A QoS-based joint user association and resource allocation scheme in ultra-dense networks

Zhiwei Si*, Gang Chuai, Weidong Gao, Jinxi Zhang, Xiangyu Chen and Kaisa Zhang

Abstract

Ultra-dense networks (UDNs) have become an important architecture for the fifth generation (5G) networks. A large number of small base stations (SBSs) are deployed to provide high-speed and seamless connections for users in the network. However, the advantage of increasing the system capacity brought by the dense distribution of SBSs comes at the cost of severe inter-cell interference. Although the user-centric virtual cell method has been proposed to solve the interference problem, some challenges have been encountered in practical applications. For example, inter-cell interference still exists to a certain extent, and the cell load may be imbalance. Hence, under the virtual cell architecture, we propose a quality of service (QoS)-based joint user association and resource allocation scheme in UDNs. In order to mitigate the interference, balance cell load and improve the system throughput, a non-convex NP-hard problem is formulated. To effectively solve this problem, we decouple the formulated problem into three sub-problems: user association, physical resource block (PRB) allocation and power allocation. First, we consider the QoS requirements of user equipment (UE) and perform user association based on the PRB estimation method. Then, based on the overlapped virtual cells constructed, we propose a graph-based PRB allocation scheme for reducing virtual inter-cell interference. Moreover, we solve power allocation sub-problem by using the difference of concave (DC) programing method. The simulation results show that our proposed scheme is superior to other schemes in terms of user rates, cell load and system throughput.

Keywords: Ultra-dense networks (UDNs), Quality of service (QoS), Virtual cells, PRB estimation method, Graph-based, DC programing, System throughput

1 Introduction

The regular heterogeneous networks consisting of macro base stations and several small base stations (SBSs) are incapable of meeting of imminent tele-traffic demands. As one of the key technologies of 5G networks, UDNs can greatly increase user communication rates [1–4]. However, a large number of deployed SBSs bring serious inter-cell interference. To reduce interference, a user-centric virtual cell clustering method is proposed [5], in which each UE is cooperatively served by a few of SBSs. By using this method, the interference signal is converted into the useful signal. Thus,
the purpose of utilizing joint transmission technology [6] to reduce interference and enhance useful signals can be achieved, thereby greatly increasing the user rates. However, under the virtual cell architecture, inter-cell interference still exists within a certain range. Due to the irregular network topology and randomly distributed UEs, the cell load may be imbalance, which will cause the degradation of the system performance. Therefore, it is necessary to seek a joint user association and resource allocation scheme under the virtual cell architecture. It is worth noting that the process of user association is equivalent to the process of virtual cell clustering in this paper.

There have been quite a few studies on coordinated multiple points (CoMP) transmission, which can effectively suppress inter-cell interference, improve system capacity and meet QoS of edge users. In [7], authors propose a two-step joint clustering and scheduling scheme. A load-aware clustering scheme is designed and solved by game theory in the first step. Based on clustering results, a fair graph-coloring based inter-cell resource scheduling can be employed at the second step to maximize the resource utilization. [8, 9] present load-aware user-centric CoMP clustering algorithms which consider trade-off between load balance and spectrum efficiency. In addition, some other studies focus on user-centric virtual cell, where each UE has some base stations associated with it to avoid low signal to interference plus noise ratio (SINR) [10, 11] propose load-aware virtual cell schemes, which can meet the user QoS requirements. Firstly, the authors find the optimal radius of the virtual cell. Then they select the appropriate activation radius based on the cell load. In [12], the authors propose a beamforming problem to maximize the sum-rate in a virtual cell network [13] formulate an energy efficiency maximization problem in the user-centric virtual cell networks by optimizing beamforming vectors and access points cluster, and considering both rate requirements and power budgets.

In recent years, some significant efforts are devoted to reducing the computational complexity of user association and resource allocation. In [14], the authors propose a cluster-based energy-saving resource allocation method in UDNs. Firstly, a modified K-means method is proposed in the clustering stage, which can reduce complexity and inter-cluster interference. And then in resource allocation stage, a two-step PRB allocation algorithm is performed and an iterative power allocation scheme is designed and solved by a non-cooperative game [15] presents the joint optimization scheme of capacity maximization and power minimization for user association and resource allocation. To reduce the complexity, the clustering scheme is adopted in the solution of the optimization problem. As an effective method to reduce the computational complexity, graph theory has been widely used in resource allocation [16–19]. [20] formulates a graph-based joint user-centric overlapped clustering and resource allocation problem, in which traffic-load and the limited number of PRBs are considered. In order to improve system throughput and reduce computational complexity, [21–25] utilize the DC programming technology, which can be used to convert a non-convex function into the difference of two convex functions, and to obtain an approximately optimal solution through an iterative method. However, the joint user association and resource allocation scheme under the virtual cell architecture, considering the use of multiple PRBs for each user to reduce interference, cell load balance, guarantee QoS requirements and improve system throughput, has not been studied in previous work.
In this paper, under the virtual cell architecture, a QoS-based joint user association and resource allocation scheme is proposed to mitigate the interference, balance cell load and improve the system throughput. It is worth mentioning that the joint problem is a non-convex NP-hard problem. To solve this problem, we propose a three-step scheme. In the first step, we design a user association scheme, in which QoS requirements and cell load is considered. In the second step, we use the modified graph theory for PRB allocation to achieve the purpose of reducing inter-cell interference. In the third step, we apply the DC programming technology to power allocation, and perform an iterative update algorithm to obtain an approximate optimal solution for convergence. Finally, simulations indicate that our scheme is proved to be better than the other schemes.

The main contributions of this paper can be summarized as follows:

1. A novel framework is proposed for jointly designing QoS-based user association and resource allocation under the virtual cell architecture. This is the first attempt to consider the use of multiple PRBs for each user to reduce interference, balance cell load, guarantee QoS requirements of UE and improve system throughput.

