This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3114629, IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2021.DOI

Efficient Resource Allocation using Distributed Edge Computing in D2D based 5G-HCN with Network Slicing

LUBNA NADEEM[1], YASAR AMIN[1], (Senior Member, IEEE), JONATHAN LOO[2], (Member, IEEE), MUHAMMAD A. AZAM [3], (Member, IEEE), KOK KEONG CHAI[4]

[1]Department of Telecommunication Engineering, University of Engineering and Technology, Taxila, 47050, Pakistan
[2]School of Computing and Engineering, University of West London, London W5 5RF, UK
[3]Technology and Innovation Research Group, School of Information Technology, Whitecliffe 6011, New Zealand
[4]School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK

ABSTRACT Fifth Generation (5G) cellular networks aim to overcome the pressing demands posed by dynamic Quality of Service (QoS) constraints, which have primarily remained unaddressed using conventional network infrastructure. Cellular networks of the future necessitate the formulation of efficient resource allocation schemes that readily meet throughput requirements. The idea of combining Device-to-Device (D2D), Mobile Edge Computing (MEC), and Network slicing (NS) can improve spectrum utilization with better performance and scalability. This work presents a spectrum efficiency optimization problem in D2D based 5G-Heterogeneous Cellular Network (5G-HCN) with NS. Owing to the shortage of resources, we propose an underlay model where macro-cell users (MUs), small-cell users (SUs), and D2D users (DUs) reuse the resources while considering the effects of interference. The goal is to maximize the average network spectrum efficiency (SE) and throughput without degrading the system performance. The problem at hand is naturally a non-convex mixed-integer non-linear programming (MINLP) problem that is intractable. Therefore, we have suggested a distributed resource allocation strategy with an edge computing (DRA-EC) approach to find the sub-optimal solution. In distributed augmented Lagrange method, each edge router located at BS will solve its problem locally, and the consensus algorithm will find the global solution using these local estimates. The central slice controller will cut the customized network slices according to the bandwidth requirements of each user type with optimized spectrum information. The simulation outcomes prove that our proposed method is near the central optimization scheme with low computational complexity. It is much better because it reduces the computational time and system overhead.

INDEX TERMS Network slicing, 5G cellular networks, Mobile edge computing, Device-to-device, Distributed optimization, Consensus algorithm.

I. INTRODUCTION

FOR 5G and beyond 5G, the elastic redesign of conventional networks is expected to shift the pattern with which human beings and machines interact. Ericsson Mobility Report [1] predicts that out of ten, every fourth user will have 5G mobile subscriptions by 2026, which is about 60% of the world's population. 5G cellular networks will have enhanced speeds and a minimum delay than the previous generations [2]. The spectrum shortage due to full occupancy of frequency bands is the main hurdle behind 5G cellular networks. The major challenge in research is assigning the resources efficiently that will optimize the spectrum utilization [3]. The scarcity in the spectrum is an outcome of inefficient and random spectrum allocation techniques. HCNs comprising macro-cells underlaid with multiple small-cells offer tremendous capability to improve the reuse of frequency and network capacity [4]. Furthermore, D2D-based communication offers a promising solution as it remains the primary method for improving the performance of 5G cellular networks.

The concept of D2D is to allow direct connection among the devices that are close enough with very little or no BS involvement. Therefore, this is an important technique to offload the BS traffic [5]. Due to proximity, the users in D2D communication have low power, and less transmission delay, resulting in increased SE and throughput. Examples

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3114629, IEEE Access

include sharing files, audio or videos, online gaming, disaster management, and traffic offloading. However, the integration of D2D in the system opens many challenges, including mode-selection, energy consumption, device discovery, interference, mobility management, and network security [6].

The exponential increase in the number of devices is a test for future research due to their diverse needs of reliability, speed, and latency [7]. For this, NS is the distinct characteristic of 5G system architecture that creates flexible logical networks over typical hardware infrastructure [8]. With the continuous progress in 5G, NS will be the future's next big thing enabling service-based customized end-to-end logical networks called network slices [9]. Slices can be allocated depending on specific QoS like throughput, latency, or reliability [10]. The core reason behind the prevalence of NS today is the reduced system capital expense (CAPEX) and operational expenditure (OPEX) because of efficient spectral utilization [11], [12]. 5G NS facilitates the customers to enjoy the processing of data and other services that have a substantial commercial perspective [13].

MEC is an evolving paradigm. It enables the central cloud computing network functionality at the edge node proximal to the end devices. MEC can provide an advantage in the effective utilization of system backhaul, computational, and storage resources. The researchers are expecting that from 2020; the edge nodes will manage about 45% of the data instead of a central cloud system [14]. MEC can optimize the network resources by processing and managing the data at the edge server before sending it to the central cloud. This will result in offloading [15] with improved system performance [16].

The latest emerging idea is to integrate D2D, NS, and MEC technology in future network designs, as proposed in our previous work [8]. In this work we have proposed a system to optimize the SE and throughput in D2D and NS based 5G-HCN by using the distributed MEC solution. This system will then send the optimized value of SE to the central network slice controller that will cut the network slices according to the data received from the edge layer. The goal of this research work is to solve the problem of resource allocation in 5G-HCN using edge computing, which is a perfect option instead of a central scheme because it will reduce the computation time and offload the system resources.

Section II discusses the related work, and Section III depicts the gaps in research and the motivation of our work. Then Section IV is about our main research contributions. Section V explains in detail the proposed system model. Further, Section VI demonstrates the problem formulation and optimization. Section VII derives the expression of the proposed technique for solving the optimization problem. Section VIII is related to simulation results and discussion. Finally, Section IX is on conclusion and future recommendations.

II. RELATED WORK

5G with HCNs design has emerged as the promising topic of research due to their enhanced capability of resource management and utilization [17]. Recently, work done in [18] provided the detailed survey on resource allocation in 5G HCNs with an explanation of existing literature, future trends, and possible challenges. The authors in [19] proposed a heterogeneous network scheme that guaranteed the QoS and fairness of all users while minimizing the interference. Similarly, the work proposed in [20] investigated the resource allocation scheme in heterogeneous networks considering the system robustness and interference efficiency. In [21], the researchers optimized the robust energy efficiency problem with security information. Many researchers have studied and proposed techniques to meet the challenges of resource allocation and managing the ever-increasing network load demands because of the spectrum scarcity [22]. Heterogeneous network deployment with a macro cell and many small cells can improve the spectrum efficiency [23]. The work in D2D communication has gained researcher's attention from the past few years with the evolution towards 5G [24]. There are two categories of D2D communications modes: licensed and unlicensed bands. The licensed frequency bands can be further divided into two types based on the frequency sharing method among the cellular users and D2D as underlay and overlay mode [25]. The proposed work is based on the underlay mode in which the cellular and D2D users share the frequency/channel resulting in interference which is the major problem in the heterogeneous environment for achieving the spectrum efficiency [26]–[28]. The work done by [29] is to increase the capacity by efficient spectrum allocation using the coalition game to overcome the interference. The research in D2D was initially on conventional networks, but the works from the past few years have shown tremendous growth of adding them in virtualized networks [30]. The study done in [31] was on resource sharing in D2D based systems, [32] investigated D2D system using joint spectrum and power assignment for both central and distributed techniques for resource allocation. Similarly, [33] proposed a semi-decentralized approach to maximize the sum rate considering interference. The MEC and D2D are essential technologies for offloading the high data rate traffic from the central network. [34] investigated MEC-D2D combination to maximize the no. of supported devices. [35] worked on D2D based 5G heterogeneous networks using MEC. The authors in [36]–[38] investigated the advantage of using MEC to improve the system performance and minimize the delay.

Many works have been done to achieve efficient radio resource management in virtual networks [39], [40]. Existing literature on NS has two main categories: (1) Infrastructure based slicing and (2) Resource spectrum based slicing. The work proposed in [41] was to maximize the profit of MVNO by joint power allocation and slice resource allocation considering the backhaul capacity and user's QoS. [42] proposed the resource (power and channel) allocation techniques to optimize the network throughput in multi-slices and multi-user

TABLE 1. List of Notations

Notation	Description	Notation	Description
$\mathcal{M}$	MBS	J	Set of SBS
U	Set of all Users	R	Set of Channel Resources
K	Set of MUE	L	Set of SUE
D	Set of D2D UE	a_k^r	Allocation Indicator for MUE
$\mathrm{a}_{j,l}^r$	Allocation Indicator for SUE	$\mathrm{a}_{d_p}^r$	Allocation Indicator for D2D
λ_k	Received SINR from MBS to MUE	$\lambda_{j,l}$	Received SINR from SBS to SUE
λ_{d_p}	Received SINR from D_T to D_R	R_k	Data Rate from MBS to MUE
$\mathrm{R}_{j,l}$	Data Rate from SBS to SUE	R_{d_p}	Data Rate from D_T to D_R
P_m	MBS Transmit Power	P_j	SBS Transmit Power
P_d	D2D User Transmit Power	δ_c	Macro-cell mode indicator
δ_j	Small-cell mode indicator	δ_k	D2D mode indicator
PL_k	Received Path loss from MBS	PL_j	Received Path loss from SBS
PL_{d_p}	Received Path loss from D_T	σ	Additive white Guassian noise
β	Sub-channel Bandwidth	Ω_{Max}^r	Maximum allowable Interference

cases. These works [41]–[43] were better in performance, but the approaches burdened the central controller because each slice must allocate resources individually. In [44], the authors proposed the virtualization framework for resource block allocation to its users using the auction-based game method but did not consider spectrum efficiency with D2D and edge computing. [45] discussed two cases of slicing; one is based on QoS to create dedicated slices depending on different services, and the second on infrastructure sharing in which the resource sharing is among the multi-tenants.

The work in [46] considered hierarchical resource allocation using NS in fog networks which overcome the issues of core layer load in which the global resource manager first assigns resources local resource manager in the slices, and these resources are then efficiently allocated to users using Stackelberg game. This scheme optimized the spectrum efficiency, but D2D and edge computing were missing. The problem of optimizing energy efficiency in the wireless heterogeneous virtualized network was proposed in [47]. However, this work focused on maximizing the revenue of infrastructure providers (InPs) and mobile virtual network operators (MVNO) instead of assigning resources by improving spectrum efficiency and adding D2D. In [48], the authors introduced a resource management technique in multi-tenant cloud-based radio access networks to perform the resource slicing considering QoS and interference, but this works also did not consider D2D communication and spectrum efficiency optimization. Similar work was proposed in [49] for resource slicing in two-tier HCNs and allocate the resources based on QoS. This scheme efficiently computed the optimum bandwidth slicing ratio in a virtualized HCN but did not use an edge computing-based approach with D2D for spectrum efficiency optimization. The work proposed in [31] investigated the centralized method of maximizing the D2D pairs sum rate without disturbing the data rate requirements of CUs using subchannel sharing, but it did not consider the HCN with NS and distributed computing. In [33], the authors proposed the semi-decentralized method of optimized resource allocation and power control with interference-aware D2D setup. However, HCN and NS with edge-based computation were missing in that work.

