Resource Allocation with Multi-Connectivity in 5G Heterogeneous Networks

Chi-Mao Chen, Jang-Ping Sheu, and Yung-Ching Kuo
Department of Computer Science, National Tsing Hua University, Taiwan
emails: s108062541@m108.nthu.edu.tw, sheujp@cs.nthu.edu.tw, and cludbuster@gmail.com

***Abstract*—This paper investigates using multi-connectivity (MC) with mixed numerology to leverage the benefits of both techniques. The scenario we consider involves users connecting to multiple base stations simultaneously, with each user being served by the desired numerology set. The goal is to maximize the utility function, considering network throughput and satisfaction rate. To achieve this goal, we present an integer non-linear programming (INLP) problem and propose a multi-connectivity resource allocation (MCRA) heuristic algorithm. Moreover, a Speed-up MCRA (SMCRA) algorithm is proposed to reduce the time complexity of the MCRA algorithm while maintaining similar performance. Simulation results demonstrate that the proposed algorithms outperform existing methods in terms of both user satisfaction and network throughput.**

***Index Terms*—5G communications, multi-connectivity, mixed numerology, resource allocation**

I. Introduction

With the increasing number of mobile devices in the future, the 5G network is expected to provide great amounts of spectrum resources to satisfy growing demands. The highly dense deployment of small cells has been regarded as a promising technique to support the macro base station (BS) under the pressure of large traffic loading. So, user equipment (UE) is more likely to be under the multiple coverages of base stations (BSs). Moreover, there are more and more UEs that can support multi-connectivity (MC), which enables them to leverage more spectrum resources and use applications with stringent quality of service (QoS) compared to LTE.

Many works have MC optimization criteria, such as transmission reliability [1], [2], network throughput [3]–[5], fairness [6], [7], resource allocation [8]–[10], and satisfaction rate [11]. Recent research [12] shows that sending duplicated packets from different access points (APs) will improve channel gains and enhance reliability. Moreover, UEs aggregating various radio resources from several BSs will leverage more spectrum resources than single connectivity to enhance the throughput. Thus, the authors [3] consider the 5G mmWave network with a beamforming technique, and the users can access multiple links to improve the throughput. [4] allows the users to switch between single and MC and aims to maximize the use of idle wireless resources. The authors propose a threshold-based algorithm to decide the connection of BSs. However, the above studies do not consider UEs' fairness or satisfaction rate. Thus UEs with lousy channel conditions are likely to be allocated fewer resources.

[6] proposes a centralized Proportional Fair (PF) scheduling scheme for Dual Connectivity (DC) in heterogeneous networks (HetNets). The PF-DC scheme outperforms the standard PF scheme and improves proportional fairness. Three heuristic association algorithms for DC are proposed to maximize PF utility. With the proposed PF scheme, DC achieves significant gains on PF utility over single connectivity and performs almost as well as the optimal PF scheme. On the other hand, the authors in [11] further consider the uplink transmission with network throughput and satisfaction rate by maximizing the utility with the sigmoid function, which incorporates the user demand into network utility. The problem is then solved by the convex optimization method.

According to the 3GPP standard, 5G networks can deal with a wide variety of services, such as enhanced Mobile Broadband (eMBB), Ultra-Reliable Low Latency Communication (URLLC), and massive Machine Type Communication (mMTC). Hence, the conventional uniform resource block (RB) structure is unsuitable for these services in 5G. To accommodate these services, 3GPP provides mixed-numerology [13], enabling users to enjoy various services. Optimization for resource allocation with mixed numerology and heterogeneous QoS requirements has been studied in [8], [14], [15]. In [14], the authors consider two types of users, latency-aware and latency-tolerant users, and formulate a resource allocation problem with throughput optimization. In [15], the authors optimize the resource and numerology allocation to achieve maximum throughput. They formulated the problem as a max-min knapsack problem and solved it by linear programming. The study in [8] aims to minimize total transmission power by joint allocating power and resource block. It divides the different numerologies into multiple bandwidth parts (BWPs) with a guard band between them to reduce the inter-numerology interference (INI) [16] in the system.

However, the studies mentioned above have not considered the scenario of MC association with different numerology resource allocations. Hence, in this work, we consider that each UE has different applications with various QoS requirements that must be served by a suitable numerology set and can be associated with multiple BSs. We aim to find the solution that maximizes the utility function considering their throughput and

This work was supported by the National Science and Technology Council, Taiwan, under grant MOST 109-2221-E-007-079-MY3.

satisfaction rate. To the best of our knowledge, we are the first to introduce mixed numerology to MC in 5G HetNets, considering the satisfaction rate and throughput. The simulation results show that our proposed MCRA and SMCRA algorithms have a better throughput and satisfaction rate than the Opportunistic Scheduling Algorithm (OSA) proposed in [17].

The remainder of this paper is organized as follows. Section II introduces the network model and formulates the objective problem. Section III describes the proposed MCRA and SMCRA algorithms. Section IV presents the simulation results, and Section V concludes the paper.

II. Network Model and Problem Formulation

A. Mixed Numerology Frame Structure

Unlike the 4G LTE system, the 5G network can accommodate various services, so it adopts flexible numerology to meet user requirements. Therefore, we define the set of numerologies as $\mathcal{N}$. Let $\mu \in \mathcal{N}$ be the numerology index, corresponding to the bandwidth of an RB $15 \times 2^{\mu}$ kHz and the time interval $1/2^{\mu}$ ms in the 5G networks. In general, μ varies from 0 to 4 [13]. The available bandwidth is partitioned into many BWPs, and each BWP is configured to specific numerology to serve the applications. We put the guard band between two adjacent BWPs to avoid the rise of INI. For each numerology μ in BS j, the BWP is divided into $\mathcal{F}_j^{\mu}$ sub-bands in the frequency domain, and in the time domain is partitioned into $\mathcal{T}_j^{\mu}$ slots. So, the number of RBs in BS j with numerology μ is given by $K_j^{\mu} = \mathcal{F}_j^{\mu} \times \mathcal{T}_j^{\mu}$. We denote the set of RBs in BS j with numerology μ by $\mathbb{B}_j^{\mu}$ and let RB $b_k^{\mu} \in \mathbb{B}_j^{\mu}$, where $k = \{1, 2, ..., K_j^{\mu}\}$.

