01 Nov 2022

Stable Matching based Resource Allocation for Service Provider's Revenue Maximization in 5G Networks

Ajay Pratap

Sajal K. Das
Missouri University of Science and Technology, sdas@mst.edu

Recommended Citation

A. Pratap and S. K. Das, "Stable Matching based Resource Allocation for Service Provider's Revenue Maximization in 5G Networks," *IEEE Transactions on Mobile Computing*, vol. 21, no. 11, pp. 4094 - 4110, Institute of Electrical and Electronics Engineers, Nov 2022.
The definitive version is available at https://doi.org/10.1109/TMC.2021.3064047

Stable Matching Based Resource Allocation for Service Provider's Revenue Maximization in 5G Networks

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

Abstract—5G technology is foreseen to have a heterogeneous architecture with the various computational capability, and radio-enabled service providers (SPs) and service requesters (SRs), working altogether in a cellular model. However, the coexistence of heterogeneous network model spawns several research challenges such as diverse SRs with uneven service deadlines, interference management, and revenue maximization of non-uniform computational capacities enabled SPs. Thus, we propose a coexistence of heterogeneous SPs and SRs enabled cellular 5G network and formulate the SPs' revenue maximization via resource allocation, considering different kinds of interference, data rate, and latency altogether as an optimization problem and further propose a distributed many-to-many stable matching based solution. Moreover, we offer an adaptive stable matching based distributed algorithm to solve the formulated problem in a dynamic network model. Through extensive theoretical and simulation analysis, we have shown the effect of different parameters on the resource allocation objectives and achieves 94 percent of optimum network performance.

Index Terms—Service provider, service requester, stable matching, 5G, IoT, smart healthcare

1 Introduction

Compared to the 4G networks, 5G communications aim at higher capacity (up to 10 Gbps), allows an increase in the number of smart device users, supports more reliable *Device-to-Device* (D2D) communication and massively deployed *Fog Access Point* (FAP) in the cellular networks [1], [2]. Moreover, 5G aims to offer lower latency, lower battery consumption, and higher data rates to satisfy the requirements of online gaming, video streaming, mobile computing, content sharing, and better implementation of *Internet of Things* (IoT) paradigm [3].

Keeping computational capacity, latency, and power constraints of IoT devices into deliberation, CISCO proposed an idea of Fog Computing, as a *Service Provider (SP)* to serve the requested services near to *Service Requester (SR)* IoT devices [4]. FAP is enabled with *computation capability*, and it can reuse the limited available *radio resources* from the cellular network to transmit the requested content to IoT devices. Moreover, FAP encapsulates not just the edge processing, but also provides the network connections needed to bring the data from IoT devices or distribute the data to IoT devices in cellular network [5]. Moreover, an IoT device can act as an SP to share the requested content to another IoT device in D2D mode [6]. An IoT device can act as SR or SP based on requirement or availability of services. However, in this work we have assumed that SPs and SRs do not change their behavior while requesting for resources underlying cellular 5G network.

Similar to the conference version of this work [7], we consider that FAP always works as an SP. SPs reuse available radio resources from the cellular 5G network to serve the IoT request. Moreover, we assume revenue as an incentive that makes SPs to provide better services to SRs. In other words, SPs aim to maximize the overall earning by serving SRs. Cellular resources are dedicated to different priorities IoT uses, so to serve a request, an SP has to follow the availability of resources and interference constraints at other SRs. An SP only be allowed to reuse the radio resource if it does not create interference to other high priority SRs. However, to serve the increasing number of IoT devices by reusing the limited available cellular resources while avoiding interference, altogether result as a challenging problem to be solved in 5G for enabling real time applications.

Let us consider a smart-health IoT network designed to serve stroke patients in a rehabilitation center. While it is necessary to continuously monitor various signals (e.g., blood pressure, heart rate and blood sugar levels) in multiple patients, there are other tasks such as fall detection (typically detected using gyroscopes, accelerometers and surveillance cameras) that play a crucial role in the avoidance of accidents during the rehabilitation period. Therefore, tasks such as fall detection take precedence over processing blood sugar readings. At the same time, the latency and bandwidth requirements for video streaming in surveillance cameras are significantly larger than those needed to communicate and process fall detection data. In other words, IoT devices typically generate heterogeneous demands (multi-priority tasks) that require diverse resource requirements (e.g., bandwidth, computational power) in the

- *Ajay Pratap is with the Department of Computer Science and Engineering, Indian Institute of Technology (Banaras Hindu University), Varanasi 221005, U.P., India. E-mail: ajay.cse@iitbhu.ac.in.*
- *Sajal K. Das is with the Department of Computer Science, Missouri University of Science and Technology, Rolla, MO 65409 USA. E-mail: sdas@mst.edu.*

Manuscript received 2 July 2020; revised 22 Feb. 2021; accepted 26 Feb. 2021. Date of publication 5 Mar. 2021; date of current version 3 Oct. 2022. (Corresponding author: Ajay Pratap.) Recommended for acceptance by I. Filippini. Digital Object Identifier no. 10.1109/TMC.2021.3064047

presence of non-identical latency constraints. In such a scenario, *Macro Base Station* (MBS) should prioritize tasks that need to be served and allocate necessary resources accordingly to different SPs, via integrating heterogeneous constraints and dynamic network environments [8].

Most of the decentralized and centralized resource allocation schemes for IoT-enabled network focused on the IoT services or task provisioning [9], [10], [11] rather than considering the actual interference, SPs revenue maximization via resource allocation procedure in 5G networks. In a typical cellular network, the optimization [12] and game theoretic approach in fog enabled network [13] are often subjected to a different application-specific domain. However, in D2D-IoT enabled fog architecture, it is essential to consider not only on-time service delivery but also the re-usability of radio resources, interference constraints, and SPs' revenue via allocating limited available resources in the network.

For the above circumstances, we formulate an optimization problem to maximize the SPs' revenue via allocating radio resources while considering various interferences different priorities SRs altogether. Due to high cardinality of the formulated problem, we first provide an algorithm for static scenario based on distributed *many-to-many stable matching* approach. We further propose an *adaptive resource allocation* algorithm based on an adoptive distributed many-to-many stable matching concept. Our proposed algorithm is distributed in a sense that all nodes (IoTs, FAPs, and MBS) take independent decisions (though they are correlated), unlike the case where only one agent (MBS or Cloud) is the central entity perform the calculation for the resource allocation procedure. Once all the nodes are discovered using control signals (e.g., beacons [14]), they exchange different messages among themselves to achieve the resource allocation objective without being solely dependent on any other central entities. However, there are several reasons for using stable matching approach, such as stable matching terminates for every given preference profile. It provides an ultimate solution in terms of stability and utility maximization, which can accurately reflect several system objectives (e.g., revenue, interference, resource constraints, deadlines) within a feasible computational complexity unlike other non-cooperative or Stackelberg game model.

Different from traditional game, in proposed matching based model each agent (IoT, FAP and MBS) selects a strategy by replication and adapt its selection for a better payoff. This strategy is different from the traditional rationality assumption used in Stackelberg or non-cooperative game to obtain Nash equilibrium where rationality implies complete information and strong computation capability of each player to calculate the best response to other players' strategies. Moreover, the proposed matching based model focuses on dynamics of strategy adaptation and reduces the amount of information exchange among the network nodes. In matching theory, different agents are free to set their objectives, prepare utility functions and accordingly negotiate with other agents to achieve stable outcome.

The contributions of this paper are summarized as:

1) Formulate SPs' revenue maximization via resource allocation while considering SR priorities and interference in 5G network as an optimization problem.
2) Solve the formulated problem efficiently by proposing a distributed *Stable Matching based Static Resource Allocation* (SMSRA) algorithm in $O(\mathbb{KNL})$ time, where $\mathbb{K}$, $\mathbb{N}$ and $\mathbb{L}$ represent the total number of SPs, available *Physical Resource Blocks* (PRBs) and set of power levels, respectively.
3) Keeping in mind the adaptive network behavior, design a distributed *Stable Matching based Dynamic Resource Allocation* (SMDRA) algorithm having $O(\mathbb{KNL})$ time complexity.
4) Prove that the SMSRA and the SMDRA schemes terminate and accord a stable resource allocation in different network scenarios within a finite time.
5) Through simulation study, demonstrate that our proposed approach outperforms existing schemes, achieving 94 percent of optimum network performance.

The rest of the paper is organized as follows. Section 2 reviews relevant work. The problem statement is introduced in Section 3. The SMSRA and the SMDRA algorithms are presented in Sections 4 and 5, respectively. Simulation results are discussed in Section 6. Section 7 offers conclusions and future research directions.

2 Background and Related Work

In recent years, 5G got huge attention to improve the data rate, maximize the revenue along with significant balance over interference scenarios and re-usability of limited available radio resources in heterogeneous network model. SPs, revenue maximization in the Fog-enabled cellular model, is one of the promising research challenges concerning resource allocation and interference management problems. However, the use of matching theory acts as an essential tool to solve the resource allocation problem in such heterogeneous network architecture [15]. In the following we review the closely related works with revenue maximization and resource allocation perspective.

2.1 SP's Revenue Maximization

In [16], authors have formulated joint radio and computational resources concerning cost performance per user as an optimization problem. Furthermore, a many-to-one matching based algorithmic model was proposed to offload the task to FAP with the help of a Cloud SP. However, this approach is not suitable for avoiding various interferences while re-using the limited available radio resources in 5G networks. In [17], authors have proposed prospect theory-based SPs' revenue maximization framework concerning 5G technology. This work does not talk about the interference constraint that can profoundly affect the densely deployed low computational IoT devices in the network. In [18], authors have described the market oriented analysis of IoT network for smart-city application. Authors have proposed a scheme based on utility model of different vendors and IoT users in the network. However, this method does not consider the actual scenario of radio resource allocation while modeling the game-theoretic based scheme.

2.2 Radio Resource Allocation

In [20], the authors have proposed a computation offloading and resource allocation for the mobile-edge computing framework. The authors have formulated a combinatorial nature of the multi-user offloading decision and proposed a sub-optimal scheme by splitting it into two parts, i.e., computation

offloading decision and resource allocation scheme. The authors have proposed two matching algorithms to enable distributed computation offloading. However, the transmit power of offloading users is found using a bisection method with approximated inter-cell interference, and the computation resources allocated for users offloading is achieved via the duality approach. In [21], a computation offloading scheme for precedence-constrained tasks in a base station-assisted D2D scenario for information-centric IoT was proposed. This work jointly aimed to minimize the weighted sum of task processing delay and resource rental fee, considering the task delay, association states, and availability of resources altogether. The authors employed Hungarian algorithm to find lightest sub-task-helper pairs by updating the cost matrix in each time interval.

In [22], authors have designed *Lagrange approximation Supple Radio controller* (LaSR), multi-connectivity scheduler that assigns radio resources to users in an OFDMA-based multi *Radio Access Technology* (RAT) while considering real time system constraints- delay to activate/deactivate heterogeneous RATs, discrete modulation, the way scheduling choices encoded onto signaling protocol (e.g., LTE/NR's DCI), and imperfect available information. However, this approach has not discussed the revenue and interference model in ultra-dense networks, where different priority users can demand non-uniform resources to heterogeneous RAT. In [23], authors have proposed graph coloring based interference mitigation technique in dense small cell networks. However, the proposed centralized approach have not discussed revenue, priority and availability of D2D link in the network.

2.3 Matching Theory in Resource Allocation

A comprehensive tutorial on the use of matching theory in limited available resources is presented in [15]. Authors of [24] provided applications and comprehensive survey on applicability of matching theory, its variant, and significant properties suitable for wireless networks. In [25], authors have first reported the concept of matching theory in general interference network. In [26], authors have presented an energy-efficient resource allocation scheme based on a one-to-one matching approach. The authors employed the Gale-Shapley algorithm for matching of macrocell UEs with D2D pairs. Based on a one-to-one matching approach, a content delivery mechanism in D2D enabled cellular model was proposed in [27]. In [28], Wang *et al*. proposed a matching based model by integrating cluster in hyper-graph theory in local storage-based D2D communication underlying cellular networks. This approach is unable to find a stable matching and needs quite a large number of iterations to converge. In [29], authors have discussed outage-aware matching game for cell selection in LTE/WLAN Multi-RAT. Users approach to RAT based on certain utility function, further RATs accept the request based on availability of resources. In [30], authors have proposed many-to-many stable matching based concept to enable the reliability level of maximum users in multi-user and multi-cellular networks. In [6], a hierarchical matching based model is presented for D2D framework in IoT-enabled network. This model considered homogeneous SR and SP framework. However, this method does not evaluate the case where more than one SR request content at the same time to an SP. In [31], the authors proposed a low-latency and reliable communication computing system for enabling mission-critical applications. Authors have formulated user-server association as a many-to-one matching game with externalities and further addressed via the notion of swap matching. In [32], a matching based resource allocation problem in a heterogeneous network was proposed. Authors have considered a multi-tier architecture where D2D devices try to reuse the spectrum of base station to improve the data rate in the system. The authors have considered the resource allocation for D2D pairs with a single resource requirement. However, in a real scenario, the resource requirement can be more than one depending on the services that are being transmitted between D2D pairs.

In [33], authors have proposed matching based concept for *Ultra-Reliable Low-Latency Communications* (URLLC) with multi users considering a single cell and small-scale fading. In [34], authors have explored the concept of URLLC in multi-connectivity scenario. Authors' developed analytical model explores that single-connectivity can perform better than multi-connectivity mode in presence of interference and competition for limited resources. Furthermore, authors have evaluated resource allocation approach based on stable matching theory to enable wireless URLLC. Authors have applied an extended many-to-one stable matching procedure by employing optimal connectivity approach for each user and optimizing the maximum number of matched resources in the network. However, priority model of different request is missing from the paper. Authors have not discussed cost and earned revenue in the proposed framework.