2. To solve the non-convex NP-hard joint user association and resource allocation problem, three independent sub-problems are formulated.

   a. A new user-centric overlapped virtual cell clustering scheme is proposed to solve the user association problem of load imbalance with considering the QoS requirements.

   b. A low-complexity PRB allocation scheme is constructed by using modified graph-based method to further mitigate virtual inter-cell interference.

   c. The DC programming method is used to power allocation and an iterative update algorithm is performed to obtain an approximate optimal solution for convergence.

3. Simulations verify that the performance of our proposed scheme is better than that of other schemes in terms of user rates, cell load and system throughput.

The rest of the paper is organized as follows. We give the system model and formulate the joint optimization problem in Sect. 2. Section 3 describes our three-step joint scheme and discuss its implementation process. We analyze the complexity of our scheme in Section IV. Section V evaluates the proposed methods and compares the performance with the other method. Finally, conclusions are drawn in the last section.

2 System model and problem formulation

The two-tier heterogeneous UDNs scenario [8] is one of the main scenarios of 5G. We consider a downlink two-tier heterogeneous UDNs that consists of a macro base station (MBS), $N$ SBSs and $K$ UEs, as is shown Fig. 1. In order to harvest densification gains and avoid cross-tier interference, the control plane (C-plane) and user plane (U-plane) split architecture in [26] is adopted in our system, where a network controller unit (NCU) is installed in MBS. The NCU takes charge of virtual cell formation, wireless radio resources allocation, and mobility management. SBSs only undertake the task of data transmission. The SBS
The index set is $\mathcal{N} = \{1, \ldots, N\}$, the UE index set is $\mathcal{K} = \{1, \ldots, K\}$. We make three assumptions: (1) SBSs and UEs are randomly distributed following Poisson point process (PPP) distribution within coverage area of MBS; (2) backhaul is ideal and channel state information (CSI) knowledge is perfect; (3) with zero-forcing (ZF) precoding, intra-cluster interference is completely eliminated.

The matrix $X$ is defined as a user association matrix, in which the elements are $x_{k,n}$, where $x_{k,n} \in \{0, 1\}$, $\forall k \in \mathcal{K}$, $\forall n \in \mathcal{N}$. $x_{k,n} = 0$ means that SBS $n$ is not accessed by UE $k$, and $x_{k,n} = 1$ means that SBS $n$ is accessed by UE $k$. $L$ represents the total number of PRBs per cell, and the PRB index set is $\mathcal{L} = \{1, \ldots, L\}$. The matrix $Y$ is defined as a PRB allocation matrix, in which the elements are $y_{k,l}$, where $y_{k,l} \in \{0, 1\}$, $\forall k \in \mathcal{K}$, $\forall l \in \mathcal{L}$. $y_{k,l} = 0$ indicates that PRB $l$ is not allocated to UE $k$, $y_{k,l} = 1$ indicates that PRB $l$ is allocated to UE $k$. $p_{n,l}$ represents the power allocated by SBS $n$ to PRB $l$, and $h_{k,n,l}$ represents the channel gain between SBS $n$ and UE $k$ on PRB $l$. $\sigma^2$ indicates the additive white Gaussian noise (AWGN).

Each PRB is assumed to have a bandwidth of $B$. The SINR of UE $k$ on PRB $l$ can be written as

$$\text{SINR}_{k,l} = \frac{\sum_{n \in \mathcal{C}_k} p_{n,l} h_{k,n,l}}{\sum_{k'=1, k' \neq k}^{K} \sum_{n' \in \mathcal{C}_{k'}} p_{n',l} h_{k,n',l} + \sigma^2}$$

(1)

where $\mathcal{C}_k$ represents the serving virtual cell cluster of UE $k$.

According to the Shannon formula, the rate of UE $k$ on PRB $l$ can be expressed as

$$R_{k,l} = y_{k,l} B \log_2 \left( 1 + \frac{\sum_{n=1}^{N} x_{k,n} p_{n,l} h_{k,n,l}}{\sum_{k'=1, k' \neq k}^{K} \sum_{n=1}^{N} x_{k',n} p_{n,l} h_{k',n,l} + \sigma^2} \right)$$

(2)
The rate of UE $k$ can be written as

$$r_k = \sum_{l=1}^{L} R_{k,l}$$  \hspace{1cm} (3)$$

The total system throughput is

$$R_{\text{total}} = \sum_{k=1}^{K} r_k$$  \hspace{1cm} (4)$$

The joint optimization problem is formulated as

$$\max_{x_{k,n},y_{k,l},p_{n,l}} R_{\text{total}}$$

s.t.

$$C_1 : 0 \leq \sum_{k=1}^{K} \sum_{l=1}^{L} x_{k,n} y_{k,l} p_{n,l} \leq P_{n}^{\text{max}}, \forall n$$

$$C_2 : 1 \leq \sum_{l=1}^{L} y_{k,l} \leq L, \forall k$$

$$C_3 : 0 \leq \sum_{k=1}^{K} \sum_{l=1}^{L} x_{k,n} y_{k,l} \leq L, \forall n$$

$$C_4 : 1 \leq \sum_{n=1}^{N} x_{k,n} \leq N_{\text{max}}, \forall k$$

$$C_5 : \sum_{l=1}^{L} y_{k,l} B \log_2 (1 + SINR_{k,l}) \geq R_{k,\text{min}}, \forall k$$

$$C_6 : x_{k,n} \in \{0, 1\}, \forall k, n$$

$$C_7 : y_{k,l} \in \{0, 1\}, \forall k, l$$

where $P_{n}^{\text{max}}$ is the maximum transmission power of SBS $n$. $N_{\text{max}}$ is the maximum number of SBSs in the virtual cell cluster. $R_{k,\text{min}}$ is the minimum required rate of UE $k$. $C_1$ indicates that the total allocated power of UEs cannot be larger than the maximum transmission power limit for each SBS. $C_2$ implies that at least one PRB is allocated to each UE. $C_3$ means that the number of PRBs used by users cannot be more than the maximum number of PRBs in each small cell (SC). $C_4$ expresses that the number of each UE accessed SBSs cannot be larger than the maximum number of SBSs in the virtual cell cluster. $C_5$ explains that the UE rate should be greater than the minimum required rate.

