Quality of service based resource allocation in D2D enabled 5G-CNs with network slicing

Nadeem, Lubna, Amin, Yasar, Loo, Jonathan ORCID: https://orcid.org/0000-0002-2197-8126, Azam, Muhammad and Kok Keong, Chai (2022) Quality of service based resource allocation in D2D enabled 5G-CNs with network slicing. Physical Communication, 52. ISSN 1874-4907

http://dx.doi.org/10.1016/j.phycom.2022.101703

Quality of Service based Resource Allocation in D2D enabled 5G-CNs with Network Slicing

Lubna Nadeem[a,*], Yasar Amin[a], Jonathan Loo[b], Muhammad A. Azam[c] and Kok Keong Chai[d]

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

ARTICLE INFO

Keywords:
Device-to-device
eMBB
mMTC
Network slicing
QoS requirements
5G cellular networks
Distributed optimization

ABSTRACT

With the rapid new advancements in technology, there is an enormous increase in devices and their versatile need for services. Fifth-generation (5G) cellular networks (5G-CNs) with network slicing (NS) have emerged as a necessity for future mobile communication. The available network is partitioned logically into multiple virtual networks to provide an enormous range of users' specific services. Efficient resource allocation methods are critical to delivering the customers with their required Quality of Service (QoS) priorities. In this work, we have investigated a QoS based resource allocation (RA) scheme considering two types of 5G slices with different service requirements; (1) enhanced Mobile Broadband (eMBB) slice that requires a very high data rate and (2) massive Machine Type Communication (mMTC) slice that requires extremely low latency. We investigated the device-to-device (D2D) enabled 5G-CN model with NS to assign resources to users based on their QoS needs while considering the cellular and D2D user's data rate requirements. We have proposed a Distributed Algorithm (DA) with edge computation to solve the optimization problem, which is novel as edge routers will solve the problem locally using the augmented Lagrange method. They then send this information to the central server to find the global optimum solution utilizing a consensus algorithm. Simulation analysis proves that this scheme is efficient as it assigns resources based on their QoS requirements. This scheme is excellent in reducing the central load and computational time.

1. Introduction

5G cellular networks broadly interpreted as International Mobile Telecommunications-2020 (IMT-2020) have revolutionized the world by reshaping how human beings and machines respond. The increasingly rapid shift is due to the growing demand for data rates, a significant no. of connected devices, and the most extensive applications worldwide [1]. 5G-CNs promise to deliver the users a dynamic range of services while considering their QoS aware constraints in traditional cellular networks. Two of the most essential 5G used cases proposed by standard organizations [2, 3], that are considered in this work: (1) enhanced Mobile Broadband (eMBB) that deliver very high data rate and capacity with fast link setup, (2) massive Machine Type Communication (mMTC) that provide support for applications having extremely low latency demand.

D2D based 5G cellular communication is known to be significant these days due to their increased benefits of providing less transmission delay, reducing system load with low power consumption improving system performance and throughput [4]. 5G network slicing is also an emerging technology of the future because it divides the available network resource into logical network slices that can be customized according to the user's service-based demands [5]. Mobile edge-based computation is a very recent concept where the load of the central controller is reduced as edge devices perform most of the control and management tasks. This also reduces the transmission delay of a system [6]. In the future, these technologies will deliver the users with real-time experience for smart online games, autonomous motor vehicles, delay-sensitive medical emergency applications, remote and intelligent services, and many more like this.

The scope of all future cellular networks design is to integrate the important 5G technologies. Combining D2D and network slicing in 5G cellular networks will provide all of the benefits in a single network design [7]. The primary goal of research for all future enhancements these days is to design an innovative system model capable of connecting with an enormous number of devices supporting a vast range of customized services with minimum system overhead and computation delay. Efficient RA in such a system is the main challenge these days, which must be addressed [8]. Network slicing allows the users to use their required QoS slice to assign resources based on their needs.

The proposed research work is based on a system model that optimizes the RA and throughput based on QoS-based constraints for two types of services (eMBB and mMTC). Distributed optimization methods are less commonly used in literature than central optimization schemes. We have proposed a mobile edge computing-based approach to solve the edge routers located near the devices to collect the required information. This information is then sent to the slice controller, assigning the network slice based on the received data. Such a system will reduce the central system's computation time and load compared to conventional optimization schemes.

Section 2 is about the recent related work, section 3 demonstrates the motivation and research gap, and section 4

lubna.nadeem@uettaxila.edu.pk (L. Nadeem)

depicts the significant contributions of this work. Section 5 is about our proposed system model, and section 6 discusses the details of problem formulation. Section 7 is about the proposed solution of the optimization problem, section 8 discusses the performance evaluation of simulation outcomes, and lastly, section 9 discusses the conclusion with future works.

2. Related Work

Previously, a lot of research has been done to exploit the efficient RA schemes in 5G cellular networks with D2D and NS individually. D2D communication technology is prominent nowadays due to its enormous benefits of very less transmission delay and high throughput, but efficient RA is essential in this case [9, 10, 11]. D2D underlay system is considered that share and reuse the resources of cellular users [12]. Most existing literature [13, 14, 15, 16] on D2D is with traditional cellular networks without using 5G and NS technology. The work done in [17] and [18] proposed resource management in D2D using 5G but lacks the NS and QoS scenario.

Many researchers have investigated the RA using NS in 5G cellular networks [19, 20, 21, 22, 23, 24] but the QoS-based RA remained ignored for many years. There are many related works, but all of them were missing in one way or the other. Like the work presented in [25] focused on minimizing the end-to-end latency of backhaul networks by virtual RA. Similarly, [26] proposed a context-aware RA scheme, and [27] investigated the delay-aware RA in fog networks. In [28, 29] the aim was to optimize the capacity, and in [30], the goal was to maximize the revenue of NS based heterogeneous cellular network. The work in [31], investigated a dynamic RA scheme and [32], proposed a resource slicing method; the aim of both of them was to maximize the system utility. All of the works mentioned above were on RA and optimization in virtualized/NS-based 5G cellular networks. Still, none considered using D2D, which improves the throughput and reduces delay. Also, no work focused on QoS-based RA for optimization that assigns the resources to users according to their service requirements.

RA considering QoS was investigated in detail in some recent works using the central approach for optimization [33, 34, 35, 24]. The QoS-based RA was exploited in [36] for vehicular networks. Similarly, in [37] for slice-aware RA using dinkelbach method and, in [38] using genetic algorithm and branch and bound method for NS based 5G heterogeneous CNs. However, these works lack the D2D based application scenario and only partially solved the problem of maximizing the system utility. All these works solved the problem of RA using centralized optimization methods, but none of them used the distributed approach, which is the major novelty of this work. In our previous work [8], we solved the problem of spectrum efficiency maximization using the distributed method in D2D enabled 5G-CNs with network slicing. Still, the QoS-based customized network slicing (mMTC and eMBB) was missing in that work.

3. Research Gaps & Motivation

In this work, we have proposed a 5G-CN with NS technology comprising D2D and capable of providing the QoS (eMBB and mMTC) based RA to users. There is no such work done previously to the best of the author's knowledge. Most of the research in the past considered either D2D or NS individually with 5G cellular networks. None of the works considered a system model with D2D application combined with NS in a 5G cellular network. Moreover, the existing literature mostly investigated the RA and optimization using central methods. Some works which have discussed the distributed approach did not solve the problem using edge computation. We have adopted the distributed edge computing technique in our proposed system, which is novel as there is no similar work done before. The edge routers near the devices will collect the information of the type of devices (MBS or D2D), channel information, and their slice type (eMBB or mMTC) requirement. With the augmented Lagrange multiplier method, every edge server will solve its optimizing problem locally. The central server or slice controller will then use these local estimates to find the global maxima or minima. QoS-based RA is the prominent approach for managing cellular networks as the main goal of all upcoming research is to design application/service-based solutions. D2D and NS combined with edge computation will achieve these goals delivering improved system performance and less cost.

