs is $\sim$ 100 meters. Importantly, since the operators typically keep the datasets of users across their base stations confidential, it is a bit difficult to acquire such real-time information of users in small cell networks. We have set a time slot duration of 20 seconds. In each time slot, the LCG evolves if a user comes or moves away from the system (following Section 5.2). Since a PDA can get a signal from more than one APs at a time, it may represent more than one node in LCG, as discussed in Fig. 2. We have taken a random priority of each PDA user between 1 and 4.

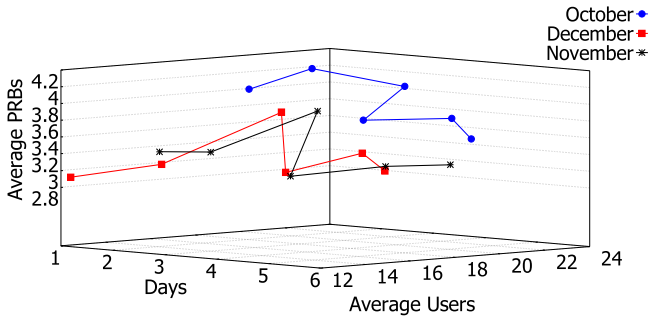

Fig. 8. Comparison among days, users, and PRBs.

Fig. 7 compares the average number of PRBs required from October 10, 2002 to October 25, 2002, and from November 10, 2002 to November 25, 2002. As observed, the average number of PRBs is higher on October 16 than other days. This is because it was the busiest day and most PDAs activated at the same time, requiring a higher number of PRBs to be allocated. Fig. 8 draws a comparison of the average number of activated PDAs and average number of required PRBs across various days. Comparison is made on the dates 1-6 of October, November, and December, 2002. The average number of required PRBs in October is more compared to those in November and December. This is due to the fact that a higher number of PDA users turned up simultaneously in October than November and December. On the other hand, the average number of required PRBs in November is almost the same as December.

7.2 Reuse Ratio Analysis

Fig. 9 compares our algorithms with two existing schemes, namely DRA [31] and INT [14]. These two schemes are chosen because of their closeness to our formulated resource allocation problem. The deployment probabilities of FAPs,

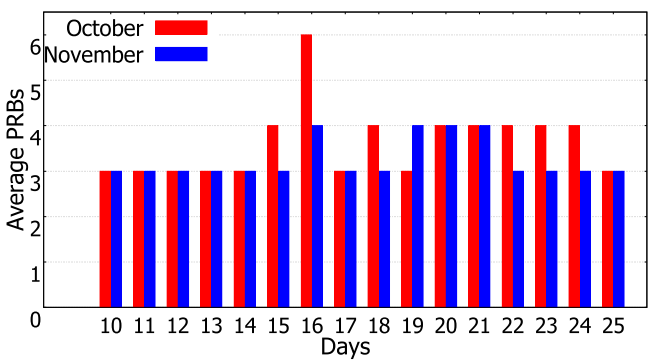

Fig. 7. Comparison between days and required PRBs.

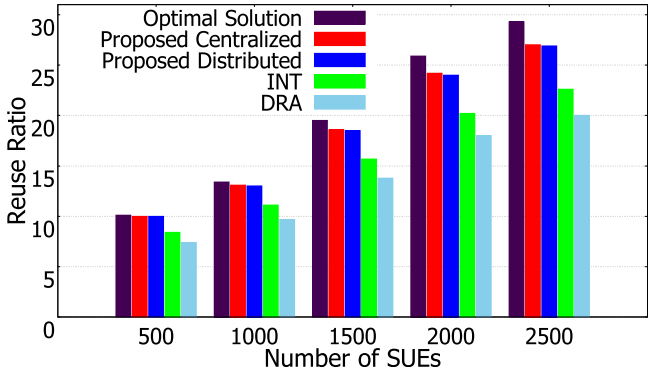

Fig. 9. Comparison between reuse ratio and number of users.

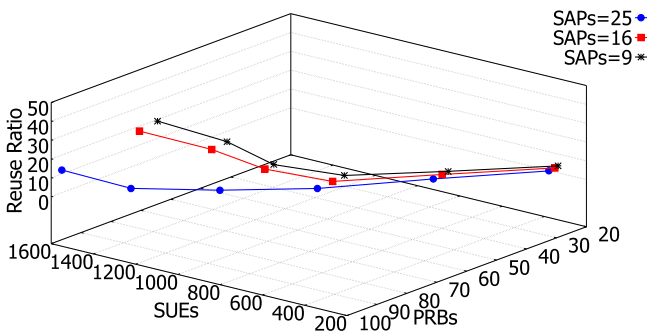

Fig. 10. Comparison among users, PRBs, and reuse ratio.