The notations and descriptions are summarized in Table I for easy reference and better understanding.

TABLE I: List of Notations

Notation	Description
$\mathbb{C}$	Set of base stations
$\mathbb{U}$	Set of users
$\mathbb{B}$	Set of RBs
$\mathcal{N}$	Set of numerologies
μ	SCS configuration
l	Maximum links the user can connect to
$x_{i,j}$	1, if user i connects to BS j; 0, otherwise
$\rho_{i,j}^{\mu,k}$	1, if RB $b_{\mu,k}$ on BS j is assigned to user i; 0, otherwise
α_i	Maximum number of RBs user i can take
β_i^{μ}	Maximum number of RBs user i with numerology μ can take
$\mathbb{B}_j^{\mu}$	Set of RBs in BS j with numerology μ
$\mathbb{N}_i$	Candidate set of numerologies for user i
$\mathbb{M}_i$	Set of RBs that assigned to user i
$\mathbb{M}_i^{\mu}$	Set of RBs that assigned to user i with numerology μ
R_i^{μ}	Total data rate of user i with numerology μ

B. System Model

We consider the downlink of two-tier HetNet with one macro BS at the center and several micro BSs and denote the set of BSs by $\mathbb{C} = \{0, 1, 2, ..., M\}$, where $j = 0$ represents the macro BS and $j = \{1, ..., M\}$ are micro BSs. The macro BS and micro BSs use different frequency resources, so there is no interference between the two different tiers of BSs. Here, the set of total RBs in the network is denoted as $\mathbb{B} = \{1, ..., N_b\}$, where $N_b = \sum_{j\in\mathbb{C}} \sum_{\mu\in\mathcal{N}} K_j^{\mu}$. The set of user equipments (UEs) is denoted by $\mathbb{U} = \{1, 2, ..., N\}$ and all of them have been equipped with multiple antennas l, which enable them to connect to at most l BSs simultaneously. The scenario of our two-tier HetNet is shown in Fig. 1.

In addition, users can use multiple services simultaneously , and each service has different latency requirements that must be met. Therefore, each service is mapped to one specific numerology according to different latency requirements to ensure its QoS. In this way, we denote the numerology set $\mathbb{N}_i \subseteq \mathcal{N}$ to represent the desired numerologies used for each user i. Lastly, due to the limitation of user capabilities [18], each user can only use at most α_i RBs to serve its applications and have at most β_i^{μ} RBs used in numerology μ. Based on our system model, we define a binary variable $x_{i,j}$ as an association indicator. If $x_{i,j} = 1$ indicates the UE i being associated with BS j; otherwise $x_{i,j} = 0$. On the other hand, we denote another binary variable $\rho_{i,j}^{\mu,k}$ to specify whether the user i occupies the RB b_k^{μ} on BS j. If $\rho_{i,j}^{\mu,k} = 1$ means the RB b_k^{μ} on BS j being used by user i; otherwise $\rho_{i,j}^{\mu,k} = 0$.

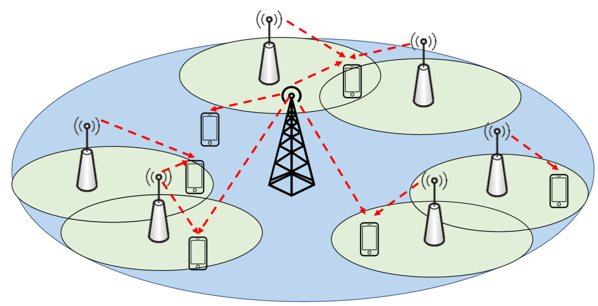


Fig. 1. Illustration of our two-tier HetNet

C. Channel Model

In our scenario, there are two types of BSs: macro and microcells. We present the Reference Symbol Received Power (RSRP) and Signal to Interference plus Noise Ratio (SINR) with different BSs. The received RSRP at the user i served by the macro cell on RB $b_k^{\mu} \in \mathbb{B}_0^{\mu}$ is given by [19]

$$rsrp_{i,0}^{\mu,k} = P_0 - L_{i,0}^{\mu,k}, \tag{1}$$

where P_0 denotes the power of RB sent from the macro cell. The $L_{i,0}^{\mu,k} = 11.4 + 28\log(dist_{i,0}) + 23\log(freq_0) + \sigma_{i,0}^{\mu,k}$ denotes the path loss from macro cell of RB b_k^{μ} to user i. The $dist_{i,0}$ denotes the distance between user i and the macro cell. The $freq_0$ is the center of the macro cell's frequency. And the $\sigma_{i,0}^{\mu,k}$ denotes the standard deviation describing large-scale signal fluctuations.

The received SINR at user i served by the macro cell on RB $b_k^{\mu} \in \mathbb{B}_0^{\mu}$ is given by

$$\gamma_{i,0}^{\mu,k} = \frac{rsrp_{i,0}^{\mu,k}}{N_0}, \tag{2}$$

where N_0 is the Additive White Gaussian Noise (AWGN). Similar to (1), the received RSRP at user i served by the microcell j on RB $b_k^\mu \in \mathbb{B}_j^\mu$, for $j \neq 0$, is expressed as

$$rsrp_{i,j}^{\mu,k} = P_j - L_{i,j}^{\mu,k}, \tag{3}$$

where $L_{i,j}^{\mu,k} = 31.4 + 20\log(dist_{i,j}) + 21\log(freq_j) + \sigma_{i,j}^{\mu,k}$ denotes the path loss from micro cell j to user i. The received SINR at user i served by the micro cell j on RB $b_k^\mu \in \mathbb{B}_j^\mu$, for $j \neq 0$ is given by

$$\gamma_{i,j}^{\mu,k} = \frac{rsrp_{i,j}^{\mu,k}}{\sum_{j' \in \mathbb{C}\setminus\{0,j\}} rsrp_{i,j'}^{\mu,k} + N_0}. \tag{4}$$