All the research works described above on 5G cellular networks with network slicing were mainly focused on increasing the overall system utility and revenue. Primarily, the researchers goal was to design such a system that would benefit business models. None of them considered improving the spectrum efficiency with D2D based system for network slicing and to solve the resource allocation problem using distributed edge based computing.

III. RESEARCH GAPS & MOTIVATION

The key motivation towards this research is that to the best of our knowledge, there is no single work on improving the spectral efficiency in 5G-HCN with D2D and NS. Furthermore, to solve such a problem using distributed MEC technique. Previous works mainly exploited the SE in these technologies individually or with two of them. Like resource management for 5G-HCN with or without D2D, with or without NS, and with or without MEC. Similarly, D2D with virtualized networks and D2D with MEC scenario. Therefore in [8], we have proposed an architecture that will integrate D2D, MEC, and NS technology to meet most of the requirements of future 5G networks. All future research focuses on designing techniques that wisely use the spectrum. D2D and network slicing are the critical enablers for achieving this for enhanced system performance and reduced cost (less hardware). The main concern is the interference management so that resource allocation will not affect the SINR (Signal to Interference and Noise Ratio) requirements of users.

IV. CONTRIBUTION

The significant contribution in this work is that no previous research considered solving SE maximization combining D2D, NS, and MEC in 5G-HCNs. This work is the first attempt to solve such a problem using distributed edge computing optimization approach called DRA-EC. Previously, the researchers have mainly proposed resource allocation by centralized methods that burden the central system and cause

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3114629, IEEE Access

a delay in transmission. Our proposed scheme solves the problem of ever-increasing spectrum demand in less time.

- To develop a system model for D2D based 5G-HCN with randomly distributed users (MUs and SUs). The proposed model has two layers: the upper and the lower layer.
- In the upper layer, firstly, a less complex and sub-optimal solution solves the base station to user equipment (BS-UE) association problem using a greedy algorithm based on achievable SNR (Signal to Noise ratio). Secondly, the D2D mode switching problem is solved using a joint distance-dependent algorithm.
- In the lower layer, we formulate the resource allocation problem using the distributed edge computing method to maximize the throughput and spectrum efficiency of the 5G HCN considering the interference and SINR constraints.
- The proposed problem is a non-convex MINLP problem transformed and decomposed into a convex problem. The problem is then solved by augmented Lagrange multiplier and consensus algorithm. The simulation results are compared with the other schemes and concluded that the computation time and load improves considerably.

V. SYSTEM MODEL

A. NETWORK LAYOUT

Fig.1 represents the detailed network architecture of our proposed scheme. In this work, we have considered a 5G-HCN comprising one macro-cell and several small cells at random locations within the coverage area. The proposed work considers the downlink scenario with a single MBS (macro base station) $\mathcal{M}$ located in the center of the cell having high transmit power and wide-area coverage. There are J SBS (small base stations) given as $J = \{1, 2 \ldots j \ldots \mathcal{J}\}$ having smaller coverage and low transmit power. The small cells coverage area is assumed to be non overlapping circles within the macro cell. Both MBS and SBSs have edge routers (MEC servers) shown in Fig.1 with local computational capabilities connected via wired backhaul link to the central edge-server of the core network. Each edge server collects the desired data and sends the information to a central server. Slice controller then processes this data through network management and orchestration to cut the desired network slices. The total number of users admitted in the network are U where $U = \{1, 2 \ldots u, \ldots \mathcal{U}\}$.

B. CHANNEL CHARACTERISTICS

The total channel bandwidth is distributed into several sub-channels, each of them occupying a band of 180kHz frequency [50]. The Noise power for the system is Additive White Gaussian Noise (AWGN) represented by σ given as:

$$\sigma = N_O * NF * \beta \tag{1}$$

N_O is noise power spectral density, NF represents noise figure, and β is the sub-channel bandwidth.

Shadowing and Rayleigh random variables are used for channel modeling to evaluate the fading in combination with the path loss models between transmitter and receiver. The distance-dependent path loss models considered are OkumuraHata model [51]:

$$PL_k = 128.1 + 37.6log(d[km])dBm \tag{2}$$

$$PL_j = 140.7 + 36.7log(d[km])dBm \tag{3}$$

$$PL_{d_p} = 148.1 + 40log(d[km])dBm \tag{4}$$

where PL_k, PL_j and PL_{d_p} is the path loss received by macro-cell, small-cell and D2D users respectively. d is the distance from the BS to its associated UE. In case of D2D, this is the distance between the D2D transmitter and receiver.

The complete system is in frequency reuse mode with MBS and SBS sharing a similar set of sub-channels denoted by $R = \{1, 2 \ldots r, \ldots \mathcal{N}\}$. However, we have assumed that the same type of UEs (user equipment) associated with one BS do not reuse the sub-channel. Therefore, alleviating the co-tier interference, but the cross-tier interference is there. For each sub-channel $r \in R$, the predefined threshold level for maximum allowable interference Ω^r_{Max} is set to protect the SINR of each UE. The CSI (channel state information) from feedback control channels decides this threshold value. The sub-channel will be assigned only if the cross-tier interference is below Ω^r_{Max}. Let P_m,P_j, and P_d is the maximum allowable transmission power of MBS, SBS, and D2D, respectively.

C. BS-UE ASSOCIATION

To solve the upper layer, the first step is the BS-UE association. The purpose of this is to associate the UEs with their respective BS (base station) (MBS or SBS), which can deliver high channel quality. The wide-band SNR received by UE u associated with BS c is estimated as follows:

$$SNR_{c,u} = \frac{P_{c,u} g_{c,u}}{\sigma^2} \tag{5}$$

where $c = 0$, if the BS is MBS and $c = 1$, if the BS is SBS, $P_{c,u}$ is the transmission power from base station to UE, $g_{c,u}$ denotes the channel gain from BS to UE and σ represent the channel noise.

For any user $u \in \{U\}$, the achievable data rate, $\mathcal{R}_{c,u}$ is given by Shannon Equation:

$$\mathcal{R}_{c,u} = \beta log(1 + SNR_{c,u}) \tag{6}$$

β is the sub-channel bandwidth.

Next, the BS-UE association problem formulation can be performed as:

$$\max_{a_{c,u}} \sum_{c \in C} \sum_{u \in U} a_{c,u} SNR_{c,u} \tag{7}$$

subject to

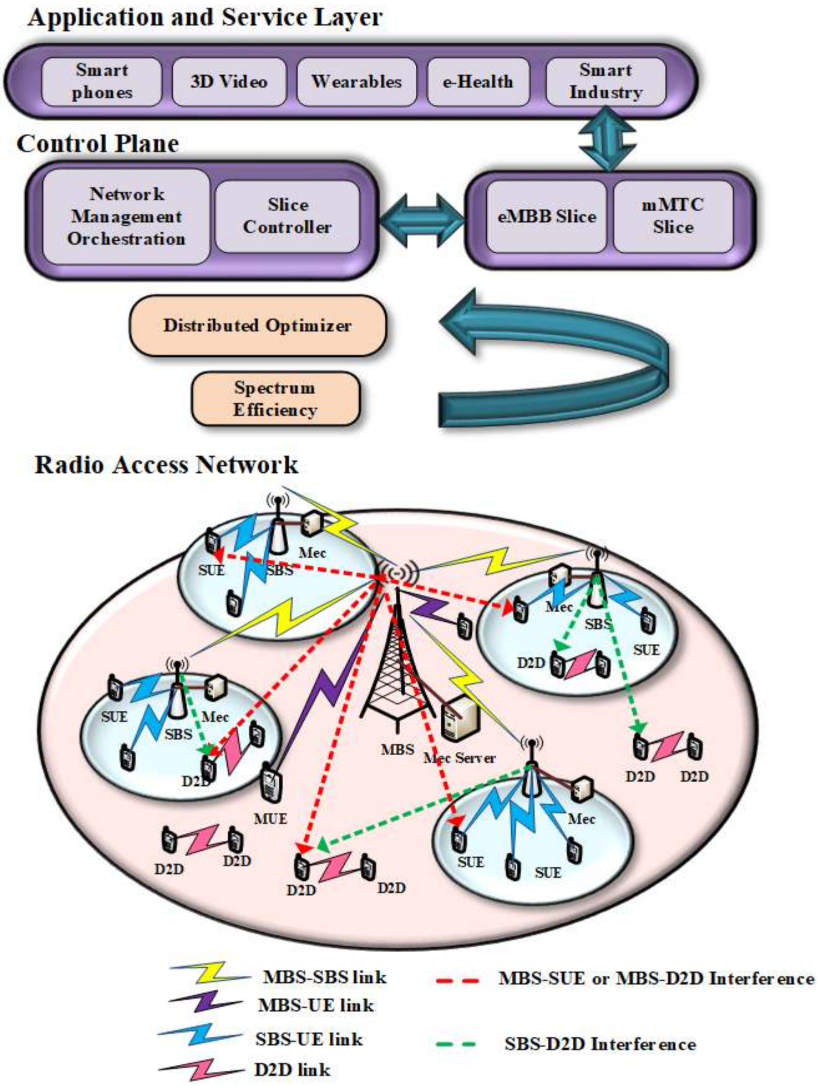

FIGURE 1. System Model

$$\sum_{u \in U} a_{c,u} \mathcal{R}_{min,u} \leq \mathcal{Q}_{tot} \text{ , } \forall c \in C, \tag{8}$$

$$\sum_{c \in C} a_{c,u} = 1, \forall u \in U \tag{9}$$

$$a_{c,u} \in \{0,1\}, \forall u \in U, \forall c \in C \tag{10}$$

where Q_{tot} in eq. (8) refers to the total available system capacity (backhaul/ fronthaul), and it ensures that the total UE data rate is bounded by its capacity on the link [48]. The constraint in (8) guarantees that the backhaul/ fronthaul link can at least carry the minimum data rate required by the associated UEs, (9) ensures that each BS can only attach one UE at a time.