4. Contributions

Our significant contribution for this research is that no existing work has proposed the QoS-based RA using distributed edge computing approach. Also, there is no current work for solving the QoS aware RA in a system model with D2D and NS in 5G-CNs. This distributed method of solving the problem reduces system load and transmission delay. The major contributions of this research are:

- To design a 5 G-based cellular network model with NS comprising D2D and cellular users. Based on the users' two service requirements, the system model comprises two network slices (1) eMBB slice and (2) mMTC slice.
- To develop problem formulation for RA to assign the users to their required service slice based on QoS constraints.
- The optimization problem is a non-convex MINLP which is intractable to compute. Therefore, we presented the distributed edge computing algorithm with a two-step process to find the optimized solution.
- In the first step, each edge router will calculate the local optimum values using the augmented Lagrangian method. In the second step, these local estimates will be combined to compute the optimum global value by a consensus algorithm.

Table 1
List of Notations

Notation	Description	Notation	Description
M	MBS	D	Set of D2D pairs
m	Macro cell user	d	D2D user
$\mathcal{I}$	Infrastructure Provider (IP)	$\mathcal{V}$	Set of MVNO
v	Single MVNO	$\mathcal{S}$	Set of Network slices
u	Slice user	j	User type (macro or D2D)
$\mathcal{X}$	eMBB slice	$\mathcal{Y}$	mMTC slice
$\mathcal{X}$m	eMBB users attached to MBS	$\mathcal{Y}$m	mMTC users attached to MBS
$\mathcal{X}$d	eMBB users in D2D mode	$\mathcal{Y}$d	mMTC users in D2D mode
P_m	MBS transmit power	P_d	D2D transmit power
σ	Noise power	$I_{m,u}$	Interference by reusing MBS channel
$r_{u,j,v}$	Distance from BS j to user u	$PL_{u,j,v}$	Pathloss from BS j to user u
$\lambda_{u,m,v}$	Received SINR by user u of MVNO v, with MBS	$\lambda_{u,d,v}$	Received SINR by user u of MVNO v, with D2D
$G_{u,m,v}$	Gain of user u from MVNO v, with MBS	$G_{u,d,v}$	Gain of user u from MVNO v, with D2D
$R_{u,m,v}$	Data rate u from MVNO v, with MBS	$R_{u,d,v}$	Data rate u from MVNO v, with D2D
$R_{thres,\mathcal{X}_j}$	Min data rate requirement for eMBB	$R_{thres,\mathcal{Y}_j}$	Min data rate requirement for mMTC
$\delta_{u,m,v}$	Binary allocation indicator for MBS	$\delta_{u,d,v}$	Binary allocation indicator for D2D
β	Sub-channel bandwidth	C	Set of edge servers

- Numerical results are analyzed by comparing them with other schemes, and it is established that our distributed system design has better average throughput with reduced system load and computation time.

5. System Model

5.1. Network Model

Fig. 1 represents our proposed D2D based 5G cellular network system model. It consists of a single Macrocell with central MBS M and several D2D pairs. The macrocell users are represented by m and D2D users by d. The set of D2D pairs can be represented as $D = \{1, 2, \ldots d\}$. This work investigates the downlink scenario with a single macrocell comprising one central MBS (macro base station) with wide area coverage and very high power. The network is assumed to be in a one-to-one spectrum sharing mode, which means that at a time, one D2D user can reuse only one type of cellular resource. Thus, there is no interference between the same type of users, but as D2D pairs are reusing the frequency resources of cellular users, cross-tier interference must be considered.

The Infrastructure Provider (IP) $\mathcal{I}$ owns, maintains, and controls all the physical resources of the system. They will then assign these resources to the Mobile Virtual Network Operators (MVNOs). Let's consider a set of MVNOs denoted by $\mathcal{V}$. Each particular MVNO, v is not aware of other MVNOs v' presence and is independent. Each MVNO is assumed to have two types of network slices $\mathcal{S}$, eMBB denoted by $\mathcal{X}$, and mMTC denoted by $\mathcal{Y}$ therefore, we have $\mathcal{S}$={$\mathcal{X}$ U $\mathcal{Y}$}, The slice user, u attached to a specific MVNO v can be categorized depending on the type of service it is using ($\mathcal{X}$ or $\mathcal{Y}$). Let N represent the overall users in the network, which is the sum of macro-cell users (MU) and D2D users (DU).

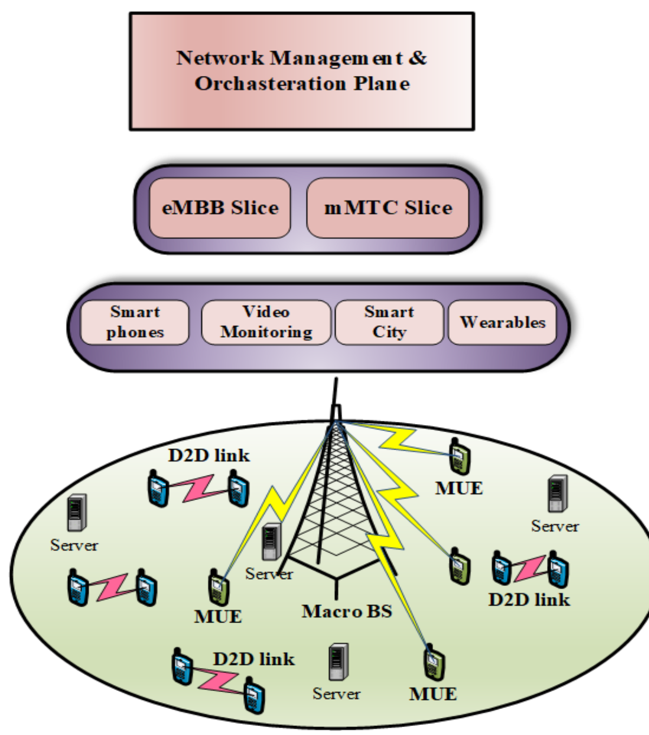


Fig. 1: System model of proposed 5G-CN

5.2. Types of Slice Users

- **Type I:** are macrocell users (associated to MBS) which can either attach to eMBB or mMTC slice. $\mathcal{X}$m represents the set of (high data rate) eMBB, and $\mathcal{Y}$m represents the set of (low latency) mMTC users, both of them attached to MVNO v, located within the area of m.

- **Type II:** are D2D type users which can either attach to eMBB or mMTC slice. $\mathcal{X}$d represents the set of (high data rate) eMBB users, and $\mathcal{Y}$d represents the set of (low latency) mMTC users both attached to MVNO v, belonging to D2D pairs d.

The network slice controller provides the central management as it has all the information required for the system, like the association of users and channel state information (CSI). The role of slice controller is two levels, i.e., network slice selection (to choose the required slice depending on the service requirement) and resource management (to optimize the resource allocation by selecting the appropriate algorithm).

5.3. Channel Characteristics

This paper proposes the virtualized bandwidth scenario controlled by a central slice controller or IP and then reassigned to MVNOs. The bandwidth β in a channel is divided into sub-channels of frequency 180kHz band each [39]. The noise power in the system denoted by σ is given as:

$$\sigma^2 = NF + N_O * \beta, \tag{1}$$

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

We assume the Rayleigh and Shadowing log-normal random variables for evaluating the effects of fading and the system path loss models, which are different for macrocell and D2D users. The distance-dependent path loss using Okumura-Hata model [40] for a user u given as:

$$PL_{u,j,v} = 128.1 + 37.6log(r_{u,j,v}[km])dBm, \tag{2}$$

if $u \in \{\mathcal{X}\text{m}, \mathcal{Y}\text{m}\}$, j∈m.

$$PL_{u,j,v} = 148.1 + 40log(r_{u,j,v}[km])dBm, \tag{3}$$

if $u \in \{\mathcal{X}\text{d}, \mathcal{Y}\text{d}\}$, j∈d.