According to Shannon's theorem, the data rates received at user i served by the BS on RB b_k^μ can be expressed by

$$r_{i,j}^{\mu,k} = W^\mu \log(1 + \gamma_{i,j}^{\mu,k})T^\mu, \tag{5}$$

where W^μ and T^μ denote RB bandwidth and time duration with numerology μ, respectively. Hence, the total data rate that the numerology μ in user i can achieve is

$$R_i^\mu = \sum_{j\in\mathbb{C}} \sum_{k\in\mathbb{B}_j^\mu} x_{i,j}\rho_{i,j}^{\mu,k} r_{i,j}^{\mu,k}. \tag{6}$$

D. Problem Formulation

Here, we present an objective function to improve the network throughput and maintain the user satisfaction rate. We formulate the resource allocation problem to maximize the network utility, incorporating the user demand and throughput into the sigmoid function. We use the sigmoid function as:

$$S(R_i^\mu, \eta_i^\mu, \delta_i^\mu) = \frac{1}{1 + e^{-\eta_i^\mu (R_i^\mu - \delta_i^\mu)}}, \tag{7}$$

where δ_i^μ is the traffic requested rate for user i with numerology μ. The η_i^μ is a parameter for balancing throughput and satisfaction rate. If η_i^μ is large, it is more like distributing resources to all users, increasing the satisfaction rate. On the other hand, if η_i^μ is small, it tends to allocate resources to users with higher data rates, reducing the satisfaction rate. According to our experiments, η_i^μ was set at $10/\delta_i^\mu$ to balance the throughput and satisfaction rate. Hence, our problem formulation is given as follows:

$$\max_{(x,\rho)} \sum_{i\in U} \sum_{\mu\in\mathcal{N}} S(R_i^\mu, \eta_i^\mu, \delta_i^\mu). \tag{8}$$

$$\sum_{j\in C} x_{i,j} \leq l \tag{9}$$

$$\sum_{i\in U} \rho_{i,j}^{\mu,k} \leq 1, \forall j \in \mathbb{C}, \forall \mu \in \mathcal{N} \tag{10}$$

$$|\mathbb{M}_i| \leq \alpha_i, \forall i \in \mathbb{U} \tag{11}$$

$$\mathbb{M}_i \subseteq \cup_{j\in C} \cup_{k\in\mathbb{N}_i} \mathbb{B}_j^k, \forall i \in \mathbb{U} \tag{12}$$

$$|\mathbb{M}_i^\mu| \leq \beta_i^\mu, \forall i \in \mathbb{U} \tag{13}$$

$$\mathbb{M}_i^\mu \subseteq \cup_{j\in C}\mathbb{B}_j^\mu, \forall i \in \mathbb{U} \tag{14}$$

Objective (8) considers the throughput and satisfaction rate with the numerology among the users. Constraint (9) forces a user to connect l BSs simultaneously at most. Constraint (10) ensures that more than one user cannot access RB b_k^μ on BS j. Constraint (11) guarantees that the number of RBs allocated to user i cannot exceed the value α_i, and $\mathbb{M}_i$ is the set of RBs assigned to user i. Constraint (12) ensures that the set of RBs assigned to user i must be in the candidate numerologies of user i. Like (11), constraint (13) guarantees that the number of RBs with numerology μ allocated to user i cannot exceed the value β_i^μ, and $\mathbb{M}_i^\mu$ denotes the set of RBs with numerology μ assigned to user i. Constraint (14) ensures that the set of RBs with numerology μ allocated to user i must be in the set of numerology μ RBs.

Since $\rho_{i,j}^{\mu,k}$ also has the information of connectivity $x_{i,j}$, we can rewrite the equation (6) as

$$R_i^\mu = \sum_{j\in\mathbb{C}} \sum_{k\in\mathbb{B}_j^\mu} \rho_{i,j}^{\mu,k} r_{i,j}^{\mu,k}, \tag{15}$$

where $x_{i,j}$ is removed from equation (6). Hence, our objective function can also rewrite as

$$\max_{(\rho)} \sum_{i\in U} \sum_{\mu\in\mathcal{N}} S(R_i^\mu, \eta_i^\mu, \delta_i^\mu). \tag{16}$$

The problem is then reduced to a resource allocation problem because determining the $\rho_{i,j}^{\mu,k}$ also decides the connectivity x_{ij}. So, we can focus on resource allocation by assigning the RBs to users rather than jointly considering the users' association and resource allocation. However, the objective function needs to decide the binary variable ρ and maximize the logarithm of users' data rate, which is non-linear. The objective function (16) is an integer non-linear programming (INLP) problem that is NP-hard. We propose a heuristic algorithm to solve the problem in the next section.

III. MC Resource Allocation (MCRA) Algorithm

This section describes the proposed two-step MCRA algorithm as the solution to the utility mentioned above maximization problem considering throughput and satisfaction rate. Because the resource allocation problem between users and RBs can be seen as a bipartite graph matching. So, the first step of MCRA is the matching-based resource allocation algorithm, which aims to maximize the network throughput. After obtaining the matching game result, an iterative greedy algorithm is used to improve the utility function.

A. Matching Game (MG) Algorithm

In the first step, we use the framework of college admissions games, also known as the many-to-one matching game. A college admissions game is used to maximize the network throughput. There are three components in this game: 1) the set of users $\mathbb{U}$ acting as colleges has the fixed admission quota corresponding to constraint (11), 2) the set of RBs $\mathbb{B}$ acting as students, and 3) preference relations for the users and RBs allowing them to build preferences over one another.