### 3 QoS-based joint user association and resource allocation scheme in UDNs

It can be observed that the objective function (5) is a non-convex NP-hard problem. Since the problem of user association and resource allocation are coupled, joint optimization problem (5) will result in extremely high computational complexity, especially in UDN scenarios. So, a novel framework is proposed for jointly designing QoS-based user association and resource allocation under the virtual cell architecture. The flowchart of the proposed user-centric overlapped clusters framework is described in Fig. 2.
We divide the problem (5) into three sub-problems: user association, PRB allocation and power allocation. The user association sub-problem is to determine the serving virtual cell cluster for each UE based on PRB estimation. A two-step graph-based approach is applied to the PRB allocation sub-problem. The DC programing method is used to power allocation sub-problem. Both of the PRB allocation and the power allocation sub-problem are to reduce virtual inter-cell interference and improve the system throughput. In this section, we will describe the three parts of the proposed scheme, and discuss implementation process for the scheme in practical systems.

3.1 User association

Our proposed user association scheme is divided into two stages: the estimation stage and the allocation stage. In the estimation stage, in order to estimate the number of PRBs required by each UE and cell load of each SC, we make two assumptions: 1) the transmit power of each SBS is equally allocated to all PRBs; 2) for each UE, the interference on each PRB comes from all SBSs except cooperative SBSs. The SINR of UE $k$ can be written as

![Flowchart of the proposed user-centric overlapped clusters framework](image)
where $h_{k,n}$ is the channel gain between SBS $n$ and UE $k$. The number of PRBs required by UE $k$ can be estimated by

$$N_k = \left\lceil \frac{R_{k,\text{min}}}{R_k} \right\rceil, \forall k \in \mathcal{K} \tag{7}$$

where $\lceil \cdot \rceil$ denotes the ceiling function that rounds up to the nearest integer. $R_k$ is the maximum achievable throughput from one PRB, it can be estimated as:

$$R_k = B \log_2 (1 + \text{SINR}_k) \tag{8}$$

The estimated load of SC $n$ is

$$\eta_n = \frac{\sum_{k \in \mathcal{N}_n} N_k}{L}, \forall n \in \mathcal{N} \tag{9}$$

where $\mathcal{N}_n$ is the set of UEs associated to SC $n$.

We define the Max Reference Signal Receiving Power (RSRP) rule: each UE associates SBSs with the first a few maximum RSRP values as its virtual cell cluster.

Here we introduce our proposed user association scheme in detail from the perspectives of UE and SBS. In the initial state, the candidate SBSs list for each UE consists of all SBSs. In the estimation stage, from the perspective of UE, $N_{\text{max}}$ SBSs are selected by each UE as its virtual cell cluster based on Max RSRP rule. From the perspective of SBS, we estimate the load and get the UEs list for each SC. In the allocation stage, from the perspective of SBS, the SCs is sorted in descending order according to the SC load. If the first SC is overloaded, we will sort the UEs of this SC in descending order based on SINR values. And the UE with the largest SINR is preferentially accepted and allocated PRBs. For other UEs we sequentially judge whether the remaining PRBs in this SC are sufficient or not. If the remaining PRBs are sufficient, the user association request is accepted, otherwise the request is rejected and UE is removed from the UE list of this SC. From the perspective of UE, the SBS of the first overloaded SC is removed from the candidate SBSs list for the UEs that are not associated to this SBS. Those UEs will reselect the cooperative SBSs according to the Max RSRP rule, and associate in turn until all the SCs are not overloaded. The specific algorithm is given in Algorithm 1.

As an example in Fig. 3, we assume that the maximum number of SBSs in the virtual cell cluster is 2, the number of PRBs for all SBSs is 16, and the number of PRBs required by each UE is 4. As shown in Fig. 3a, UE1, ..., UE6 select cooperative SBSs based on the Max RSRP rule respectively. In Fig. 3b, according to the Max SINR rule, UE1, ..., UE4 is preferentially accepted, and the UE5 is rejected by the SBS1. At this time, the cooperative SBSs cluster of UE1 and UE2 are {SBS1, SBS2}, and the cooperative SBSs cluster of UE3 and UE4 are {SBS3, SBS4}. SBS1 is removed in candidate SBSs list of UE5 and UE6. UE5 and UE6 re-associate the SBSs according to the Max
RSRP rule, thereby obtaining Fig. 3c. So far, the user-centric virtual cells are formed for each UE.

Algorithm 1: User association based on PRB estimation

Initialization:

1) the virtual cell cluster size $N_{\text{vcell}}$;
2) the number of initial PRBs used by the SC: $N_{\text{vcell}}^\text{init} = 0$, $\forall c \in C$;

while ($N \neq \emptyset$ )

for each $k \in K$ do

Find cooperative SBs cluster of UE $k$ based on Max RSRP rule.
Calculate $\text{SINR}_k$, $N_k$, and $R_k$.

end for

for each $\tau \in \tau$ do

Calculate SC load $\rho_\tau$, and obtain the UE list of the SC $\tau$.

end for

Find $\tau_\text{max} = \arg\max_{\tau \in \tau} \rho_\tau$.

if SC load $\rho_{\tau_\text{max}} > 1$

Sort UEs in descending order according to SINR value

$\lambda_k = \{k_{\text{max}}, \ldots, k_{\text{min}}\}$

where $K_{\text{max}}$ is the number of UE in the SC $\tau$.