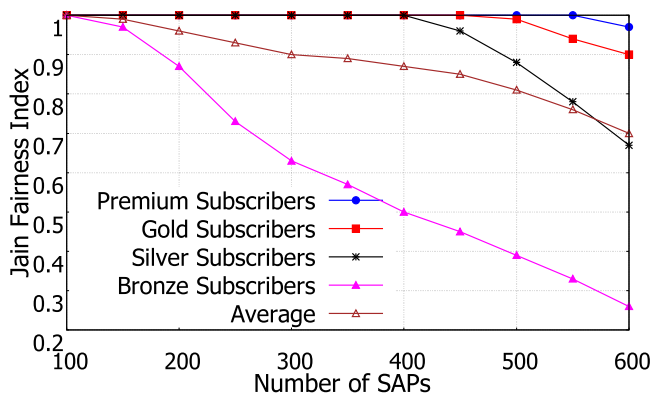

Fig. 11. Jain fairness index for different priority subscribers.

PAPs, and MAPs are considered as 0.3, 0.4, and 0.3, respectively. The reuse ratio of the proposed algorithms is better than the DRA [31] and INT [14] schemes by 26 percent and 16 percent, respectively on an average. We have studied an equal ratio of different priority subscribers in the network. The improvement is expected because, in the DRA algorithm, the resource allocation is done per femtocell basis without considering user density. We obtained better performance than the INT algorithm because as in [14], users are deployed within the femtocells region, and when these users are deployed within disparate areas, very less chance exists for interference due to the non-uniformity among femtocells, picocells, and microcells. The obtained results demonstrate that the proposed distributed algorithm performs the same as the centralized algorithm due to the localized nature of the network interference objective function. We have also compared our findings with the optimal solution obtained using the "ILOG CPLEX" solver [49]. We considered the same set of users and SAPs for comparing the average reuse ratio obtained from the proposed schemes and ILOG CPLEX solver. Fig. 9 shows that the performance of our proposed centralised and distributed schemes are quite close to the optimal value, even with increased number of users.

We compare the total number of users, total number of SAPs, minimum number of required PRBs, and reuse ratio in Fig. 10. The deployment probabilities of FAPs, PAPs, and MAPs are considered as 0.2, 0.4, and 0.4, respectively. We observe from the results that the number of SAPs plays a significant role in the reuse ratio of PRBs. The reuse ratio of 25 SAPs is less than that of 16 and 9 SAPs for the same number of users in the network. This is because an increase in SAPs causes higher interference in the network within the same region, thus resulting in a lower reuse ratio of PRBs.

7.3 Fairness Analysis

We evaluate user-level and cell-level fairness as follows:

7.3.1 User-Level Fairness

Let $\varpi^{a}_{u,n} \in \{\alpha^{m}_{i,n}, \beta^{p}_{j,n}, \gamma^{f}_{k,n}\}$ be a binary variable that represents an association of user $u \in U$ with SAP $a \in \{M, P, F\}$ at an allocated PRB $n \in N$. This variable is likely be an outcome of the proposed algorithms as follows:

$$\varpi^{a}_{u,n} = \begin{cases} 1, & \text{if algorithm allocates a PRB } n \text{ to link } (u,a) \\ 0, & \text{otherwise.} \end{cases} \tag{33}$$

We further define the Resource Allocation Ratio (RAR) for a user u as $RAR(u) = \sum_{a\in\{M,P,F\}} \sum_{n\in N} \varpi^{a}_{u,n}/d_u$. To estimate the user-level fairness, we define the premium, gold, silver, and bronze subscribers as $U_{\Re_1}$, $U_{\Re_2}$, $U_{\Re_3}$ and $U_{\Re_4}$, respectively such that $U = U_{\Re_1} + U_{\Re_2} + U_{\Re_3} + U_{\Re_4}$. We further consider the ratio of premium, gold, silver, and bronze subscribers in the network as $\Re_1, \Re_2, \Re_3$ and $\Re_4$, respectively, where $\Re_1 + \Re_2 + \Re_3 + \Re_4 = 1$, and $\Re_1, \Re_2, \Re_3, \Re_4 \geq 0$. Thus, the user-level fairness for $U_{\Re_i}$ users is given by [50]:

$$JF(U_{\Re_i}) = \frac{\left(\sum_{u\in U_{\Re_i}} RAR(u)\right)^2}{\left(|U|\cdot\Re_i\right)\cdot\sum_{u\in U_{\Re_i}} \left(RAR(u)\right)^2}. \tag{34}$$

A dense number of SAPs is deployed in the network to estimate the user-level fairness, as shown in Fig. 11. We have assumed that a maximum of 20 SUEs can be associated with an SAP. $\Re_1, \Re_2, \Re_3$ and $\Re_4$ are set at 0.3, 0.3, 0.2 and 0.2, respectively. We execute the centralized algorithm to measure the user-level fairness for different subscribers in the network. We have further shown the average fairness for different subscribers in the network. From Fig. 11, we conclude that the premium subscribers always receive better fairness compared to other subscribers. However, with an increase in the number of SAPs, the user-level fairness decreases due to the reduction in the reuse ratio of PRBs and increase in the network density and interference.