2.4 Shortcomings of Existing Methods

IDifferent from the traditional cellular networks, the main challenge to analyze the performance of the underlying SPs model is that SPs are often deployed in random order, and the severe interference may drastically deteriorate the performance of different SRs. The existing D2D network model is not directly applicable in D2D-IoT, and FAP enabled heterogeneous SPs' network because IoT devices have limited battery power, and they are much concerned about latency while requesting for services to SPs. However, some SPs can serve more than one SR at a time; therefore, they may need an increased number of radio resources compared to other SPs that serve just one SR. Specifically, to the best of our knowledge, there does not exist any work concerning SPs' revenue maximization via resource allocation with an actual characterization of SRs and interferences in 5G networks.

3 Problem Statement

We consider 5G cellular model that contains a set of SPs and SRs, working underlying MBS. Two types of IoT users are considered, one which provides requested services (i.e., SP) and second, which requests services (i.e., SR). IoT users can form a D2D connection, given that they are within the proximity range to each other. We consider FAPs as SPs in our model, which delivers the requested services to IoT users. An FAP offers services to SR if both lie within the proximity range. However, an SP serves respective SR given that this service complies with interference and availability of radio resources.

IoT users request for services and SPs contest for required 5G radio resources available at MBS to serve the respective

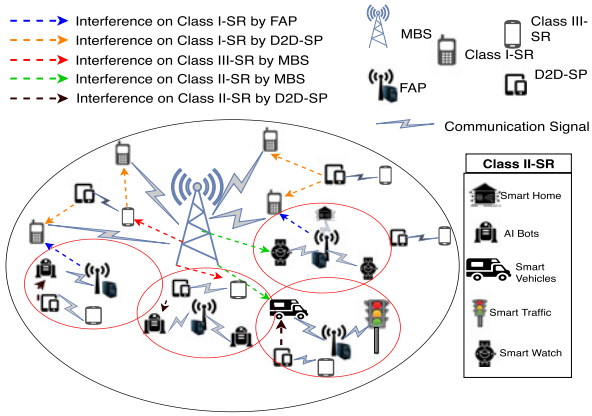


Fig. 1. System model.

services. We have assumed that the association of SRs with SPs has been completed with the help of control signals before serving the request.[1] SPs re-use the limited available radio resources (i.e., PRBs) to serve the respective services. However, an PRB is the smallest unit that can be assigned to a device in the 5G network. It refers to 0.5 ms time slot and 180 kHz frequency band [35]. Due to the limited availability of PRBs, SPs re-use them while serving a request. Moreover, the re-usability of limited available PRBs creates a severe problem of interference at different SRs (as shown in Fig. 1), and this phenomenon leads to a requirement of an efficient resource allocation mechanism in IoT-enabled 5G network.

3.1 System Model

Let the sets of SPs, SRs and PRBs are represented as $\mathbb{K} = \{1, \ldots, \kappa\}$, $\mathbb{S} = \{1, \ldots, S\}$, and $\mathbb{N} = \{1, \ldots, N\}$, respectively. SP selects an appropriate power from a finite power level set $\mathbb{L} = \{1, \ldots, L\}$. However, power level L banks on device density within the network. Furthermore, SPs select an appropriate set of *PRB-power* level combination (we refer as a *resource* throughout this paper) to transmit the requested services to respective SRs given that the selected resource comply with different interference constraints.

Different services may have different *Quality of Service* (QoS) provisioning based on requested data size, required data rate, and latency, as shown in Table 1 [19]. In this work, we consider that QoS for a given SR is dependent on requested service deadline and required radio resources.[2] Specifically, let each service request belongs to one *Service Class* (SC), and SP aims to allocate radio resources, to serve the following three different SCs[17]. (Class I): high-rate and delay-sensitive communications; (Class II): ultra-reliable and low-latency communications; and (Class III): low-rate and delay-tolerant communications.

Depending on the criticality of services, the priority levels of different SCs are arranged as Class I $>$ Class II $>$ Class III. This categorisation is also applicable in *Wireless Body Area Networks* (WBANs) domain where health data is ranked into three different classes such as emergency call, vital health data, and regular health data [36], [37]. As a consequence, emergency call have given the highest priority in resource allocation procedure compared to other two service classes. Moreover, SPs earn revenue by serving the different classes of services in various domains [38]. To maximize the profit, SP has to accomplish the SR's service request. Each SC has a deadline d associated with it and this deadline has to be met while serving the request. In other words, service accomplishment is necessary condition to earn revenue. Depending on the SC and accomplishment of services, SP earns higher figures. Without loss of generality, we formulate a function between revenue earned by SP $K \in \mathbb{K}$ and SC as follows:

TABLE 1
QoS Provisioning [19]

SC	Services	Data Rate [kbps]	Required No. of PRBs	Delay [ms]
Class I	Real-time gaming	128∼384	5∼13	50
	Live streaming	128∼700	5∼24	100
Class II	IP multimedia signaling	128∼384	5∼13	100
	Conventional video	64∼700	3∼24	150
Class III	File sharing	8∼3400	1∼110	300
	Web	8∼3400	1∼110	300

$$M(SC_K) = \begin{cases} R_1, & \text{if } SC_K \in \text{Class I;} \\ R_2, & \text{if } SC_K \in \text{Class II;} \\ R_3, & \text{if } SC_K \in \text{Class III.} \end{cases} \tag{1}$$

We have considered a relation among earned revenues as $R1 > R2 > R3$ depending on the priorities of requests and QoS levels. Particularly, for the sake of simplicity we have considered the values of $R1$, $R2$ and $R3$ as 200, 180 and 140 units, respectively [17]. The aim of these values is to prioritize the resource allocation procedure for higher paying SRs compared to lower paying ones'.

In order to serve SRs, there is a need to allocate the desired number of resources between SPs and SRs. In this regard, we have assumed that an SP can ask q_K number of resources to accomplish the different service requests generated by SRs. For each PRB, there exists a predefined threshold of maximum aggregated interference enforced to high priority SC requester. Let MBS operates at fixed power $P_M^n > 0, \forall n \in \mathbb{N}$. The transmit power vector of SP $K \in \mathbb{K}$ over the PRBs is given by $\mathbb{P} = [P_K^1, P_K^2, \ldots, P_K^N]^T$. $P_K^n > 0$, if PRB n is allocated to SP K, $P_K^n = 0$ otherwise. We introduce a binary variable decision, $X_K^{(n,l)}$, which checks the allocation of power levels and PRBs to an SP K. $X_K^{(n,l)} = 1$, if SP K transmits at PRB n and power level $l = P_K^n$, otherwise $X_K^{(n,l)} = 0$ (description of notations is given in Table 2). However, multiple SPs can transmit on the same PRB by selecting appropriate power levels. Moreover, different applications need different data rate for achieving the QoS criteria as shown in Table 1. Furthermore, required data rates of ECG, Pulse-oximeter, gyroscope insulin actuator, temperature sensor and accelerometer are 71-288 kbps, 16 bps, 1600 bps, 120 bps, and 35 kbps, respectively to achieve the desired QoS levels in WBANs domain [39]. Following the required QoS criterion of different applications we estimate the desired data rate between SP and SR to deliver the requested task. Let $I_K \in \mathbb{S}$ be an SR

1. The session setup and synchronization among the various devices can be done by synchronization or reference signal, i.e., beacons [14].

2. Identification of the QoS as a parameter is absolutely a rough approximation. However, a multidimensional QoS model definitely represents a complete approach. Unfortunately, this solution requires a broader characterization of service types that is out of the scope of our work in its current form.

TABLE 2
Description of Notations

Variables	Description
$\mathbb{K}$	Set of SPs
κ	Number of SPs
$\mathbb{S}$	Set of SRs
S	Number of SRs
$\mathbb{N}$	Set of PRBs
N	Number of PRBs
$\mathbb{L}$	Power level set
L	Number of power levels
$M(SC_K)$	Function between SP and SC
R_i	Earned revenue by serving SC_i
P_M^n	Power level of MBS
$\mathbb{P}$	Transmit power vector of SPs
P_K^n	Binary variable of power level
$X_K^{(n,l)}$	Binary variable of resource allocation
I_K	SR that receive services from SP K
$\mathbb{R}_{I_K}$	Achievable data rate of I_K
$\mathbb{B}$	Bandwidth of a PRB
$\Gamma_{I_K}^{(n)}$	SINR
N_{ther}	Thermal noise
G_{K,I_K}^n	Link gain between K and I_K
G_{K',I_K}^n	Interference gain between K' and I_K
$\mathbb{I}^{(n)}$	Interference experienced by Class I SR on the nth PRB
$\mathbb{J}^{(n)}$	Interference experienced by Class II SR on the nth PRB
$\mathbb{I}_{max}^{(n)}$	Interference threshold on Class I SR
$\mathbb{J}_{max}^{(n)}$	Interference threshold on Class II SR
Δ_{I_K}	Content size of I_K
τ_{I_K}	Service latency of I_K
q_K	Maximum demand of resources
Th_{dr}	Data rate threshold
d_{I_K}	Deadline threshold
ε_{elec}	Energy spent in transmitter electronic circuitry
e_{amp}	Energy spent in amplifiers
E_{Tx}	Transmission energy consumption
E_{Rx}	Reception energy consumption

IoT-device that receives service from SP $K \in \mathbb{K}$. Thus, the achievable data rate of an IoT I_K corresponding to SP K (using *Shannon-Hartley* theorem [40]) can be computed as follows:

$$\mathbb{R}_{I_K} = \sum_{n\in\mathbb{N}} \sum_{l\in\mathbb{L}} X_K^{(n,l)} \mathbb{B}\ log_2\Big(1 + \Gamma_{I_K}^{(n)}\Big), \tag{2}$$

where, $\mathbb{B}$ be bandwidth corresponding to an PRB[3] and $\Gamma_{I_K}^{(n)}$ be *Signal-to-Interference-Plus-Noise Ratio* (SINR),[4] which is computed as follows:

$$\Gamma_{I_K}^{(n)} = \frac{G_{K,I_K}^n P_K^n}{\sum_{K'\in\mathbb{K},K'\neq K} \sum_{l'\in\mathbb{L}} X_{K'}^{(n,l')} G_{K',I_K}^n P_{K'}^n + \sigma^2}, \tag{3}$$

where, $\sigma^2 = N_{ther}\mathbb{B}$ and, N_{ther} denotes thermal noise. G_{K,I_K}^n be link gain between SP K and SR I_K. However, G_{K',I_K}^n is the interference gain between any other SP K' and SR I_K. The interference model is also applicable in intra-WBANs application where sharing of the same resources by excessive patients causes severe interference, incurring low transmission rates [37]. Moreover, limited resources result in slow transmission rates, and dramatically increasing interference cost in OFDMA model. Thus, based on the above Eqs. (2) and (3) we can conclude that the achievable data rate between any SP K and SR I_K also depends upon the choice of other SPs' resources that has not a null link gain with SP K. We use the concept of reference IoT user to calculate the aggregated interference on the nth PRB [44]. Aggregated interference experienced by Class I SR on the nth PRB is written as follows:

$$\mathbb{I}^{(n)} = \sum_{K'\in\mathbb{K},K'\neq K} \sum_{l'\in\mathbb{L}} X_{K'}^{(n,l')} G_{K',m_{K'}^*}^n P_{K'}^n, \tag{4}$$

where, $m_{K'}^* = \underset{m}{\operatorname{argmax}}\ G_{K',m}^n$, $\forall m \in$ Class I. In other words, for any PRB n, the interference caused by SP K' is determined by the highest gain between SP K' and SR of Class I, i.e., Class I SR $m_{K'}^*$, who is the mostly affected by the SP K'. Accordingly, we can write the interference experienced by Class II SR at the nth PRB as follows:

$$\mathbb{J}^{(n)} = \sum_{K'\in\mathbb{K},K'\neq K} \sum_{l'\in\mathbb{L}} X_{K'}^{(n,l')} G_{K',s_K^*}^n P_{K'}^n, \tag{5}$$

where, $s_K^* = \underset{s}{\operatorname{argmax}}\ G_{K',s}^n$, $\forall s \in$ Class II. Interference caused by underlay SPs to Class I and Class II SRs should satisfy a predefined maximum interference threshold $\mathbb{I}^{(n)} < \mathbb{I}_{max}^{(n)}$ and $\mathbb{J}^{(n)} < \mathbb{J}_{max}^{(n)}$, $\forall n \in \mathbb{N}$. However, in this work specifically we have assumed that the maximum threshold values are known to each underlay-tier through feedback control signal [45]. Moreover, if an IoT user I_K requests a content file of size Δ_{I_K} to SP K then the service latency can be calculated as follows:

$$\tau_{I_K} = \frac{\Delta_{I_K}}{\mathbb{R}_{I_K}}. \tag{6}$$

In the real scenario, PRBs are allocated in groups to achieve the desired QoS in the different applications (see, Table 1). For example, LTE uses *Downlink Control Indicator* (DCI) to fulfil the signaling constraints [46]. Moreover, use of DCI in the *Physical Downlink Control Channel* (PDCCH) enables information such that which PRB carry data for which user [22], [46]. As depicted in Fig. 2 of [22], the resource allocation in LTE's DCI can be done in three ways: type 0, 1 and 2. In type 0, consecutive PRBs are grouped into *Resource Block Groups* (RBGs) and the scheduler takes the same action for all the PRBs in a group. Type 1 groups RBGs into subsets using standard modulo and users get individual PRBs allocated within one subset. However, in type 2 any number of virtually continuous PRBs can be allocated via an offset/length pair. Furthermore, these PRBs can either be physically continuous or distributed by standard perpetuation functions [22]. However, in any cases, the scheduler is constrained to assign the PRBs in the groups.

Inspired by type 2 modeling of DCI format, in this work we have assumed that an SP K may need q_K number of PRBs to fulfill the required service demand. In other words, an SP can be assigned with an RBG consisting of q_K number of PRBs. However, the PRBs in an RBG need not to be in continuous

3. We have considered the fixed bit-rate for the same PRB at different users [35]. However, different MIMO configuration that yield different data rates for the same PRB [41], [42] can also be taken into consideration by modifying respective parameters.

4. For the sake of simplicity, we assume that the channel exhibits flat fading. However, this can be extended to frequency-selective fading [43] and fast fading [29] channels as well.

order. In our model, to select the most appropriate q_K PRBs in the group, SP follows the utility function given in Section 4.2.