for $k' = 1: K_{\text{max}}$ do

if $N_{k'}^\text{init} + \sum_{c \in C} N_{c} > L$

Obtain the UE list of SC $\tau_{k'}$

$\lambda_{k'} = \{k_{\text{max}}, \ldots, k_{\text{min}}\}$

end if

end for

for each $k \in C_{k'}$ do

Find cooperative SBs cluster of UE $k$ based on Max RSRP rule.

end for

for each $\tau \in \tau$ do

Calculate $N_{\text{vcell}}^\text{init}$, and obtain the UE list of SC $\tau$.

end for

Remove SC $\tau_j$ from UE candidate list

$N \leftarrow N \setminus \{k_j\}$

Remove already associated UEs from the list of UEs that are not associated

$K_{\text{max}} \leftarrow K_{\text{max}} / C_{k'}$

else

break

end if

end while
Now we can obtain the user association matrix $X$ based on Algorithm 1. And the clustering result of virtual cells is $\{C_1, C_2, ..., C_K\}$.

### 3.2 PRB allocation

After the UEs are associated to the SBSs, the next step is to solve the user-centric PRB allocation problem. Let $F = XX^T - N_{\text{max}}I_K$, then the matrix $F$ is the overlapped indicator matrix of the virtual cell cluster $C_i$ and $C_j$ (where $i, j \in \{1, ..., K\}$). The element $f_{ij}$ represents the number of overlapped SBSs in the virtual cell cluster $C_i$ and $C_j$.

The graph coloring method has been widely used in resource allocation for decades to reduce the computational complexity. In the graph theory, each UE represents a vertex and the edge between the two vertices indicates interference between the two UEs. It can be easily seen that the underlying PRB allocation sub-problem can be converted into a graph coloring problem, but this graph coloring problem cannot be directly solved by existing graph-based method. In [20], each UE is only allocated to one PRB, but in our proposed scheme each UE can be allocated to multiple PRBs. Thus, the previous graph coloring method cannot be used in our proposed scheme directly. In order to use the graph coloring theory, we extend the graph coloring scheme in [20] for PRB allocation. The specific process is as follows.

#### 3.2.1 Graph construction

The graph is constructed as $G = (V, E)$, where $V$ is the set of vertices $\{C_1, C_2, ..., C_K\}$ corresponding to the virtual cell clusters $\{C_1, C_2, ..., C_K\}$, and $E$ is the edge connecting any two vertices. $d(C_k)$ represents the degree of the vertex $C_k$, which is equal to the number of all edges associated with the vertex $C_k$. We construct edges based on matrix $F$. If $f_{ij} > 1$, an edge between the cluster $C_i$ and the cluster $C_j$ is formed.

We can get the number of PRBs used by each SC $N_{\text{used}}^n$, $\forall n \in \mathcal{N}$ by Algorithm 1. We assume that the total transmitted power of each SBS is allocated equally to the PRBs used by each SC, ie. $P_{\text{max}}^n / N_{\text{used}}^n$, $\forall n \in \mathcal{N}$. The interference of UE $k$ on PRB $l$ can be written as

$$I_{k,l} = \sum_{k'=1}^{K} \sum_{k' \neq k} \sum_{n' \in C_{k'}} y_{k',l} \frac{P_{\text{max}}^{n'}}{N_{\text{used}}^{n'}} h_{k,n',l}$$  \hspace{1cm} (10)
3.2.2 Graph coloring

The vertex with the maximum degree should be colored preferentially. The reasons are as follows: (1) the higher the vertex degree is, the larger the number of vertices adjacent to the vertex are. This means that the number of clusters overlapped SBSs is larger; (2) the orthogonal PRBs need to be allocated among the clusters with the same serving SBS; (3) since the number of PRBs are insufficient, PRBs need to be reused. If the vertex with the maximum degree is colored, we can choose more PRBs to other vertices with lower degrees. To solve this problem conveniently, our method is divided into two stages: the sorting stage and coloring stage.

In the sorting stage, we search for the vertex $C^*_1$ with the highest degree in the $V$ firstly. We define the set $D_{C_k}$ as the set of the vertices adjacent to the vertex $C_k$, where $C_k \in V, \forall k \in K$. Then, we find the set $D_{C^*_1}$ based on matrix $F$. The degree of all vertices in $D_{C^*_1}$ is reduced by one. This operation is performed on the remaining vertices in the $V$ to obtain vertices $C^*_2, C^*_3, ..., $ until all vertices are sequentially placed into empty set $S$.

$$S = \{C^*_1, ..., C^*_K\}$$  \hspace{2cm} (11)

In the coloring stage, we sequentially label the PRBs required by each vertex in the set $S$, and get the set $L_S$

$$L_S=\{L_C^{C^*_1}_1, ..., L_N^{C^*_1}_{C^*_1}, L_C^{C^*_2}_1, ..., L_N^{C^*_2}_{C^*_2}, ..., L_C^{C^*_K}_1, ..., L_N^{C^*_K}_{C^*_K}\}$$  \hspace{2cm} (12)

The corresponding natural number set is $L_S = \{1, ..., \sum_{k=1}^K N^*_k\}$. The subset $\{L_C^{C^*_1}_1, ..., L_N^{C^*_K}_{C^*_K}\}$ of $L_S$ represents the set of $N^*_k$ PRBs required by vertex $C^*_k$. Each PRB is assumed have a specific color in the $G$.

We assign the colors to the elements in the set $L_S$ in turn. For the first $L$ elements of the set $L_S$, we randomly assign $L$ different colors to them and update matrix $Y$. For other elements of the set $L_S$, we firstly find the vertex $C^*_k$ corresponding to the element. Then, we find the set $D_{C^*_k}$ based on matrix $F$. Thirdly, we choose the PRB with the minimum interference that is not used by the vertices of the set $D_{C^*_k}$, and assign the color to vertex $C^*_k$. Finally, we update the interference on this PRB and matrix $Y$. 