We obtain a better average Jain (see Eq. (34)) fairness index of the proposed algorithms compared to the existing schemes [31], [14], as shown in Fig. 12. We have considered that a maximum of five SUEs can be associated with an SAP, and there is an equal ratio of different priority

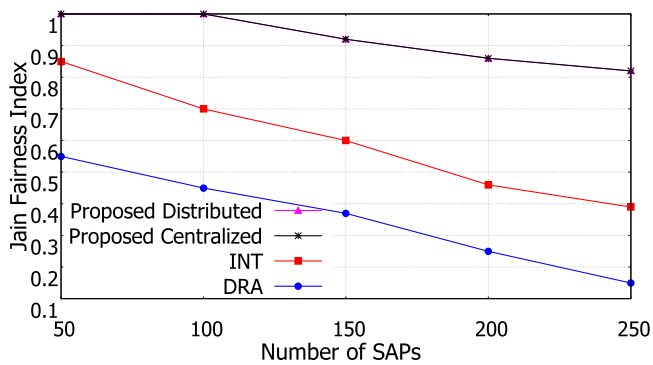

Fig. 12. User-level fairness comparison.

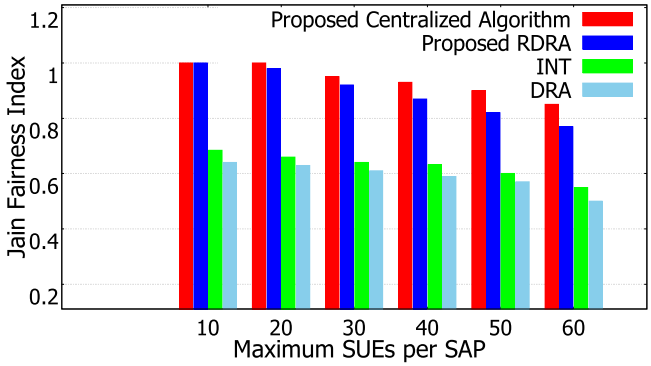

Fig. 13. SAP-level fairness comparison.

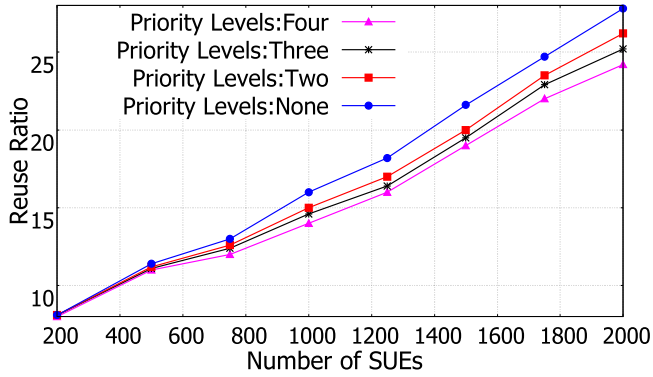

Fig. 14. Reuse ratio analysis at different priority levels.

subscribers in the network. With an increase in SAPs, the user-level fairness decreases due to an increase in the interference. Moreover, improved fairness compared to DRA is due to the fact that resource allocation in DRA is done at the cell level, whereas in our algorithms, it is done at both user and cell levels. Similarly, better fairness than the INT model can be explained as follows. In the INT model, although the resource allocation is executed at the user-level, while allocating the resources to nodes, it does not consider the actual priority of the users. While the INT uses the same algorithmic steps for different priority users, the improvement of our proposed algorithms over INT is due to the re-usability of unused PRBs by interfering high priority users, which allows a better distribution of resources among the remaining lower priority users. Furthermore, the obtained fairness from the proposed centralized and distributed algorithms are the same due to the analogous localized nature of fairness and network objective functions.

7.3.2 SAP-Level Fairness

Based on the above Eqs. (33) and (34) we define the Resource Allocation Ratio (RAR) for an SAP a as, $RAR(a) = \sum_{u\in U}\sum_{n\in N}\varpi^{a}_{u,n}/D_a$. Accordingly, the average SAP-level fairness based on Jain fairness index is defined as [50]:

$$JF(SAP) = \frac{(\sum_{a\in\{M,P,F\}} RAR(a))^2}{|\{M\cup P\cup F\}|\sum_{a\in\{M,P,F\}}(RAR(a))^2}. \quad (35)$$

We considered a total of 30 SAPs with different available resources. The resource demand depends on the number of users connected with the respective SAPs. Fig. 13 shows results for an average of all the SAPs while considering the same ratio of different priority users. Compared with existing works [31], [14], we obtained better SAP-level fairness. This is because unlike existing works, we always keep track of total available resources at respective SAPs and their demands while allocating resources to each node.

7.4 Analysis of Priority Levels

In this subsection, we study the impact of the average reuse ratio and fairness on the different priority levels.

Fig. 14 examines the impact of different priority levels over the reuse ratio. The deployment probabilities of FAPs, PAPs, and MAPs are set at 0.33, 0.33, and 0.34, respectively. We have considered an equal ratio of different priority subscribers. From Fig. 14, we conclude that with an increase of priority levels, the reuse ratio decreases on an average. This is because our proposed algorithm first assigns PRB to high priority users within the one-hop view, and as per the interference constraint, the same PRB can not be assigned to any two nodes within one-hop. Thus, if a PRB n is assigned to any node i, then it can not be further reassigned to another node in τ_i; hence, this reduces the usage of PRB n in the network by τ_i. In other words, with an increase of priority levels, we restrict the usage of PRBs by the one-hop view of each high priority node, thus decreasing the average reuse ratio of PRBs.