$PL_{u,j,v}$ is the received path loss by the slice user, u, of MVNO v belonging to either MBS m or D2D pair d ($j \in m, d$) respectively. Similarly, $r_{u,j,v}$ represents the range/distance of the user u from either MBS m or for D2D pairs. This is the range between the D2D transmitter and the receiver.

The total allowed transmit power of MBS and D2D is denoted by P_m and P_d, respectively. Our system model presumed that no two macrocell users would share the same frequency channel. Similarly, no two D2D users will share the same frequency resource. However, D2D users can reuse the macro-cell user's channel, considering macro-cell users' minimum data rate requirement. Therefore, cross-tier interference must be considered while assigning resources to the users.

6. Problem Formulation

The received SINR $\lambda_{u,m,v}$ by user u of MVNO v, having network slice services $\{\mathcal{X}\text{m}, \mathcal{Y}\text{m}\}$ in the coverage location of MBS m can be formulated as:

$$\lambda_{u,m,v} = \frac{P_m G_{u,m,v}}{P_d G_{u,d} + \sigma^2}, \tag{4}$$

where $G_{u,m,v}$ represents the channel gain from MBS m to slice user u attached to MVNO v and σ^2 is noise. $G_{u,d}$ is the gain received by macrocell users from D2D users using the same frequency resource.

The received SINR $\lambda_{u,d,v}$ by user u of MVNO v, having network slice services $\{\mathcal{X}\text{d}, \mathcal{Y}\text{d}\}$ in D2D mode d is given:

$$\lambda_{u,d,v} = \frac{P_d G_{u,d,v}}{P_m G_{u,m} + \sigma^2}, \tag{5}$$

where $G_{u,d,v}$ is the channel gain received from D2D transmitter d to slice user u attached to MVNO v and $G_{u,m}$ is the gain received by D2D user from macrocell user using the same frequency resource.

Therefore, data rate can be computed considering the above equations. The total achievable data rate $R_{u,m,v}$ by user u of MVNO v, having network slice services $\{\mathcal{X}\text{m}, \mathcal{Y}\text{m}\}$ in the coverage area of MBS m is given:

$$\mathcal{R}_{u,m,v} = \beta log(1 + \lambda_{u,m,v}), \tag{6}$$

$$\mathcal{R}_{u,m,v} = \beta log(1 + \frac{P_m G_{u,m,v}}{P_d G_{u,d} + \sigma^2}). \tag{7}$$

Similarly, the total achievable data rate $R_{u,d,v}$ by user u of MVNO v, having network slice services $\{\mathcal{X}\text{d}, \mathcal{Y}\text{d}\}$ in D2D mode d is given:

$$\mathcal{R}_{u,d,v} = \beta log(1 + \lambda_{u,d,v}), \tag{8}$$

$$\mathcal{R}_{u,d,v} = \beta log(1 + \frac{P_d G_{u,d,v}}{P_m G_{u,m} + \sigma^2}). \tag{9}$$

β in equation (8) and (9) is the sub-channel bandwidth which is 180KHz.

Table 2
QoS Parameters [37, 42]

Parameter	Values
$L_{v,\mathcal{Y}}$	4 kbit
$L_{v,\mathcal{X}}$	10 kbit
η	e^{-3}
T_{max}	100 ms
$\Delta_{v,\mathcal{Y}}$	5 packets/s
$\Delta_{v,\mathcal{X}}$	10 packets/s

6.1. Data Rate Model

The data packets for MU arrive periodically, and therefore, their data rate requirement must be assured in a deterministic way. For eMBB, the packet arrival rate is defined as the periodic method with rate $\Delta_{v,\mathcal{X}_j}$ (packet/sec), and a constant size of $L_{v,\mathcal{X}_j}$ (bits). The minimum data rate $R_{thres,\mathcal{X}_j}$ requirement for each eMBB is given as [37]:

$$\mathcal{R}_{thres,\mathcal{X}_j} = \Delta_{v,\mathcal{X}_j} L_{v,\mathcal{X}_j} \tag{10}$$

6.2. Delay Model

The IP plans the network slicing depending on the system service requirements and network functions. The MVNO may install multiple network slices providing similar services for different users in the network. The network slices are created to deliver the specific QoS requirement to the group of users, and here in this work, we evaluate two cases of network slices. In the mMTC slice case, the minimum achievable data rate $R_{u,m,v}$ must be equal to or more than the effective threshold data rate R_{thres}. The characteristic of mMTC traffic is that it is randomly generated, so the QoS, in this case, is evaluated statistically by using the effective bandwidth theory. This theory ensures the data packet transmission delay bound probability. Delay is defined as the time needed by the mMTC packet from BS to when the user receives it. The effective bandwidth depends on the system latency boundary violation probability η. The maximum allowable delay boundary threshold denoted as T_{max}, the average data rate of arrival for packets is $\Delta_{v,\mathcal{Y}_j}$ and the packet size for data is $L_{v,\mathcal{Y}_j}$ for mMTC users. Therefore, $R_{thres,\mathcal{Y}_j}$ is given as [41, 42]:

$$\mathcal{R}_{thres,\mathcal{Y}_j} = \frac{L_{v,\mathcal{Y}_j} log(\eta)}{T \ln(1 - \frac{log(\eta)}{\Delta_{v,\mathcal{Y}_j}})}. \tag{11}$$

The resource allocation here is based on the data rate requirements of each type of device, i.e., for eMBB user, it is from equation (10) and for mMTC user, from equation (11). The scope of this work is to assign the resources to both types of devices based on their QoS service requirements.

6.3. Optimization problem

In this work we have proposed the QoS based resource allocation system model in D2D enable 5G-CN with network slicing. The optimization problem $\mathcal{P}1$ is given as:

$$\max \sum_{v \in V} \Big(\sum_{u \in (X_m, Y_m)} \delta_{u,m,v} R_{u,m,v} + \sum_{u \in (X_d, Y_d)} \delta_{u,d,v} R_{u,d,v} \Big), \tag{12}$$

subject to:

$$\delta_{u,m,v} R_{u,m,v} \geq \mathcal{R}_{thres,\mathcal{X}_m}, \forall\, u \in \mathcal{X}_m, \forall\, v \in V, \tag{13}$$

$$\delta_{u,m,v} R_{u,m,v} \geq \mathcal{R}_{thres,\mathcal{Y}_m}, \forall\, u \in \mathcal{Y}_m, \forall\, v \in V, \tag{14}$$

$$\delta_{u,d,v} \left[R_{u,m,v} - \mathcal{R}_{thres,\mathcal{X}_d} \right] \geq 0, \forall\, u \in \mathcal{X}_d, \forall\, v \in V, \tag{15}$$

$$\delta_{u,d,v} \left[R_{u,m,v} - \mathcal{R}_{thres,\mathcal{Y}_d} \right] \geq 0, \forall\, u \in \mathcal{Y}_d, \forall\, v \in V, \tag{16}$$

$$\sum_{j \in (m,d)} \delta_{u,j,v} = 1, \forall\, u \in (\mathcal{X}_j, \mathcal{Y}_j), \forall\, v \in V, \tag{17}$$

$$\delta_{u,m,v} \in \{0, 1\}, \forall\, u \in (\mathcal{X}_m, \mathcal{Y}_m), \forall\, v \in V, \tag{18}$$

$$\delta_{u,d,v} \in \{0, 1\}, \forall\, u \in (\mathcal{X}_d, \mathcal{Y}_d), \forall\, v \in V, \tag{19}$$

$\mathcal{P}1$ here in (12) is the maximization of the total throughput which is the sum of throughput of both types of slice users considering the data rate requirements of users. The main objective is maximize the overall system throughput in such a way that each type of user is assigned to its required QoS based slice (eMMB and mMTC). Depending on the data rate requirement, the user will be assigned to that slice and then the system performance is evaluated by throughput analysis. Equation (10) and (11) depicts how threshold data rate can be calculated for both services that will be used as the parameter to choose either the device belong to eMMB or mMTC slice. The resulting average throughput is maximized when assigning resources to users with the type of service they demand.