Algorithm 2: Graph-Based User-Centric PRB Allocation.

for each $k \in K$ do
    $d(C_k)$. 
end for

for each $k \in K$ do
    Find $C_k^* = \arg \max_{C_k} d(C_k)$ Let $S = \emptyset$.
    $V = V \setminus \{C_k^*\}$, $S = S \cup \{C_k^*\}$
    Find $D_{C_k^*}$.
    for each $C_j \in D_{C_k^*}$ do
      $d(C_j) = d(C_j) - 1$.
    end for
end for

Get $S = \{C_1^*, \ldots, C_k^*\}$.

Get the set $L_0$
$$L_0 = \left[ E_1^1, \ldots, E_N^1 \right]$$

Get the set $L_1$
$$L_1 = \left[ \frac{1}{N_1^1}, \ldots, \frac{1}{N_k^1} \right]$$

for $n = 1:L$ do
  Assign color $n$ to the $n$-th element of $L_0$.
end for

for $n = L + 1: \left| V \right|$ do
  Let $C_{n,\text{new}}$ denote the vertex corresponding to the number $n$.
  Find $C_{n,\text{new}}$ and $D_{C_{n,\text{new}}}$.
  for $i = 1:L$ do
    if PRB $i$ is not used by the set $D_{C_{n,\text{new}}}$
      Calculate interference on PRB $i$.
    end if
  end for
  Find the PRB with the minimum interference, and assign the color of this PRB to $C_{n,\text{new}}$
end for
Now we can obtain the PRB allocation matrix \( Y \) based on Algorithm 2.

### 3.2.3 Power allocation

After forming the user-centric overlapped virtual cell cluster and allocating PRB to each UE, we will solve the user-centric power allocation problem. Relying on the user association matrix \( X \) and the PRB allocation matrix \( Y \), the problem (5) is converted into the problem (13)

\[
\max_{p_{n,l}} R_{\text{total}}
\]

\[
\text{s.t.}
\]

\[
C1 : 0 \leq \sum_{k=1}^{K} \sum_{l=1}^{L} x_{k,n} y_{k,l} p_{n,l} \leq P_n^{\max}, \forall n
\]

\[
C5 : \sum_{l=1}^{L} y_{k,l} B_2 \log_2 (1 + \text{SINR}_{k,l}) \geq R_{k,\min}, \forall k
\]

where

\[
R_{\text{total}} = \sum_{k=1}^{K} \sum_{l=1}^{L} y_{k,l} B \log_2 \left( 1 + \frac{\sum_{n=1}^{N} x_{k,n} p_{n,l} h_{k,n,l}}{\sum_{k'=1,k' \neq k}^{K} \sum_{n=1}^{N} x_{k',n} p_{n,l} h_{k,n,l} + \sigma^2} \right)
\]

We can observe that the objective function of (13) is not concave [23]. However, it has a special structure that we can utilize. The specific utilization process is as follows.

We define \( f(p) = \sum_{k=1}^{K} \sum_{l=1}^{L} f_{k,l}(p) \), and \( g(p) = \sum_{k=1}^{K} \sum_{l=1}^{L} g_{k,l}(p) \), where

\[
f_{k,l}(p) = y_{k,l} B \log_2 \left( \sum_{n=1}^{N} x_{k,n} p_{n,l} h_{k,n,l} \right)
\]

\[
g_{k,l}(p) = y_{k,l} B \log_2 \left( \sum_{k'=1,k' \neq k}^{K} \sum_{n=1}^{N} x_{k',n} p_{n,l} h_{k,n,l} + \sigma^2 \right)
\]

and \( p \in P \), \( P \) denotes the feasible set spanned by constraints \( C1 \) and \( C5 \). Then,

\[
f(p) = \sum_{k=1}^{K} \sum_{l=1}^{L} \sum_{n=1}^{N} x_{k,n} p_{n,l} h_{k,n,l} + \sum_{k'=1,k' \neq k}^{K} \sum_{n=1}^{N} x_{k',n} p_{n,l} h_{k,n,l} + \sigma^2
\]

\[
g(p) = \sum_{k=1}^{K} \sum_{l=1}^{L} \sum_{n=1}^{N} x_{k,n} p_{n,l} h_{k,n,l} + \sigma^2
\]

\( f(p) \) and \( g(p) \) are obviously two concave functions. Thus, utilizing the structure of objective function, the DC programming method [21–25] can be applied to convert the objective function of (13) into \( f(p) - g(p) \). In the similar manner, \( C5 \) can be written as
In DC programming, we can start from a feasible initial point and solve the optimization problem iteratively. In order to solve the convex problem, let $\tau$ denote the iteration number. At the $\tau$-th iteration, we employ the first-order Taylor approximation for $g(p)$ and $g_k,l(p)$, then

$$
\frac{\partial g(p^{(\tau)})}{\partial p} = B \ln 2 \cdot \sum_{k=1}^{K} \sum_{l=1}^{L} \frac{y_{k,l}x_{k,n,l}h_{k,n,l}}{\sum_{k=1}^{K} \sum_{l'=1}^{L} x_{k',n,l'}h_{k,n,l'} + \sigma^2}
$$

Hence, by substituting $g(p^{(\tau)})$ and $\nabla g(p^{(\tau)})$ into the optimization problem (13), the problem (13) can be written as

$$
\max_p f(p) - g(p^{(\tau)}) + \nabla g(p^{(\tau)})(p - p^{(\tau)})
$$

s.t.

$$
C_1 : 0 \leq \sum_{k=1}^{K} \sum_{l=1}^{L} x_{k,n,l}y_{k,l}p_{n,l} \leq p_{\max}, \forall n
$$

$$
C'_5 : \sum_{l=1}^{L} (f_{k,l}(p) - \nabla g_k,l(p^{(\tau)})(p - p^{(\tau)})) \geq R_{k,\min}, \forall k
$$

In order to use the DC programming method to solve the power allocation subproblem, we need to prove the following three propositions.