Fig. 15 compares the effect of different priority levels on the average user-level fairness. The deployment probabilities of FAP, PAP, and MAP are set at 0.33, 0.33, and 0.34, respectively. We have considered an equal ratio of different priority subscribers. The obtained results conclude that with an increase in the priority level, there is a corresponding decrease in the average user-level fairness. This is because with the rise of priority levels, the reuse ratio decreases (see Fig. 14), thus reducing the number of users assigned with valid PRBs in the network.

7.5 Performance Analysis

In this subsection, we analyze the computational and message complexity of our proposed algorithms.

7.5.1 Execution Time

Table 3 compares the time complexity of our proposed schemes with the existing schemes, INT [14] and DRA [31]. The INT algorithm [14] iterates over all vertices by picking up a maximal independent set (MIS) and assigning the lowest available resource to this set. To find an MIS, the algorithm needs to check all edges present in the LCG, and in a complete graph the number of edges is $O(\mathbb{N}^2)$. Thus, the

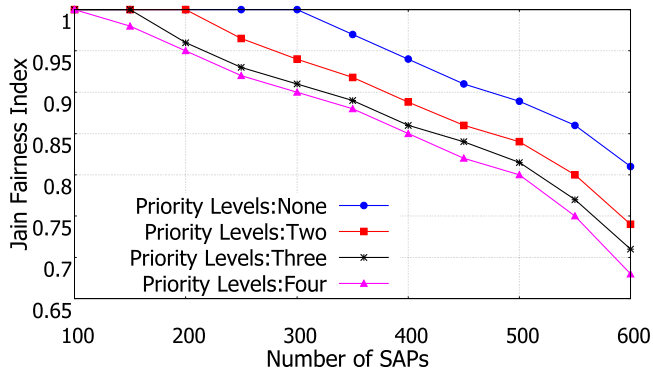

Fig. 15. Fairness analysis at different priority levels.

TABLE 3
Time Complexity Comparison

Algorithm	Time Complexity
INT [14]	$O(N\mathbb{N}^3)$
DRA [31]	$O(N\mathbb{S})$
Centralized (Proposed)	$O(\mathbb{N}^2 + N\mathbb{N}\Delta)$
Distributed (Proposed)	$O(\Delta)$
RDRA (Proposed)	$O(\Delta)$ (Expected)

worst-case time complexity of INT is given by $O(N\mathbb{N}^3)$. The DRA scheme needs to hash the resources of each cell iteratively. Keeping the worst-case scenario in mind, we assumed that all SAPs $\mathbb{S} = |\{M \cup P \cup F\}|$ could utilize the available PRBs pairs. In case of collision, each SAP may try all PRBs, resulting in a time complexity of $O(N\mathbb{S})$. It is evident that the asymptotic time complexity of the proposed centralized scheme is higher than that of DRA, but the proposed scheme provides a significant performance improvements over the DRA. In other words, proposed centralized algorithm takes more number of iterations compared to the DRA but gives better performance in terms of reuse ratio (Fig. 9) and fairness (Figs. 12 and 13).

We have compared the execution time of INT, DRA, and our proposed algorithms in Fig. 16. Our results were obtained on an Intel Core 2 Duo 3.0 GHz with 8 GB RAM under Windows 7 Enterprise. We observe that with an increase in the number of users, the total execution time increases. This is because with the rise of users, the total number of nodes increases, which takes more time to assign valid PRBs in the network. The rate of growth in the execution time for the number of users in RDRA is much smaller than any other algorithm. Unlike centralized and distributed algorithms, in RDRA, each node selects the PRB probabilistically once the rank assignment is done, resulting in faster termination of the RDRA algorithm.

7.5.2 Message Overhead, Delay, and Throughput

For evaluating the proposed distributed algorithm, we consider the following performance metrics: *(i) Message Overhead:* the average number of message transmissions; *(ii) Execution Delay:* the average time required for a user to acquire a valid PRB; and *(iii) Throughput:* the number of users acquiring valid PRB.

As the message overhead and execution delay depend on the LCG topology. We define *Interfering Edge Density* (IED) of the generated topology as the ratio between actual interfering edges in LCG to the maximum possible interfering edges, i.e., $IED = |E|/\binom{\mathbb{N}}{2}$. To generate different IED based LCGs, we varied the set of SAPs and number of SUEs within the network. As the throughput and IED increase, there is a corresponding increase in the execution time and message overhead, as shown in Fig. 17. The reason is that with an increase in IED for a given set of nodes, there is a corresponding increase in interference among wireless links, and; this leads to a higher message overhead and delay to assign valid PRBs to each SUE.