The constraint in (13) and (15) guarantees that the minimum data rate requirement of eMBB users must be ensured for both MUs and DUs in the system. The constraints in (14) and (16) guarantees that for all users with mMTC case $\mathcal{Y}_m$ or $\mathcal{Y}_d$, the achievable data rate in both types should be by no means less than their threshold data rate $R_{thres,\mathcal{Y}_j}$. Besides, $\delta_{u,m,v}$ and $\delta_{u,d,v}$ are the user association indicators for MUs or DUs respectively. The constraint in (17) guarantees that at the moment, the Type I (macro cell) user will either be associated with eMBB or mMTC slice. Similarly, the Type II (D2D) user can only be associated with one of the slices, eMBB or mMTC, at a time. The constraints in (18) and (19) are the binary indicator variables which can be either 0 or 1.

7. Proposed Solution

By examining the problem $\mathcal{P}1$ in (12) closely, it can be concluded that the given problem is a non-convex MINLP problem which is NP-hard and difficult to solve directly.

Proposition 1: The given problem (12) is non-convex and NP-hard.

Proof: Refer to Appendix A.

Proposition 2: The given problem (12) is a MINLP.

Proof: Refer to Appendix B.

Therefore, we have to convert this problem into a convex form that can be solved using different mathematical tools. In this system model, the solution size becomes unavoidably large as it has two different types of users (MUs and DUs) and two different types of service requirements (eMBB and mMTC). Hence, the distributed method of solving is proposed as given in Algorithm 1. Recently, the distributed approach of solving the optimization problem has gained importance as the prominent research topic [43, 44] with Lagrange method for solving the problem has been verified in [45, 46].

The original optimization problem $\mathcal{P}1$ is transformed to distribute it into separate parts, and the solution of each part is computed in parallel. We can now distribute (12) into two sub-problems based on two types of users in the system (MUs and DUs). Let a and b in (20) represent two sub-problems for macro-cell and D2D pairs case, respectively. a denotes the equation for calculating the data rate of macro-cell users and b for the data rate of D2D users. This division aims to solve the problem distributively, reducing the time of computation and system load. Every edge server/router is responsible for calculating its own local optimum value.

$$\underbrace{\sum_{u\in(X_m,Y_m)} R_{u,m,v}}_{\text{a}} + \underbrace{\sum_{u\in(X_d,Y_d)} R_{u,d,v}}_{\text{b}}, \tag{20}$$

$$f(\delta,a,b) = f(\delta_{u,m,v},a) + f(\delta_{u,d,v},b), \tag{21}$$

$$f(\delta_{u,m,v},a) = \sum_{v\in V}\Big(\sum_{u\in(X_m,Y_m)} \delta_{u,m,v}R_{u,m,v}\Big), \tag{22}$$

$$f(\delta_{u,d,v},b) = \sum_{v\in V}\Big(\sum_{u\in(X_d,Y_d)} \delta_{u,d,v}R_{u,d,v}\Big). \tag{23}$$

Let us consider that our system comprises of C edge nodes or servers at various locations that perform the task of local distributively. We first transform the constraints in (13), (14), (15) and (16) to decompose them. The new constraints are now given in (24), (25), (26) and (27) as:

$$\delta_{u,m,v}R_{u,m,v} \geq \frac{\mathcal{R}_{thres,\mathcal{X}_m}}{C}, \forall\, u\in\mathcal{X}_m, \forall\, v\in V, \tag{24}$$

$$\delta_{u,m,v}R_{u,m,v} \geq \frac{\mathcal{R}_{thres,\mathcal{Y}_m}}{C}, \forall\, u\in\mathcal{Y}_m, \forall\, v\in V, \tag{25}$$

$$\delta_{u,d,v}\Big[R_{u,m,v} - \frac{\mathcal{R}_{thres,\mathcal{X}_d}}{C}\Big] \geq 0, \forall\, u\in\mathcal{X}_d, \forall\, v\in V, \tag{26}$$

$$\delta_{u,d,v}\Big[R_{u,m,v} - \frac{\mathcal{R}_{thres,\mathcal{Y}_d}}{C}\Big] \geq 0, \forall\, u\in\mathcal{Y}_d, \forall\, v\in V. \tag{27}$$

Each edge server C will solve its optimization problem locally, taking into view the QoS requirement of users. The given problem can easily be solved by finding the local approximate values of (24), (25), (26) and (27). These values will then be used to compute the global maxima or minima. An augmented Lagrangian multiplier method is known as one of the best approaches due to its fast convergence rate. The constraints from (17) - (19) are global constraints, and we will solve them using a consensus algorithm for obtaining the combined agreement on local optimum values [47, 48].

Proposition 3: The given augmented Lagrangian optimization method, when used with the consensus approach, is generally faster in convergence in comparison to other distributed algorithms for optimization.

Proof: Refer to Appendix C.

Remarks: By analyzing our algorithm theoretically, it is verified that it converges at a faster rate of $O(1/k)$ with continuous-state errors, which are controllable using numerous consensus algorithm steps.

7.1. Local Variables Optimization

Each edge server/router will compute approximates for its local optimum value. The partial Lagrangian equation for MU can be represented by:

$$\sum_{v\in V}\sum_{u\in(X_m,Y_m)} L(a_{u,m,v},\eta1_{u,m,v},\eta2_{u,m,v}), \tag{28}$$

$$\begin{aligned} &= \sum_{v\in V}\sum_{u\in(X_m,Y_m)} \delta_{u,m,v}R_{u,m,v} \\ &+\eta1_{u,m,v}\left(\delta_{u,m,v}R_{u,m,v} - \frac{\mathcal{R}_{thres,\mathcal{X}_m}}{C}\right), \\ &+\eta2_{u,m,v}\left(\delta_{u,m,v}R_{u,m,v} - \frac{\mathcal{R}_{thres,\mathcal{Y}_m}}{C}\right) \end{aligned} \tag{29}$$

Algorithm 1 Distributed Resource Allocation Algorithm

Initialize:
R_m=0, R_d=0, $X_j \leftarrow$eMBB, $Y_j \leftarrow$mMTC
where j∈[m,d], u=$u_m + u_d$

for all $m \in M$ **do**
for all $d \in D$ **do**

Step-1: Calculate Data rate $R_{u,m,v}$ for Macro cell users.
$R_m = R_{u,m,v}$
Step-2: Calculate Data rate $R_{u,d,v}$ for D2D users.
$R_d = R_{u,d,v}$
Step-3: Find Threshold rate requirement for eMBB X_j slice.
$R_{thres,X_j} \leftarrow$From Eq.(10)

Step-4: Find Threshold rate requirement for mMTC Y_j slice.
$R_{thres,Y_j} \leftarrow$From Eq.(11)

[Solve Local Variables]: Augmented Lagrange method
Step-5: Slice for Macro cell users:
if $R_{u,m,v} \geq R_{thres,X_j}$ **then**
Set $u \in X_m$
else if
$R_{u,m,v} \geq R_{thres,Y_j}$ **then**
Set $u \in Y_m$ **and update**
end if

Step-6: Slice for D2D users:
if $R_{u,d,v} \geq R_{thres,X_j}$ **then**
Set $u \in X_d$
else if
$R_{u,d,v} \geq R_{thres,Y_j}$ **then**
Set $u \in Y_d$ **and update**
end if

[Solve Global Variables]: Consensus method
Step-7:while Each Macro cell user can only attach to either eMBB or mMTC at a time **do**
$u_m \leftarrow X_m$orY_m
end while
Step-8:while Each D2D user can only attach to either eMBB or mMTC at a time **do**
$u_d \leftarrow X_d$orY_d
end while
Step-9: The optimized results are sent to slice controller.