**Proposition 1** The approximation of (19) gives a tight lower bound for the objective function of (13).

**Proof** Since $g(p^{(\tau)})$ is a concave function, due to the first-order condition of the concave functions [27], we have.

$$
g(p) \leq g(p^{(\tau)}) + \nabla g(p^{(\tau)})(p - p^{(\tau)})
$$

From (23), we can conclude that $f(p) - g(p) \geq f(p^{(\tau)}) - g(p^{(\tau)}) - \nabla g(p^{(\tau)})(p - p^{(\tau)})$. When $p = p^{\tau}$, the equality holds which shows the tightness of the lower bound.

**Proposition 2** The approximation of (19) results in a sequence of improved solutions for the problem of (13).
Proof: The objective function of (13) in the $\tau$-th iteration is $f(p^{(\tau)}) - g(p^{(\tau)})$. We have:

\[
\begin{align*}
&f(p^{(\tau+1)}) - g(p^{(\tau+1)}) \\
&\quad \geq f(p^{(\tau+1)}) - g(p^{(\tau)}) - \nabla g(p^{(\tau)}) (p^{(\tau+1)} - p^{(\tau)}) \\
&\quad \geq f(p^{(\tau)}) - g(p^{(\tau)}) - \nabla g(p^{(\tau)}) (p^{(\tau)} - p^{(\tau)}) \\
&\quad = f(p^{(\tau)}) - g(p^{(\tau)})
\end{align*}
\]

where the inequality $(a)$ follows from the convexity of $g(p)$, and the inequality $(b)$ follows from the fact that $p^{(\tau+1)}$ is the optimal solution of problem (22) at the $\tau + 1$-th iteration.

Thus, the objective function of (13) takes larger values as iterations continue.

**Proposition 3** Proposition 3: The approximation of (19) has a tight bound for the objective problem of (13).

**Proof** The objective function of (13) in the $\tau$-th iteration is $f(p^{(\tau)}) - g(p^{(\tau)})$. Obviously, $f(p^{(\tau)}) - g(p^{(\tau)})$ is a continuous function, and the interval formed by $C'$1 and $C'$2 is a closed interval. Using the closed interval set theorem (theorem 4.15 in [28]), proposition 3 is proved.

Hence, the objective value of $f(p^{(\tau)}) - g(p^{(\tau)})$ is finite and monotonically increasing sequences with upper bounds. According to the nature of theorem 3.14 in [28], the DC approximation always converges.

Notice that function (22) is a concave function since it is the addition of a line and a concave function, and at the same time $C'1$ and $C'5$ are also convex. The proposed iterative DC algorithm for solving the problem in (22) is presented in Algorithm 3. We initialize $p^0$ to a column vector with $NL$ elements, where the fix power allocation of each PRB in our algorithm 2 is used as the initial power $p^0$. 
4 Computational complexity

Here, we evaluate the complexity of our proposed scheme. The complexity of Algorithm 1 is $O(KN^2N_{\text{max}} + NK^2N_{\text{max}})$. For comparison, the complexity of exhaustive search for the optimal solution of user association is $O\left(\prod_{k=1}^{K} \sum_{n=0}^{N_{\text{max}}} C_n^m\right)$. The complexity of Algorithm 2 (PRB allocation algorithm) is $O(NKL^2N_{\text{max}}^2 + L^2)$, where the complexity of graph construction algorithm is $O(NKL^2N_{\text{max}}^2)$, and the complexity of PRB allocation algorithm is $O(L^2)$. As a comparison, the complexity of exhaustive search algorithm for the optimal solution is $O(K^L)$ and the complexity of SA algorithm or RA algorithm is $O(L)$. The complexity of the optimal exhaustive search increases exponentially as the number of UEs increases. Hence, compared with the optimal exhaustive search solution, our proposed solution efficiently reduces the complexity in UDNs.

5 Simulation results and discussion
5.1 System simulation

In this section, we aim to characterize the performance of the proposed framework under different conditions via numerical simulations. We consider one MBS coverage with a circle of 300 m radius. SBSs and UEs are randomly distributed following PPP distribution with density $\lambda_M$ and $\lambda_K$. By default, we set $\lambda_M = 50$ SBS/km$^2$, $\lambda_K = 350$ UE/km$^2$ and $N_{\text{max}} = 3$. We consider three different rate requirements of UEs: 256kbps, 512kbps and 1024kbps, where each user rate requirement is randomly simulated. The other simulation parameters are listed in Table 1. We consider Rayleigh fading to models the channels and assume that the path loss model of each SBS is given by [29]
Fig. 4 SC load of Max RSRP user association with 350 UE/km² and 50 SBS/km²

Fig. 5 SC load of algorithm 1 user association with 350 UE/km² and 50 SBS/km²
Fig. 6 SC load of Max RSRP user association versus the different density of SBSs

Fig. 7 SC load of algorithm 1 user association versus the different density of SBSs
Fig. 8 SC load of Max RSRP user association versus the different density of UEs.

Fig. 9 SC load of algorithm 1 user association versus the different density of UEs.
5.2 Analysis of the simulation results

Figures 4 and 5 show the changes of the load of each SC caused by different the size of virtual cell cluster based on the two different user association schemes under the conditions of 350 UE/km$^2$ and 50 SBS/km$^2$. It can be seen that as the cluster size increases, the load of the SC increases under the two different user association schemes. The reasons are as follows: as the size of the SBS cluster increases, from the perspective of the UE, more SBSs become the cooperative SBSs of UE, from the perspective of SBS, more UEs are associated to SBS. Figure 4 shows the user association results based on the Max RSRP rule. Larger cluster size may improve RSRP but may also reduce total throughput. When $N_{\text{max}}$ is not less than 4, the SCs are overloaded because of insufficient PRBs. However, due to interference issue, the maximum cluster size of SBSs or CoMP will restrict to 3 in most cellular system. Thus, we set $N_{\text{max}} = 3$ by default. Figure 5 gives us the results of user association rule based on Algorithm 1. Compared with the Max RSRP rule, Algorithm 1 eliminates the overloaded SC within a certain range.