Each user (RB) defines its preferred RBs (users) in a matching game. Two sets of players in a matching game are users $i \in \mathbb{U}$ and RBs $b \in \mathbb{B}$. Let the preference relation be denoted as $\succ$. The expression $b \succ_i b'$ implies that user i prefers

Algorithm 1 Many-to-One Matching Game Algorithm

Input:
Data rate table $D[1..N][1..N_b]$ and RSRP table $S[1..N][1..N_b]$

Output:
Matching table $T[1..N_b]$

```
1: Initial: A[1..N] = ∅, M[1..N] = ∅, T[1..N_b] = −1
2: Build user preference list p_i and RB preference list p_b
   based on the tables D with (17) and table S with (18),
   respectively.
3: while any T[b] = -1 and p_b is not empty do
4:     Store each unmatched RB in list A[i] if the first
       element in p_b is user i and then remove the user
       i from p_b.
5:     for i = 1 to N do
6:         User i matches most α_i preferred RBs in A[i] ∪
           M[i] by p_i under constraints (9), (11), (12), (13).
7:         Recalculate the data rate of each RB by (5).
8:         Sets T[b] = i if user i matches RB b; otherwise,
           T[b] = -1, ∀b ∈ A[i] ∪ M[i].
9:         Updates the matched RBs in list M[i].
10:    end for
11: end while
```

to take RB b rather than RB $b^{'}$. A similar expression $i \succ_b i^{'}$ implies that RB b prefers to serve the user i rather than user $i^{'}$. Let p_i be the preference list of user i, and p_b be the preference list of RB b. In the preference list p_i of user i, the preference relation can be expressed as

$$b \succ_i b^{'} \Longleftrightarrow r_{ib} \ > \ r_{ib'}, \tag{17}$$

where r_{ib} ($r_{ib'}$) is the data rate of user i with RB b ($b^{'}$). Equation (17) indicates that user i prefers RB b to $b^{'}$ as r_{ib} is greater than $r_{ib'}$. Since MG aims to maximize the total throughput among users, we use the data rate of RB as a preference relation for users. Similar to (17), the RBs also determine their preference for users. In the preference list p_b of RB b, the preference relation can be expressed as

$$i \succ_b i^{'} \Longleftrightarrow rsrp_{ib} \ > \ rsrp_{i'b}, \tag{18}$$

where $rsrp_{ib}$ ($rsrp_{i'b}$) is the RSRP of user i ($i^{'}$) with RB b. Equation (18) indicates that RB b prefers user i to $i^{'}$ as the $rsrp_{ib}$ is greater than $rsrp_{i'b}$. This is because we tend to associate the users close to the BSs.

The pseudo-code of our matching game algorithm is given in Algorithm 1. The input tables $D[1..N][1..N_b]$ and $S[1..N][1..N_b]$ are the input of data rate and RSRP between users and RBs, respectively. We define the list $A[i]$ to store the RBs b that want to match user i and the list $M[i]$ to keep the RBs that already match user i. The output of the matching table $T[1..N_b]$ records each of the N_b RBs serving one of the N users. Each table entry $T[1..N_b]$ is initially set to -1 as empty. In line 2, we sort the preference list p_i for each user i and p_b for each RB b based on the preferences (17) and (18), respectively. The time complexity of sorting the preference lists is $O(NN_blog(N_b) + N_bNlog(N))$ = $O(NN_blog(NN_b))$.

In line 4, if there is any unmatched RB b with the preferred users in p_b, RB b will try to match the most preferred user i in p_b. This time complexity is $O(N_b)$ since there are N_b RBs. We save the RB b in list $A[i]$, and the user i will be removed from p_b to prevent RB b from matching user i again. In line 6, the users will select the most α_i (constraint (11)) preferred RBs by comparing their preferences in the list of $A[i] \cup M[i]$ with the order of p_i under constraints (9), (12), and (13). Note that user i may prefer the RBs in $A[i]$ to those in $M[i]$ because strong RSRP (RB to a user) does not guarantee a high data rate due to the interference of the RBs used in other RBs with the same frequency. The data rate must be recalculated by equation (5) according to the current matching result (line 7). The time complexity of line 6 is $O(N_b)$ since there are at most N_b RBs to be compared by each user. In line 8, we update the matching table $T[b]$ to i if RB b matches user i; otherwise, we set it to -1. In line 9, we update the list $M[i]$ for each user i with the newly matched RBs. The algorithm will terminate when all RBs are matched, or the preference list of unmatched RB is empty.

Since the RBs will be matched in constant times, our time complexity in lines 3-9 is $O(N_b + N_b)$ = $O(N_b)$. The total MG time complexity is $O(NN_blog(NN_b) + N_b)$ = $O(NN_blog(NN_b))$. Our MG algorithm only considers the total throughput. However, objective (16) wants to maximize the utility considering the throughput and satisfaction rate. So, in the following subsection, we will propose the Iterative Greedy algorithm to maximize the network utility based on the result of the MG algorithm.

B. Iterative Greedy (IG) Algorithm

Here, we introduce the *Iterative Greedy* (IG) algorithm to enhance (16) by considering the satisfaction rate. After executing the MG, each RB will belong to one of the users and obtain a network utility NU by equation (16). The main idea of the IG algorithm is to iteratively explore different resource allocations by moving each resource block (RB) to different users and evaluating the new network utility after the move. The IG considers all possible RB-user assignments and selects the assignment with the largest network utility improvement, subject to the given constraints. Assume the RB i moves to user j can produce the largest network utility NU' for all RBs in $\mathbb{B}$ and users in $\mathbb{U}$, and if $NU^{'}$ - $NU >$ threshold ϵ, we will move the RB i to the user j. The IG algorithm will repeat the above procedure until the utility improvement is smaller than the threshold ϵ. The IG algorithm is described as follows.