Algorithm 1 BS-UE Association (Greedy Algorithm)

Set $\mathcal{C}_{rem} = \mathcal{C}$, $\mathcal{Q}_{rem,c} = \mathcal{Q}_{tot,c}$, $a_{c,u} = 0$,Γ_c=0
forall u $\in U$ **do**
  for all $c \in C$ **do**
  Step-1: Calculate $SNR_{c,u}$
    $\Gamma_c = SNR_{c,u}$
  end for
  Find $c = argmax_{i \in \mathcal{C}_{rem}} \Gamma_i$
  while $\mathcal{Q}_{rem,c} \geq \mathcal{R}_{min,c}$ **do**
    if $\mathcal{Q}_{rem,c} \geq \mathcal{R}_{min,c}$ **then**
      Set $a_{c,u} = 0$ **and update**
      $\mathcal{Q}_{rem,c} = \mathcal{Q}_{rem,c} - \mathcal{R}_{min,c}$
    else
      $\mathcal{C}_{rem} = \mathcal{C}_{rem}/[c]$
    end if
  end while
  Step-2: Find the user association with cell,
    SNR$_{\max} \sum \sum a_u \Gamma_{c,u}$
end

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3114629, IEEE Access

Proposition 1: The optimization problem (BS-UE association) depicted in (7) is NP-complete 0-1 multiple-knapsacks [52].
Proof: Refer to Appendix A.

The dynamic methods for programming are not efficient for a multiple-knapsack case because of larger computational complexity [48]. To find the feasible sub-optimal solution, we have proposed a greedy method as in Algorithm 1. The UEs are associated with that particular BS which provides them with the highest SNR. For that purpose, we compute the achievable SNR from all the base stations to that particular user. The BS that delivers the highest value of SNR will be nominated to assess the (8). If it satisfies, the user will associate itself with the desired BS. Otherwise, it will select the next highest value of SNR from BS to UE and repeat the process for all the users. We assumed that every UE must be attached to the BS. $Q_{rem,c}$ in algorithm 1 represent the remaining system capacity. It is basically the variable in which the updated value is stored. Initially, $Q_{rem,c}$ = Q_{tot} which means that the remaining system capacity is the same as the total available capacity. Once the process of BS-UE association begins, the value of $Q_{rem,c}$ starts decreasing, and the value is updated in this variable. This process continues until the full system capacity is utilized or all the users are associated with their respective BS. The constraint (8) must fulfill for the algorithm to run smoothly.

The BS-UE association will result in two types of UEs. The set of $\mathcal{K}$ macro-cell users (MUE) $K = \{1, 2 \ldots k \ldots \mathcal{K}\}$ associated with MBS and the set of $\mathcal{L}$ small-cell users (SUE) $L = \{1, 2 \ldots l \ldots \mathcal{L}\}$ also associated with their respective SBS.

D. D2D MODE SWITCHING

The next step is to determine the possible D2D pairs in the system. This is executed by D2D mode switching as in Algorithm 2. $\mathcal{U}_{tot}$ refers to the total users in the system before D2D mode switching. $\mathcal{U}_{tot} = \mathcal{U}_{mbs} + \mathcal{U}_{sbs}$ which means total users in the system are equal to no. of MBS and SBS users respectively. $\mathcal{R}_{tot}$ refers to the total no. of available resources (channels) in the system and $\mathcal{R}_{tot} < \mathcal{U}_{tot}$ means that the total available resources are less than total users in our model.

D2D mode selection will switch particular users from cellular to D2D mode based on distance criteria. Using the Euclidean distance calculations [53], the algorithm calculates the distance of each user u from all the other users and generates the list of users that are equal to, or below certain threshold D2D distance d_{thres}. If the two users are in the range of d_{thres}, the resource is available and their link gain is higher only then they are eligible to switch to D2D mode; otherwise, they will remain in cellular mode. The value of d_{thres} is set initially when the system model is defined and it is not fixed and can be varied for analysis. The set of D DUs is given as: $D = \{1, 2, \ldots d_p, \ldots d_q, \ldots \mathcal{D}\}$.

Algorithm 2 D2D Mode (Distance-dependent Algorithm)

```
Set U_tot = U_mbs + U_sbs
R_tot = R ∋ R_tot < U_tot
Step-1: Find the proximal distance d between devices
for i = 1 : U_tot
   for j = 1 : U_tot
      d_u = |U_i − U_j|
end
  end
  if(d_u ≤ d_thres)
    if(resource r is available)
       if(D2D mode link gain is higher ) then
          the UE U will select the D2D mode
       else
          the UE U will select the cellular mode
       end
     end
  end
```

After the D2D mode switching, we now have three types of users (1) MUE, (2) SUE, and (3) DUE.

VI. PROBLEM FORMULATION

This section formulates the resource allocation problem for efficiently assigning spectrum to each user considering their interference and QoS requirements as all the users reuse the spectrum resources [54]. The respective resource allocation indicator functions are given as:

$$a_k^r = \begin{cases} 1 & if\ MUE\ k\ in\ MBS\ is\ assigned\ r \\ 0 & otherwise \end{cases} \tag{11}$$

$$a_{j,l}^r = \begin{cases} 1 & if\ SUE\ l\ in\ SBS\ j\ is\ assigned\ r \\ 0 & otherwise \end{cases} \tag{12}$$

$$a_{d_p}^r = \begin{cases} 1 & if\ D2D\ user\ d_p\ is\ assigned\ r \\ 0 & otherwise \end{cases} \tag{13}$$

The SINR λ_k^r received by k_{th} MUE associated with MBS using r_{th} sub-channel is given as:

$$\lambda_k^r = \frac{a_k^r P_m g_{m,k}^r}{\Omega_k^r + \sigma^2} \tag{14}$$

$$\Omega_k^r = \sum_{s=1 \to J} P_s g_{s,k}^r + \sum_{\forall_d} P_d g_{d,k}^r \tag{15}$$

$$\lambda_k^r = \frac{a_k^r P_m g_{m,k}^r}{\sum_{s=1 \to J} P_s g_{s,k}^r + \sum_{\forall_d} P_d g_{d,k}^r + \sigma^2} \tag{16}$$

where Ω_k^r represents the interference experienced by MUE. P_m, P_j and P_{d_p} is the transmit power of MBS, SBS and D2D user respectively. $g^r_{m,k}$ is the channel gain from MBS to MUE k, $g^r_{s,k}$, $g^r_{d,k}$ is the gain from all SBS and D2D transmitters using same sub-channel r, σ^2 is noise.

Similarly, the SINR $\lambda^r_{j,l}$ received by l_{th} SUE associated with j_{th} SBS using r_{th} sub-channel is given as:

$$\lambda^r_{j,l} = \frac{a^r_{j,l} P_j g^r_{j,l}}{\Omega^r_{j,l} + \sigma^2} \tag{17}$$

$$\Omega^r_{j,l} = P_m g^r_{m,l} + \sum_{s=1\to L \ s\neq j} P_s g^r_{s,l} + \sum_{\forall_d} P_d g^r_{d_p,l} \tag{18}$$

$$\lambda^r_{j,l} = \frac{a^r_{j,l} P_j g^r_{j,l}}{P_m g^r_{m,l} + \sum_{s=1\to L \ s\neq j} P_s g^r_{s,l} + \sum_{\forall_d} P_d g^r_{d_p,l} + \sigma^2} \tag{19}$$

where, $\Omega^r_{j,l}$ represents the interference experienced by SUE. $g^r_{j,l}$ is the channel gain from SBS to SUE l, $g^r_{m,l}$, $g^r_{s,l}$, $g^r_{d_p,l}$ is the channel gain from MBS, other SBS and D2D transmitter to SUE l using resource r.

And the acheivable SINR $\lambda^r_{d_p}$ of d_{pth} D2D user using r_{th} sub-channel is given as:

$$\lambda^r_{d_p} = \frac{a^r_{d_p} P_d g^r_{d_p,d_r}}{\Omega^r_{d_p} + \sigma^2} \tag{20}$$

$$\Omega^r_{d_p} = P_m g^r_{m,d_p} + \sum_{s=1\to J} P_s g^r_{s,d_p} + \sum_{\forall_{d_q}} P_{d_q} g^r_{d_q,d_p} \tag{21}$$

$$\lambda^r_{d_p} = \frac{a^r_{d_p} P_d g^r_{d_p,d_r}}{P_m g^r_{m,d_p} + \sum_{s=1\to J} P_s g^r_{s,d_p} + \sum_{\forall_{d_q}} P_{d_q} g^r_{d_q,d_p} + \sigma^2} \tag{22}$$

where, $\Omega^r_{d_p}$ represents the interference experienced by DUE. $g^r_{d_p,d_r}$ is the channel gain between D2D pair, g^r_{m,d_p}, g^r_{j,d_p} is the channel gain from MBS and SBS to d_{pth} D2D user using same sub-channel r.

Thus the total achievable throughput by MUE from MBS, SUE from j_{th} SBS and DUE respectively is given below:

$$\mathcal{R}^r_k = \beta log(1 + \lambda^r_k) \tag{23}$$

$$\mathcal{R}^r_{j,l} = \beta log(1 + \lambda^r_{j,l}) \tag{24}$$

$$\mathcal{R}^r_{d_p} = \beta log(1 + \lambda^r_{d_p}) \tag{25}$$

β in equation (23), (24) and (25) is the same sub-channel bandwidth for MBS and SBS which is 180KHz.

The spectrum efficiency (SE) in each case can be represented in equation (26), (27) and (28):

$$\mathcal{SE}^r_k = \mathcal{R}^r_k/\beta \tag{26}$$

$$\mathcal{SE}^r_j = \mathcal{R}^r_j/\beta \tag{27}$$

$$\mathcal{SE}^r_{d_p} = \mathcal{R}^r_{d_p}/\beta \tag{28}$$

The decision variable δ_{d_p} indicates if any user will select the D2D mode or not.

$$\delta_{d_p} = \begin{cases} 1 & if\ any\ user\ d_p\ uses\ the\ D2D\ mode \\ 0 & otherwise \end{cases} \tag{29}$$

The decision variable δ_c represents if any user is in cellular mode.

$$\delta_c = \begin{cases} 1 & if\ cellular\ user\ c\ is\ assigned,\ c \in \{\mathcal{M}, \mathrm{J}\} \\ 0 & otherwise \end{cases} \tag{30}$$

In case the small cell user is considered, δ_c transforms into δ_j while for macrocell user it is δ_k.