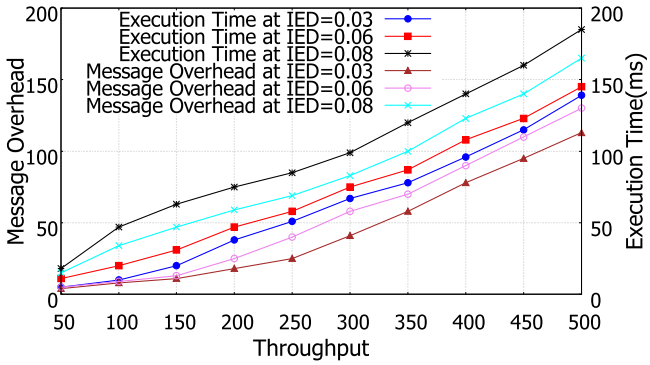

Fig. 17. Throughput, message overhead and execution delay.

Fig. 18 shows the *trade-off* between throughput and IED at a different set of available PRBs. We fixed the number of SUEs at 800 and varied the number of SAPs to estimate the throughput at different IEDs. Fig. 18 shows that the throughput decreases with an increase of IED at a fixed number of PRBs. However, with increased PRBs, there is a corresponding increase in the throughput. We also show the trade-off between IED and throughput for the INT and DRA schemes. From the results, we conclude that the existing schemes follow the same trade-off pattern between the throughput and IED as our proposed methods, which exhibit better performance due to higher reuse ratio (see Fig. 9), thus assigning valid PRBs to more users.

7.5.3 Number of Rounds

In Fig. 19; the deployment probabilities of FAP, PAP, and MAP are considered as 0.5, 0.2, and 0.3, respectively. The results demonstrate that the randomized algorithm takes fewer rounds than the distributed algorithm; the former algorithm terminates much faster than the latter because nodes select PRBs probabilistically in early rounds. More-

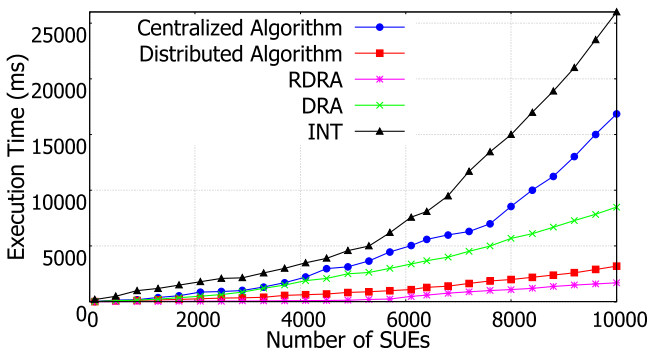

Fig. 16. Execution time comparison.

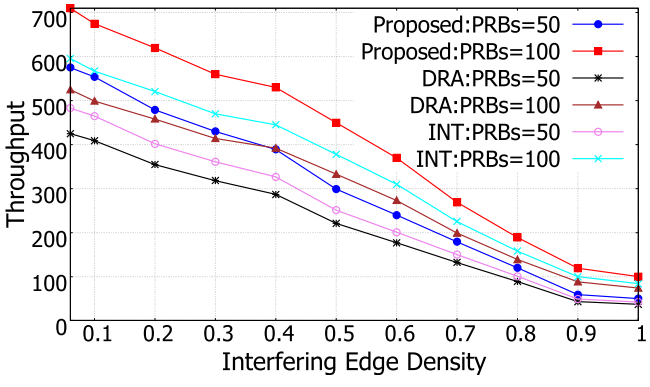

Fig. 18. Comparison among throughput, IED, and number of PRBs.

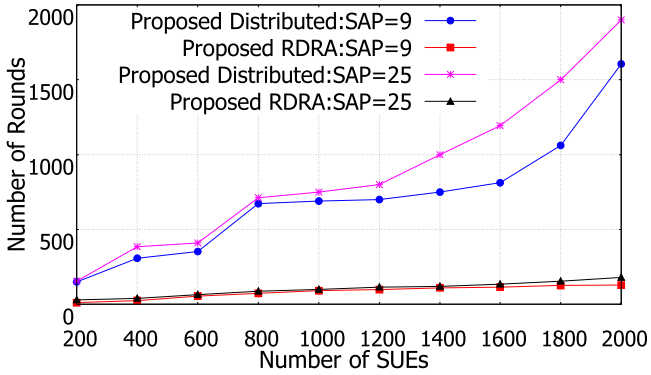

Fig. 19. Comparison between users and rounds.

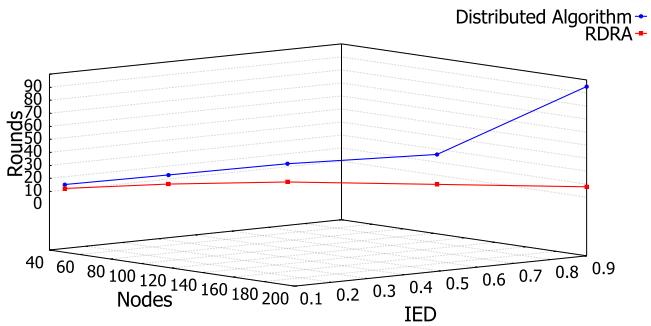

Fig. 20. Comparison among rounds, nodes and IED.

over, with increased number of SAPs, the total number of nodes increases, thus requiring a higher number of rounds to assign valid PRBs to the same number of users.