Let $U(b, k)$ denote the network utility if moving an RB b from its original assigned user to a new user k. Note that the RB b and user k must satisfy the constraints (9), (11), (12), and (13). Let (b^*, k^*) = $\arg\max_{b \in \mathbb{B}, k \in \mathbb{U}} U(b, k)$. Then b^* is the RB index, and k^* is the user index, which can produce the largest network utility if we move RB b^* to user k^*. The time complexity to find b^* and k^* is $O(NN_b)$ because we need to check N_b RBs with N users. Assume RB b^* originally belonged to user j. Let $NU^{'}$ = $U(b^*, k^*)$. In each round, if

$NU^{'}$-$NU > \epsilon$, we move the RB b^* from user j to k^*. If the numerology of RB b^* is μ, we update the data rate of R_j^{μ} and $R_{k^*}^{\mu}$ corresponding to equation (6) and set $NU = NU^{'}$. The above steps will repeat until the utility enhancement is less than the threshold ϵ. To ensure the efficiency of MCRA, we limit the maximum number of rounds to N_b. Thus, the time complexity of MCRA is $O(N{N_b}^2)$.

C. Speed-Up MC Resource Allocation (SMCRA) Algorithm

In this subsection, we introduce the SMCRA algorithm to reduce the time complexity of IG in the MCRA algorithm. Let RB_{μ} denote the RB with numerology μ and UT_i^{μ} denote the utility of user i with numerology μ. Let UT_{min} = $\min_{\mu \in \mathcal{N}, i \in \mathbb{U}} UT_i^{\mu}$ and u_{min} be the user with the UT_{min}. The main idea of the algorithm is that if the numerology of the UT_{min} is μ, we take an RB_{μ} with the least data rate from some users and allocate the RB_{μ} to u_{min}. In this way, we can enhance the network utility of the system by improving the UT_{min} for fairness or satisfaction rate.

To efficiently retrieve the resource blocks with the lowest data rate for numerology $\hat{\mu}$ in each user, we construct a min-heap $\mathcal{H}_i^{\mu}$ for each user i with numerology μ. The root of $\mathcal{H}_i^{\mu}$ is the least data rate of RB_{μ} in user i, and the corresponding RB is denoted as rb_i^{μ}. We also maintain a min-heap $\mathcal{U}$ for the utility of each user i with its numerology μ. The root of $\mathcal{U}$ is UT_{min} and the corresponding user and numerology of UT_{min} are u_{min} and $\hat{\mu}$, respectively. Our goal is to remove some RBs from some users to u_{min} to improve the fairness or satisfaction rate. However, we will also decrease the users' data rate sacrificing their RBs with the numerology $\hat{\mu}$. To reduce the decreased data rate of the sacrificed user, we only consider sacrificing the RB with the minimum data rate $rb_i^{\hat{\mu}}$, $\forall i \in \mathbb{U}$.

Let $Utility(rb_i^{\hat{\mu}}, u_{min})$ denote the network utility if moves $rb_i^{\hat{\mu}}$ from user i to user u_{min}. Note that users u_{min} and i must satisfy the constraints (9), (11), (12), and (13). Let $rb_{i^*}^{\hat{\mu}} = \arg\max_{i \in \mathbb{U}} Utility(rb_i^{\hat{\mu}}, u_{min})$. Let $NU^* = Utility(rb_{i^*}^{\hat{\mu}}, u_{min})$. If NU^* - $NU > 0$, we extract the $rb_{i^*}^{\hat{\mu}}$ from $\mathcal{H}_{i^*}^{\hat{\mu}}$ and insert it to $\mathcal{H}_{u_{min}}^{\hat{\mu}}$ for the movement of RB $rb_{i^*}^{\hat{\mu}}$ from user i^* to user u_{min}. Also, we will increase the data rate of $R_{u_{min}}^{\hat{\mu}}$ and decrease the data rate of $R_{i^*}^{\hat{\mu}}$. Then, the utility of $UT_{u_{min}}^{\hat{\mu}}$ and $UT_{i^*}^{\hat{\mu}}$ in $\mathcal{U}$ will update accordingly. Note that if NU^* - $NU \le 0$, we will extract the UT_{min} from $\mathcal{U}$ because there is no RB $rb_i^{\hat{\mu}}$, $\forall i \in \mathbb{U}$, to enhance the network utility by increasing the data rate of $R_{u_{min}}^{\hat{\mu}}$. The iteration will stop when the $\mathcal{U}$ is empty.

The pseudo-code of the SMCRA algorithm is given in Algorithm 2. Line 1 initiates the algorithm by constructing min-heaps $\mathcal{U}$, and $\mathcal{H}_i^{\mu}$ for each user i and its numerology μ. Line 3 calculates the network utility NU by equation (7). Line 4 initialize NU^* = 0 and $rb_{i^*}^{\hat{\mu}}$ is set to -1 as a sentinel value in line 5. The user with the minimum utility, u_{min}, along with its associated numerology, $\hat{\mu}$, is obtained from the root of $\mathcal{U}$ in line 6. In line 9, we get the least data rate of resource block $rb_i^{\hat{\mu}}$ with the numerology $\hat{\mu}$ in user i. Lines 7-16 encompass a loop that iterates through each user i in the set $\mathbb{U}$. Within this loop, we evaluate whether a specific RB assignment is feasible and whether it enhances the overall utility.