Figures 6 and 7 depict the SC load versus the different density of SBSs under user association rule based on Max RSRP and algorithm 1 respectively. Figures 8 and 9 show the

$$PL = 36.7 \log_{10}(d) + 22.7 + 26 \log_{10}(f_c)$$  (25)
**Fig. 11** The system throughput versus the density of UEs

**Fig. 12** The system throughput versus the density of SBSs
SC load versus the different density of UEs under user association rule based on Max RSRP and algorithm 1 respectively. It is clearly observed that the SC load is balanced by Algorithm 1, and SBSs are reselected as cooperative SBSs cluster for UEs that cannot associate to the overload SBSs.

Figures 10 show that the system throughput of the proposed user association based on Algorithm 1 is higher than that of the user association based on Max RSRP rule. This is because that the SC load is considered by Algorithm 1. The UEs of overloaded SC are transferred to the non-overloaded SC, which effectively guarantees user QoS requirements. Through Figs. 4, 5 and 10, we have a conclusion that the optimal number of $N_{\text{max}}$ is 3 considering the trade-off between load balance and system throughput.

In Figs. 11 and 12, we compare Algorithm 2 with other algorithm such as random allocation (RA) [30], sequential allocation (SA) [30] and uniform frequency reuse (UFR) [31] on user-centric clusters by Algorithm 1. Figure 11 depicts the system throughput versus the different density of UEs under 50 SBS/km$^2$. We observe in Fig. 11 that the system throughput is increased for all the solutions when the UEs become denser. However, the performance of our proposed Algorithm 2 outperforms that of other algorithms in mitigating the virtual inter-cell interference and improving system throughput.

Next, Fig. 12 describes system throughput versus the different density of UEs under 350 UE/km$^2$. We can observe that when the density of SBSs is less than 50 the system throughput increases. And then the system throughput becomes stable. This is because

![Fig. 13 The system throughput evolution in different density of UEs](image)
when the density of the SBCs is small, only a few UEs can be served by the SBSs due to the limited number of PRBs. As SBSs become denser, the UEs can be served by more cooperative SBSs and the system throughput also increases. From Figs. 11 and 12, we can conclude that our proposed algorithm 2 performance is superior to that of other algorithms.

We validate the convergence of our power allocation scheme by examining the evolution of system throughput in iterations. Figure 13 describes the system throughput of different the density of UEs when the density of SBSs is 50 SBS/km². And Fig. 14 shows the system throughput of different the density of SBSs when the density of UEs is 350
UE/km². We can see that the throughput can reach a stable state in less than 12 iterations. In Figs. 11 and 12, due to the density and the locations of UEs (or SBSs) are different, the final convergent values of system throughput in different density of UEs (or SBSs) are unequal. This shows that the performance of power allocation scheme using the DC programing method is better than that of the original fixed power allocation scheme in terms of system throughput.

6 Conclusion

In this paper, we propose a novel QoS-based joint user association and resource allocation scheme under the virtual cell architecture in a downlink two-tier heterogeneous UDNs. To mitigate interference, balance cell load, guarantee QoS requirements of UE and improve system throughput, a non-convex NP-hard problem is formulated, and this joint problem is decoupled into the three independent sub-problems. To effectively solve these sub-problems, we propose three schemes: a new user-centric overlapped virtual cell clustering scheme, a low-complexity PRB allocation scheme and a power allocation scheme using the DC programing method. Simulation results confirm that our proposed schemes are better than existing schemes in terms of user rates, cell load and system throughput.

Abbreviations

UDNs: Ultra-dense networks; 5G: The fifth generation; SBS: Small base station; QoS: Quality of service; PRB: Physical resource block; UE: User equipment; DC: The difference of concave; CoMP: Coordinated multiple points; SINR: Signal to interference plus noise ratio; MBS: Macro base station; C-plane: The control plane; U-plane: The user plane; NCU: Network controller unit; PPP: Poisson point process; CSI: Channel state information; ZF: Zero-forcing; AWGN: The additive white Gaussian noise; SC: Small cell; RA: Random allocation; SA: Sequential allocation; UFR: Uniform frequency reuse.

Authors’ contributions

ZWS conceived the study and performed the simulation experiments. JXZ wrote the paper. XYC, KSZ, WDG and GC reviewed and edited the manuscript. All authors read and approved the final manuscript.

Funding

This research was supported by the National Science and Technology Major Project of the People’s Republic of China (Grant No. 2018ZX03001029-004).

Competing interests

The authors declare that they have no competing interests.