$$= \sum_{v\in V} \sum_{u\in(X_m,Y_m)} \delta_{u,m,v}\ \beta log(1+\frac{P_m G_{u,m,v}}{P_d G_{u,d}+\sigma^2}) \\ +\eta 1_{u,m,v}\left(\delta_{u,m,v}R_{u,m,v}-\frac{\mathcal{R}_{thres,\mathcal{X}_m}}{C}\right). \\ +\eta 2_{u,m,v}\left(\delta_{u,m,v}R_{u,m,v}-\frac{\mathcal{R}_{thres,\mathcal{Y}_m}}{C}\right) \tag{30}$$

Similarly, the partial Lagrangian equation for D2D can be represented by:

$$\sum_{v\in V} \sum_{u\in(X_d,Y_d)} L(b_{u,d,v},\eta 1_{u,d,v},\eta 2_{u,d,v}), \tag{31}$$

$$= \sum_{v\in V} \sum_{u\in(X_d,Y_d)} \delta_{u,d,v}R_{u,d,v} \\ +\eta 1_{u,d,v}\left(\delta_{u,d,v}\left[R_{u,m,v}-\frac{\mathcal{R}_{thres,\mathcal{X}_d}\right]}{C}\right), \\ +\eta 2_{u,d,v}\left(\delta_{u,d,v}\left[R_{u,m,v}-\frac{\mathcal{R}_{thres,\mathcal{Y}_d}\right]}{C}\right) \tag{32}$$

$$= \sum_{v\in V} \sum_{u\in(X_d,Y_d)} \delta_{u,d,v}\ \beta log(1+\frac{P_d G_{u,d,v}}{P_m G_{u,m}+\sigma^2}) \\ +\eta 1_{u,d,v}\left(\delta_{u,d,v}\left[R_{u,m,v}-\frac{\mathcal{R}_{thres,\mathcal{X}_d}\right]}{C}\right). \\ +\eta 2_{u,d,v}\left(\delta_{u,d,v}\left[R_{u,m,v}-\frac{\mathcal{R}_{thres,\mathcal{Y}_d}\right]}{C}\right) \tag{33}$$

In above equations, $\eta 1_{u,m,v}$, $\eta 2_{u,m,v}$, $\eta 1_{u,d,v}$ and $\eta 2_{u,d,v}$ are variables of lagrangian multipliers.

The overall Partial Lagrangian is the sum of (30) and (33) can be represented by:

$$\sum_{v\in V} \sum_{u\in(X_j,Y_j)} L(a,b,\eta 1_{u,j,v},\eta 2_{u,j,v}), \tag{34}$$

where $j \in \{m,d\}$

$$L(a,b,\eta 1_{u,m,v},\eta 2_{u,m,v},\eta 1_{u,d,v},\eta 2_{u,d,v}) = \\ \sum_{v\in V}\Bigg[\sum_{u\in(X_m,Y_m)} \delta_{u,m,v}\ \beta log(1+\frac{P_m G_{u,m,v}}{P_d G_{u,d}+\sigma^2}) \\ +\eta 1_{u,m,v}\left(\delta_{u,m,v}R_{u,m,v}-\frac{\mathcal{R}_{thres,\mathcal{X}_m}}{C}\right) \\ +\eta 2_{u,m,v}\left(\delta_{u,m,v}R_{u,m,v}-\frac{\mathcal{R}_{thres,\mathcal{Y}_m}}{C}\right), \\ +\sum_{u\in(X_d,Y_d)} \delta_{u,d,v}\ \beta log(1+\frac{P_d G_{u,d,v}}{P_m G_{u,m}+\sigma^2}) \\ +\eta 1_{u,d,v}\left(\delta_{u,d,v}\left[R_{u,m,v}-\frac{\mathcal{R}_{thres,\mathcal{X}_d}\right]}{C}\right) \\ +\eta 2_{u,d,v}\left(\delta_{u,d,v}\left[R_{u,m,v}-\frac{\mathcal{R}_{thres,\mathcal{Y}_d}\right]}{C}\right)\Bigg] \tag{35}$$

$$f(\eta 1_{u,j,v},\eta 2_{u,j,v}) = \begin{cases} \max \sum_{u\in(X_m,Y_m)}\ L(a_{u,m,v},\eta 1_{u,m,v},\eta 2_{u,d,v}) \\ +\max \sum_{u\in(X_d,Y_d)}\ L(b_{u,d,v},\eta 1_{u,d,v},\eta 2_{u,d,v}) \\ subject\ to:\ \sum_{j\in(m,d)}\delta_{u,j,v}=1, \end{cases} .$$

$$(36)$$

$$j \in \{m, d\}, \forall_v \in V, \forall u \in (\mathcal{X}_j, \mathcal{Y}_j).$$

7.2. Global Variables Optimization

Each agent must know about the global variables for a system with multiple agents. By using the consensus algorithm, these global variables reach a mutual agreement after having information of local variables approximates [47]. In the proposed system (17) - (19) are global constraints. Let us transform them using closed expression as follows:

$$\max_{p_1, p_2 \cdots p_c} \sum_{k=1}^{c} e_k^T p_k, \tag{37}$$

where all edge servers are denoted by $k = 1, 2 \ldots c$, p_k represents the decision vector variable for particular edge server/router and e_k depicts the local cost. The coupled or global constraint in (36) can be represented in closed expression as:

$$\sum_{i=1}^{c} A_k z_k \leq 1, \tag{38}$$

$$l_i(k) = \sum_{u \in U_i(k)} \{a_j^i(k) \gamma_j(k)\}, \tag{39}$$

$$z_i(k+1) \leftarrow \arg\ max_{x_i \in vert(z_i)} \, (e_i^T \ + \ l_i(k)^T A_i) z_i, \tag{40}$$

$$\varphi_i(k) = max_{uU_i(k)} \{\rho_i(k)\}, \tag{41}$$

$$\bar{\zeta}_i(k+1) = max \ [\bar{\zeta}_i(k), \ A_i z_i(k+1)], \tag{42}$$

$$\underline{\zeta_i}(k+1) = min \ [\underline{\zeta_i}(k), \ A_i z_i(k+1)], \tag{43}$$

$$\rho_i(k+1) = max[Q_i(k), \rho\{\bar{\zeta}_i(k+1) - \underline{\zeta_i}(k+1)\}]. \tag{44}$$

8. Performance Evaluation

We consider a single cell D2D based 5G-CN that comprises one central MBS and several macro-cell users and D2D pairs randomly located within the geographical area of the cell. The slice users are assumed to have a subscription from 3 MVNOs over the MBS coverage area of 1000m × 1000m, and the range of D2D is 80m. The details of channel characteristics and the path-loss of users considered are described in Section 5.3. The simulation parameters and their values used for analysis are summarized in Table 3. The values of QoS parameters used for eMBB and mMTC are shown in Table 2. All the simulations are performed on MATLAB software's latest version. For smoothness of curves, we use Monte-Carlo simulations of 10,000 iterations. The laptop's specifications for this work are Intel(R) Core i5-6200U CPU @ 2.40GHz 16GB RAM having 64-bit Windows 10 operating system.

Table 3
Performance Parameters

Parameter	Measurement
MBS coverage area	1000X1000 m^2
Bandwidth	180 kHz
MBS transmit power	43 dBm
D2D transmit power	23 dBm
D2D allowable range	80 m
Noise spectral density	-174 dBm/Hz
Shadowing standard deviation	10 dB
Noise figure	9 dB
MBS path loss	128.1+37.6$log(d[km])$
D2D path loss	148.1+40$log(d[km])$

We compare and evaluate our simulation results using the following five methods:

1. **QARA-DA**:(Quality-Aware Resource Allocation using Distributed Algorithm), which is our proposed scheme using distributed edge computing method for finding the solution.
2. **QARA-CA**: (Quality-Aware Resource Allocation using Centralized Algorithm) where the resource optimization of our proposed scheme is performed using a centralized algorithm.
3. **LARA-GA**: (Latency-Aware Resource Allocation using Genetic Algorithm) to optimize the resource allocation problem with this scheme [38].
4. **LARA-BnB**: (Latency-Aware Resource Allocation using Branch and Bound Method) to optimize the resource allocation problem with this scheme [38].
5. **SARA-DM**: (Slice-Aware Resource Allocation Using Dinkelbach method) to optimize the resource allocation problem with this scheme [37].