Algorithm 2 Speed-Up MC Resource Allocation Algorithm

Input:
Data rate table $D[1..N][1..N_b]$

1: Construct min-heap of users' application utility $\mathcal{U}$ and min-heap of RBs based on data rate for each users application $\{\{\mathcal{H}_i^{\mu}\}_{i=1}^{N}\}_{\mu=0}^{2}$
2: **while** $\mathcal{U} \neq \emptyset$ **do**
3: $\quad NU \leftarrow \sum_{i=1}^{N} \sum_{\mu=0}^{2} S(R_i^{\mu}, \eta_i^{\mu}, \delta_i^{\mu})$
4: $\quad NU^* \leftarrow 0$
5: $\quad rb_{i^*}^{\hat{\mu}} \leftarrow -1$
6: $\quad (,u_{min}, \hat{\mu}) \leftarrow \mathcal{U}$.getMin()
7: $\quad$ **for** $i \in \mathbb{U}$ **do**
8: $\quad\quad$ **if** $\mathcal{H}_i^{\hat{\mu}} = \emptyset$ **then continue**
9: $\quad\quad (,rb_i^{\hat{\mu}}) \leftarrow \mathcal{H}_i^{\hat{\mu}}$.getMin()
10: $\quad\quad$ **if** do $Utility(rb_i^{\hat{\mu}}, u_{min})$ violates the constraints (9), (11), (12) and (13) **then continue**
11: $\quad\quad NU^{'} \leftarrow Utility(rb_i^{\hat{\mu}}, u_{min})$
12: $\quad\quad$ **if** $NU^{'} > NU^*$ **then**
13: $\quad\quad\quad NU^* \leftarrow NU^{'}$
14: $\quad\quad\quad rb_{i^*}^{\hat{\mu}} \leftarrow rb_i^{\hat{\mu}}$
15: $\quad\quad$ **end if**
16: $\quad$ **end for**
17: $\quad$ **if** $NU^* - NU > 0$ **then**
18: $\quad\quad \mathcal{H}_{i^*}^{\hat{\mu}}$.extract()
19: $\quad\quad \mathcal{H}_{u_{min}}^{\hat{\mu}}$.insert($D[u_{min}][rb_{i^*}^{\hat{\mu}}]$, $rb_{i^*}^{\hat{\mu}}$)
20: $\quad\quad$ \\ update $R_{u_{min}}^{\hat{\mu}}$, reorder the new $UT_{u_{min}}^{\hat{\mu}}$ in $\mathcal{U}$
21: $\quad\quad R_{u_{min}}^{\hat{\mu}} \leftarrow (R_{u_{min}}^{\hat{\mu}} + D[u_{min}][rb_{i^*}^{\hat{\mu}}])$
22: $\quad\quad \mathcal{U}$.extract()
23: $\quad\quad \mathcal{U}$.insert($UT_{u_{min}}^{\hat{\mu}}$, u_{min}, $\hat{\mu}$)
24: $\quad\quad$ \\ update $R_{i^*}^{\hat{\mu}}$ and reorder the new $UT_{i^*}^{\hat{\mu}}$ in $\mathcal{U}$
25: $\quad\quad R_{i^*}^{\hat{\mu}} \leftarrow (R_{i^*}^{\hat{\mu}} - D[i^*][rb_{i^*}^{\hat{\mu}}])$
26: $\quad\quad \mathcal{U}$.decreaseKey($UT_{i^*}^{\hat{\mu}}$, i^*, $\hat{\mu}$)
27: $\quad$ **else**
28: $\quad\quad \mathcal{U}$.extract()
29: $\quad$ **end if**
30: **end while**

Line 18 removes the resource block with numerology $\hat{\mu}$, which has the least data rate, from user i^*. This resource block is assigned to user u_{min} with line 19. Line 21 updates the total data rate of user u_{min} with numerology $\hat{\mu}$. Line 22 removes the root from heap $\mathcal{U}$ corresponding to the outdated utility value of u_{min} with numerology $\hat{\mu}$. A newly updated utility of user u_{min} with numerology $\hat{\mu}$ is inserted into the heap $\mathcal{U}$ in line 23. Line 25 updates the total data rate of user i^* with numerology $\hat{\mu}$. Finally, line 26 decreases the utility of user i^* with numerology $\hat{\mu}$ in the heap $\mathcal{U}$. Conversely, if the new network utility is smaller than the original utility, compared in line 12, we remove the root of heap $\mathcal{U}$ in line 28. The algorithm iteratively executes the while loops until $\mathcal{U}$ becomes empty. The number of iterations is close to $O(N_b)$. So, the time complexity of SMCRA is $O(NN_b log(NN_b) + N_b(N + logN))$

$= O(NN_b log(NN_b))$.

IV. Simulation Results

A. Simulation Setting

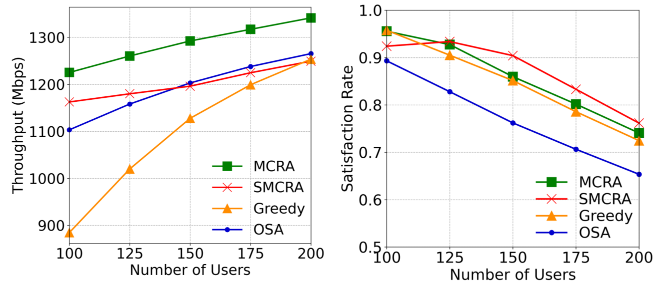
In our simulations, the map area is 1000 m $\times$ 1000 m. A macro BS is located at the center of the map, and six micro BSs are located in three clusters, as shown in Fig. 1. Two-thirds of the users are distributed in the three clusters, and the rest are randomly distributed in the macro cell coverage. Table II shows the simulation parameters. Our experiments compare the proposed algorithms MCRA and SMCRA to the candidate algorithm OSA [17] and the greedy algorithm. The OSA allocates the RBs to users by greedily picking up the best association strategy with users and RBs under the network constraints in each iteration. Hence, the time complexity of OSA is $O(N{N_b}^2)$, the same as the MCRA. The greedy algorithm assigns the RBs to users with the data rate demand ratio $(r_{i,j}^{\mu,k}/R_i^{\mu})$ from high to low. If a user is satisfied, it will not assign more RBs to the user. So, it aims to satisfy the users as more as possible. The time complexity of the greedy algorithm is $O(NN_b logNN_b)$, the same as the SMCRA.