A. OPTIMIZATION PROBLEM

Objective is to maximise the average system throughput by efficiently allocating spectrum to all the users considering interference. The optimization problem $\mathcal{P}1$ is given as:

$$\mathcal{P}1 : \max_{\delta,a} \sum_{r\in R} \Big(\sum_{k\in K} \delta_k \mathcal{R}^r_k + \sum_{j\in J} \sum_{l\in L} \delta_j \mathcal{R}^r_{j,l} + \sum_{d_p\in D} \delta_{d_p} \mathcal{R}^r_{d_p} \Big) \tag{31}$$

subject to:

$$\sum_{k\in K} a^r_k \leq 1, \forall r \in R \tag{32}$$

$$\sum_{l\in L} a^r_{j,l} \leq 1, \forall r \in R, \forall j \in J \tag{33}$$

$$\sum_{d_p\in D} a^r_{d_p} \leq 1, \forall r \in R, \forall d_p \in D \tag{34}$$

$$\Omega^r_k \delta_k \leq \Omega^r_{max}, \forall_r \in R, \forall_k \tag{35}$$

$$\Omega^r_{j,l} \delta_j \leq \Omega^r_{max}, \forall_r \in R, \forall_j\ j \in J \tag{36}$$

$$\Omega^r_{d_p} \delta_{d_p} \leq \Omega^r_{max}, \forall_r \in R, \forall_{d_p}\ d_p \in D \tag{37}$$

$$\delta_k \in \{0, 1\}, k \in K \tag{38}$$

$$\delta_j \in \{0, 1\}, j \in J \tag{39}$$

$$\delta_{d_p} \in \{0, 1\}, d_p \in D \tag{40}$$

The problem $\mathcal{P}1$ in (31) is a maximization of the overall system throughput. The constraint in (32), (33) and (34) ensures that one type of user can reuse up to one channel resource and one channel resource can be reused by at most one user type. The constraint (35), (36) and (37) limits the interference and guarantees that the interference experienced by each user type reusing sub-channel r must be below maximum allowable value. Therefore, it must meet each type of user (MUE, SUE, and DUE). The constraint (38), (39), and (40) is mode selection indicators that can be either 0 or 1 if the user is in cellular or D2D mode.

VII. PROPOSED TECHNIQUE

In order to examine the complexity of $\mathcal{P}1$ in (31), we considered:

Proposition 2: The optimization problem in (31) is NP hard and difficult to solve in direct way.

Proof: Refer to Appendix B.

$\mathcal{P}1$ is a non-convex MINLP (mixed-integer non-linear) programming problem that is computationally problematic. Therefore, finding a solution for such a problem is unfeasible [55], [56]. To find the solution, we must convert it into a convex optimization problem. For our system model, the size of the solution becomes significant with the increase in no. of small cells and users. Therefore, we have proposed the distributed optimization method (DRA-EC) called the distributed resource allocation using edge computing. Recently, the distributed optimization-based solutions have emerged as the highly prevalent research area [57], [58]. Furthermore, to solve such a problem using Lagrange multiplier has been proved in [59], [60].

Therefore, we transformed the original problem $\mathcal{P}1$ to make it separable, and multiple local copies of global variables are defined. We can now divide the optimization problem (31) into three subproblems based on three types of users in the system. Let x, y, and z represent three subproblems: macro-cell, small cell, and D2D case, respectively. x represents the data rate equation for a macro cell, y for a small cell, and z for D2D users. The idea of dividing this is that we are solving the optimization problem in a distributive manner. Each edge router located at each type of base station will solve its optimization problem locally.

$$\underbrace{\sum_{r\in R}\Bigg(\sum_{k\in K}\delta_k\mathcal{R}_k^r}_{\text{x}}+\underbrace{\sum_{j\in J}\sum_{l\in L}\delta_j\mathcal{R}_{j,l}^r}_{\text{y}}+\underbrace{\sum_{d_p\in D}\delta_{d_p}\mathcal{R}_{d_p}^r\Bigg)}_{\text{z}} \tag{41}$$

$$f(\delta,x,y,z)=f(\delta_k,x)+f(\delta_j,y)+f(\delta_{d_p},z) \tag{42}$$

$$f(\delta_k,x)=\sum_{r\in R}\sum_{k\in K}\delta_k\mathcal{R}_k^r \tag{43}$$

$$f(\delta_j,y)=\sum_{r\in R}\sum_{j\in J}\sum_{l\in L}\delta_j\mathcal{R}_{j,l}^r \tag{44}$$

$$f(\delta_{d_p},z)=\sum_{r\in R}\sum_{d_p\in D}\delta_{d_p}\mathcal{R}_{d_p}^r \tag{45}$$

Consider that we have E edge servers located at each BS used to solve the distributed optimization problem. Hence the overall problem (31) become divided into E sub-problems.

When we decompose the problem, the constraints from (35) - (37) become local constraints which means that each edge router will solve its optimization problem locally considering these interference constraints. These local constraints are now transformed as below:

$$\Omega_k^r\delta_k\leq\frac{\Omega_{max}^r}{E},\forall_r\in R,\forall_k\, k\in K \tag{46}$$

$$\Omega_j^r\delta_j\leq\frac{\Omega_{max}^r}{E},\forall_r\in R,\forall_j\, j\in J \tag{47}$$

$$\Omega_{d_p}^r\delta_{d_p}\leq\frac{\Omega_{max}^r}{E},\forall_r\in R,\forall_{d_p}\, d_p\in D \tag{48}$$

E denotes the edge router. Each edge router located at each base station will solve its optimization problem considering the QoS requirement of users.

The constraints from (32) - (34) become global constraints and can be defined as consensus constraint.

Let $u\in\{k,l,d_p\}$.

$$\text{subject to}\sum_{u\in U}a_u^r\leq 1,\forall_r\in R,\forall_u\in U \tag{49}$$

In our solution, we first find the local approximate of the previous global constraint functions (35) - (37). An augmented Lagrangian multiplier method is used for each local constraint to find the local maxima. Then the consensus-based algorithm is used with finite steps among iterations to gain a joint agreement on these local approximates [61], [62].

Proposition 3: The augmented Lagrangian multiplier method with consensus algorithm converges at a faster rate practically as compared to other distributed techniques.

Proof: Refer to Appendix C.

Remarks: The theoretical analysis confirms that the given algorithm approach converges at a rate of $O(1/k)$ and produces a steady-state error that is manageable by various consensus-based steps.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3114629, IEEE Access

A. SOLVING LOCAL VARIABLE

From above discussion, we get the local constraints from (46) - (48) solved by each edge server or agent individually. The edge servers will optimize the problem to find local maxima. Following augmented langrangian steps, the partial lagrangian for MUE can be represented by:

$$\sum_{r\in R} L_r(x_k^r, \eta_k^r) \tag{50}$$

$$\sum_{r\in R} L_r(x_k^r, \eta_k^r) = \sum_{k\in K} a_k^r \ \beta log(1+\lambda_k^r) + \eta_k^r \left(\delta_k \Omega^r - \frac{\Omega^r_{max}}{E}\right) \tag{51}$$

Partial Lagrangian for SUE can be represented by:

$$\sum_{r\in R} L_r(y_{j,l}^r, \eta_l^r) \tag{52}$$

$$\sum_{r\in R} L_r(y_{j,l}^r, \eta_l^r) = \sum_{j\in J}\sum_{l\in L} a_{j,l}^r \ \beta log(1+\lambda_{j,l}^r) + \eta_l^r \left(\delta_j \Omega^r - \frac{\Omega^r_{max}}{E}\right) \tag{53}$$

Partial Lagrangian for D2D users can be represented by:

$$\sum_{r\in R} L_r(z_{d_p}^r, \eta_{d_p}^r) \tag{54}$$

$$L_r(z_{d_p}^r, \eta_{d_p}^r) = \sum_{d_p\in D} a_{d_p}^r \ \beta log(1+\lambda_{d_p}^r) + \eta_{d_p}^r \left(\sum_{d_p\in D} \delta_{d_p} \Omega^r - \frac{\Omega^r_{max}}{E}\right) \tag{55}$$

where η_k^r, η_l^r and $\eta_{d_p}^r$ are Lagrangian multipliers.

Overall Partial Lagrangian can be represented by:

$$\sum_{r\in R} L_r(x, y, z, \eta_u^r), \tag{56}$$

where $u \in \{k, l, d_p\}$

$$\text{L}_r(x, y, z, \eta_u^r) = \sum_{r\in R}\Big[\sum_{k\in K} a_k^r \ \beta log(1+\lambda_k^r) + \eta_k^r \left(\delta_k \Omega^r - \frac{\Omega^r_{max}}{E}\right) + \sum_{j\in J}\sum_{l\in L} a_{j,l}^r \ \beta log(1+\lambda_{j,l}^r) + \eta_l^r \left(\delta_j \Omega^r - \frac{\Omega^r_{max}}{E}\right) + \sum_{d_p\in D} a_{d_p}^r \ \beta log(1+\lambda_{d_p}^r) + \eta_{d_p}^r \left(\sum_{d_p\in D} \delta_{d_p} \Omega^r - \frac{\Omega^r_{max}}{E}\right)\Big]$$
(57)

$$f(\eta_u^r) = \begin{cases} \max_{a_k^r} \ L_r(x_k^r, \eta_k^r) + \max_{a_{j,l}^r} \ L_r(y_{j,l}^r, \eta_l^r) \\ + \max_{a_{d_p}^r} \ L_r(z_{d_p}^r, \eta_{d_p}^r) \\ subject\ to: \ \sum_{u\in U} a_u^r \le 1 \end{cases} \tag{58}$$

$$u \in \{k, l, d_p\} \ \forall_r \in R, \ \forall_k \in K, \ \forall_l \in L, \ \forall_{d_p} \in D$$

B. SOLVING GLOBAL VARIABLE

In a multi-agent system, each agent must have the information of global constraint. The global constraint agrees upon the consensus by using the local estimates as calculated above [61]. For our system, the global constraint 49) is coupled, and all agents must cooperatively determine it by administering mutual consensus.