Fig. 20 shows the termination rates of the distributed and randomized algorithms against the number of nodes and IED. The termination rate of RDRA is much faster than the distributed algorithm. The improvement is significant when the number of nodes increases. This is due to the fact that most of the nodes in the RDRA algorithm select their requested PRBs in the early rounds.

8 Conclusion and Future Works

In this paper, we have designed efficient resource allocation algorithms by minimizing interference and jointly maximizing resource reuse ratio, user-level, and cell-level fairness in 5G networks. In particular, we proposed three algorithms—centralized, distributed, and randomized—and analyzed their performance theoretically and experimentally via simulation. The performance comparison of our algorithms with two existing schemes, DRA and INT, respectively demonstrates 26 percent and 16 percent improvement in terms of average reuse ratio.

Our future work will aim to address interference and fairness in dynamic scenarios that will enable the users to connect with multiple SAPs simultaneously and account for advanced techniques, such as cooperative communications and interference management in heterogeneous 5G networks. Deriving approximation ratios of our proposed sub-optimal algorithms and further trade-off analysis with existing schemes are other future works that we will focus on. Finally, we plan to formulate the resource allocation as an online optimization problem and propose efficient algorithms for small cell-enabled 5G and beyond 5G networks.

Acknowledgments

The authors would like to thank the Associate Editor and anonymous reviewers for insightful comments that helped them improve the quality of the manuscript significantly. This work is partially supported by NSF Grants under Award numbers CNS-1818942, CCF-1725755, CNS-1545037, and CNS-1545050.

References

[1] G. De La Roche, A. Valcarce, D. López-Pérez, and J. Zhang, "Access control mechanisms for femtocells," *IEEE Commun. Magazine*, vol. 48, no. 1, pp. 33–39, Jan. 2010.
[2] P. Xia, V. Chandrasekhar, and J. G. Andrews, "Open versus closed access femtocells in the uplink," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3798–3809, Dec. 2010.
[3] Z. Du, Y. Sun, W. Guo, Y. Xu, Q. Wu, and J. Zhang, "Data-driven deployment and cooperative self-organization in ultra-dense small cell networks," *IEEE Access*, vol. 6, pp. 22839–22848, 2018.
[4] A. Pratap, R. Misra, and S. K. Das, "Resource allocation to maximize fairness and minimize interference for maximum spectrum reuse in 5G cellular networks," in *Proc. 19th IEEE Int. Symp. World Wireless Mobile Multimedia Netw.*, 2018, pp. 1–9.
[5] N. Brännström, E. G. Lundin, and P. Lundh, "Measurement based QoS adaptation," US Patent 9,867,069, Jan. 9 2018.
[6] S. Zhang and J. Wang, "Method and network entity for QoS control," US Patent App. 16/046,204, Nov. 15 2018.
[7] Y.-S. Liang, W.-H. Chung, G.-K. Ni, Y. Chen, H. Zhang, and S.-Y. Kuo, "Resource allocation with interference avoidance in OFDMA femtocell networks," *IEEE Trans. Veh. Technol.*, vol. 61, no. 5, pp. 2243–2255, Jun. 2012.
[8] S. Uygungelen, G. Auer, and Z. Bharucha, "Graph-based dynamic frequency reuse in femtocell networks," in *Proc. IEEE 73rd Veh. Technol. Conf.*, 2011, pp. 1–6.
[9] Y. Yang, B. Bai, and W. Chen, "Spectrum reuse ratio in 5G cellular networks: A matrix graph approach," *IEEE Trans. Mobile Comput.*, vol. 16, no. 12, pp. 3541–3553, Dec. 2017.
[10] T. Maksymyuk, M. Brych, and A. Masyuk, "Fractal geometry based resource allocation for 5G heterogeneous networks," in *Proc. 2nd Int. Scientific-Practical Conf. Problems Infocommunications Sci. Technol.*, 2015, pp. 69–72.
[11] X. Lagrange, "Multitier cell design," *IEEE Commun. Magazine*, vol. 35, no. 8, pp. 60–64, Aug. 1997.
[12] E. Pateromichelakis, M. Shariat, A. U. Quddus, and R. Tafazolli, "Graph-based multicell scheduling in OFDMA-based small cell networks," *IEEE Access*, vol. 2, pp. 897–908, 2014.
[13] F. Zhao, W. Ma, M. Zhou, and C. Zhang, "A graph-based QoS-aware resource management scheme for OFDMA femtocell networks," *IEEE Access*, vol. 6, pp. 1870–1881, 2018.
[14] Z. Lu, T. Bansal, and P. Sinha, "Achieving user-level fairness in open-access femtocell-based architecture," *IEEE Trans. Mobile Comput.*, vol. 12, no. 10, pp. 1943–1954, Oct. 2013.
[15] K. Lee, O. Jo, and D.-H. Cho, "Cooperative resource allocation for guaranteeing intercell fairness in femtocell networks," *IEEE Commun. Lett.*, vol. 15, no. 2, pp. 214–216, Feb. 2011.
[16] B.-G. Kim, J.-A. Kwon, and J.-W. Lee, "Subchannel allocation for the OFDMA-based femtocell system," *Comput. Netw.*, vol. 57, no. 17, pp. 3617–3629, 2013.
[17] Y. L. Lee, J. Loo, T. C. Chuah, and A. A. El-Saleh, "Fair resource allocation with interference mitigation and resource reuse for LTE/LTE-A femtocell networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8203–8217, Oct. 2016.
[18] X. Chen, Z. Feng, and D. Yang, "An energy-efficient macro-micro hierarchical structure with resource allocation in OFDMA cellular systems," in *Proc. 7th Int. Conf. Commun. Netw.*, 2012, pp. 519–523.
[19] D. Cao, S. Zhou, and Z. Niu, "Optimal base station density for energy-efficient heterogeneous cellular networks," in *Proc. IEEE Int. Conf. Commun.*, 2012, pp. 4379–4383.
[20] F. Richter and G. Fettweis, "Cellular mobile network densification utilizing micro base stations," in *Proc. IEEE Int. Conf. Commun.*, 2010, pp. 1–6.
[21] H.-S. Jo, C. Mun, J. Moon, and J.-G. Yook, "Interference mitigation using uplink power control for two-tier femtocell networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 4906–4910, Oct. 2009.