Fig. 2 demonstrates the effect of increasing the average packet arrival rate for slice users with average system throughput. We suppose that the different slice users (eMBB and mMTC) have the same average packet arrival rate. It is noticed that the system throughput improves with the increase in packet arrival rate for all the schemes. However, our proposed scheme QARA-DA has a high throughput value due to D2D devices and edge-based computation. Further, our proposed distributed (QARA-DA) optimization scheme results are very close to the centralized (QARA-CA) approach.

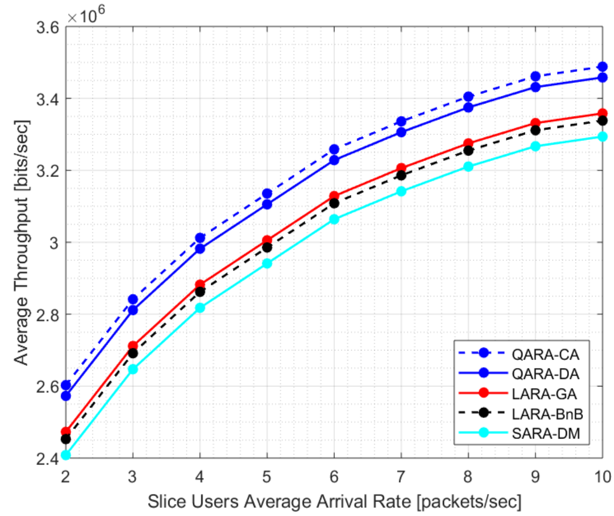


Fig. 2: Impact of average packet arrival rate on the system throughput

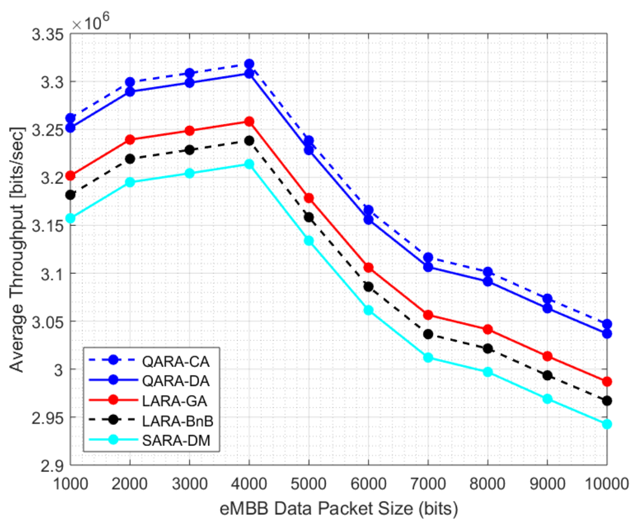


Fig. 3: Impact of eMBB data packet size on the total system throughput

Fig. 3 shows the effect of increasing the size of a data packet for the eMBB slice on the system average throughput. It can be observed that the system average throughput increases by increasing the packet size up to about 4000 bits, and after that, it starts decreasing. This is due to the limited network bandwidth capacity. Our proposed scheme QARA-DA outperforms in throughput compared to all other schemes. The results of QARA-DA are very close to the QARA-CA scheme.

Fig. 4 demonstrates the evaluation of our proposed distributed scheme (QARA-DA) with the central scheme (QARA-CA). It shows the plot of time the algorithm takes in computation with the no. of iterations. By examining the plot, it is observed that the time taken in QARA-DA is much less than QARA-CA because the latter has to manage all the computation and processing centrally. Therefore, the QARA-CA scheme burdens the system while our proposed scheme QARA-DA offloads the central system due to performing the tasks in parallel with edge computing servers distributively.

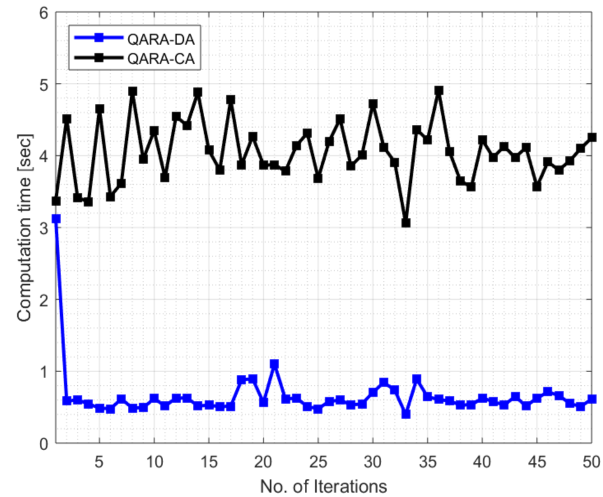


Fig. 4: Impact on computational time of algorithm with no. of iterations

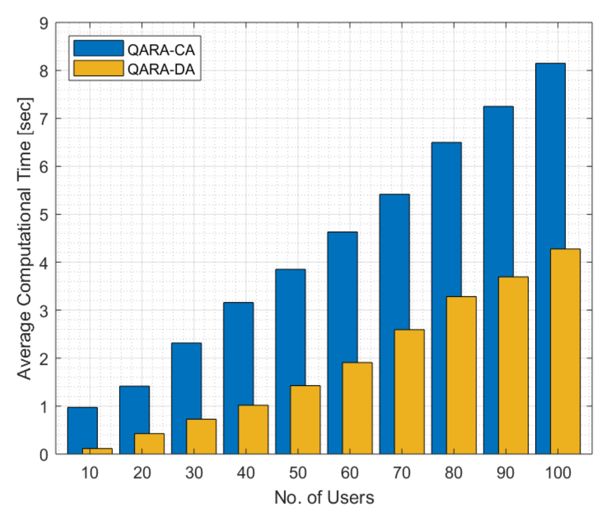


Fig. 5: Impact on average computational time of algorithm vs. no. of users

Fig. 5 illustrates the comparison of the proposed distributed scheme (QARA-DA) with the central scheme (QARA-CA). It depicts the analysis of average computational time taken by both algorithms with the increase in no. of users. The results are computed using 50 no. of iterations. It is observed that the average computational time increases with the increase in no. of users for both QARA-CA and QARA-DA. However, the time taken in QARA-DA is much less than QARA-CA since the latter performs complete handling and processing tasks centrally. In this way, our proposed scheme QARA-DA is better in performance because it reduces the system load and delays in computing the tasks due to parallel processing. The bit error rate is also reduced as compared to the central scheme.

8.1. Algorithm Performance

In this work, we have theoretically and analytically evaluated the performance of our proposed scheme using both central (QARA-CA) and distributed (QARA-DA) algorithms. By viewing the results of Fig. 2 and Fig. 3, it can be concluded that the results of QARA-DA are very close to QARA-CA, which is the central optimization method. We claimed that our proposed distributive scheme QARA-DA is better in performance as compared to QARA-CA because it uses an edge computing method where edge devices/servers calculate most of the tasks which were previously done by central or slice controller as evident from Fig. 4 and Fig. 5. For QARA-CA, the central controller collects all channel and interference information from the base station, complex as the network conditions change continuously. In practical systems, this information exchange has to occur many times, overloading the network processing resulting in poor performance. However, in QARA-DA, the computation load is considerably low due to less involvement of the central system. The QARA-DA scheme is efficient as it reduces the system load and overhead. Each edge router is responsible for completing its computation task and collecting channel information locally. This will minimize the computation time/ delay as the tasks are now parallel and distributed. The slice controller will then assign the slices based on the information received from each edge router that will offload the central system and reduce time. Although the throughput results in QARA-CA are slightly better than QARA-DA, which is obviously due to central management, this method burdens the system. It will utilize more time in execution with fewer chances of errors than the distributed edge method.