3.2 Problem Formulation

The aim of this work is to maximize SPs' revenue by allocating proper resources at SP levels, in order to minimize the different interferences in the network. Moreover, increasing the number of resource-allocated SPs in the system, a higher level of revenue can be achieved while complying with a specific set of constraints, as stated in the following:

$$\mathbf{P}: \max_{X_K^{(n,l)}} \sum_{K\in\mathbb{K}} \sum_{n\in\mathbb{N}} \sum_{l\in\mathbb{L}} X_K^{(n,l)} M(SC_K), \tag{7}$$

Subject to the constraints

$$\sum_{l\in\mathbb{L}} \sum_{n\in\mathbb{N}} X_K^{(n,l)} \leqslant q_K \quad ,\forall K \in \mathbb{K}, \tag{8}$$

$$\mathbb{R}_{I_K} > Th_{dr} \quad ,\forall K \in \mathbb{K}, \tag{9}$$

$$\tau_{I_K} < d_{I_K} \quad ,\forall K \in \mathbb{K}, \tag{10}$$

$$\mathbb{I}^{(n)} < \mathbb{I}_{max}^{(n)} \quad ,\forall n \in \mathbb{N}, \tag{11}$$

$$\mathbb{J}^{(n)} < \mathbb{J}_{max}^{(n)} \quad ,\forall n \in \mathbb{N}, \tag{12}$$

$$X_K^{(n,l)} \in \{0,1\} \quad ,\forall n \in \mathbb{N}, \forall K \in \mathbb{K}, \forall l \in \mathbb{L}. \tag{13}$$

The objective of the formulated problem **P** is to maximize the revenue of SPs via allocating resources, subject to the set of *constraints* given in Eqs. (8), (9), (10), (11), (12), and (13). Eq. (8) shows that an SP can ask q_K number of resources based on the number of requested services and the required data rate to deliver them. Since an SP (specifically, FAP) can serve more than one SR at a time, thus it can demand more than one resource to be allocated. However, if an SP has to send a large video clip within a certain deadline, then that SP can also bid for more number of resources. Eq. (9) shows that the data rate for serving an SR should be at least the minimum threshold value to deliver the requested content successfully. Eq. (10) signifies that the service latency between any SP and SR should be less than the threshold value. Eq. (11) indicates a predefined interference threshold value $\mathbb{I}_{max}^{(n)}$ at service Class I requester by limiting the reusability of resources at other SPs. Eq. (12) shows the aggregated interference at service Class II requester by other SPs serving request of the Class III category SRs. Eq. (13) shows a binary variable for PRB and power level alignment.

A centralized solution of the formulated problem **P** is strongly computational hard especially for the large set of $\mathbb{K}$, $\mathbb{N}$ and $\mathbb{L}$. The complexity to solve the problem using centralized exhaustive search is of $O\big((\mathbb{N}\mathbb{L})^{\mathbb{K}}\big)$ even if each SP selects a single resource i.e., $q_K = 1, \forall K \in \mathbb{K}$ [7]. Due to high dimensionality and ill-posed nature of the formulated problem (7), (8), (9), (10), (11), (12), and (13), and the fact that Eq. (10) may not be satisfied for each class of services, this work proposes a complete framework to provide a sub-optimal solution based on the interaction among different agents (namely SPs, PRBs, and power levels) using a distributed stable matching approach in the following section.

4 Static Resource Allocation

The resource allocation mechanism involves many decision-making processes. The confirmation of matching request depends on the preference profiles, i.e., each resource and underlay SP holds preference over the opposite set. The outcome of the matching yields a mutually beneficial assignment between resources and SPs' revenue. In heterogeneous SRs and SPs enabled 5G networks, preference could be based on parameters like *Channel Quality Information* (CQI) and achievable SINR between SR and SP [47]. In fact, the assignment of PRBs and power-level are coupled and motivated us to treat the (PRB, power-level) pairs as one individual entity. All possible combinations of these two types of resources can be enumerated and mapped the SP sets to resource pair sets. However, this process of enumerating and mapping can be under the assistance of MBS, which is responsible for the control signal communication with SRs and SPs. An appropriate matching model that offers this structure is the many-to-many stable matching approach, where SPs will be assigned with various resources and vice-versa. In this section, we first introduced the mapping of the proposed problem with many-to-many matching solution in Section 4.1 followed by utility and preference model in Section 4.2. The proposed algorithm is introduced in Sections 4.3 and further, an illustrative example and analysis of proposed scheme are given in Sections 4.4 and 4.5, respectively.

4.1 Basic Property of Matching Theory

Matching is defined as an assignment of resources to underlay SPs forming the set $\{K, n, l\} \in \mathbb{K} \times \mathbb{N} \times \mathbb{L}$. Each underlay SP can be assigned with multiple PRBs, and multiple SPs can transmit on the same PRB to improve the resource utilization. This scheme naturally falls to many-to-many matching theory.

Definition 1. *A matching φ is defined as a function, i.e., $\varphi : \mathbb{K} \times \mathbb{N} \times \mathbb{L} \rightarrow \mathbb{K} \times \mathbb{N} \times \mathbb{L}$ such that,*

$$\begin{cases} i) \;\; \varphi(K) \in \mathbb{N} \times \mathbb{L} \;\; and \;\; |\varphi(K)| \in \{1, 2, \ldots, q_K\} \\ ii) \;\; \varphi(n) \in \mathbb{K} \times \mathbb{L} \cup \{\phi\} \;\; and \;\; |\varphi(n)| \in \{1, 2, \ldots, \mathbb{K}\} \end{cases},$$

where $\varphi(K) = \{n, l\} \Leftrightarrow \varphi(n) = \{K, l\}, \forall n \in \mathbb{N}, K \in \mathbb{K}, l \in \mathbb{L}$ and $|\varphi(.)|$ denotes the cardinality of matching-outcome $\varphi(.)$.

The above definition implies that φ results in a many-to-many matching between SPs and PRBs. The value of $\varphi(n) = \phi$ indicates that some PRB $n \in \mathbb{N}$ may be unused by any underlay SPs under the matching φ. *The outcome of our proposed algorithm determines a set of allocated PRB and corresponding power level, e.g., $\varphi \equiv \mathbb{X}$, where,*

$$\mathbb{X} = \left[X_1^{(1,1)}, \ldots, X_1^{(1,L)}, \ldots, X_1^{(N,1)}, \ldots, X_{|\mathbb{K}|}^{(N,L)}\right]^T. \tag{14}$$

In the following subsection, we define the utility function for each agent.

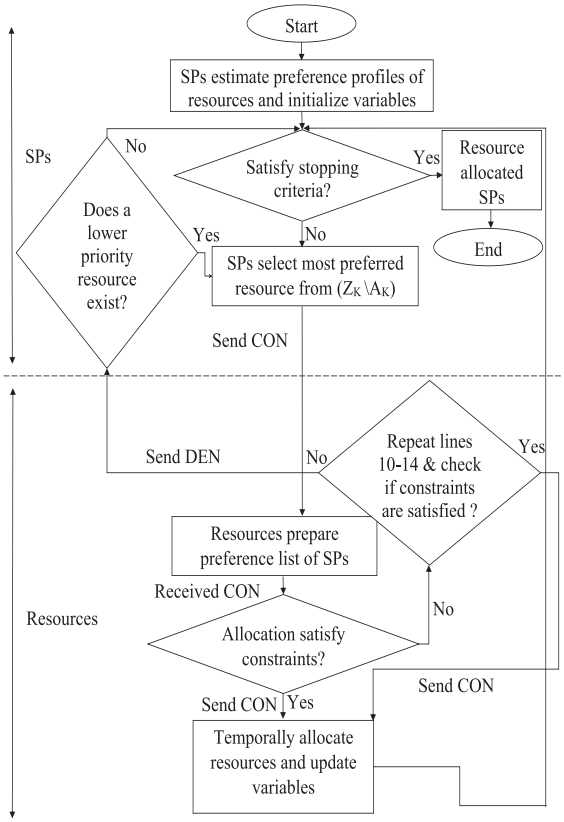


Fig. 2. Flow chart of proposed algorithm.

4.2 Utility and Preference Model

Let the achievable SINR of an SR $I_K \in \mathbb{S}$ on resource (n, l) is denoted as the parameter $\gamma_{I_K}^{(n,l)} \triangleq \Gamma_{I_K|P_K^n=l}^n$, where $\Gamma_{I_K}^n$ is defined in above Eq. (3). We express the data rate as a function of SINR. In particular, let $R(\gamma_{I_K}^{(n,l)}) = \mathbb{B}\ log_2(1 + \gamma_{I_K}^{(n,l)})$ be the achievable data rate of an SP K over resource (n, l). The utility of an SP for a resource (n, l) is ascertained by; (i) achievable data rate for a given (n, l), and (ii) cost function that represents the aggregated interference caused to SRs on that PRB. However, inspired by the work [16] the data rate instead of service delay is considered because actual delay is strongly related to the requested data size to be transmitted from SP to SR.[5] Therefore, data rate is a more fair measurement for utility than the delay value if comparing with other SRs. Thus, the utility of SP $K \in \mathbb{K}$ at resource (n, l) is estimated as follows:

$$\vartheta_K^{(n,l)} = w_1 R\left(\gamma_{I_K}^{(n,l)}\right) - w_2\left(\mathbb{I}^{(n)} - \mathbb{I}_{max}^{(n)}\right) - w_3\left(\mathbb{J}^{(n)} - \mathbb{J}_{max}^{(n)}\right), \tag{15}$$

where w_1, w_2, and w_3 are the biasing factors that can be envisaged according to the status of PRBs to be prioritized for resource allocation among different SCs. Inspired by the existing works [49], [50], the units of w_1, w_2 and w_3 have been respectively selected to have the inverse units of the first (data rate) and, the second and third (interference) components of the utility function. In this case the utility would end up being unitless. Moreover, the numeric values of w_1, w_2 and w_3 are highly subjective to the network parameters and how much significance operators want to give to each component. Particularly, we have considered numeric values of biasing factors such as $w_1 > w_2 > w_3$, that means, any SPs can reuse an PRB if they do not violate the interference constraint at Class I requester, and SPs which serve Class III requester, are allowed to re-use the PRBs if they do not violate the interference constraint at Class II requester based on our problem setting. The Eq. (15) asserts that the PRB can not be re-used by the lower priority classes if this re-usability severely affect the interference at the higher priority classes (e.g., in cases $\mathbb{I}^{(n)} > \mathbb{I}_{max}^{(n)}$ and $\mathbb{J}^{(n)} > \mathbb{J}_{max}^{(n)}$). Moreover, the preference ordering of different SCs is mentioned in Section 3. The preference profile of an underlay SP K on the sets $\mathbb{N}$ and $\mathbb{L}$ can be obtained using the concept of utility function. The preference profile of SP K is defined as a vector of linear order $\mathfrak{P}_K(\mathbb{N}, \mathbb{L}) = [\vartheta_K^{(n,l)}]_{n\in\mathbb{N}, l\in\mathbb{L}}$. The notion $(n_1, l_1) \succeq_K (n_2, l_2)$ denotes that SP K prefers transmission alignment (n_1, l_1) to (n_2, l_2) and this results $\vartheta_K^{(n_1,l_1)} > \vartheta_K^{(n_2,l_2)}$. Similarly, we can write the preference of PRBs over SPs and power levels as $\mathfrak{P}_n(\mathbb{K}, \mathbb{L}) = [\vartheta_K^{(n,l)}]_{K\in\mathbb{K}, l\in\mathbb{L}}$. The notion $(K_1, l_1) \succeq_n (K_2, l_2)$ denotes that PRB n prefers alignment (K_1, l_1) to (K_2, l_2) and this results $\vartheta_{K_1}^{(n,l_1)} > \vartheta_{K_2}^{(n,l_2)}$. Hence, in-order to solve the resource allocation problem in heterogeneous IoT-enabled fog networks, there is need to keep utility value for monitoring the dynamic behavior of entities. Thus, in the following subsection we propose a distributed *Stable Matching based Static Resource Allocation* (SMSRA) algorithm to solve the formulated problem based on above described utility function.

The proposed SMSRA algorithm is applicable in the domain where SPs monitor continuous data request from static SRs. For instant, there is need of continuous monitoring of stroke patients in a rehabilitation centre along with various readings received from sensors mounted on the patient's body. To enable the monitoring, a set of allocated resources is needed between the SPs and the sensors. Since, this kind of applications needs a fixed number of resources for continuous monitoring. Thus, the above continuous monitoring kind of applications where SRs and SPs are static and need a fixed number of radio resources, can be allocated with the resources by applying the following SMSRA algorithm.

4.3 Proposed SMSRA Algorithm

The proposed SMSRA exploits the concept of a many-to-many mapping, which is a modified version of Gale Shapley's one-to-one matching algorithm [51]. However, our proposed many-to-many model adds new contribution in the matching theory as follows: Unlike the pure many-to-many stable matching, our approach additionally combines resource allocation with leveraging individual PRB and power level selection for each SP and optimizing the total earned revenue in the networks. Required number of resources at each SP, preference profile of SPs, and PRBs are the inputs to the algorithm. A list of resource-allocated SPs is the output of the algorithm. The SMSRA works in three phases; lines 1–3 initialize all the parameters that will be used in the update phase (lines 4-30). Line 31 allocates the resources to each SP based on the final matching obtained from the update phase. Two kinds of messages are sent, such as the CON message to request the highest priority entity and DEN message to reject a request. Each

5. Identifying the actual delay due to packet loss or holding it in the buffer due to unavailability of resources will definitely give a complete approach [48]. However, this solution requires a broader characterization of service delays that is out of scope of our work in its current form.