Received: 7 January 2020  Accepted: 7 December 2020

Published online: 06 January 2021

References

1. J.G. Andrews et al., What will 5G be? IEEE J. Sel. Areas Commun. 32(6), 1065–1082 (2014)
2. A. Gotsis, S. Stefanatos, A. Alexiou, Ultra-dense networks: the new wireless frontier for enabling 5G access. IEEE Veh. Technol. Mag. 11(6), 71–78 (2016)
3. M. Kamel, W. Hamouda, A. Youssef, Ultra-dense networks: a survey. IEEE Commun. Surv. Tutorials. 18(4), 2522–2545 (2016)
4. D. Lopez-Perez, M. Ding, H. Claussen, A. Jafari, Towards 1 Gbps/UE in cellular systems: understanding ultra-dense small cell deployments. IEEE Commun. Surv. Tutorials. 17(4), 2078–2101 (2015)
5. S. Chen, F. Qin, B. Hu, X. Li, Z. Chen, User-centric ultra-dense net-works for 5G: challenges, methodologies, and directions. IEEE Wireless Commun. 23(2), 78–85 (2016)
6. V. Garcia, Y. Zhou, J. Shi, Coordinated multipoint transmission in dense cellular networks with user-centric adaptive clustering. IEEE Trans. Wirel. Commun. 13(8), 4297–4308 (2014)
7. L. Liu, Y. Zhou, V. Garcia, L. Tian, J. Shi, Load aware joint CoMP clustering and inter-cell resource scheduling in heterogeneous ultra-dense cellular networks. IEEE Trans. Veh. Technol. 67(3), 2741–2755 (2018)
8. S. Bassoy, M. Jaber, M.A. Imran, P. Xiao, Load aware self-organizing user-centric dynamic CoMP clustering for 5G networks. IEEE Access 4, 2895–2906 (2016)
9. S. Bassoy, M. Jaber, M.A. Imran, S. Yang, R. Tafazolli, A Load-aware clustering model for coordinated transmission in future wireless networks. IEEE Access 7, 2169–3536 (2019)
10. Q. Liu, G. Chuai, W. Gao, K. Zhang, Load-aware user-centric virtual cell design in ultra-dense network, 2017 IEEE Conference on Computer Communications (INFOCOM WSHPS), Atlanta, GA, May 2017, pp. 619–624, doi: https://doi.org/10.1109/INFOCOMW.2017.8116448.

11. Q. Liu, G. Chuai, W. Gao, K. Zhang, Fuzzy Logic-based virtual cell design in ultra-dense network. Eura J. Wirel. Commun. Netw. 2018, 87 (2018)

12. H.L. Kim, S. Chong, Virtual cell beamforming in cooperative networks. IEEE J. Sel. Areas Commun. 32(6), 1126–1138 (2014)

13. J. Shi, H. Xu, Z. Yang, M. Chen, Energy efficient beamforming for user-centric virtual cell networks. IEEE Trans. Green Commun. Netw. 3(3), 575–590 (2019)

14. L. Liang, W. Wang, Y. Jia, S. Fu, A cluster-based energy-efficient resource management scheme for ultra-dense networks. IEEE Access 4, 6823–6832 (2016)

15. G. Zhang, F. Ke, Y. Peng, C. Zhang, User access and resource allocation in full-duplex user-centric ultra-dense heterogeneous networks, 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates, pp. 1–6, Dec. 2018, doi: https://doi.org/10.1109/GLOCOM.2018.8648065.

16. L. Lu, D. He, G.Y. Li, X. Yu, Graph-based robust resource allocation for cognitive radio networks. IEEE Trans. Signal Process. 63(14), 3825–3836 (2015)

17. Z. Zhou, K. Ota, M. Dong, C. Xu, Energy-efficient matching for resource allocation in D2D enabled cellular networks. IEEE Trans. Veh. Technol. 66(6), 5256–5268 (2016)

18. T. Yang, R. Zhang, X. Cheng, L. Yang, Graph coloring based resource sharing (GCRS) scheme for D2D communications underlaying full-duplex cellular networks. IEEE Trans. Veh. Technol. 66(8), 7506–7517 (2017)

19. Y. Meng, J. Li, H. Li, M. Pan, A transformed conflict graph-based resource-allocation scheme combining interference alignment in OFDMA femtocell networks. IEEE Trans. Veh. Technol. 64(10), 4728–4737 (2015)

20. Y. Lin, R. Zhang, C. Li, L. Yang, L. Hanzo, Graph-based joint user-centric overlapped clustering and resource allocation in ultradense networks. IEEE Trans. Veh. Technol. 67(5), 4440–4453 (2018)

21. H.H. Kha, H.D. Tuan, H.H. Nguyen, Fast global optimal power allocation in wireless networks by local D.C. programming. IEEE Trans. Wireless Commun. 11(2), 510–515 (2012)

22. B. Khamidehi, A. Rahmati, M. Sabbaghian, Joint sub-channel assignment and power allocation in heterogeneous networks: An efficient optimization method. IEEE Commun. Lett. 20(12), 2490–2493 (2016)

23. Y. Liu, X. Li, F.R. Yu, H. Ji, H. Zhang, V.C.M. Leung, Grouping and coordinating among access points in user-centric ultra-dense networks with non-orthogonal multiple access. IEEE J. Sel. Areas Commun. 35(10), 2295–2311 (2017)

24. J. Peng, J. Zeng, X. Su, B. Liu, H. Zhao, A QoS-based cross-tier cooperation resource allocation scheme over ultra-dense HetNets. IEEE Access 7, 27086–27096 (2019)

25. C.-H. Fang, P.R. Li, K.-T. Feng, Joint interference cancellation and resource allocation for full-duplex cloud radio access networks. IEEE Trans. Wireless Commun. 18(6), 3019–3033 (2019)

26. H. Ibrahim et al., Mobility-aware modeling and analysis of dense cellular networks with C-plane/U-plane split architecture. IEEE Trans. Commun. 64(11), 4879–4894 (2016)

27. S. Boyd, L. Vandenberghe, Convex Optimization (Cambridge Univ, Cambridge, 2004).

28. W. Rudin Principles of Mathematical Analysis. 3rd Edition, McGraw Hill, and Copied by McGraw Hill Education (India) 2013, New Deli, India, 1976

29. Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced. Document ITU-R M.2135.1, ITU-R, Geneva, Switzerland, Dec. 2009.

30. W. Wang and X. Liu, List-coloring based channel allocation for open-spectrum wireless networks, in VTC-2005-Fall 2005 IEEE 62nd Vehicular Technology Conference, 2005, Dallas, TX, USA, pp. 690–694, Sept. 2005, doi: https://doi.org/10.1109/VETECE.2005.1558001.

31. F. Jin, R. Zhang, L. Hanzo, Fractional frequency reuse aided twin-layer femtocell networks: Analysis, design and optimization. IEEE Trans. Wireless Commun. 61(5), 2074–2085 (2013)

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.