[22] V. Chandrasekhar and J. G. Andrews, "Spectrum allocation in tiered cellular networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3059–3068, Oct. 2009.
[23] M. Necker and M. C. Necker, "A graph-based scheme for distributed interference coordination in cellular OFDMA networks," in *Proc. IEEE Veh. Technol. Conf.*, 2008, pp. 713–718.
[24] L. G. Garcia, K. I. Pedersen, and P. E. Mogensen, "Autonomous component carrier selection: Interference management in local area environments for LTE-advanced," *IEEE Commun. Magazine*, vol. 47, no. 9, pp. 110–116, Sep. 2009.
[25] A. Hatoum, N. Aitsaadi, R. Langar, R. Boutaba, and G. Pujolle, "FCRA: Femtocell cluster-based resource allocation scheme for OFDMA networks," in *Proc. IEEE Int. Conf. Commun.*, 2011, pp. 1–6.
[26] E. Pateromichelakis, M. Shariat, A. Quddus, M. Dianati, and R. Tafazolli, "Dynamic clustering framework for multi-cell scheduling in dense small cell networks," *IEEE Commun. Lett.*, vol. 17, no. 9, pp. 1802–1805, Sep. 2013.
[27] Y. Wang, K. Zheng, X. Shen, and W. Wang, "A distributed resource allocation scheme in femtocell networks," in *Proc. IEEE 73rd Veh. Technol. Conf.*, 2011, pp. 1–5.
[28] A. Hatoum, R. Langar, N. Aitsaadi, R. Boutaba, and G. Pujolle, "Cluster-based resource management in OFDMA femtocell networks with QoS guarantees," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2378–2391, Jun. 2014.
[29] S. D. Ganni, A. Pratap, and R. Misra, "Distributed algorithm for resource allocation in downlink heterogeneous small cell networks," in *Proc. 7th ACM Int. Workshop Mobility Interference MiddleWare Manage. HetNets*, 2017, pp. 1–6.
[30] A. Pratap and R. Misra, "Random graph coloring-based resource allocation for achieving user level fairness in femtocellular LTE-A networks," *Wireless Pers. Commun.*, vol. 98, no. 2, pp. 1975–1995, 2018.
[31] K. Sundaresan and S. Rangarajan, "Efficient resource management in OFDMA femto cells," in *Proc. 10th ACM Int. Symp. Mobile Hoc Netw. Comput.*, 2009, pp. 33–42.
[32] G. Huang and J. Li, "Interference mitigation for femtocell networks via adaptive frequency reuse," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2413–2423, Apr. 2016.
[33] E. E. Tsiropoulou, P. Vamvakas, G. K. Katsinis, and S. Papavassiliou, "Combined power and rate allocation in self-optimized multi-service two-tier femtocell networks," *Comput. Commun.*, vol. 72, pp. 38–48, 2015.
[34] E. E. Tsiropoulou, P. Vamvakas, and S. Papavassiliou, "Joint customized price and power control for energy-efficient multi-service wireless networks via S-modular theory," *IEEE Trans. Green Commun. Netw.*, vol. 1, no. 1, pp. 17–28, Mar. 2017.
[35] D. Wubben et al., "Benefits and impact of cloud Computing on 5G signal processing: Flexible centralization through cloud-RAN," *IEEE Signal Process. Magazine*, vol. 31, no. 6, pp. 35–44, Nov. 2014.
[36] A. Pratap, R. Singhal, R. Misra, and S. K. Das, "Distributed randomized *k*-Clustering based PCID assignment for ultra-dense femtocellular networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 29, no. 6, pp. 1247–1260, Jun. 2018.
[37] "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall Description; Stage 2," TS 36.300 V10.5.0," 2011. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2430, Accessed: Oct. 18, 2019.
[38] J. Padhye, S. Agarwal, V. N. Padmanabhan, L. Qiu, A. Rao, and B. Zill, "Estimation of link interference in static multi-hop wireless networks," in *Proc. 5th ACM SIGCOMM Conf. Internet Meas.*, 2005, pp. 28–28.
[39] D. Niculescu, "Interference map for 802.11 networks," in *Proc. 7th ACM SIGCOMM Conf. Internet Meas.*, 2007, pp. 339–350.
[40] M. Y. Arslan, K. Sundaresan, S. V. Krishnamurthy, and S. Rangarajan, "FERMI: A femtocell resource management system forinterference mitigation in OFDMA networks," in *Proc. 17th Annu. Int. Conf. Mobile Comput. Netw.*, 2011, pp. 25–36.
[41] D. S. Johnson and M. R. Garey, *Computers and Intractability: A Guide to the Theory of NP-completeness*, New York, NY, USA: WH Freeman and Company, 1979.
[42] F. Kuhn, "Local multicoloring algorithms: Computer a nearly-optimal TDMA schedule in constant time," in *Proc. 26th Int. Symp. Theoretical Aspects Comput. Sci.*, 2009, pp. 613–624.
[43] F. Kuhn and R. Wattenhofer, "On the complexity of distributed graph coloring," in *Proc. 25th ACM Symp. Principles Distrib. comput.*, 2006, pp. 7–15.
[44] "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (version 10.4.0), 3GPP Std. TS 36.213," 2011. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2427, Accessed: Oct. 18, 2019.
[45] M. Baker, "Uplink transmission procedures," *LTE - The UMTS Long Term Evolution: From Theory to Practice*, 2nd Ed., Hoboken, NJ, USA: Wiley, pp. 407–420, 2011.
[46] M. McNett and G. M. Voelker, "Access and mobility of wireless PDA users," *ACM SIGMOBILE Mobile Comput. Commun.*, vol. 9, no. 2, pp. 40–55, 2005.
[47] Y. Guan, Y. Xiao, H. Feng, C.-C. Shen, and L. J. Cimini, "MobiCacher: Mobility-aware content caching in small-cell networks," in *Proc. IEEE Global Commun. Conf.*, 2014, pp. 4537–4542.
[48] K. Poularakis and L. Tassiulas, "Code, cache and deliver on the move: A novel caching paradigm in hyper-dense small-cell networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 3, pp. 675–687, Mar. 2017.
[49] "CPLEX optimizer," 2012. [Online]. Available: https://www.ibm.com/analytics/cplex-optimizer, Accessed: Oct. 18, 2019
[50] R. Jain, D.-M. Chiu, and W. Hawe, "A Quantitative measure of fairness and discrimination for resource allocation in shared systems," Digital Equipment Corporation, Maynard, MA, Tech. Rep. DEC-TR-301, 1984.