Algorithm 1. Stable Matching Based Static Resource Allocation

Input: Preference profiles $\mathfrak{P}_K(\mathbb{N},\mathbb{L}), \forall K \in \mathbb{K}$ and $\mathfrak{P}_n(\mathbb{K},\mathbb{L}), \forall n \in \mathbb{N}$, q_K
Output: $\mathbb{X} = [X_1^{(1,1)}, \ldots, X_1^{(1,L)}, \ldots, X_1^{(N,1)}, \ldots, X_{|\mathbb{K}|}^{(N,L)}]^T$

Initialization:
1: Each SP estimates CQI
2: $\mathfrak{P}_K(\mathbb{N},\mathbb{L})$ and $\mathfrak{P}_n(\mathbb{K},\mathbb{L})$ are preference profiles prepared by each SP $K \in \mathbb{K}$ and PRB $n \in \mathbb{N}$ based on Eq. (15).
3: $Q_K \leftarrow \phi$, $A_K \leftarrow \phi$, $B_K \leftarrow \phi$, $Z_K \leftarrow \mathfrak{P}_K(\mathbb{N},\mathbb{L})$ /*Initialize variables*/

Update:
4: **While** ($\exists K$ with $|Q_K| < q_K$) or ($\exists n$ with $X_K^{(n,l)} = 0, \forall K \in \mathbb{K}, \forall l \in \mathbb{L}$ and $\mathfrak{P}_n(\mathbb{K},\mathbb{L}) \neq 0$) **do** /*Until allocated*/
5: $A_K \leftarrow MostPreferred(Z_K \backslash A_K)$ /*Select most preferred resources*/
6: **Forall** $(n_{mp}, l_{mp}) \in A_K$ **do** $send(CON, (n_{mp}, l_{mp}))$ /* Send CON to most preferred resources*/
7: When MBS receives $(CON, (n_{mp}, l_{mp}))$ from SP K
8: **If** $\mathbb{I}^{(n_{mp})} \geq \mathbb{I}_{max}^{(n_{mp})}$ or $\mathbb{J}^{(n_{mp})} \geq \mathbb{J}_{max}^{(n_{mp})}$ **then** /*Interference constraints */
9: **Repeat**
10: $(K_{lp}, l_{lp}) \leftarrow$ least preferred SP of n_{mp} at power level l_{lp} /*Select least preferred resource*/
11: $send(DEN, (K_{lp}, l_{lp}))$ /*Send DEN to least preferred resource*/
12: $\mathfrak{P}_{n_{mp}}(\mathbb{K},\mathbb{L}) \leftarrow \mathfrak{P}_{n_{mp}}(\mathbb{K},\mathbb{L}) \backslash (K_{lp}, l_{lp})$, $\mathfrak{P}_{K_{lp}}(\mathbb{N},\mathbb{L}) \leftarrow \mathfrak{P}_{K_{lp}}(\mathbb{N},\mathbb{L}) \backslash (n_{mp}, l_{mp})$ /*Update preference lists*/
13: Update $\mathbb{I}^{(n_{mp})}$ and $\mathbb{J}^{(n_{mp})}$ based on Eqs. (4) and (5) /*Update interference values*/
14: **Until** Eqs. (9), (10), (11), and (12) are satisfied /* Validate set of constraints*/
15: **Else** $send(CON, (n_{mp}, l_{mp}))$ to K /*Send CON to SP K*/
16: **While** $Z_K \neq \phi$ /*Repeat until preference list of SP K becomes empty*/ **do**
17: When SP K receives $(msg, (n_{mp}, l_{mp}))$ message from MBS
18: **If** $msg = CON$ **then** $B_K \leftarrow B_K \cup (n_{mp}, l_{mp})$ /*Upon receive CON, update set B_K*/
19: **If** $msg = DEN$ **then**
20: $Z_K \leftarrow Z_K \backslash (n_{mp}, l_{mp})$ /*Upon receive DEN, remove resource from set Z_K*/
21: **If** $(n_{mp}, l_{mp}) \in A_K$ **then** /*Check if most preferred resource is in set A_K*/
22: $A_K \leftarrow A_K \backslash (n_{mp}, l_{mp})$, $(n'_{mp}, l'_{mp}) \leftarrow MostPreferred(Z_K \backslash A_K)$, $A_K \leftarrow A_K \cup (n'_{mp}, l'_{mp})$ /*Update the sets*/
23: send (CON, (n'_{mp}, l'_{mp})) /*Send CON to next preferred resource*/
24: **If** $\exists (n_{mp}, l_{mp}) \in (A_K \backslash Q_K) \cap B_K$ **then** /*If temporarily allocated resources are available*/
25: $Z_K \leftarrow Z_K \backslash (n_{mp}, l_{mp})$, $B_K \leftarrow B_K \backslash (n_{mp}, l_{mp})$, $Q_K \leftarrow Q_K \cup (n_{mp}, l_{mp})$ /*Update the sets*/
26: **If** $|Q_K| = q_K$ **then** /*Quota for resources*/
27: **Forall** $(n_{lp}, l_{lp}) \in Z_K$ **do** send (DEN,(n_{lp}, l_{lp})) /*Send DEN to lower priority resources*/
28: $Z_K \leftarrow \phi$ /*Reset the preference profile of SP K*/
29: **Forall** $(n_{mp}, l_{mp}) \in Q_K$ **do** $X_K^{(n_{mp}, l_{mp})}$ = 1 /*Set the binary variable as 1 for all resources available in the set Q_K*/
30: Update $\mathfrak{P}_K(\mathbb{N},\mathbb{L})$ and $\mathfrak{P}_n(\mathbb{K},\mathbb{L})$ based on $\mathbb{J}_K^{(n)}$, $\mathbb{I}_K^{(n)}$ /* Update preference lists based on interferences */

Allocation:
31: For each underlay SP, allocate resources (i.e., PRB and power) based on result obtained from above update phase.

message $msg \in \{CON, DEN\}$ has the following formats $msg < SID, DID, (n, l) >$, where, SID, DID, n and l represent source node ID, destination node ID, PRB and power level, respectively. In 5G, *Physical Uplink Control Channel* (PUCCH) [52] and *Physical Downlink Control Channel* [53] can be exploited for exchanging the messages between MBS and SPs. Working flow of the proposed algorithm is also shown in Fig. 2.

In *initialization phase*, preference profiles of SPs and PRBs are estimated. MBS prepares a preference list of SPs on behalf of PRBs. Four-set of variables (A_K, B_K, Q_K, Z_K) is used to demonstrate Algorithm 1. In line 3, A_K represents a set of PRBs to whom SP K has sent a connection request. B_K keeps the information of PRBs, which approaches SP K with a CON request, and Q_K keeps list of allocated resources. Z_K is initialized with a preference profile of each SP K.

The *update phase* is repeated until un-allocated pairs of PRBs and SPs exist (line 4). SP K selects the most preferred resource from set Z_K and sends a CON message for allocation (lines 5-6).

Upon receiving the CON message, MBS executes lines 7-15. If the most preferred PRB receives CON message from an SP that has interference more than or equal to the predefined threshold value, then MBS sends a CON message to respective SP and removes all its least preferred SPs *(based on biasing factor w)* that cause interference to this SP K; and accordingly updates the preference profile of PRBs (lines 8-14). However, if SP K does not violate any constraints, then MBS sends a CON message to respective SP K (line 15).

When an SP receives a message from MBS, it executes the set of actions given in lines 16-29. If received message is CON, then the PRB is added to set B_K (line 18). However, if an SP receives a DEN message, it sends a new connection request to next possible resource and removes the respective PRB from the sets A_K and Z_K (lines 19-23).

If there exists a resource that is not allocated to an SP K, and accepts a connection request, then that resource is assigned to the SP K (i.e., added to the set Q_K and removed from sets B_K and Z_K) as shown in lines 24-25. Lines 26-28, explain that if q_K resources are allocated to an SP K, then that SP sends a denial request to the remaining resources in its preference profile and, consequently, the while loop terminates ($Z_K \leftarrow \phi$). SP K is assigned with the resources of set Q_K (line 29). Each SP and MBS updates its preference profile based on current iteration and allocated resources (lines 29-30). Since the preference list is updated in an iterative manner, thus the proposed algorithm ends up at a local stable matching.

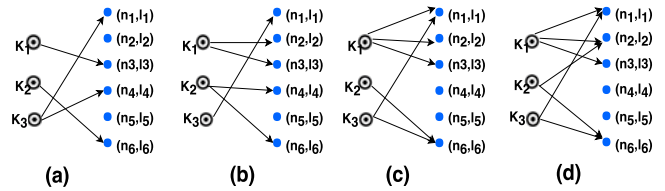


Fig. 3. Illustration of SMSRA. (a) initial allocation, (b) first iteration, (c) second iteration and (d) third iteration.

TABLE 3
Preference Profile of SPs

K_1	K_2	K_3
$(n_2, l_2) : 1 : Yes$	$(n_4, l_4) : 1 : No$	$(n_1, l_1) : 1 : Yes$
$(n_3, l_3) : 2 : Yes$	$(n_6, l_6) : 2 : Yes$	$(n_5, l_5) : 2 : No$
$(n_1, l_1) : 3 : Yes$	$(n_2, l_2) : 3 : Yes$	$(n_6, l_6) : 3 : Yes$
$(n_6, l_6) : 4 : No$	$(n_1, l_1) : 4 : Yes$	$(n_2, l_2) : 4 : No$
$(n_5, l_5) : 5 : Yes$	$(n_3, l_3) : 5 : No$	$(n_3, l_3) : 5 : No$
$(n_4, l_4) : 6 : Yes$	$(n_5, l_5) : 6 : Yes$	$(n_4, l_4) : 6 : No$

4.4 Illustrating the SMSRA Algorithm

Let there be three SPs $\{K_1, K_2, K_3\}$ and six resources $\{(n_1, l_1), (n_2, l_2), (n_3, l_3), (n_4, l_4), (n_5, l_5), (n_6, l_6)\}$ in a system. The required number of connections at each SP is: $\{q_{K_1}, q_{K_2}, q_{K_3}\}$ where, $q_{K_1} = 3$, $q_{K_2} = 2$ and $q_{K_3} = 2$.

Let at some iteration an allocated pairs resulted as $\{\mathbb{X}_{K_1}^{(n_3,l_3)}, \mathbb{X}_{K_2}^{(n_6,l_6)}, \mathbb{X}_{K_3}^{(n_4,l_4)}, \mathbb{X}_{K_3}^{(n_1,l_1)}\}$ (Fig. 3a) and in the same iteration, we obtained the preference and constraints of SPs; and preference profile of resources such as given in Tables 3 and 4, respectively.

In the given Table 3, *Yes* represents that a particular SP and resource combination satisfy the constraints, and *No* represents that they do not satisfy the constraints as per the above allocation. For the sake of simplicity, we have assumed that the above relations (i.e., *Yes* and *No*) do not change with the choice of other SPs. As SP K_1 and resource (n_3, l_3) comply with constraints, hence this allocation is matched. In the same way, K_2 and (n_6, l_6) are matched because this allocation also follows the constraints. As K_3 and (n_4, l_4) do not abide by constraints, the resource (n_4, l_4) looks for other lower-priority SP compared to K_3 but there is no other SP in lower priority list. Thus, MBS sends a DEN message to SP K_3 on behalf of resource (n_4, l_4). Hence, SP K_3 and resource (n_4, l_4) do not paired up. However, K_3 and (n_1, l_1) follow constraints, hence these two agents are paired up.

In the next iteration each SP sends a CON message to higher priority unallocated resource. Thus, a temporary allocation results as $\{\mathbb{X}_{K_1}^{(n_2,l_2)}, \mathbb{X}_{K_1}^{(n_3,l_3)}, \mathbb{X}_{K_2}^{(n_4,l_4)}, \mathbb{X}_{K_2}^{(n_6,l_6)}, \mathbb{X}_{K_3}^{(n_1,l_1)}\}$ (Fig. 3b). The allocation $\mathbb{X}_{K_2}^{(n_4,l_4)}$ does not follow the constraints; and as (n_4, l_4) is not assigned to any other lower priority SP, hence MBS sends DEN message for this allocation. In the next iteration the temporary resource assignment results such as $\{\mathbb{X}_{K_1}^{(n_2,l_2)}, \mathbb{X}_{K_1}^{(n_1,l_1)}, \mathbb{X}_{K_1}^{(n_3,l_3)}, \mathbb{X}_{K_2}^{(n_6,l_6)}, \mathbb{X}_{K_3}^{(n_1,l_1)}, \mathbb{X}_{K_3}^{(n_6,l_6)}\}$, (Fig. 3c). Quotas of K_1 and K_3 are fulfilled and this particular allocation also followed the constraints criteria. Since K_2 still needs one more assignment thus it sends CON request to MBS. Hence, in the next iteration K_2 sends the CON request to its higher priority resource (n_2, l_2). This allocation follows the constraints criteria. Thus, by the end of third iteration a valid match is found such as $\{\mathbb{X}_{K_1}^{(n_1,l_1)}, \mathbb{X}_{K_1}^{(n_2,l_2)}, \mathbb{X}_{K_1}^{(n_3,l_3)}, \mathbb{X}_{K_2}^{(n_2,l_2)}, \mathbb{X}_{K_2}^{(n_6,l_6)}, \mathbb{X}_{K_3}^{(n_1,l_1)}, \mathbb{X}_{K_3}^{(n_6,l_6)}\}$ as shown in Fig. 3d.

TABLE 4
Preference Profile of Resources

(n_1, l_1)	(n_2, l_2)	(n_3, l_3)	(n_4, l_4)	(n_5, l_5)	(n_6, l_6)
$K_3 : 1$	$K_2 : 1$	$K_1 : 1$	$K_1 : 1$	$K_2 : 1$	$K_3 : 1$
$K_1 : 2$	$K_1 : 2$	$K_3 : 2$	$K_2 : 2$	$K_3 : 2$	$K_2 : 2$
$K_2 : 3$	$K_3 : 3$	$K_2 : 3$	$K_3 : 3$	$K_1 : 3$	$K_1 : 3$

4.5 Analysis of SMSRA Algorithm

This section explains the theoretical analysis of the proposed SMSRA algorithm. Particularly, termination, stability, and time complexity of the proposed SMSRA algorithm are explained in the following.