TABLE II: Simulation Parameters

Parameter	Value
Number of Macro BS	1
Macro BS transmit power (P_0)	43 dBm
Micro BS transmit power	30 dBm
Macro BS transmit frequency ($freq_0$)	2 GHz
Micro BS transmit frequency	6 GHz
Noise power (N_0)	-174 dBm/Hz
Macro BS bandwidth	40 MHz
Micro BS bandwidth	20 MHz
Scheduling time	1 msec
Maximum connectivity l	3
Threshold ϵ	10^{-4}
Maximum number of RBs for user i (α_i)	[10, 20]
Maximum number of RBs for user i with numerology μ (β_i^{μ})	[10, 12]
Demand of user user i with numerology μ (δ_i^{μ})	[3, 8] Mbps
Number of RBs for mix numerology in 1 sub-frame duration in Macro BS	1×72 RBs for μ = 0, 2×36 RBs for μ = 1, 4×18 RBs for μ = 2
Number of RBs for mix numerology in 1 sub-frame duration in Micro BS	1×36 RBs for μ = 0, 2×18 RBs for μ = 1, 4×9 RBs for μ = 2

B. Varying Number of UEs

1) Total throughput: The total throughput with various numbers of users is shown in Fig. 2(a). The total throughput is calculated as $\sum_{i=1}^{N}\sum_{\mu\in\mathbb{N}_i} R_i^{\mu}$. In Fig. 2(a), as the number of UEs grows, the total throughput becomes higher because more users are close to BSs, which gives rise to more data rates in RBs. Besides, our MCRA outperforms the OSA with an increasing number of users. The performance of SMCRA is lower than MCRA since SMCRA has fewer choices than MCRA when moving RBs to improve the utility. However, it has lower time complexity achieving comparable performance to OSA. The greedy algorithm achieves the lowest throughput since it aims to satisfy the request of users. Thus, more RBs may be assigned to unsatisfied users far from the BSs, causing a drop in throughput when the number of users is small.



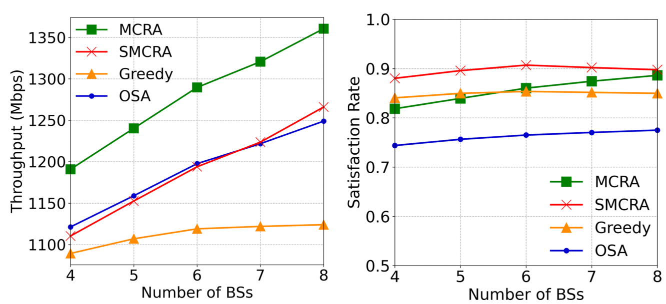
(a) (b)

Fig. 2. Network performance versus the number of users in our network scenario (a) total throughput (b) satisfaction rate

2) Satisfaction rate: Fig. 2(b) shows the satisfaction rate over the different numbers of users. The satisfaction rate among users can be calculated by $\sum_{i=1}^{N}\sum_{\mu\in\mathbb{N}_i} e_i^{\mu}/n$, where e_i^{μ} is a binary variable that indicates the request from user i with numerology μ is satisfied or not. $n = \sum_{i\in\mathbb{U}} |\mathbb{N}_i|$ indicates the total number of requests with different numerologies in the system. In Fig. 2(b), our MCRA and SMCRA outperform the OSA with a 10% satisfaction rate as the number of users grows. The SMCRA achieves the highest satisfaction rate since it enhances utility by first serving the user with minimum utility. So, the SMCRA serves the unsatisfied user first. On the other hand, the MCRA improves the utility by picking up the best result among all RBs movements. This may induce MCRA to maintain higher throughput but a lower user satisfaction rate. Although the satisfaction rate of the greedy algorithm is comparable to our MCRA, it has the lowest network throughput since more RBs are assigned to unsatisfied users far from BSs.

C. Varying Number of BSs

1) Total throughput: In the subsequent simulations, we maintain a fixed user count of 150 while varying the number of base stations from 4 to 8. We observe the resulting changes in throughput and satisfaction rate. Note that the BSs are evenly distributed among three clusters whenever possible. Otherwise, any remaining BSs will be randomly assigned to any clusters.



(a) (b)

Fig. 3. Network performance versus the number of BSs in our network scenario (a) total throughput (b) satisfaction rate

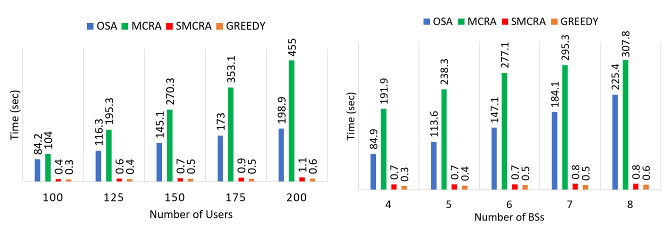
The total throughput with various BSs is evaluated in Fig 3(a). The throughput increases as BSs grow because more RBs can serve users. Besides, the MCRA outperforms the

OSA in our network scenario, and the SMCRA uses less time complexity to achieve a close result with OSA. The greedy algorithm achieves the lowest throughput since it assigns more RBs than other algorithms to users far from BSs.

2) Satisfaction rate: Fig. 3(b) shows the satisfaction rate over different numbers of BSs. As the number of BSs grows, the satisfaction rate increases since more RBs can be allocated to users. Our MCRA and SMCRA have better satisfaction rates than OSA. Since the OSA only greedily finds the best association for the objective function in each iteration, it may not have a better association for throughput, which causes a lower satisfaction rate. Although the satisfaction rate of the greedy algorithm is comparable to MCRA, it has the lowest network throughput.