Consider the reformulation of (53) in closed form as follows:

$$\max_{y_1, y_2 \dots y_e} \sum_{i=1}^{e} c_k^T y_k \tag{59}$$

where for all edge servers $k = 1, 2 \dots e$, y_k is decision vector of each edge server and c_k is it local cost. While the coupled constraint in (58) can be expressed in closed form as:

$$\sum_{i=1}^{e} A_k x_k \le 1 \tag{60}$$

$$l_i(k) = \sum_{j\in N_i(k)} \{a_j^i(k)\gamma_j(k)\} \tag{61}$$

$$x_i(k+1) \leftarrow \arg\ max_{x_i \in vert(X_i)} \ (c_i^T \ + \ l_i(k)^T A_i) x_i \tag{62}$$

$$\varphi_i(k) = max_{j\in N_i(k)}\{\rho_i(k)\} \tag{63}$$

$$\bar{\zeta}_i(k+1) = max \ [\bar{\zeta}_i(k), \ A_i x_i(k+1)] \tag{64}$$

$$\underline{\zeta_i}(k+1) = min \ [\underline{\zeta_i}(k), \ A_i x_i(k+1)] \tag{65}$$

$$\rho_i(k+1) = max[Q_i(k), \rho\{\bar{\zeta}_i(k+1) - \underline{\zeta_i}(k+1)\}] \tag{66}$$

The results of the optimization problem give us optimum throughput and SE values through the distributed method. Each edge router solves its resource optimization problem, like for MUE, SUE, and D2D users. These resulting values are then used by the slice controller, specializing in creating and deleting slices. The slice orchestrator will then cut the network slices according to the network requirements and optimum spectrum allocations.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3114629, IEEE Access

C. COMPUTATIONAL COMPLEXITY ANALYSIS

The computational complexity of any algorithm can be calculated depending on the overall number of flops it takes to execute the process. The work in [63] explains in detail the representation of a flop by simple floating-point operations. Each operation in the algorithm has its subsequent number of flops: multiplication, addition, division, logical operator, assignment operator, matrix multiplication, Etc. To compute the complexity of our proposed scheme DRA-EC, we first have to determine the number of iterations, variables, and constraints. As explained in Section IV our system is divided into two layers. The upper layer complexity can be analyzed from Algorithm 1 (BS-UE association), whose complexity can be counted as $\mathcal{O}(|\mathcal{C}||\mathcal{U}|)$ and from Algorithm 2 (D2D mode switching), the complexity is: $\mathcal{O}(|\mathcal{R}||\mathcal{U}|)$. The lower layer involves the resource allocation method using augmented Lagrange multipliers defined as $\mathcal{S}$ and $\mathcal{T}$, respectively. The complexity of this during each iteration can be computed for channel assignment as $\mathcal{O}(|\mathcal{K}||\mathcal{M}||\mathcal{N}|)$ operations. Accordingly, the updates in Lagrange multipliers need $\mathcal{O}(|\mathcal{K}||\mathcal{M}||\mathcal{N}|)$ operations according to (51), (53) and (54). Here, T is a polynomial function for sum iterations $\mathcal{O}(\mathcal{T}(|\mathcal{K}||\mathcal{M}||\mathcal{N}|)^2)$. Hence the total computational complexity of the proposed algorithm is counted as $\mathcal{O}(\mathcal{S}\mathcal{T}(|\mathcal{K}||\mathcal{M}||\mathcal{N}|)^2 + 3|D|)$.

D. CONVERGENCE ANALYSIS

The convergence of an algorithm depends on the number of iterations it takes for the output to become closer and closer to a particular estimate or limit. For iterative algorithms, there is a separate error in each stage, and the algorithm's goal is to minimize the error. The algorithm converges when the error becomes smaller and smaller value. The global optimum of an algorithm is when the system has the least possible error. Such an algorithm converges to the optimum global value.The computation of feasible region or global maxima is challenging for resource allocation problems in cellular networks. At the same time, it is significant to analyze in order to guarantee the algorithm's convergence. We proposed the DRA-EC scheme, the process of updating the spectrum allocation and Lagrange multiplier is repeated until the lowest possible error, and the algorithm converges.

Proof: Please see the reference [64] for the proof of algorithm convergence.

Our proposed distributed algorithm converges at a faster rate than the central algorithm because it is tough to compute the global maxima of a single complex central system with different types of users and their needs. Such a system will take a much longer time to converge, or in some cases, it may diverge. Therefore, we have distributed this complex system so that edge routers present at each type (macro, small and D2D) will compute their local maxima first. All these tasks are done in parallel, and the values obtained from local maxima are then analyzed to find the global maxima. Thus this algorithm converges faster with fewer chances of errors.

TABLE 2. Simulation Parameters

Parameter	Value
Macro/Small/UE distribution	Central/Random/Random
Macro-cell radius	1000 m
Small-cell radius	200 m
No. of Small-cell	4
No. of UE	100
No. of RB	20
RB Bandwidth	180 kHz
Threshold distance D2D	30 m
Transmit power MBS	43 dBm
Transmit power SBS	30 dBm
Transmit power D2D	23 dBm
Antenna gain	5 dB
Noise spectral density	-174 dBm/Hz
Noise figure	9 dB
Shadowing standard deviation	10 dB
Path loss model for MBS	128.1+37.6$log(d[km])$
Path loss model for SBS	140.7+36.7$log(d[km])$
Path loss model for D2D	148.1+40$log(d[km])$

VIII. SIMULTAION AND RESULTS

A. SIMULATION PARAMETERS

In this work, we have considered a single cell 5G-HCN that consists of one MBS located in the center with four SBS randomly distributed within the cell. We assumed that the total no. of users in the system is 200. Each BS can be associated with a maximum of 40 UEs. The coverage radius of MBS and SBS is 1000m × 1000m and 200 m, respectively. The channel is assumed to be with zero-mean and unit variance. The path-loss models considered are discussed in Section V-B [51]. The summary of performance parameters used in simulations are in Table 2.

B. RESULTS AND DISCUSSION

This subsection demonstrates the theoretical expressions numerically to evaluate the simulation results. The simulations are performed on MATLAB latest version using 1000 Monte-Carlo simulations. We used Intel(R) Core(TM) i5-6200U CPU @ 2.40GHz 16GB RAM with 64-bit Windows 10 operating system. The proposed setup is for maximizing the overall system throughput and spectral efficiency.

Initially, we evaluate the performance of the proposed DRA-EC scheme in 5G-HCN without using D2D users. Fig. 2 is the graphical representation of average throughput vs. no. of cellular users in the system. It shows that the average system throughput value for both MBS and all SBSs increases with the increase in the no. of users, and this value is highest for MBS users because of its high transmission power (43 dBm). The transmit power of all the SBSs is the same (30 dBm); that is why their average received throughput values appear identical but actually, they are different. By zooming the plot, we can conclude that values are very close, with a slight difference.

Similarly, Fig. 3 illustrates the comparison of average SE achieved vs. no. of cellular users in each BS (MBS and SBS). It is clear from the plot that the average SE curve for both MBS and all SBSs increases with the increase in the no. of users. As explained above, the higher value of MBS is due to the more significant value of MBS transmit power compared to SBSs. The portion of the graph is zoomed to show that there is a difference in values of each SBS, but they are very close.

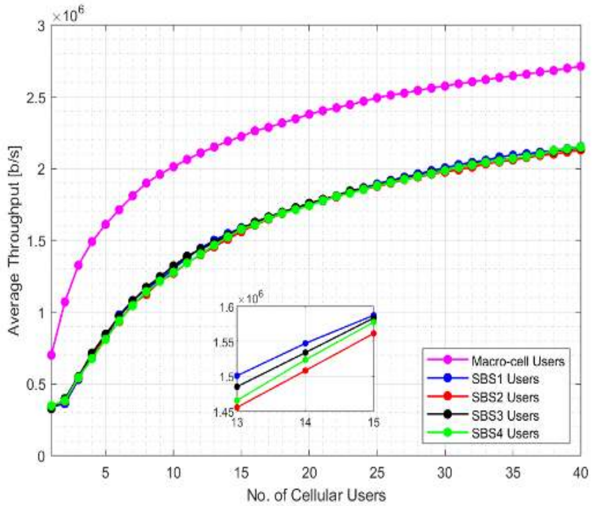

FIGURE 2. Avg. Throughput vs. Cellular users

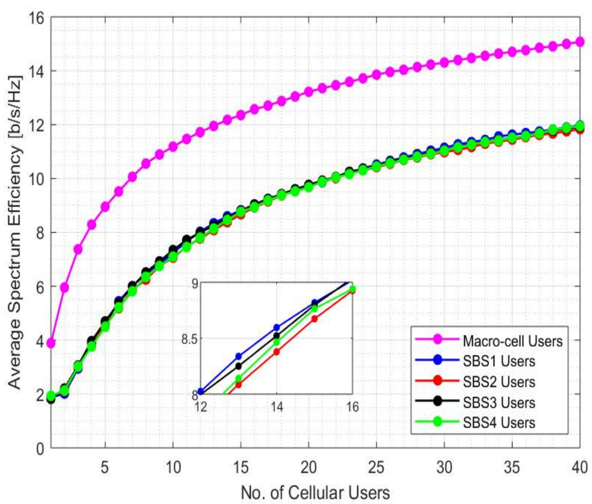

FIGURE 3. Avg. SE vs. Cellular users

Fig. 4 represents the analysis of the proposed DRA-EC scheme with D2D users. In Fig. 4, we have compared the effect on the average throughput by increasing the no. of D2D pairs in each BS. We varied the no. of D2D pairs allowed in each BS from 2 to 10 and then computed the average received throughput in each case. The rest of the parameters are kept the same as in Table 2. We observed that the system throughput and efficiency increase by adding the D2D pairs to the cellular network. Further, we concluded that the trend increases with the number of D2D pairs in the system because more users can reuse the spectrum. It is noted that the cellular users must meet the minimum SINR criteria when increasing the number of D2D pairs.