Lemma 1. *SMSRA scheme terminates for every SP $K \in \mathbb{K}$.*

Proof. SMSRA terminates for SP K when MBS replies for each resource with either a connection request or a denial one, i.e., $Z_K = \phi$ or when q_K number of resources are allocated to each SP, i.e., $Q_K = q_K$, $\forall K \in \mathbb{K}$. The algorithm does not terminate only when the SPs keep on resending the CON request in the future for the resources to whom it has previously received the DEN message. Now, if we prove by contradiction that this dependency cannot exist, it would be enough for us to prove that the algorithm always terminates. In the proposed algorithm, the allocated pairs are formed, if and only if the allocation of SP K and resources (n, l) does not violate constraints mentioned in Eqs. (9), (10), (11), and (12). Otherwise, to make a connection, the most preferred resource removes the least preferred SP from its preference profile until a specific threshold value is satisfied. Since an SP removes the respective resource from its preference list once it receives the DEN message and it does not resend CON message for that particular resource in the future; thus, it contradicts the fact that the mentioned dependency would be formed. So, the proposed algorithm always converges. □

The term stability in the matching φ means that neither SPs nor resources prefer to alter the allocation obtained in φ. Formally, we define stability as follows:

Definition 2. *Matching φ is individually rational if there does not exist any SP or resource which would prefer to remain unallocated than to be matched with their current allocation.*

Definition 3. *A matching φ is blocked by a pair of any SP and any resource which would both prefer to be matched with each other than with their current allocation, i.e., $(n, l) \succeq_K \varphi(K)$ and $K \succeq_{(n,l)} \varphi(n, l)$.*

Definition 4. *Matching φ is stable if it is individually rational and is not blocked by any pair of SPs and resources.*

The above definitions can be fathomed with WBAN application where local devices (e.g., SPs) collect data from the sensors mounted on the body of patients [37]. Further, local devices report the data to *Mobile Edge Computing* (MEC) server by selecting the appropriate radio resources while satisfying different constraints. The matching

Algorithm 2. Stable Matching Based Dynamic Resource Allocation

Input: Initial resource allocated vector, $\mathbb{X} = [X_1^{(1,1)}, \ldots, X_1^{(1,L)}, \ldots, X_1^{(N,1)}, \ldots, X_{|\mathbb{K}|}^{(N,L)}]^T$

Phase 1: Execution at SP

Case 1: Arrival of new SP

1: Estimate $\mathfrak{P}_K(\mathbb{N}, \mathbb{L}) \leftarrow$ preference of newly arrived SP K based on CQI and Equation (15); MBS updates $\mathbb{K} \leftarrow K \cup \mathbb{K}$ and $\mathfrak{P}_n(\mathbb{K}, \mathbb{L})$, q_K

2: $Q_K \leftarrow \phi$, $A_K \leftarrow \phi$, $B_K \leftarrow \phi$, $R_K \leftarrow \phi$, $Z_K \leftarrow \mathfrak{P}_K(\mathbb{N}, \mathbb{L})$

3: **While** ($|Z_K - A_K - B_K - R_K| \neq 0$) or ($|Q_K| < q_K$) **do**

4: $(n_{mp}, l_{mp}) \leftarrow MostPreferred(Z_K - A_K - B_K - R_K)$

5: $send(CON, (n_{mp}, l_{mp}))$ message to MBS and update $A_K \leftarrow (n_{mp}, l_{mp})$

6: When SP K receives $(msg, (n_{mp}, l_{mp}))$ message from MBS /*After execution of Phase 2*/

7: **If** $msg = CON$ **then** $B_K \leftarrow B_K \cup (n_{mp}, l_{mp})$

8: **If** $msg = REL$ and $(n_{mp}, l_{mp}) \in Z_K \backslash R_K$ **then** $send(CON, (n_{mp}, l_{mp}))$ to MBS and $A_K \leftarrow A_K \cup (n_{mp}, l_{mp})$

9: **If** $msg = DEN$ **then** $R_K \leftarrow R_K \cup (n_{mp}, l_{mp})$

10: **If** $\exists (n_{mp}, l_{mp}) \in (A_K \backslash Q_K) \cap B_K$ **then** $Q_K \leftarrow Q_K \cup (n_{mp}, l_{mp})$

11: **If** $|Q_K| = q_K$ **then** $\forall (n_{mp}, l_{mp}) \in Q_K$ **do** $X_K^{(n_{mp}, l_{mp})} = 1$ and $\forall (n_{lp}, l_{lp}) \in \{Z_K \backslash Q_K\}$ $send(DEN, (n_{lp}, l_{lp}))$

12: **End While**

Case 2: Release of Resource or Departure of SP

13: A SP K, $send(DEN, (n_{mp}, l_{mp}))$ message to release allocated resource (n_{mp}, l_{mp}) and reset $X_K^{(n_{mp}, l_{mp})} = 0$.

Case 3: Change in Preference

14: When SP K, receives $(REL, (n_{mp}, l_{mp}))$ message from MBS

15: **If** $\vartheta_K^{(n_{mp}, l_{mp})} > \vartheta_K^{(n_{pre}, l_{pre})}$ **then** SP K, $send(CON, (n_{mp}, l_{mp}))$ message

16: Upon receive of $(CON, (n_{mp}, l_{mp}))$ message from MBS

17: SP K $send(DEN, (n_{pre}, l_{pre}))$ message and reset $X_K^{(n_{pre}, l_{pre})} = 0$, and $X_K^{(n_{mp}, l_{mp})} = 1$

Phase 2: Execution at MBS

18: When MBS receives $(msg, (n_{mp}, l_{mp}))$ from SP K

19: **If** $msg = CON$ **then**

20: **If** Eqs. (9), (10), (11), and (12) are not satisfied **then** $(send(DEN, (n_{mp}, l_{mp}))$ to SP K /*Validation of constraints*/

21: **Else** $send(CON, (n_{mp}, l_{mp}))$ to SP K

22: **If** $msg = DEN$**then** update $\mathbb{K} \leftarrow \mathbb{K} \backslash K$, $\mathfrak{P}_n(\mathbb{K}, \mathbb{L})$ and $send(REL, (n_{mp}, l_{mp}))$ to all $K \in \mathbb{K}$

Output: Updated resource allocated vector $\mathbb{X} = [X_1^{(1,1)}, \ldots, X_1^{(1,L)}, \ldots, X_1^{(N,1)}, \ldots, X_{|\mathbb{K}|}^{(N,L)}]^T$.

between local devices and resources can be obtained by applying SMSRA algorithm. The resultant outcome φ forms stable matching between devices and resources within feasible polynomial computational time complexity as proven in the following Theorems 1 and 2, respectively.

Theorem 1. *SMSRA results in stable resource allocation.*

Proof. Let us prove the theorem by contradiction saying that there does not exist stable allocation. Let (n, l) be in higher priority order of SP K but they are not matched in φ. Thus, based on this assumption, (K, n, l) will block φ. Let, $(n, l) \succeq_K (n', l')$ and $\varphi(K) = (n', l')$. This signifies that (n, l) is not assigned to SP K and (K, n', l') is a valid better assignment. Thus, (K, n, l) will never block φ and this contradicts the assumption. Thus, matching φ results in a valid assignment. □

Theorem 2. *Time complexity of SMSRA scheme is $O(\mathbb{KNL})$.*

Proof. The preference profiles of resources and SPs can be sorted out using any sorting method in time complexity of $O(\mathbb{NL}\log \mathbb{NL})$ and $O(\mathbb{KL}\log \mathbb{KL})$, respectively (line 5 and line 30). The update section executes until all the SPs are allocated to at most q_K number of resources. Since, the update section terminates in a finite time (proved in above Lemma 1) and the total input length of the algorithm is equal to summation of both preference profiles, i.e., $\mathfrak{P}_K(\mathbb{N}, \mathbb{L})$ and $\mathfrak{P}_n(\mathbb{K}, \mathbb{L})$ result as $\sum_{K=1}^{\mathbb{K}} |\mathfrak{P}_K(\mathbb{N}, \mathbb{L})| + \sum_{n=1}^{\mathbb{N}} |\mathfrak{P}_n(\mathbb{K}, \mathbb{L})| = 2\mathbb{KNL}$=$O(\mathbb{KNL})$. Using the suitable data structure (as shown in [54]), it can be proved that the update section linearly depends upon $\mathbb{K}$, $\mathbb{N}$ and $\mathbb{L}$. Lines 8-15 will take $O(\mathbb{KNL})$ iteration in the worst case scenario. Similarly, lines 16-28 are linearly related to $\mathbb{KNL}$ because the maximum number of iterations could increase to $\mathbb{KNL}$, i.e., the total input length. Therefore, the overall time complexity of proposed SMSRA is $O(\mathbb{KNL})$. □

The proposed SMSRA works for the static deployment of networks. However, the SMSRA does not describe the dynamic scenario of the system where an SP arrives/leaves or change its preference order to select the most preferred resources. Thus, in the following, we propose a dynamic approach where an SP can leave/join or change the preferences of its currently allocated resources, which in turn disturb the result of the SMSRA algorithm.

5 Dynamic Resource Allocation

The proposed dynamic approach exploits the three specific cases of the network, such as when an SP joins, leaves, or changes its current allocated resources. In the case of joining, newly arrived SP prepares a preference list for resources, and MBS develops a preference list on behalf of

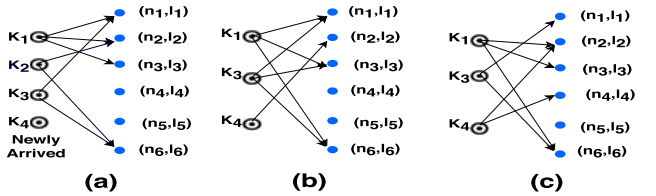


Fig. 4. Example, (a) initial allocation, (b) first iteration, and (c) second iteration.

TABLE 5
Preference Profile of SPs

K_1	K_3	K_4
$(n_2, l_2) : 1 : Yes$	$(n_1, l_1) : 1 : Yes$	$(n_2, l_2) : 4 : Yes$
$(n_6, l_6) : 2 : Yes$	$(n_5, l_5) : 2 : No$	$(n_4, l_4) : 2 : Yes$
$(n_3, l_3) : 3 : Yes$	$(n_6, l_6) : 3 : Yes$	$(n_3, l_3) : 3 : Yes$
$(n_1, l_1) : 4 : Yes$	$(n_2, l_2) : 4 : No$	$(n_1, l_1) : 1 : No$
$(n_5, l_5) : 5 : Yes$	$(n_3, l_3) : 5 : No$	$(n_6, l_6) : 5 : No$
$(n_4, l_4) : 6 : Yes$	$(n_4, l_4) : 6 : No$	$(n_5, l_5) : 6 : No$

resources regarding newly-arrived SP. In the case of departure, MBS deletes the respective SP from the resources' preference list. However, when an SP updates its preference list, no update takes place at the resources' preference list level, but various interferences may transform radically. We propose a distributed *Stable Matching based Dynamic Resource Allocation* (SMDRA) algorithm to solve the above-stated challenges in the following.

The SMDRA algorithm uses five-set of variables $(Q_K, A_K, B_K, R_K, Z_K)$ and three kinds of messages (CON, DEN, REL), to execute the resource allocation procedure. Here, A_K represents a set of resources to whom SP K has sent a connection request, B_K keeps the information of resources which approached SP K with the CON message, Q_K keeps the list of allocated resources and R_K stores resources which sends DEN message to SP K. Z_K is initialized with the preference profile $\mathfrak{P}_K(\mathbb{N}, \mathbb{L})$. REL message has the same format as CON and DEN messages depicted in the SMSRA algorithm. SMDRA works in two phases. Phase 1 executes at SP level (lines 1-17) and Phase 2 (lines 18-22) at the MBS level. Phase 1 is further divided into three cases. Case 1, case 2, and case 3 exploit the arrival of new SP, departure of SP, or release of resources and change in the preference list of SPs, respectively.

Newly arrived SP prepares a preference list of resources based on Eq. (15), and MBS updates the total number of SPs and the preference list of resources, accordingly (line 1). Variables Q_K, A_K, B_K, R_K, are initialized to ϕ and Z_K as the preference list of SP K (line 2). The set $(|Z_K - A_K - B_K - R_K|)$ contains the set of resources that were either never approached and rejected by an SP K or never rejected SP K.

If the number of resources in set Q_K is less than q_K or the set $(|Z_K - A_K - B_K - R_K|)$ is not empty, then lines 4-12 will execute. The most preferred resource (n_{mp}, l_{mp}) is chosen from the set $(|Z_K - A_K - B_K - R_K|)$, sent CON request, and this is added to the set A_K. After sending connection request, SP waits for response from MBS (Phase 2).

Upon receiving messages from SP K, MBS takes the ultimate decision by sending the respective replies. If MBS gets a connection request for resource (n_{mp}, l_{mp}), then it verifies if this particular allocation violates the constraints. If so, then MBS sends a DEN message to respective SP; otherwise, it sends a CON message for this allocation (lines 19-21). However, if MBS receives a denial message for resource (n_{mp}, l_{mp}), then it sends a REL message to inform the existing SPs. Moreover, this REL message is sent to notify the availability of resources.

When an SP K receives a message from MBS, it executes lines 6-11 of the SMDRA algorithm. If the received message is CON, then it updates B_K variable. If SP K receives a $(REL, (n_{mp}, l_{mp}))$ message and this particular resource is also available in its preference list then SP K sends CON message and updates the set A_K. If SP K receives $(DEN, (n_{mp}, l_{mp}))$ message from MBS then it adds the resource (n_{mp}, l_{mp}) in rejected set R_K. However, if selected resource (n_{mp}, l_{mp}) is available in sets A_K and B_K but not included in the set Q_K then it is added to set Q_K. Moreover, if the number of allocated resources to SP K becomes its required number of resources q_K, then all Q_K resources get allocated to SP K, and SP K sends DEN message for other lower-priority resources in the set $(Z_K \backslash Q_K)$.

An SP K releases resource either upon completion of content delivery or if it moves away from the network (Case 2). However, in these two cases, SP K sends $(DEN, (n_{mp}, l_{mp}))$ message to release the resource (n_{mp}, l_{mp}) and reset the allocation $X_K^{n_{mp}, l_{mp}}$.