9. Conclusion and Future Works

In this work, we address the problem of assigning resources based on the user's QoS requirement in a D2D enabled 5G-CN. We considered two significant 5G used cases, eMBB, and mMTC, for two types of network slices in the system. The optimization problem maximizes the average system throughput considering customized services' delay and data rate constraints. Distributed edge computing optimization approach is used to find the solution by dividing the problem into local and global constraints. The Lagrange method is used to find the local estimates, and these values are then sent to find the global maxima using a consensus algorithm. The simulation results show that our proposed scheme is better in performance than other schemes. Further, using the edge computation approach for finding the solution takes less time in computation and offloads the central system. In the future, this work can be extended to improve the capacity of the backhaul system, using application-based scenarios of the proposed system like content caching and working on edge nodes capacity improvement. Adding the concepts of deep reinforcement learning and the internet of things will further improve this system in the future.

A. Proof of Proposition 1

By considering the equation (12) subject to (13) - (18), it is clearly visible that the proposed problem function and all its given constraints are smooth. The throughput equations $R_{u,m,v}$ and $R_{u,d,v}$ are non-linear and logarithmic form functions with non-convex constraints (13) - (16). Hence, (12) is a smooth , non-linear and non-convex optimization problem. Such category of programming scenarios have been declared as generally NP-hard and are troublesome in computation [49]. Also the proposed problem is a QoS based throughput optimization problem which has been justified to be NP-hard previously [50], further this work [51] prove the NP-hardness of problem $\mathcal{P}1$.

■

B. Proof of Proposition 2

The definition of Mixed integer non-linear programming (MINLP) problem is the field of optimization with integer and continuous variables having non-linear functions in the objective function and the constraints. MINLP has been phenomenal over the past few years because of their applications in the areas of engineering with contributions in analytics, algorithms and computations. The example of general form of a MINLP can be represented as:

$$\min f(x, y) \tag{45}$$

subject to:

$$G_i(x, y) = 0 \forall\ i \in F \tag{46}$$

$$G_i(x, y) \leq 0 \forall\ i \in I \tag{47}$$

$$x \in X\ and y \in Y\ [0, 1] \tag{48}$$

Here, the $G_i(x, y)$ is a constraint and F and I belongs to both equality and inequality values. The objective functions f and C_i are continuous with few smoothness properties.

The given problem (12) is a MINLP because it has integer user association variables from equation (18-19) and continuous variables in the constraints from equations (13-16). Similarly, the objective function has nonlinear functions due to the fractional equations and variables of data rate equations. Thus, the given resource allocation problem is considered in the category of MINLP.

■

C. Proof of Proposition 3

The theoretical analysis of the multi-agent constrained-based optimization problem with the distributed method was proposed by [47]. In this work, the authors verified through evaluation that the augmented Lagrangian algorithm has a fast convergence rate compared to other distributed optimization methods. This convergence depends upon the separation of given constraints to global and local constraints. At first, the local optimum is calculated, and these approximate values are then utilized to compute the global optimum by consensus approach. This method is useful to achieve an agreement or consensus for a multi-agent system with several distributed optimization methods. This will ensure the system's reliability for finding the solution of such problems with multiple nodes [47].

■

References

[1] A. Ghosh, A. Maeder, M. Baker, D. Chandramouli, 5G evolution: A view on 5G cellular technology beyond 3GPP release 15, IEEE Access 7 (2019) 127639–127651.

[2] P. Popovski, K. F. Trillingsgaard, O. Simeone, G. Durisi, 5G wireless network slicing for eMBB, uRLLC, and mMTC: A communication-theoretic view, IEEE Access 6 (2018) 55765–55779.

[3] G. A. Akpakwu, B. J. Silva, G. P. Hancke, A. M. Abu-Mahfouz, A survey on 5g networks for the internet of things: Communication technologies and challenges, IEEE access 6 (2017) 3619–3647.

[4] Z. Zhou, M. Dong, K. Ota, J. Wu, T. Sato, Energy efficiency and spectral efficiency tradeoff in device-to-device (D2D) communications, IEEE Wireless Communications Letters 3 (5) (2014) 485–488.

[5] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, J. Folgueira, Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges, IEEE Communications Magazine 55 (5) (2017) 80–87.

[6] Z. Ning, X. Wang, J. Huang, Mobile edge computing-enabled 5G vehicular networks: Toward the integration of communication and computing, ieee vehicular technology magazine 14 (1) (2018) 54–61.

[7] L. Nadeem, M. A. Azam, Y. Amin, M. A. Al-Ghamdi, K. K. Chai, M. F. N. Khan, M. A. Khan, Integration of D2D, network slicing, and MEC in 5g cellular networks: Survey and challenges, IEEE Access 9 (2021) 37590–37612.

[8] L. Nadeem, Y. Amin, J. Loo, M. A. Azam, K. K. Chai, Efficient resource allocation using distributed edge computing in D2D based 5G-HCN with network slicing, IEEE Access 9 (2021) 134148–134162. doi:10.1109/ACCESS.2021.3114629.

[9] S. Jayakumar, S. Nandakumar, A review on resource allocation techniques in D2D communication for 5G and B5G technology, Peer-to-Peer Networking and Applications 14 (1) (2021) 243–269.

[10] P. K. Mishra, S. Pandey, S. K. Biswash, Efficient resource management by exploiting D2D communication for 5G networks, IEEE Access 4 (2016) 9910–9922. doi:10.1109/ACCESS.2016.2602843.

[11] P. K. Mishra, S. Pandey, S. K. Udgata, S. K. Biswash, Device-centric resource allocation scheme for 5G networks, Physical Communication 26 (2018) 175–184.

[12] Y. Shi, E. Alsusa, M. W. Baidas, Joint DL/UL decoupled cell-association and resource allocation in D2D-underlay hetnets, IEEE Transactions on Vehicular Technology 70 (4) (2021) 3640–3651. doi:10.1109/TVT.2021.3067269.

[13] T. Huynh, T. Onuma, K. Kuroda, M. Hasegawa, W.-J. Hwang, Joint downlink and uplink interference management for device to device communication underlaying cellular networks, IEEE Access 4 (2016) 4420–4430.

[14] M. I. Khan, M. M. Alam, Y. L. Moullec, E. Yaacoub, Throughput-aware cooperative reinforcement learning for adaptive resource allocation in device-to-device communication, Future Internet 9 (4) (2017) 72.

[15] R. AliHemmati, B. Liang, M. Dong, G. Boudreau, S. H. Seyedmehdi, Power allocation for underlay device-to-device communication over multiple channels, IEEE Transactions on Signal and Information Processing over Networks 4 (3) (2017) 467–480.

[16] M. Hasan, E. Hossain, Distributed resource allocation in D2D-enabled multi-tier cellular networks: An auction approach, in: 2015 IEEE International Conference on Communications (ICC), IEEE, 2015, pp. 2949–2954.

[17] S. Bayat, A. Khalili, S. Zargari, M. R. Mili, Z. Han, Multi-objective resource allocation for D2D and enabled MC-NOMA networks by tchebycheff method, IEEE Transactions on Vehicular Technology 70 (5) (2021) 4464–4470. doi:10.1109/TVT.2021.3068855.

[18] M. R. Mili, A. Khalili, N. Mokari, S. Wittevrongel, D. W. K. Ng, H. Steendam, Tradeoff between ergodic energy efficiency and spectral efficiency in D2D communications under rician fading channel, IEEE Transactions on Vehicular Technology 69 (9) (2020) 9750–9766. doi:10.1109/TVT.2020.3002296.