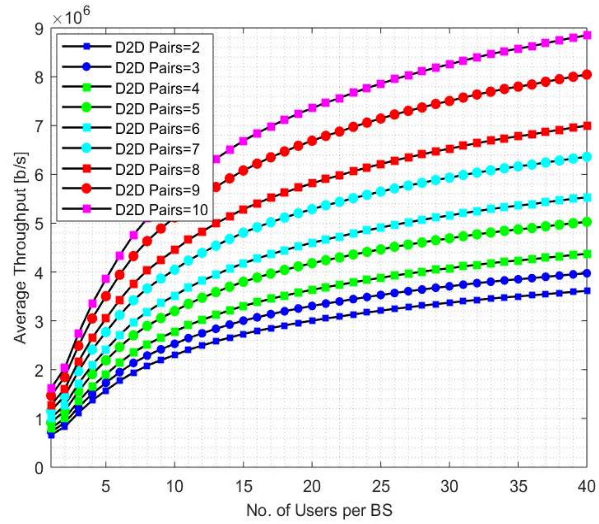

FIGURE 4. Avg. Throughput vs. No. of users with D2D

We compared our proposed DRA-EC scheme with other centralized and distributed methods to evaluate the performance.

1) **RA-CO**: Resource allocation with central optimization. We performed a simulation analysis of our proposed system model using central optimization in this scheme. The rest of the parameters are the same as with distributed.
2) **RA-SD [33]**: This scheme is on resource allocation using a semi-distributed method for D2D based cellular networks.
3) **JSPA-CO [31]**: Joint spectrum and power allocation scheme to observe the system performance by using a centralized optimization scheme.
4) **JSPA-DO [31]**: The decentralized approach to compare the system (joint spectrum and power allocation) performance in D2D based system.
5) **RO**: Random optimization scheme is designed by randomly assigning resources without optimization technique.

Fig. 5 demonstrates the comparison of average throughput with no. of users. We considered six different schemes for evaluation. It is concluded from the results that our proposed scheme DRA-EC with D2D users is near to the RA-CO. The average throughput of both these schemes is highest compared to others due to heterogeneous and D2D based network design. The throughput value decreases when DRA-EC and

RA-CO schemes are considered without D2D. The curve for JSPA-CO has the lowest value because it is based on a single macrocell instead of a heterogeneous network design. The significance of this figure is that the heterogeneous cellular network setup with macro-cells, small cells, and D2D is most efficient in managing the spectrum.

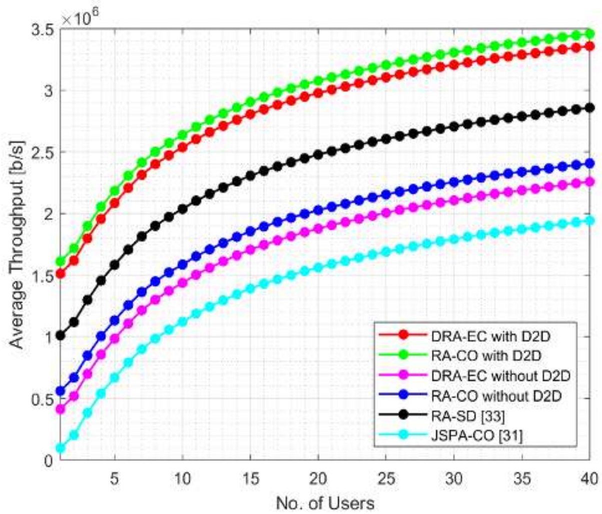

FIGURE 5. Avg. Throughput vs No. of Users

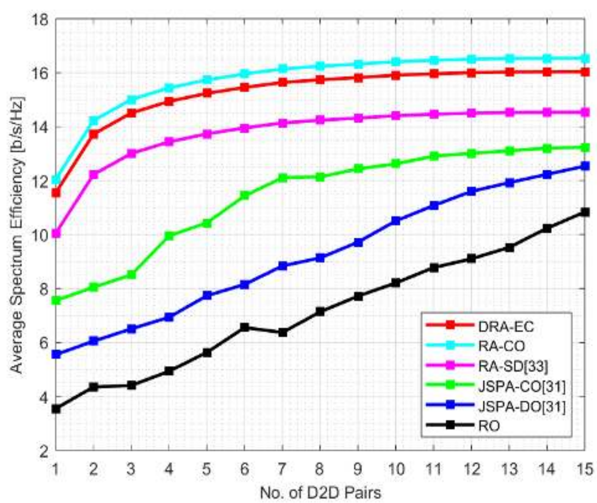

FIGURE 6. Avg. SE vs. No. of D2D pairs

Fig. 6 depicts the analysis of average achievable SE vs. no. of D2D pairs. We compared our results with different schemes as mentioned earlier for evaluation. It is concluded that our proposed system outperforms in performance and SE compared to all the other methods except RA-CO. The SE [b/s/Hz] values in the DRA-EC scheme are very close to the central optimization, which is fair enough. However, our proposed distributed system (DRA-EC) performs better than the central (RA-CO) due to its computational efficiency. All the calculations will be done in parallel by edge computing in each BS and then sent to the central controller in the distributed scheme. This will save time and reduce the load on the central system. On the other hand, the central optimization method will take more time in computation, and the entire load is on a central system.

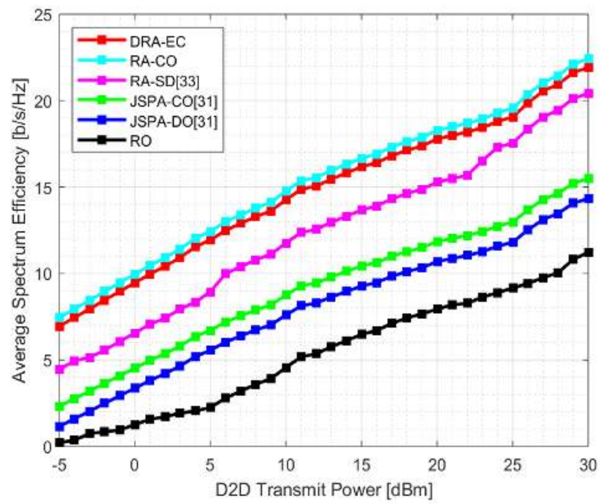

FIGURE 7. Avg. SE vs. D2D transmit power

Fig. 7 demonstrates the relation of average SE achieved vs. D2D transmit power. The results of the proposed system are compared with other schemes. We varied the value of transmit power from -5dBm to 30dBm, and the rest of the parameters are the same as in Table 2. The SE for all the cases increases with the increase in D2D transmit power, and the value is highest for both proposed schemes DRA-EC and RA-CO. The results of distributed optimization are very close to the central optimization scheme. However, the power should be increased in a controlled manner because above a specific value, it will start interfering with the cellular users and will degrade their performance requirements.

Fig. 8 represents the analysis of computational time taken by both the schemes DRA-EC and RA-CO. The plot shows the trend of algorithm computation time vs. no. of iterations. It can be concluded that the time taken in DRA-EC is significantly less than RA-CO because the latter performs all the processing tasks centrally. This will burden the system by increasing computation complexity. On the other hand, in DRA-EC, each base station has its edge computing server to compute the tasks in parallel. Thus the optimization problem is solved in a distributed manner so that it will take less time.

C. PERFORMANCE EVALUATION

We analyzed in detail the performance gap of both algorithms RA-CO (central) and DRA-EC (distributed) theoretically and analytically. The primary significance of our proposed distributed scheme (DRA-EC) is that its performance results are very close to the central optimization, taking less computation time than other methods. From the results in Fig. 6, Fig. 7 and Fig. 8 we have proved that our proposed

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3114629, IEEE Access

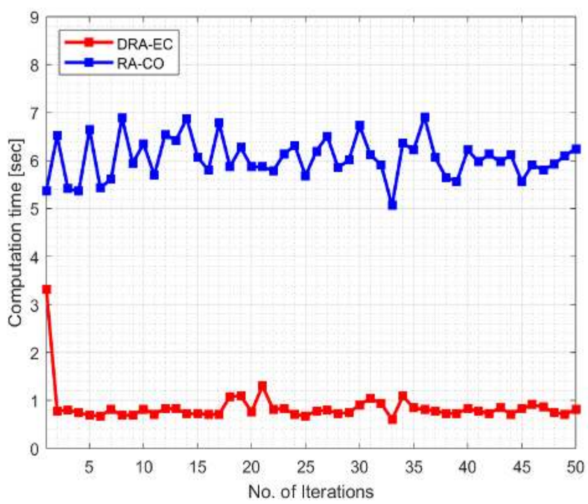

FIGURE 8. Algorithm Computation Time vs. No. of Iterations

scheme is better in performance as compared to others. In RA-CO, the central slice controller gathers data like CSI and interference from BS before the resource allocation. This process is very complex for networks with constantly varying CSI, and when this information exchange has to repeat many times, that will burden the network. All the processing performed is central, which will take a longer time to execute and will slow down the system performance. On the other hand, the computational load in our proposed (DRA-EC) scheme is substantially less. Each BS has an edge router in this scheme that collects the CSI (interference) at the edge layer and optimizes its resource allocation problem locally. The computational tasks are distributed in parallel, and the information needed at each edge router is only from its respective link. The central controller will then cut the required optimized network slice according to the results received from each edge server. This method will save much time and offload the core network, which will considerably improve system performance. In the RA-CO approach, the system must know the interference information among D2D pairs and cellular users (MUs and SUs) to D2D receivers. This data collection is very complicated practically and will involve much overhead, while this data is collected locally only in DRA-EC. The iterations might cause additional overhead, which is pretty low as the BS is only interested in data collection and broadcasting.

IX. CONCLUSION AND FUTURE RECOMMENDATIONS

This work presents a distributed scheme with resource (spectrum) optimization in D2D based under-laying 5G-HCN with network slicing. The complete system reuses the spectrum while considering the effect of interference upon allocating resources to each type of user (MUE, SUE, and DUE). We have proposed a multi-layered framework for our model comprising upper and lower layers. The upper layer solves the UE-BS association problem using a greedy algorithm and a D2D mode switching problem using a distance-dependent algorithm. The lower layer solves the spectrum optimization problem of a complete system with the DRA-EC approach. The proposed scheme distinguishes itself from existing models because distributed edge computing devices solve the underlying resource optimization problem. This approach reduces the load of a central controller and minimizes the computational time. The results of SE and throughput are very close as in the centralized process which is good enough. After optimization, these results can be used by central slice controller for efficient utilization of spectrum by slicing the network according to system demands. Our proposed scheme can be extended to function at the core layer of the network to relate backhaul capacity with spectrum efficiency and develop techniques for efficient network slicing. In the future, we can create application-based scenarios for D2D and MEC platforms such as content caching and information-centric networks. Furthermore, we can formulate new intelligent schemes of resource allocation in 5G networks leveraging machine learning and deep reinforcement learning techniques.