Case 3 describes the scenario of change in the preference list. An SP changes its preference list based on the availability of better resources compared to the presently allocated ones'. Upon receiving REL message from MBS, SP compares the utility values of presently allocated resource (n_{pre}, l_{pre}) and newly released resource (n_{mp}, l_{mp}). If there is an advantage in the utility value with newly released resource then SP K sends $(CON, (n_{mp}, l_{mp}))$ message to MBS. Upon receiving the $(CON, (n_{mp}, l_{mp}))$ message from MBS SP K releases its currently allocated resource and forms a pair with newly released resource (n_{mp}, l_{mp}).

5.1 An Illustrative Example of SMDRA Algorithm

We employed Fig. 3d as an input to SMDRA scheme. Based on Tables 3 and 4, SMSRA results a final allocation (see Fig. 3d) such as $\{\mathbb{X}_{K_1}^{(n_1,l_1)}, \mathbb{X}_{K_1}^{(n_2,l_2)}, \mathbb{X}_{K_1}^{(n_3,l_3)}, \mathbb{X}_{K_2}^{(n_2,l_2)}, \mathbb{X}_{K_2}^{(n_6,l_6)}, \mathbb{X}_{K_3}^{(n_1,l_1)}, \mathbb{X}_{K_3}^{(n_6,l_6)}\}$. We demonstrate all three cases i.e., arrival, departure and change in preference order of SPs. Let in the above allocation (Fig. 3d), a new SP K_4 arrives (say, $q_{K_4} = 2$) and K_2 *departs from the network* (Fig. 4a). As K_2 departs from the network so it sends DEN message and resets $\mathbb{X}_{K_2}^{(n_2,l_2)} = 0$, and $\mathbb{X}_{K_2}^{(n_6,l_6)} = 0$. Let, there be no change in the preference lists due to new arrival but K_1 receives better proposal for released (n_6, l_6) resource as shown in updated Tables 5 and 6. SPs K_1 and K_4 both send a CON message. K_1 sends $(CON, (n_6, l_6))$ message to MSB and, as this allocation follows the constraints hence, MBS sends CON message to SP K_1. Thus, K_1 sends DEN message to lower preference resource (n_1, l_1) and forms a pair with (n_6, l_6). In the same iteration SP K_4 sends CON message to the most preferred resource (n_2, l_2). Since, this allocation follows the constraints thus they are paired-up, i.e., $\mathbb{X}_{K_4}^{(n_2,l_2)} = 1$ (Fig. 4b). Since, still $Q_K \leq 2$ thus, SP K_4 chooses the next resource, i.e., (n_4, l_4) and sends CON message. This allocations follows the

TABLE 6
Preference Profile of Resources

(n_1, l_1)	(n_2, l_2)	(n_3, l_3)	(n_4, l_4)	(n_5, l_5)	(n_6, l_6)
$K_3 : 1$	$K_4 : 1$	$K_4 : 1$	$K_1 : 1$	$K_3 : 1$	$K_3 : 1$
$K_4 : 2$	$K_1 : 2$	$K_1 : 2$	$K_4 : 2$	$K_1 : 2$	$K_1 : 2$
$K_1 : 3$	$K_3 : 3$	$K_3 : 3$	$K_3 : 3$	$K_4 : 3$	$K_4 : 3$

constraints, hence, receives a CON message for it and gets paired-up i.e., $\mathbb{X}_{K_4}^{(n_4,l_4)} = 1$. Thus, the final allocation is resulted $\{\mathbb{X}_{K_1}^{(n_2,l_2)}, \mathbb{X}_{K_1}^{(n_6,l_6)}, \mathbb{X}_{K_1}^{(n_3,l_3)}, \mathbb{X}_{K_3}^{(n_1,l_1)}, \mathbb{X}_{K_3}^{(n_6,l_6)}, \mathbb{X}_{K_4}^{(n_4,l_4)}, \mathbb{X}_{K_4}^{(n_2,l_2)}\}$ as shown in Fig. 4c.

5.2 Analysis of SMDRA Algorithm

We demonstrate termination, stability and time complexity of the proposed SMDRA algorithm in the following:

Lemma 2. *SMDRA scheme terminates for every SP $K \in \mathbb{K}$.*

Proof. For the dynamic case, any changes may cause some allocation to be abrogated and reissued on SPs. As three cases are mentioned in the SMDRA algorithm to monitor the arrival of new SPs, departure of SPs, or release of resources and change in the preference list of SPs, thus we prove that any of the above-stated changes takes a finite amount of time in the following.

Case 1: Newly arrived SP prepares a preference list and sends a CON message to the most preferred resource. The algorithm does not terminate only if the newly arrived node keeps on posting a CON message to whom it has previously received the DEN message. We prove by contradiction that this dependency does not occur in the proposed SMDRA algorithm. Upon receiving a CON request, MBS corroborates the constraints (Eqs. (9), (10), (11), and (12)). If constraints are not fulfilled, then MBS sends a DEN message; otherwise, it sends a CON message to respective SP. Since an SP adds the respective resource in rejected list R_K and it does not resend CON message for that resource in the future. Thus, it contradicts the fact that the mentioned dependency would be formed. So, the proposed algorithm always converges for case 1.

Case 2: An SP releases the allocated resources before its departure, and MBS updates the total number of available SPs in the network; consequently, this scenario takes a constant number of iterations to converge.

Case 3: This case elaborates the scenario of any changes in the preference list. Upon receiving REL message from MBS, an SP checks if the newly released resource adds in the utility. If the recently released resource results in more utility then the SP requests for it and upon receiving CON message, it releases the allocated resources and gets assigned with the newer ones'. This phenomenon takes a finite number of iteration. Hence case 3 terminates in a fixed amount of repetitions.

Thus, the proposed SMDRA terminates for every SP $K \in \mathbb{K}$ in a finite number of iterations. □

Theorem 3. *SMDRA results in stable resource allocation.*

Proof. Let (n_{mp}, l_{mp}) be in a higher priority order of SP K but they are not matched in φ. This scenario arises upon release of resource (n_{mp}, l_{mp}). However, upon release of any resource, MBS informs each SP and consequently each SP estimates utility value. If (n_{mp}, l_{mp}) is higher in priority then it results in better utility compared to allocated resource (n_{pre}, l_{pre}). If so, then SP K sends CON message and upon getting approval from MBS, it releases resource (n_{pre}, l_{pre}) and matches with the higher priority resource (n_{mp}, l_{mp}). Hence there will never be blocking in the final matching. Thus, matching φ results in a stable assignment. □

Theorem 4. *Time complexity of SMDRA scheme is $O(\mathbb{KNL})$.*

Proof. The arrival of new SP prepares the preference list of all possible resources, i.e., $\mathbb{N} \times \mathbb{L}$, and these resources can be sorted in time complexity of $O(\mathbb{NL}\log \mathbb{NL})$. However, let us consider the worst-case scenario where all $\mathbb{K}$ SPs arrive at a time and request for resources. In this scenario, worst-case time complexity to execute the *case 1* be the summation of time taken by SPs and MBS. Hence, the worst-case time complexity in case 1 be summation of both preference profiles, i.e., $\mathfrak{P}_K(\mathbb{N}, \mathbb{L})$ and $\mathfrak{P}_n(\mathbb{K}, \mathbb{L})$ to be $\sum_{K=1}^{\mathbb{K}} |\mathfrak{P}_K(\mathbb{N}, \mathbb{L})| + \sum_{n=1}^{\mathbb{N}} |\mathfrak{P}_n(\mathbb{K}, \mathbb{L})| = 2\mathbb{KNL}$=$O(\mathbb{KNL})$. In the *case 2*, an SP releases the resource ($O(1)$), and MBS deletes the respective SP from preference profile ($O(\mathbb{K})$). Thus, this phenomenon takes $O(\mathbb{K})$ time complexity in the worst-case scenario. *Case 3* works upon the release of resources. A respective update takes place at the MBS level if the released resource adds utility in the current allocated resource. However, in the worst-case, all SPs send CON request for newly released resources; hence it ends up in case 1 for which we have already proved time complexity as $O(\mathbb{KNL})$. Thus, the worst-case time complexity of the proposed SMDRA algorithm is $O(\mathbb{KNL})$. □

6 Performance Study

In this section, we present the performance of the proposed algorithms through simulation experiments. SPs and SRs are randomly deployed underlying MBS. We have used a distance-dependent transmission model, such as given in [55]. Simulation environment is shown in Table 7. The total number of PRBs is considered as 100 for 20 MHz bandwidth, thermal noise as -174 dBm/Hz. Operational power of MBS, is considered as 46 dBm. The distance between the D2D-IoT pair is assumed 10 m. The path loss between D2D-IoT and SR connected with FAP, MBS and SR connected with FAP, FAP and other SR are considered as $\varrho_o(l) = 15.3 + 40 \times log(R) + Ef_s + \gamma_r$, path loss between FAP and SR connected with FAP, MBS and D2D-IoT, D2D-IoT and other SR are considered as $\varrho_i(l) = 38.46 + 20 \times log(R) + Ef_k$ and path loss between any two SRs connected with FAPs is considered as $\varrho_m(l) = 148 + 40 \times log(R) + Ef_q$. The outdoor wall loss γ_r is set to 2 dB. Variable R denotes the distance between any two devices. The values of Ef_s, Ef_k and Ef_q are set at 6 dB, 10 dB and 6 dB, respectively. The deployment probabilities of Class I, Class II, and Class III SRs are denoted as p_m, p_f, and p_d, respectively. Moreover, delay requirements, data size, and the required number of resources are determined by specific service request types.

We have compared our findings with the existing works [16], [34] and the optimal solution obtained using the "ILOG CPLEX" solver [56]. We have denoted *Student-Project Allocation* (SPA) and *Single-Connectivity* (SC) for implying the proposed algorithms in [16] and[34], respectively. We have

TABLE 7
System Specification

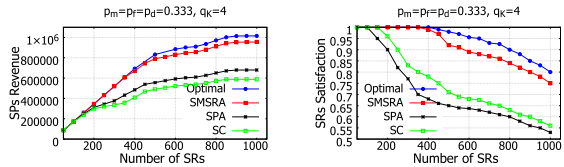
Parameters	Details
Cellular layout	One MBS
Modulation scheme	64 QAM
Carrier frequency	2 GHz
Macro Cell radius	1 KM
q_K	[1-15]
Total PRBs, N	100
Bandwidth	20 MHz
PRB bandwidth	180 kHz
MBS transmit power	46 dBm
Transmission power of SR-IoTs	20 dBm
Thermal noise density	-174 dBm/Hz
Outdoor Wall Loss	2 dB
Proximity Distance	10 meters
w_1, w_2 **and** w_3	0.5, 0.3 and 0.2
Power-level Set $\mathbb{L}$	{20, 24, 30, 36, 42} dBm
Dara rate bound Th_{dr}	[0.002-2] Mbps
R_1, R_2 **and** R_3	200, 180 and 140
Latency bound d_{I_K}	[0.5-200] seconds
Constant propagation loss exponent β	2
Distance between MBS and SPs ϖ	{0.01-1}KM
Energy spent in trans. elec. circ. ε_{elec}	50 nJ/bit
Energy spent in amplifiers e_{amp}	100 pJ/bit/m^2

applied the same framework to compare all algorithms. The preference models and executions of the algorithms are done based on the respective approaches given in the works. Specifically, both existing approaches have applied Many-to-One matching concept. However, SC has considered quota at user level and SPA at SP and resource levels while finding the stable matching. Particularly, we have compared SC model given in [34] due to fact that SC assignment constitutes an upper bound compared to other Multi-Connectivity assignment as proven in Theorem 1 of [34]. Moreover, to compare our proposed algorithm with others methods, we have assumed that the SPs collect request from SRs and respectively approach for resources. We have compared with the existing approaches only when the deployment probabilities or number of SRs (or UEs) in different classes are equally likely in-order to make a fair observation.

6.1 Experimental Results Based on SMSRA Algorithm

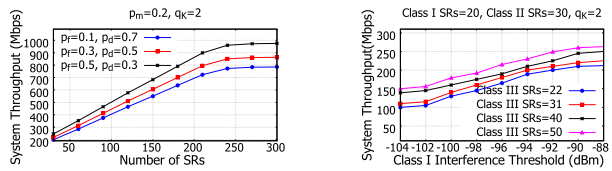
We explained the simulation results based on our proposed SMSRA algorithm for static network scenario.

Revenue and Satisfaction Analysis. We plot the relation between SPs' revenue and the number of SRs in the network. We set the interference threshold values as one dBm, deployment probabilities as $p_m = p_f = p_d = 0.333$ and $q_K = 4$. From Fig. 5a, we can see that even in a dense network scenario, our proposed SMSRA reaches 94 percent to the optimal solution compared to 67 and 58.91 percent of SPA [16] and SC [34] on an average. The reason for the better performance of the proposed SMSRA scheme compared to existing works is that SMSRA facilitates the re-usability of limited resources and D2D communication paradigm between IoT devices that consequently results in a better revenue earned by SPs in the network. However, the reason for getting better revenue from SPA compared to SC is that,



(a) SPs revenue vs SRs. (b) SRs and satisfaction. .

Fig. 5. Revenue and satisfaction analysis.



(a) Throughput vs SRs. (b) Throughput vs $\mathbb{I}_{max}$.

Fig. 6. Throughput analysis.

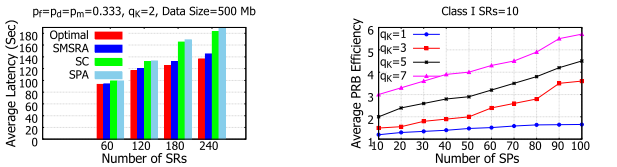
SPA always gives priority to higher paying users while constructing preference list for SPs unlike SC, which just focuses on the SINR value.

We have compared the SRs' satisfaction with respect to the ratio of SRs whose actual service latency meet their requirements. Apparently, there is a trade-off between a number of SRs and SRs' satisfaction, as shown in Fig. 5b. The initial satisfaction ratio of all approaches is one, and after that, the existing schemes drops faster than the other two methods. Moreover, SC performs better than SPA scheme due to fact that while allocating resource to SRs, SC allocates minimum required resources to each SR unlike SPA which primarily focuses on revenue along with minimum required latency. From Figs. 5a and 5b, we conclude that our proposed algorithm not only consider SPs revenue but also each IoT user's satisfaction while allocating resources in the network.