D. Execution Time

In this subsection, we show the execution time of our algorithms and baselines in Fig. 4(a) and Fig. 4(b). Our experiments run on a desktop with Intel Core i5-9400 and RAM with 16 GB. In Fig. 4(a), we run the algorithms with one macro station and six micro stations, and the $N_b = 864$. We show the execution time over different numbers of users. When the number of users grows, they spend more time achieving their goals. The MCRA and OSA spend more time than SMCRA and Greedy since the time complexity of MCRA and OSA is $O(N{N_b}^2)$, which is larger than the SMCRA and Greedy with the time complexity $O(NN_b log NN_b)$. The SMCRA and Greedy algorithms spend less than one second to achieve their goals. In Fig. 4(b), we run the algorithms with 150 users and show the execution time over different numbers of BSs. The simulation result is similar to Fig. 4(a).



(a) (b)

Fig. 4. Execution time versus (a) the number of users (b) the number of BSs.

V. Conclusion

This paper investigates the resource allocation problem in two-tier HetNets with MC and mixed-numerology. Our objective focuses on the network throughput and satisfaction rate of users. We propose an MCRA algorithm to solve the resource allocation problem. Besides, an SMCRA algorithm is proposed to reduce the time complexity of the MCRA algorithm with comparable performance. Simulation results show that the performance of the MCRA algorithm is better than the OSA and greedy algorithms. In addition, the proposed SMCRA algorithm has comparable performance on network throughput compared to the baselines but has a better satisfaction rate and lower time complexity than the baselines.

References

[1] P. Schulz, A. Traßl, A. N. Barreto, and G. Fettweis, "Efficient and reliable wireless communications via multi-connectivity using rateless codes in single- and multi-user scenarios," *IEEE Transactions on Wireless Communications*, vol. 20, no. 9, pp. 5714–5729, 2021.

[2] M.-T. Suer, C. Thein, H. Tchouankem, and L. Wolf, "Adaptive multi-connectivity scheduling for reliable low-latency communication in 802.11be," in *IEEE Wireless Communications and Networking Conference*, pp. 102–107, Austin, TX, USA, 2022.

[3] R. Liu, G. Yu, and G. Y. Li, "User association for ultra-dense mmwave networks with multi-connectivity: A multi-label classification approach," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1579–1582, 2019.

[4] X. Ba and Y. Wang, "Load-Aware Cell Select Scheme for Multi-Connectivity in Intra-Frequency 5G Ultra Dense Network," *IEEE Communications Letters*, vol. 23, no. 2, pp. 354–357, 2019.

[5] L.-H. Shen, C.-Y. Su, and K.-T. Feng, "CoMP Enhanced Subcarrier and Power Allocation for Multi-Numerology Based 5G-NR Networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 5, pp. 5460–5476, 2022.

[6] P. K. Taksande, P. Chaporkar, P. Jha, and A. Karandikar, "Proportional fairness through dual connectivity in heterogeneous networks," in *IEEE Wireless Communications and Networking Conference*, Seoul, South Korea, 2020.

[7] D. Csercsik and E. Jorswieck, "Preallocation-based combinatorial auction for efficient fair channel assignments in multi-connectivity networks," *IEEE Transactions on Wireless Communications*, 2023.

[8] P. K. Korrai, E. Lagunas, A. Bandi, S. K. Sharma, and S. Chatzinotas, "Joint Power and Resource Block Allocation for Mixed-Numerology-Based 5G Downlink Under Imperfect CSI," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1583–1601, 2020.

[9] X. Cai, A. Chen, L. Chen, and Z. Tang, "Joint optimal multi-connectivity enabled user association and power allocation in mmwave networks," in *IEEE Wireless Communications and Networking Conference*, Nanjing, China, 2021.

[10] T.-Y. Chen, C.-H. Wang, J.-P. Sheu, G.-S. Lee, and D.-N. Yang, "Resource Allocation for the 4G and 5G Dual-Connectivity Network with NOMA and NR," in *IEEE International Conference on Communications*, pp. 3784–3789, Seoul, South Korea, 2022.

[11] K. Alexandris, C.-Y. Chang, N. Nikaein, and T. Spyropoulos, "Multi-connectivity resource allocation with limited backhaul capacity in evolved lte," in *IEEE Wireless Communications and Networking Conference*, Barcelona, Spain, 2018.

[12] E. J. Khatib, D. A. Wassie, G. Berardinelli, I. Rodriguez, and P. Mogensen, "Multi-Connectivity for Ultra-Reliable Communication in Industrial Scenarios," in *IEEE 89th Vehicular Technology Conference*, Kuala Lumpur, Malaysia, 2019.

[13] 3GPP, "5G; NR; Physical channels and modulation," Technical Specification (TS) 38.121, 3GPP, 10 2018. Version 15.3.0.

[14] L. You, Q. Liao, N. Pappas, and D. Yuan, "Resource Optimization With Flexible Numerology and Frame Structure for Heterogeneous Services," *IEEE Communications Letters*, vol. 22, no. 12, pp. 2579–2582, 2018.

[15] L. Marijanovic, S. Schwarz, and M. Rupp, "Multi-user resource allocation for low latency communications based on mixed numerology," in *IEEE 90th Vehicular Technology Conference*, Honolulu, Hawaii, USA, 2019.

[16] X. Zhang, L. Zhang, P. Xiao, D. Ma, J. Wei, and Y. Xin, "Mixed Numerologies Interference Analysis and Inter-Numerology Interference Cancellation for Windowed OFDM Systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7047–7061, 2018.

[17] K. Alexandris, C.-Y. Chang, and N. Nikaein, "Utility-Based Opportunistic Scheduling Under Multi-Connectivity With Limited Backhaul Capacity," *IEEE Networking Letters*, vol. 1, no. 2, pp. 80–83, 2019.

[18] 3GPP, "NR; User Equipment (UE) radio access capabilities," Technical Specification (TS) 38.306, 3GPP, 07 2020. Version 16.1.0.

[19] S. Sun, T. S. Rappaport, S. Rangan, T. A. Thomas, A. Ghosh, *et al.*, "Propagation Path Loss Models for 5G Urban Micro- and Macro-Cellular Scenarios," in *IEEE 83rd Vehicular Technology Conference*, Nanjing, China, 2016.