.

APPENDIX A "PROOF OF PROPOSITION 1"

The 0-1 multiple knapsacks are defined to be as a combinatorial optimization problem in mathematics as follows:

Definition: Let us assume that $\mathcal{V}$ and $\mathcal{W}$ are the sets of items and knapsacks, respectively. Each item $i \in \mathcal{V}$ has a weight of y_i and provides a non-negative profit value of z_i, whereas each knapsack $j \in \mathcal{W}$ has a capacity of Y_j. The solution to a given problem is to fill all the knapsacks with the available items. This technique will maximize the overall profit for each knapsack without surpassing the total capacity.

By evaluating the above definition for Eq. (7), we get variables mapping in terms of our problem as:

$\mathcal{V} = \mathcal{U}_a$, $\mathcal{W} = \mathcal{C}$, $i = u$, $j = c$, $y_i = \mathcal{R}_{min,u}$, $Y_j = \mathrm{Q}_{max}$, $z_i = \Gamma_{c,u}$.

APPENDIX B "PROOF OF PROOF OF PROPOSITION 2"

Let us suppose that the possible (a_k), $(a_{j,l})$, (a_{d_p}) and (δ) are given.

If we consider Eq. (31) subject to (32) - (40). Obviously, the proposed function and all its constraints are smooth. The throughput equations R_k, $R_{j,l}$ and R_{d_p} are non-linear logarithmic functions with non-convex constraints (35) - (37). Therefore,(31) is a smooth, non-linear and non-convex optimization problem. These types of programming problems have been demonstrated as commonly NP-hard and are computationally troublesome [55]. Moreover, the problem can be regarded as a throughput (Data-rate) maximization problem which has already been proved to be NP-hard in [56] and [65] that further justified the NP-hardness of (31).

APPENDIX C "PROOF OF PROOF OF PROPOSITION 3"

[61] investigated the solution of multi-agent constrained problem using the distributed algorithm. They proved that

the method of augmented Lagrangian multiplier method converges at a faster rate practically as compared to other distributed techniques [62]. The convergence relies on the separability of the global constraints into local constraints. Once we find the local maxima/minima, then to achieve the global maxima, we use the consensus-based algorithm. It is a method to attain an agreement on a particular value between various distributed procedures. Consensus is built in a multi-agent system to ensure the reliability of a system. Finding Solutions for such types of problems are significant in distributed and multiple nodes systems as proved in [61]. Similarly, our proposed approach can solve this following these steps.

REFERENCES

[1] M. Report, “"ericsson mobility report november 2020",” https://www.ericsson.com/, Nov 2020.

[2] A. Saddoud, W. Doghri, E. Charfi, and L. C. Fourati, “5g radio resource management approach for multi-traffic iot communications,” Computer Networks, vol. 166, p. 106936, 2020.

[3] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, “Network slicing in 5g: Survey and challenges,” IEEE Communications Magazine, vol. 55, no. 5, pp. 94–100, 2017.

[4] Y. Niu, C. Gao, Y. Li, L. Su, D. Jin, Y. Zhu, and D. O. Wu, “Energy-efficient scheduling for mmwave backhauling of small cells in heterogeneous cellular networks,” IEEE Transactions on Vehicular Technology, vol. 66, no. 3, pp. 2674–2687, 2016.

[5] M. Li and H.-L. Tsai, “Design and evaluation of a hybrid d2d discovery mechanism in 5g cellular networks,” in 2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN). IEEE, 2018, pp. 641–643.

[6] M. Kamruzzaman, N. I. Sarkar, J. Gutierrez, and S. K. Ray, “A mode selection algorithm for mitigating interference in d2d enabled next-generation heterogeneous cellular networks,” in 2019 International Conference on Information Networking (ICOIN). IEEE, 2019, pp. 131–135.

[7] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, “Network slicing for 5g with sdn/nfv: Concepts, architectures, and challenges,” IEEE Communications Magazine, vol. 55, no. 5, pp. 80–87, 2017.

[8] L. Nadeem, M. A. Azam, Y. Amin, M. A. Al-Ghamdi, K. K. Chai, M. F. N. Khan, and M. A. Khan, “Integration of d2d, network slicing, and mec in 5g cellular networks: Survey and challenges,” IEEE Access, vol. 9, pp. 37 590–37 612, 2021.

[9] Q. Ye and W. Zhuang, “Distributed and adaptive medium access control for internet-of-things-enabled mobile networks,” IEEE Internet of Things Journal, vol. 4, no. 2, pp. 446–460, 2016.

[10] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, “5g wireless network slicing for embb, urllc, and mmtc: A communication-theoretic view,” IEEE Access, vol. 6, pp. 55 765–55 779, 2018.

[11] M. Jiang, M. Condoluci, and T. Mahmoodi, “Network slicing management & prioritization in 5g mobile systems,” in European Wireless 2016; 22th European Wireless Conference. VDE, 2016, pp. 1–6.

[12] X. Xu, H. Zhang, X. Dai, Y. Hou, X. Tao, and P. Zhang, “Sdn based next generation mobile network with service slicing and trials,” China Communications, vol. 11, no. 2, pp. 65–77, 2014.

[13] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, “Network slicing and softwarization: A survey on principles, enabling technologies, and solutions,” IEEE Communications Surveys & Tutorials, vol. 20, no. 3, pp. 2429–2453, 2018.

[14] Z. Ning, X. Wang, and J. Huang, “Mobile edge computing-enabled 5g vehicular networks: Toward the integration of communication and computing,” ieee vehicular technology magazine, vol. 14, no. 1, pp. 54–61, 2018.

[15] T. X. Tran, A. Hajisami, P. Pandey, and D. Pompili, “Collaborative mobile edge computing in 5g networks: New paradigms, scenarios, and challenges,” IEEE Communications Magazine, vol. 55, no. 4, pp. 54–61, 2017.

[16] K. Zhang, Y. Mao, S. Leng, Q. Zhao, L. Li, X. Peng, L. Pan, S. Maharjan, and Y. Zhang, “Energy-efficient offloading for mobile edge computing in 5g heterogeneous networks,” IEEE access, vol. 4, pp. 5896–5907, 2016.

[17] D. Gutierrez-Estevez, Ö. Bulakci, M. Ericson, A. Prasad, E. Pateromichelakis, J. Belschner, P. Arnold, and G. Calochira, “Ran enablers for 5g radio resource management,” in 2017 IEEE Conference on Standards for Communications and Networking (CSCN). IEEE, 2017, pp. 1–6.

[18] Y. Xu, G. Gui, H. Gacanin, and F. Adachi, “A survey on resource allocation for 5g heterogeneous networks: current research, future trends and challenges,” IEEE Communications Surveys & Tutorials, 2021.

[19] Y. Xu, H. Xie, and R. Q. Hu, “Max-min beamforming design for heterogeneous networks with hardware impairments,” IEEE Communications Letters, vol. 25, no. 4, pp. 1328–1332, 2020.

[20] Y. Xu, G. Gui, T. Ohtsuki, H. Gacanin, B. Adebisi, H. Sari, and F. Adachi, “Robust resource allocation for two-tier hetnets: An interference-efficiency perspective,” IEEE Transactions on Green Communications and Networking, 2021.

[21] Y. Xu, H. Xie, C. Liang, and F. R. Yu, “Robust secure energy efficiency optimization in swipt-aided heterogeneous networks with a non-linear energy harvesting model,” IEEE Internet of Things Journal, 2021.

[22] H. Wang, S.-H. Leung, and R. Song, “Uplink area spectral efficiency analysis for multichannel heterogeneous cellular networks with interference coordination,” IEEE Access, vol. 6, pp. 14 485–14 497, 2018.

[23] Z. Tan, X. Li, F. R. Yu, L. Chen, H. Ji, and V. C. Leung, “Joint access selection and resource allocation in cache-enabled hcns with d2d communications,” in 2017 IEEE wireless communications and networking conference (WCNC). IEEE, 2017, pp. 1–6.

[24] A. A. Algedir and H. H. Refai, “Energy efficiency optimization and dynamic mode selection algorithms for d2d communication under hetnet in downlink reuse,” IEEE Access, vol. 8, pp. 95 251–95 265, 2020.

[25] J. Iqbal, M. A. Iqbal, A. Ahmad, M. Khan, A. Qamar, and K. Han, “Comparison of spectral efficiency techniques in device-to-device communication for 5g,” IEEE Access, vol. 7, pp. 57 440–57 449, 2019.

[26] K. Zia, N. Javed, M. N. Sial, S. Ahmed, A. A. Pirzada, and F. Pervez, “A distributed multi-agent rl-based autonomous spectrum allocation scheme in d2d enabled multi-tier hetnets,” IEEE Access, vol. 7, pp. 6733–6745, 2019.

[27] H. Min, W. Seo, J. Lee, S. Park, and D. Hong, “Reliability improvement using receive mode selection in the device-to-device uplink period underlaying cellular networks,” IEEE Transactions on Wireless Communications, vol. 10, no. 2, pp. 413–418, 2010.

[28] B. Zhang, X. Mao, J.-L. Yu, and Z. Han, “Resource allocation for 5g heterogeneous cloud radio access networks with d2d communication: A matching and coalition approach,” IEEE Transactions on Vehicular Technology, vol. 67, no. 7, pp. 5883–5894, 2018.

[29] A. Ramezani-Kebrya, M. Dong, B. Liang, G. Boudreau, and S. H. Seyedmehdi, “Joint power optimization for device-to-device communication in cellular networks with interference control,” IEEE Transactions on Wireless Communications, vol. 16, no. 8, pp. 5131–5146, 2017.

[30] G. Sun, H. Al-Ward, G. O. Boateng, and W. Jiang, “Content-aware d2d caching for reducing visiting latency in virtualized cellular networks.” TIIS, vol. 13, no. 2, pp. 514–535, 2019.

[31] W. Zhao and S. Wang, “Resource sharing scheme for device-to-device communication underlaying cellular networks,” IEEE Transactions on Communications, vol. 63, no. 12, pp. 4838–4848, 2015.

[32] R. Yin, C. Zhong, G. Yu, Z. Zhang, K. K. Wong, and X. Chen, “Joint spectrum and power allocation for d2d communications underlaying cellular networks,” IEEE Transactions on Vehicular Technology, vol. 65, no. 4, pp. 2182–2195, 2015.