System Throughput. In Fig. 6a, we have shown a comparison between the number of SRs and system throughput. Parameters are set as $q_K = 2$, p_m =0.2, $\mathbb{I}_{max}$= -95 dBm and $\mathbb{J}_{max}$= -98 dBm. As we can see from the result that with an increase of p_f, the total throughput of the system increases. On the other hand, it can be concluded, with an increase in the ratio of p_f/p_d total system throughput increases. The reason behind this finding is that SRs connected with FAPs achieve higher data rates compared to SR and SP IoT devices connected in D2D mode.

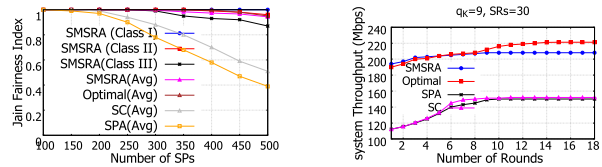
In Fig. 6b, we have shown a relation between $\mathbb{I}_{max}$ and system throughput at $\mathbb{J}_{max}$=-98dBm. The values of q_K, Class I, and Class II requesters are set at 2, 20, and 30, respectively. We can observe that with an increase in the threshold value, system throughput increases. The reason behind this is that with a rise in the threshold value, the re-usability of resources increases. On the other hand, with an increase of Class III SRs, total system throughput increases at a fixed interference threshold value.

Analysis of Latency and PRB Efficiency. In Fig. 7a, we compare the service latency concerning different SRs and respective data sizes at $\mathbb{I}_{max}$= -95 dBm, $\mathbb{J}_{max}$= -98 dBm. We have considered that an SR can request to SP for services of maximum 500 Mb size. Based on the achievable data rate between SP and SR devices, we calculated the service latency, a ratio of requested file size to achievable data rate (



(a) Service latency vs SRs. (b) SPs vs PRB efficiency.

Fig. 7. Analysis of latency and PRB efficiency.



(a) Fairness comparison. (b) Rounds vs throughput.

Fig. 8. Fairness and convergence analysis.

Eq. (6)). From the simulation result we conclude that our proposed re-usability based heuristic performs better than other existing approaches and close to optimal solution. However, SC achieves slight better average latency compared to SPA due to fact that SC allocates resources in iterative way and users with lower quota of resources get their requirement fulfilled at the early unlike SPA which depends on earned revenue of SPs as a primary goal.

In Fig. 7b, we have shown a comparison between the total number of underlay SPs and average PRB efficiency considering different values of q_K at $\mathbb{I}_{max}$=-95 dBm, $\mathbb{J}_{max}$=-98 dBm. PRB efficiency is defined as $\alpha_n = \sum_{i \in K, j \in \mathbb{P}} X_i^{(n,j)}$. As can be seen, with an increase in resource demand of SPs average PRB efficiency increases. The reason is due to an increase in the re-usability of PRBs. The same PRB is reused by many SPs with an increase in resource demand while adhering to interference constraint.

Fairness and Convergence Analysis. Regarding the given results, it is important to assess if resources are fairly distributed between SRs. In Fig. 8a, we have shown Jain's fairness index calculation as the average for all the networks by setting $p_m = p_f = p_m = 1/3$. However, Jain fairness index is a well known metric for evaluation of equatability of resource allocation mechanism, spanning from 1 (best case i.e., equal share) to $1/S$, where S is the number of SRs in the system [57]. We distinguish between different SRs and then show the average for all of that. Although the existing approaches apply the same mechanism for lower priority SRs allocation, the improvement over existing is due to the fact that the relocation of unused resources by rejected high priority SPs allows better distribution of that at remaining lower priority SPs.

In Fig. 8b, we have shown conversant of the SMSRA and compared with the optimal method and existing approaches at $\mathbb{I}_{max}$=-95 dBm, $\mathbb{J}_{max}$=-98 dBm. We set the number of SPs at 5. We observed the system throughput achieved by optimal method is high compared to SMSRA method. However, the proposed SMSRA approach converges in less number of rounds compared to any other methods. Moreover, the average number of iterations becomes constant after a certain round; this is because the different entities receive the most preferred resources while avoiding interferences.

Energy Consumption due to Message Overhead. As discussed in the above Section 4.3, $msg \in \{CON, DEN\}$ has following format $msg < SID, DID, (n, l) >$ for communication between MBS and SPs. Total number of *SID=DID=total numbers of SPs and MBSs*. We have assumed the deployment of SPs underlying a single MBS. Thus, the total number of required bits $\nu = \lceil \log_2(\kappa+1) \rceil + \lceil \log_2(\kappa+1) \rceil + \lceil \log_2(N) \rceil + \lceil \log_2(L) \rceil$ for the msg. Energy consumption for transmission and reception of ν bits of msg over a distance ϖ is given by $E_{Tx} = \nu \times (\varepsilon_{elec} + e_{amp} \times \varpi^{\beta})$ and $E_{Rx} = \nu \times \varepsilon_{elec}$, respectively [58], [59]. Here,

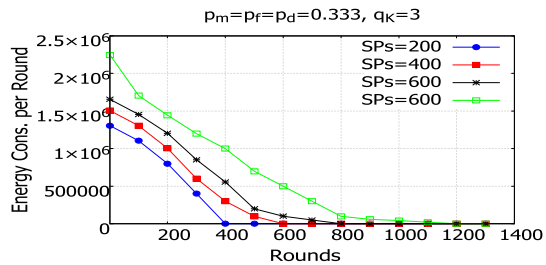


Fig. 9. Energy consumption per round (in nJ).

ε_{elec}= 50 nJ/bit, e_{amp}= 100 pJ/bit/m^2 and β= 2 represent energy spent in transmitter electronic circuitry, energy spent in amplifiers, and constant propagation loss exponent, respectively [59]. Distance ϖ between SPs and MBS is randomly selected between {0.01-1} KM.

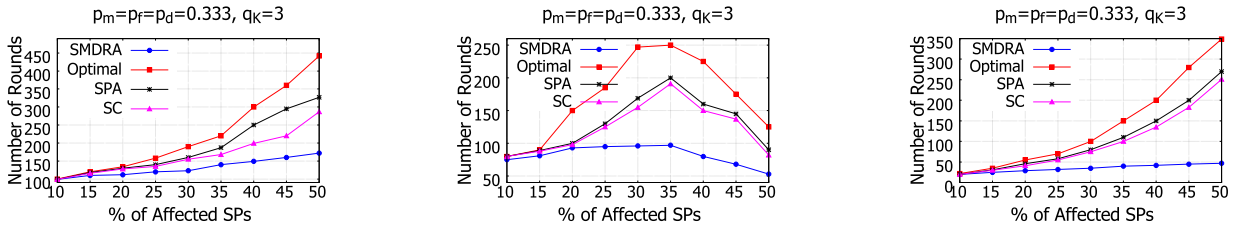
In Fig. 9, we have compared energy consumption per round due to message overhead. Outcome concludes that with increase of SPs, total energy consumption increases. The reasons behind this finding is that with increase of SPs different interferences and competition for selection of limited resources increases, consequently more messages overhead needed which results in higher energy consumption. However, with increase of rounds energy consumption decreases and further reaches to zero. This is because, after a certain number of rounds all SPs select the most appropriate resources (Lemma 1) and message overhead becomes zero in subsequent rounds.

6.2 Experimental Results Based on SMDRA Algorithm

In this subsection, we explained the results obtained based on SMDRA algorithm for dynamic networks. We have considered the scenarios where SPs join/leave or change their preferences in the network.

Re-Convergence Analysis. We examine the proposed SMDRA in rounds. Three cases are considered for execution, such as join, leave, and change in the preference of SPs, as shown in Fig. 10. We varied the number of affected SPs from 10 to 50 percent in our simulation model. We have considered a dense network of 1500 SPs. Deployment probabilities are set as $p_f = p_d = p_m = 0.333$. The maximum resource demand q_K for each SP is set randomly between 1 and 3.

In the join operation (Fig. 10a), an SP arrives and prepares the list of available resources, and tries the best one by sending the CON message as described in the SMDRA algorithm. With an increase in the percentage of newly arrival nodes, the total number of convergence rounds increases. The reason for this finding is that with an increase in the number of SPs, the total number of competitors increases for limited available resources, and consequently, the total number of



(a) Joining of SPs. (b) Leaving of SPs. (c) Preference change of SPs.

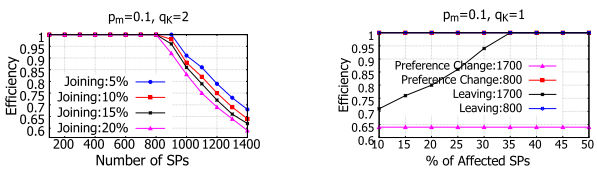
Fig. 10. Re-convergence analysis of different schemes.

rounds increases. However, the number of required rounds in our proposed heuristic is far less than optimal, SPA and SC schemes due to distributed localized nature of SMDRA approach for selection of appropriate resources unlike other centralized schemes which search all possible pairing.

In the case of departure (Fig. 10b), an SP sends a DEN message to inform the MBS regarding the release of allocated resources. Upon availability of resources, MBS sends a REL message to SPs. Moreover, release of the resources becomes an attractive point for all SPs to improve the utility, and consequently, the number of rounds increases. However, leave operation poses a challenge because the allocated resources to the departing SPs may be at a higher priority list of other SPs. Noticed through the simulation of SMDRA, optimal and existing works, when the substantial number of SPs depart (more than 35 percent), the remaining SPs get the most appropriate resources in early rounds. The reason is that, with the departure of SPs, remaining SPs have more available choices for resources than established ones.

In the case of a change in preference (Fig. 10c), an SP releases the allocated resources, and that gets assigned with the newer one. However, change in preference of a large number of SPs affects a higher number of SPs to get it assigned the most referred resources. Thus, with an increase of affected SPs, the number of required rounds to re-converse the SMDRA algorithm increases. However, this effect completes in early rounds using SMDRA algorithm compared to optimal, SPA and SC schemes because SPs are already assigned with valid resources, and MBS sends a CON message to the most appropriate ones. Moreover, compared to existing works, our proposed algorithm selects most appropriate pairing while avoiding interference and keeping a wide searching range for minimizing number of rounds among already configured PRBs as well as updated preference list unlike other centralized approaches.

Efficiency of SMDRA Algorithm. We estimates the efficiency of proposed SMDRA algorithm as follows:



(a) Efficiency vs SPs. (b) Efficiency vs affected SPs.

Fig. 11. Efficiency analysis of proposed SMDRA.

$$Efficiency = \frac{\sum_{K\in\mathbb{K}}\sum_{n\in\mathbb{N}}\sum_{l\in\mathbb{L}} X_K^{(n,l)}}{\sum_{K\in\mathbb{K}} q_K}. \tag{16}$$

We compared the joining of new SPs concerning efficiency in Fig. 11a. We varied the percentage of newly joined SPs from 5 to 20 percent and the number of existing SPs from 200 to 1,400 in the network. From the outcome, we conclude that with a higher percentage of newly joined SPs, efficiency goes down. Moreover, with an increase of SPs, efficiency decreases. In other words, there is a trade-off between efficiency and the number of SPs. The reason for this trade-off is that increase of SPs creates interference to other SRs, consequently efficiency decreases.

In Fig. 11b, we have shown a correlation between the percentage of affected SPs and efficiency concerning preference change and departure of SPs. We varied the rate of affected SPs by fixing the total number of SPs at 800 and 1700. From the outcome, we conclude that efficiency does not depend upon the percentage of affected SPs. However, an increase of SPs could result in a decrease of efficiency, as found in the cases of 800 and 1700 SPs. The reason for this finding is that with an increase in SPs, interference increases, and consequently, efficiency decreases. Moreover, the departure of SPs may help to improve efficiency, as shown in a case where 1700 SPs leave the network. The reason is that when SPs leave the system, the MBS reassigns the released resources to other waiting SPs, consequently efficiency increases. However, if all the SPs are already assigned with valid resources, then there does not occur any change at efficiency level concerning leave and change in the preference of SPs.

7 Conclusion and Future Works

In this work, we proposed many-to-many stable matching based resource allocation algorithms for service providers' revenue maximization in 5G wireless (heterogeneous) networks, while achieving a guaranteed level of quality for the requested services. Each SP/resource is free to build its preference list according to the utility function. We have found that the proposed algorithms terminated and succeeded in maximizing the throughput of the network. Our proposed approach outperforms state-of-the-art schemes and achieves 94 percent of the optimal value.

The current form of the manuscript has considered only three priority classes. However, n-level of priority classes is more practical model to be considered in our future work. Moreover, we would like to formulate the resource allocation in IoT enabled fog architecture as an online

optimization problem where SRs and SPs can switch their behaviours while requesting for resources and serving to requests underlying cellular 5G networks. Furthermore, the dynamic mapping of varying interference threshold and revenue model is another dimension need to be explored. The proposed scheme has considered single connectivity with fixed modulation scheme and uniform data rate per PRB. However, to fulfill the requirement of highly dense upcoming network, there is need for multi-connectivity [22] framework with adaptive modulation and coding scheme having heterogeneous MIMO configuration [43] at SPs and SRs levels to achieve the higher data rate between them. Identification of multidimensional Quality of Experience (QoE) as a sole parameter while maximizing SP's revenue is an interesting topic to be explored. Deriving approximation ratio and finding the trade-off with existing schemes are other directions where we'll be focusing on. In our future work, we'll jointly optimize the radio resource allocation and cache placement in an IoT-enabled 5G network. Finally, exciting paths of future research would be to develop a variation of the proposed scheme that can guarantee latency and freshness of data constraints [37] for the Internet of Medical Things in WBAN underlying cellular 5G networks.

Acknowledgments

The authors would like to thank the Associate Editor and anonymous reviewers for insightful comments that helped us improve the quality of the manuscript significantly. The work of Ajay Pratap was supported in part by the Science and Engineering Research Board (SERB), Government of India under Grant SRG/2020/000318. The work of Sajal K. Das was supported in part by the NSF under Grant CNS-1818942, CCF-1725755, and SCC-IRG 1952045.