Ajay Pratap received the BTech degree in computer science and engineering from Uttar Pradesh Technical University Lucknow, India, in 2011, the MTech degree in computer science and engineering from IIIT Bhubaneswar, India, in 2014, and the PhD degree in computer science and engineering from the Indian Institute of Technology (IIT) Patna, India, in July 2018. He is a postdoctoral researcher in the Department of Computer Science at the Missouri University of Science and Technology, Rolla, MO, USA. His research interests include QoS level issues at the MAC layer, resource allocation, and algorithm design for next-generation advanced wireless networks. His current work is related to HetNet, Small Cells, Fog Computing, IoT, and D2D communication underlaying cellular 5G and beyond 5G networks.

Rajiv Misra received the MTech degree in computer science and engineering from the Indian Institute of Technology (IIT) Bombay, Mumbai, Maharashtra, India, and the PhD degree in mobile computing from IIT Kharagpur, Kharagpur, West Bengal, India, in 2009. He is currently an associate professor with the Department of Computer Science and Engineering, IIT Patna, Patna, Bihar India. His research interests include distributed algorithms for wireless networks and 5G, edge computing, and distributed machine learning. He is a senior member of the IEEE.

Sajal K. Das is a professor of computer science and Daniel St. Clair Endowed Chair at the Missouri University of Science and Technology, where he was the chair of the Computer Science Department during 2013-2017. He is also an International Visiting Professor and SERB-sponsored VAJRA Faculty at IIT Kharagpur, India, and Satish Dhawan Visiting Chair Professor at the Indian Institute of Science, Bangalore. His research interests include wireless sensor networks, mobile and pervasive computing, cyber-physical systems and IoTs, smart environments, cloud computing, cyber security, and biological and social network. He has published more than 700 papers in high quality journals and refereed conference proceedings. He holds five US patents and coauthored four books. He is a recipient of 10 Best Paper Awards, the IEEE Computer Society's Technical Achievement Award for pioneering contributions to sensor networks, and the University of Missouri System President's Award for Sustained Career Excellence. He is the founding editor-in-chief of Elsevier's *Pervasive and Mobile Computing*, and associate editor of the *IEEE Transactions on Dependable and Secure Computing*, *IEEE Transactions on Mobile Computing*, and *ACM Transactions on Sensor Networks*. He is an IEEE Fellow.