[33] M. Klügel and W. Kellerer, “Semi-decentralized interference aware scheduling in d2d-enabled cellular networks,” IEEE Access, vol. 8, pp. 132 285–132 301, 2020.

[34] Y. He, J. Ren, G. Yu, and Y. Cai, “Joint computation offloading and resource allocation in d2d enabled mec networks,” in ICC 2019-2019 IEEE International Conference on Communications (ICC). IEEE, 2019, pp. 1–6.

[35] A. Omran, L. Sboui, B. Rong, H. Rutagemwa, and M. Kadoch, “Joint relay selection and load balancing using d2d communications for 5g hetnet mec,” in 2019 IEEE International Conference on Communications Workshops (ICC Workshops). IEEE, 2019, pp. 1–5.

[36] U. Saleem, Y. Liu, S. Jangsher, X. Tao, and Y. Li, “Latency minimization for d2d-enabled partial computation offloading in mobile edge comput-

ing," IEEE Transactions on Vehicular Technology, vol. 69, no. 4, pp. 4472–4486, 2020.

[37] Y. Li, G. Xu, K. Yang, J. Ge, P. Liu, and Z. Jin, "Energy efficient relay selection and resource allocation in d2d-enabled mobile edge computing," IEEE Transactions on Vehicular Technology, 2020.

[38] U. Saleem, Y. Liu, S. Jangsher, Y. Li, and T. Jiang, "Mobility-aware joint task scheduling and resource allocation for cooperative mobile edge computing," IEEE Transactions on Wireless Communications, 2020.

[39] S. Schaller and D. Hood, "Software defined networking architecture standardization," Computer standards & interfaces, vol. 54, pp. 197–202, 2017.

[40] V.-G. Nguyen, A. Brunstrom, K.-J. Grinnemo, and J. Taheri, "Sdn/nfv-based mobile packet core network architectures: A survey," IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1567–1602, 2017.

[41] T. LeAnh, N. H. Tran, D. T. Ngo, and C. S. Hong, "Resource allocation for virtualized wireless networks with backhaul constraints," IEEE Communications Letters, vol. 21, no. 1, pp. 148–151, 2016.

[42] S. Parsaeefard, R. Dawadi, M. Derakhshani, and T. Le-Ngoc, "Joint user-association and resource-allocation in virtualized wireless networks," IEEE Access, vol. 4, pp. 2738–2750, 2016.

[43] K. Wang, H. Li, F. R. Yu, and W. Wei, "Virtual resource allocation in software-defined information-centric cellular networks with device-to-device communications and imperfect csi," IEEE Transactions on Vehicular Technology, vol. 65, no. 12, pp. 10 011–10 021, 2016.

[44] L. Gao, P. Li, Z. Pan, N. Liu, and X. You, "Virtualization framework and vcg based resource block allocation scheme for lte virtualization," in 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring). IEEE, 2016, pp. 1–6.

[45] T. M. Ho, N. H. Tran, L. B. Le, Z. Han, S. A. Kazmi, and C. S. Hong, "Network virtualization with energy efficiency optimization for wireless heterogeneous networks," IEEE Transactions on Mobile Computing, vol. 18, no. 10, pp. 2386–2400, 2018.

[46] Y. Sun, M. Peng, S. Mao, and S. Yan, "Hierarchical radio resource allocation for network slicing in fog radio access networks," IEEE Transactions on Vehicular Technology, vol. 68, no. 4, pp. 3866–3881, 2019.

[47] M. Richart, J. Baliosian, J. Serrat, and J.-L. Gorricho, "Resource slicing in virtual wireless networks: A survey," IEEE Transactions on Network and Service Management, vol. 13, no. 3, pp. 462–476, 2016.

[48] Y. L. Lee, J. Loo, T. C. Chuah, and L.-C. Wang, "Dynamic network slicing for multitenant heterogeneous cloud radio access networks," IEEE Transactions on Wireless Communications, vol. 17, no. 4, pp. 2146–2161, 2018.

[49] Q. Ye, W. Zhuang, S. Zhang, A.-L. Jin, X. Shen, and X. Li, "Dynamic radio resource slicing for a two-tier heterogeneous wireless network," IEEE Transactions on Vehicular Technology, vol. 67, no. 10, pp. 9896–9910, 2018.

[50] 3GPP, "Evolved universal terrestrial radio access (e-utra); physical channels and modulation," 2017.

[51] J. S. Seybold, Introduction to RF propagation. John Wiley & Sons, 2005.

[52] D. S. Johnson, "The np-completeness column: an ongoing guide," Journal of Algorithms, vol. 6, no. 3, pp. 434–451, 1985.

[53] L. Liberti and C. Lavor, Euclidean Distance Geometry. Springer, 2017.

[54] E. LTE, "Evolved universal terrestrial radio access (e-utra); radio frequency (rf) requirements for lte pico node b (3gpp tr 36.931 version 9.0. 0 release 9)," 2011.

[55] K. G. Murty and S. N. Kabadi, "Some np-complete problems in quadratic and nonlinear programming," Tech. Rep., 1985.

[56] C. Kai, H. Li, L. Xu, Y. Li, and T. Jiang, "Joint subcarrier assignment with power allocation for sum rate maximization of d2d communications in wireless cellular networks," IEEE Transactions on Vehicular Technology, vol. 68, no. 5, pp. 4748–4759, 2019.

[57] P. Yi, Y. Hong, and F. Liu, "Distributed gradient algorithm for constrained optimization with application to load sharing in power systems," Systems & Control Letters, vol. 83, pp. 45–52, 2015.

[58] Y. Zhao, Y. Liu, G. Wen, and G. Chen, "Distributed optimization for linear multiagent systems: Edge-and node-based adaptive designs," IEEE Transactions on Automatic Control, vol. 62, no. 7, pp. 3602–3609, 2017.

[59] M. Li, "Generalized lagrange multiplier method and kkt conditions with an application to distributed optimization," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 66, no. 2, pp. 252–256, 2018.

[60] G. Ruan, H. Zhong, J. Wang, Q. Xia, and C. Kang, "Neural-network-based lagrange multiplier selection for distributed demand response in smart grid," Applied Energy, vol. 264, p. 114636, 2020.

[61] Y. Zhang and M. M. Zavlanos, "A consensus-based distributed augmented lagrangian method," in 2018 IEEE Conference on Decision and Control (CDC). IEEE, 2018, pp. 1763–1768.

[62] N. Chatzipanagiotis, D. Dentcheva, and M. M. Zavlanos, "An augmented lagrangian method for distributed optimization," Mathematical Programming, vol. 152, no. 1, pp. 405–434, 2015.

[63] C. Van Loan and G. Golub, "Matrix computations, vol. 3," 2012.

[64] D. P. Bertsekas, Constrained optimization and Lagrange multiplier methods. Academic press, 2014.

[65] Y.-F. Liu and Y.-H. Dai, "On the complexity of joint subcarrier and power allocation for multi-user ofdma systems," IEEE transactions on Signal Processing, vol. 62, no. 3, pp. 583–596, 2013.

LUBNA NADEEM obtained her BS and MS in Telecommunication Engineering from the University of Engineering and Technology (UET) Taxila, Pakistan, in 2011 and 2013, respectively. She is currently doing a Ph.D. and associated with Agile Creative Technologies for Smart Electromagnetic Novel Applications (ACTSENA) Research Group, UET Taxila. She is currently working as a Research Associate in Communication Systems Research Group, School of Electronic Engineering and Computer Science (EECS) Queen Marry University of London (QMUL), UK. She is working as a Lecturer with the Telecommunication Engineering Department, UET Taxila, since March 2014. Her main research interests include 5G Wireless communication, Radio Resource Allocation and Optimization in Cellular Networks, Internet of Things with Machine Learning, and Deep Learning.

YASAR AMIN received the BSc degree in Electrical Engineering with specialization in Telecommunication. His MSc is in Electrical Engineering with a specialization in System on Chip Design. Also, his Ph.D. is in Electronic and Computer Systems from Royal Institute of Technology (KTH), Sweden, with the research focus on printable green RFID antennas for embedded sensors. He is currently a Professor and chairman of the Telecommunication Engineering Department, University of Engineering and Technology Taxila, Pakistan. His research interests include the next-generation mobile communication systems, millimeter-wave and terahertz antenna array, implantable, wearable electronics in microwave applications. He has authored or co-authored more than 100 international technical papers in conferences and journals.

JONATHAN LOO (A.K.A. KOK KEONG LOO) received his M.Sc. degree in Electronics (with Distinction) and the Ph.D. degree in Electronics and Communications from the University of Hertfordshire, Hertfordshire, U.K., in 1998 and 2003, respectively. He is currently a Chair Professor in Computing and Communication Engineering at the School of Computing and Engineering, University of West London, U.K. His research interests include machine learning and AI, natural language and image processing, IoT/cyber-physical systems, cyber-security, information-centric networking, wireless/mobile systems, and communication networks. He has successfully graduated 18 Ph.D. students and has co-authored over 350 journal and conference papers in the areas mentioned above.

MUHAMMAD AWAIS AZAM received his B.Sc. degree in Computer Engineering from the University of Engineering and Technology (UET), Taxila, Pakistan and won Gold Medal in 2006, his M.Sc. degree in Wireless Networks (with distinction) from Queen Mary University, London, the UK in 2008 and his Ph.D. degree in Pervasive and Ubiquitous Computing from London, the UK in 2012. He is currently working as a Head of School IT at Whitecliffe. There he leads a research team of M.Sc. and Ph.D. students in pervasive and ubiquitous computing. His research interest includes network architecture, IoT, network security, embedded systems, wireless communications, opportunistic networks, and recommender systems.

KOK KEONG CHAI (MICHAEL) Kok Keong Chai (Michael) received the B.Eng. (Hons), M.Sc., and Ph.D. degrees, in 1998, 1999, and 2007, respectively. He joined the School of Electronic Engineering and Computer Science (EECS), Queen Mary University of London (QMUL), in August 2008. He is currently a Reader of the Internet of Things and the Queen Mary Director of Joint Programme with the Beijing University of Posts and Telecommunications (BUPT) and a Communication Systems Research Group member, QMUL. His current research interests include sensing and prediction in distributed smart grid networks, smart energy charging schemes, applied blockchain technologies, dynamic resource management, wireless communications, and medium access control (MAC) for M2M communications and Networks. He has authored more than 65 technical journals and conference papers in the areas mentioned above.