[19] R. Su, D. Zhang, R. Venkatesan, Z. Gong, C. Li, F. Ding, F. Jiang, Z. Zhu, Resource allocation for network slicing in 5G telecommunication networks: A survey of principles and models, IEEE Network 33 (6) (2019) 172–179.

[20] F. Song, J. Li, C. Ma, Y. Zhang, L. Shi, D. N. K. Jayakody, Dynamic virtual resource allocation for 5G and beyond network slicing, IEEE Open Journal of Vehicular Technology 1 (2020) 215–226.

[21] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, V. C. Leung, Network slicing based 5G and future mobile networks: mobility, resource management, and challenges, IEEE communications magazine 55 (8) (2017) 138–145.

[22] S. A. Kazmi, T. N. Dang, N. H. Tran, M. Bennis, C. S. Hong, Radio resource management techniques for 5G verticals, 5G Verticals: Customizing Applications, Technologies and Deployment Techniques (2020) 119–136.

[23] M. A. Kamal, H. W. Raza, M. M. Alam, M. M. Su'ud, et al., Resource allocation schemes for 5G network: A systematic review, Sensors 21 (19) (2021) 6588.

[24] Y. Sun, M. Peng, S. Mao, S. Yan, Hierarchical radio resource allocation for network slicing in fog radio access networks, IEEE Transactions on Vehicular Technology 68 (4) (2019) 3866–3881. doi:10.1109/TVT.2019.2896586.

[25] W. Li, Y. Zi, L. Feng, F. Zhou, P. Yu, X. Qiu, Latency-optimal virtual network functions resource allocation for 5G backhaul transport network slicing, Applied Sciences 9 (4) (2019) 701.

[26] D. Wu, Z. Yang, H. Wang, B. Yang, R. Wang, UCRA: A user-centric context-aware resource allocation for network slicing, in: 2020 International Conference on Computing, Networking and Communications (ICNC), 2020, pp. 808–814. doi:10.1109/ICNC47757.2020.9049713.

[27] T. Dang, M. Peng, in: 2018 IEEE International Conference on Communications Workshops (ICC Workshops.

[28] S. O. Oladejo, O. E. Falowo, 5g network slicing: A multi-tenancy scenario, in: 2017 Global Wireless Summit (GWS), IEEE, 2017, pp. 88–92.

[29] S. O. Oladejo, O. E. Falowo, Profit-aware resource allocation for 5G sliced networks, in: 2018 European Conference on Networks and Communications (EuCNC), IEEE, 2018, pp. 43–9.

[30] C. Liang, F. R. Yu, H. Yao, Z. Han, Virtual resource allocation in information-centric wireless networks with virtualization, IEEE Transactions on Vehicular Technology 65 (12) (2016) 9902–9914.

[31] Z. Jian, W. Muqing, M. Ruiqiang, W. Xiusheng, Dynamic resource sharing scheme across network slicing for multi-tenant C-RANs, in: 2018 IEEE/CIC International Conference on Communications in China (ICCC Workshops), IEEE, 2018, pp. 172–177.

[32] Q. Ye, W. Zhuang, S. Zhang, A.-L. Jin, X. Shen, X. Li, Dynamic radio resource slicing for a two-tier heterogeneous wireless network, IEEE

Transactions on Vehicular Technology 67 (10) (2018) 9896–9910.
[33] T. Dang, M. Peng, Delay-aware radio resource allocation optimization for network slicing in fog radio access networks, in: 2018 IEEE International Conference on Communications Workshops (ICC Workshops), IEEE, 2018, pp. 1–6.
[34] Y. Hao, Y. Jiang, M. S. Hossain, A. Ghoneim, J. Yang, I. Humar, Data-driven resource management in a 5G wearable network using network slicing technology, IEEE Sensors Journal 19 (19) (2019) 8379–8386. doi:10.1109/JSEN.2018.2883976.
[35] Y. Han, X. Tao, X. Zhang, S. Jia, Delay-aware resource management for multi-service coexisting LTE-D2D networks with wireless network virtualization, IEEE Transactions on Vehicular Technology 69 (7) (2020) 7339–7353. doi:10.1109/TVT.2020.2990402.
[36] Y. Chen, Y. Wang, M. Liu, J. Zhang, L. Jiao, Network slicing enabled resource management for service-oriented ultra-reliable and low-latency vehicular networks, IEEE Transactions on Vehicular Technology 69 (7) (2020) 7847–7862.
[37] J. Ma, C. Pan, C. Yin, X. Li, Slice-aware resource management in SDN enabled heterogeneous cellular networks, in: 2019 IEEE/CIC International Conference on Communications in China (ICCC), IEEE, 2019, pp. 869–874.
[38] S. O. Oladejo, O. E. Falowo, Latency-aware dynamic resource allocation scheme for multi-tier 5G network: A network slicing-multitenancy scenario, IEEE Access 8 (2020) 74834–74852. doi:10.1109/ACCESS.2020.2988710.
[39] 3GPP, Evolved universal terrestrial radio access (E-UTRA); physical channels and modulation (2017).
[40] J. S. Seybold, Introduction to RF propagation, John Wiley & Sons, 2005.
[41] D. Wu, R. Negi, Effective capacity: a wireless link model for support of quality of service, IEEE Transactions on wireless communications 2 (4) (2003) 630–643.
[42] S. O. Oladejo, O. E. Falowo, Latency-aware dynamic resource allocation scheme for 5G heterogeneous network: A network slicing-multitenancy scenario, in: 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), IEEE, 2019, pp. 1–7.
[43] P. Yi, Y. Hong, F. Liu, Distributed gradient algorithm for constrained optimization with application to load sharing in power systems, Systems & Control Letters 83 (2015) 45–52.
[44] Y. Zhao, Y. Liu, G. Wen, G. Chen, Distributed optimization for linear multiagent systems: Edge-and node-based adaptive designs, IEEE Transactions on Automatic Control 62 (7) (2017) 3602–3609.
[45] M. Li, Generalized lagrange multiplier method and KKT conditions with an application to distributed optimization, IEEE Transactions on Circuits and Systems II: Express Briefs 66 (2) (2018) 252–256.
[46] G. Ruan, H. Zhong, J. Wang, Q. Xia, C. Kang, Neural-network-based lagrange multiplier selection for distributed demand response in smart grid, Applied Energy 264 (2020) 114636.
[47] Y. Zhang, M. M. Zavlanos, A consensus-based distributed augmented lagrangian method, in: 2018 IEEE Conference on Decision and Control (CDC), IEEE, 2018, pp. 1763–1768.
[48] N. Chatzipanagiotis, D. Dentcheva, M. M. Zavlanos, An augmented lagrangian method for distributed optimization, Mathematical Programming 152 (1) (2015) 405–434.
[49] K. G. Murty, S. N. Kabadi, Some np-complete problems in quadratic and nonlinear programming, Tech. rep. (1985).
[50] C. Kai, H. Li, L. Xu, Y. Li, T. Jiang, Joint subcarrier assignment with power allocation for sum rate maximization of D2D communications in wireless cellular networks, IEEE Transactions on Vehicular Technology 68 (5) (2019) 4748–4759.
[51] Y.-F. Liu, Y.-H. Dai, On the complexity of joint subcarrier and power allocation for multi-user OFDMA systems, IEEE transactions on Signal Processing 62 (3) (2013) 583–596.

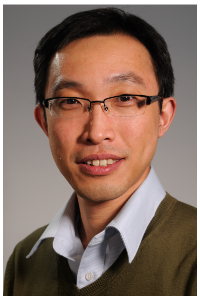
Lubna Nadeem obtained her BS and MS in Telecommunication Engineering from the University of Engineering and Technology (UET) Taxila, Pakistan, in 2011 and 2013, respectively. She is currently doing a Ph.D. and associated with Agile Creative Technologies for Smart Electromagnetic Novel Applications (ACTSENA) 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) 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.