References

[1] M. Aazam, K. A. Harras, and S. Zeadally, "Fog Computing for 5G tactile industrial Internet of Things: QoE-aware resource allocation model," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 3085–3092, May 2019.

[2] B. M. ElHalawany, R. Ruby, and K. Wu, "D2D communication for enabling Internet-of-Things: Outage probability analysis," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2332–2345, Mar. 2019.

[3] I. B. de Almeida, L. L. Mendes, J. J. Rodrigues, and M. A. da Cruz, "5G waveforms for IoT applications," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2554–2567, Third Quarter 2019.

[4] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proc. 1st Ed. MCC Workshop Mobile Cloud Comput.*, 2012, pp. 13–16.

[5] A. Pratap, R. Gupta, V. S. Siddhardh Nadendla, and S. K. Das, "On maximizing task throughput in IoT-enabled 5G networks under latency and bandwidth constraints," in *Proc. IEEE Int. Conf. Smart Comput.*, 2019, pp. 217–224.

[6] B. Wang, Y. Sun, S. Li, and Q. Cao, "Hierarchical matching with peer effect for low-latency and high-reliable caching in social IoT," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 1193–1209, Feb. 2019.

[7] A. Pratap, S. Singh, S. Satapathy, and S. K. Das, "Maximizing joint data rate and resource efficiency in D2D-IoT enabled multi-tier networks," in *Proc. IEEE 44th Conf. Local Comput. Netw.*, 2019, pp. 177–184.

[8] A. Pratap, R. Gupta, V. S. S. Nadendla, and S. K. Das, "Bandwidth-constrained task throughput maximization in IoT-enabled 5G networks," *Pervasive Mobile Comput.*, vol. 69, 2020, Art. no. 101281.

[9] S. Jošilo and G. Dán, "Decentralized algorithm for randomized task allocation in fog computing systems," *IEEE/ACM Trans. Netw.*, vol. 27, no. 1, pp. 85–97, Feb. 2019.

[10] X. Li, Y. Liu, H. Ji, H. Zhang, and V. C. Leung, "Optimizing resources allocation for fog computing-based Internet of Things networks," *IEEE Access*, vol. 7, pp. 64907–64922, 2019.

[11] S. F. Abedin, M. G. R. Alam, S. M. A. Kazmi, N. H. Tran, D. Niyato, and C. S. Hong, "Resource allocation for ultra-reliable and enhanced mobile broadband IoT applications in fog network," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 489–502, Jan. 2019.

[12] R. Zhang, M. Wang, X. Shen, and L.-L. Xie, "Probabilistic analysis on QoS provisioning for Internet of Things in LTE-A heterogeneous networks with partial spectrum usage," *IEEE Internet Things J.*, vol. 3, no. 3, pp. 354–365, Jun. 2016.

[13] S. F. Abedin, A. K. Bairagi, M. S. Munir, N. H. Tran, and C. S. Hong, "Fog load balancing for massive machine type communications: A game and transport theoretic approach," *IEEE Access*, vol. 7, pp. 4204–4218, 2018.

[14] G. Lee, W. Saad, and M. Bennis, "An online optimization framework for distributed fog network formation with minimal latency," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2244–2258, Apr. 2019.

[15] Y. Gu, W. Saad, M. Bennis, M. Debbah, and Z. Han, "Matching theory for future wireless networks: Fundamentals and applications," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 52–59, May 2015.

[16] Y. Gu, Z. Chang, M. Pan, L. Song, and Z. Han, "Joint radio and computational resource allocation in IoT fog computing," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7475–7484, Aug. 2018.

[17] R. Fantacci and B. Picano, "When network slicing meets prospect theory: A service provider revenue maximization framework," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3179–3189, Mar. 2020.

[18] Y. Zhang, Z. Xiong, D. Niyato, P. Wang, and Z. Han, "Information trading in Internet of Things for smart cities: A market-oriented analysis," *IEEE Netw.*, vol. 34, no. 1, pp. 122–129, Jan./Feb. 2020.

[19] Y.-S. Liang, W. Chung, G. Ni, I. Chen, H. Zhang, and S. Kuo, "Resource allocation with interference avoidance in OFDMA femtocell networks," *IEEE Trans. Veh. Technol.*, vol. 61, no. 5, pp. 2243–2255, Jun. 2012.

[20] Q.-V. Pham, T. Leanh, N. H. Tran, B. J. Park, and C. S. Hong, "Decentralized computation offloading and resource allocation for mobile-edge computing: A matching game approach," *IEEE Access*, vol. 6, pp. 75868–75885, 2018.

[21] J. Xie, Y. Jia, Z. Chen, Z. Nan, and L. Liang, "D2D computation offloading optimization for precedence-constrained tasks in information-centric IoT," *IEEE Access*, vol. 7, pp. 94888–94898, 2019.

[22] L. Diez, A. Garcia-Saavedra, V. Valls, X. Li, X. Costa-Perez, and R. Agüero, "LaSR: A supple multi-connectivity scheduler for multi-RAT OFDMA systems," *IEEE Trans. Mobile Comput.*, vol. 19, no. 3, pp. 624–639, Mar. 2020.

[23] H. Zhang, K. Yang, S. Zhang, and O. A. Dobre, "Subchannel allocation based on clustered interference alignment for differentiated data streams in dense small cell networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 14049–14054, Nov. 2020.

[24] S. Bayat, Y. Li, L. Song, and Z. Han, "Matching theory: Applications in wireless communications," *IEEE Signal Process. Mag.*, vol. 33, no. 6, pp. 103–122, Nov. 2016.

[25] E. A. Jorswieck, "Stable matchings for resource allocation in wireless networks," in *Proc. 17th Int. Conf. Digit. Signal Process.*, 2011, pp. 1–8.

[26] Z. Zhou, K. Ota, M. Dong, and C. Xu, "Energy-efficient matching for resource allocation in D2D enabled cellular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5256–5268, Jun. 2017.

[27] C. Xu, C. Gao, Z. Zhou, Z. Chang, and Y. Jia, "Social network-based content delivery in Device-to-Device underlay cellular networks using matching theory," *IEEE Access*, vol. 5, pp. 924–937, 2016.

[28] L. Wang *et al.*, "Hypergraph-based wireless distributed storage optimization for cellular D2D underlays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2650–2666, Oct. 2016.

[29] A. I. Abdulshakoor, M. Anany, and M. M. Elmesalawy, "Outage-aware matching game approach for cell selection in LTE/WLAN multi-RAT HetNets," *Comput. Netw.*, vol. 183, 2020, Art. no. 107596.

[30] M. Simsek, T. Hößler, E. Jorswieck, H. Klessig, and G. Fettweis, "Multiconnectivity in multicellular, multiuser systems: A matching-based approach," *Proc. IEEE*, vol. 107, no. 2, pp. 394–413, Feb. 2019.

[31] C.-F. Liu, M. Bennis, M. Debbah, and H. V. Poor, "Dynamic task offloading and resource allocation for ultra-reliable low-latency edge computing," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4132–4150, Jun. 2019.
[32] S. Shamaei, S. Bayat, and A. M. A. Hemmatyar, "Interference management in D2D-enabled heterogeneous cellular networks using matching theory," *IEEE Trans. Mobile Comput.*, vol. 18, no. 9, pp. 2091–2102, Sep. 2019.
[33] T. Hößler, M. Simsek, and G. P. Fettweis, "Matching-based resource allocation for multi-user URLLC in unlicensed frequency bands," in *Proc. 16th Int. Symp. Wireless Commun. Syst.*, 2019, pp. 102–106.
[34] T. Hößler, P. Schulz, E. A. Jorswieck, M. Simsek, and G. P. Fettweis, "Stable matching for wireless URLLC in multi-cellular, multi-user systems," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5228–5241, Aug. 2020.
[35] A. Pratap, R. Misra, and S. K. Das, "Maximizing fairness for resource allocation in heterogeneous 5G networks," *IEEE Trans. Mobile Comput.*, vol. 20, no. 2, pp. 603–619, Feb. 2021.
[36] K. Zhang *et al.*, "PHDA: A priority based health data aggregation with privacy preservation for cloud assisted WBANs," *Inf. Sci.*, vol. 284, pp. 130–141, 2014.
[37] Z. Ning *et al.*, "Mobile edge computing enabled 5G health monitoring for Internet of medical Things: A decentralized game theoretic approach," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 463–478, Feb. 2021.
[38] C. Yi, Z. Zhao, J. Cai, R. L. de Faria, and G. M. Zhang, "Priority-aware pricing-based capacity sharing scheme for beyond-wireless body area networks," *Comput. Netw.*, vol. 98, pp. 29–43, 2016.
[39] S. Misra and S. Sarkar, "Priority-based time-slot allocation in wireless body area networks during medical emergency situations: An evolutionary game-theoretic perspective," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 2, pp. 541–548, Mar. 2015.
[40] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, pp. 3–55, Jan. 2001.
[41] G. Y. Li, J. Niu, D. Lee, J. Fan, and Y. Fu, "Multi-cell coordinated scheduling and MIMO in LTE," *IEEE Commun. Surveys Tuts*, vol. 16, no. 2, pp. 761–775, Second Quarter 2014.
[42] Y. Yang, B. Bai, and W. Chen, "Spectrum reuse ratio in 5G cellular networks: A matrix graph approach," *IEEE Trans. Mobile Comput.*, vol. 16, no. 12, pp. 3541–3553, Dec. 2017.
[43] K.-K. Wong, R. D. Murch, and K. B. Letaief, "Optimizing time and space MIMO antenna system for frequency selective fading channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 7, pp. 1395–1407, Jul. 2001.
[44] K. Son, S. Lee, Y. Yi, and S. Chong, "REFIM: A practical interference management in heterogeneous wireless access networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1260–1272, Jun. 2011.
[45] P. Gope, "LAAP: Lightweight anonymous authentication protocol for D2D-aided fog computing paradigm," *Comput. Secur.*, vol. 86, pp. 223–237, 2019.
[46] G. Ku and J. M. Walsh, "Resource allocation and link adaptation in LTE and LTE advanced: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1605–1633, Third Quarter 2015.
[47] A. Virdis, C. Vallati, G. Nardini, G. Tanganelli, G. Stea, and E. Mingozzi, "D2D communications for large-scale fog platforms: Enabling direct M2M interactions," *IEEE Veh. Technol. Mag.*, vol. 13, no. 2, pp. 24–33, Jun. 2018.
[48] M. Irazabal, E. Lopez-Aguilera, I. Demirkol, and N. Nikaein, "Dynamic buffer sizing and pacing as enablers of 5G low-latency services," *IEEE Trans. Mobile Comput.*, early access, Aug. 17, 2020, doi: 10.1109/TMC.2020.3017011.
[49] P. Semasinghe, E. Hossain, and K. Zhu, "An evolutionary game for distributed resource allocation in self-organizing small cells," *IEEE Trans. Mobile Comput.*, vol. 14, no. 2, pp. 274–287, Feb. 2015.
[50] M. Hasan and E. Hossain, "Distributed resource allocation in D2D-enabled multi-tier cellular networks: An auction approach," in *Proc. IEEE Int. Conf. Commun.*, 2015, pp. 2949–2954.
[51] D. Gale and L. S. Shapley, "College admissions and the stability of marriage," *Amer. Math. Monthly*, vol. 69, no. 1, pp. 9–15, 1962.
[52] L. Kundu, G. Xiong, and J. Cho, "Physical uplink control channel design for 5G new radio," in *Proc. IEEE 5G World Forum*, 2018, pp. 233–238.
[53] K. Takeda, H. Xu, T. Kim, K. Schober, and X. Lin, "Understanding the heart of the 5G air interface: An overview of physical downlink control channel for 5G new radio," *IEEE Commun. Standards Mag.*, vol. 4, no. 3, pp. 22–29, Sep. 2020.
[54] G. O'Malley, "Algorithmic aspects of stable matching problems," PhD dissertation, Dept. Inform. Math. Sci., Univ. Glasgow, Glasgow, U.K., 2007.
[55] Z. Bai, "Evolved universal terrestrial radio access (E-UTRA); physical layer procedures," 3GPP, Sophia Antipolis, Technical Specification 36.213 v. 11.4. 0, 2013. Accessed: Mar. 13, 2021. [Online]. Available: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjRqMzu96zvAhW57XMBHaEIB48QFjAAegQIARAF&url=https%3A%2F%2Fwww.etsi.org%2Fdeliver%2Fetsi_ts%2F136200_136299%2F136213%2F11.04.00_60%2Fts_136213v110400p.pdf&usg=AOvVaw0kfUxTCXVkJrGzHCdL8r4a
[56] CPLEX optimizer. Accessed: Mar. 15, 2020. [Online]. Available: https://www.ibm.com/analytics/cplex-optimizer
[57] R. Jain, D.-M. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," DEC Res., Hudson, MA, USA, Tech. Rep. TR-301, vol. 38, 1984.
[58] M. Naeem *et al.*, "Distributed gateway selection for M2M communication in cognitive 5G networks," *IEEE Netw.*, vol. 31, no. 6, pp. 94–100, Nov./Dec. 2017.
[59] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. 33rd Annu. Hawaii Int. Conf. Syst. Sci.*, 2000, pp. 10–pp.

Ajay Pratap (Member, IEEE) received the PhD degree in computer science and engineering from the Indian Institute of Technology Patna, India, in July 2018. He is currently an assistant professor with the Department of Computer Science and Engineering, Indian Institute of Technology (BHU) Varanasi, India. Before joining IIT (BHU), he was associated with the Department of Computer Science and Engineering, National Institute of Technology Karnataka (NITK) Surathkal, India, as an assistant professor from December 2019 to May 2020. He worked as a postdoctoral researcher with the Department of Computer Science, Missouri University of Science and Technology, Rolla, Missouri, from August 2018 to December 2019. His research interests include cyber-physical systems, IoT-enabled smart environments, statistical learning, algorithm design for next-generation advanced wireless networks, mobile computing and networking, applied graph theory, and game theory. His current work is related to HetNet, small cells, fog computing, IoT, and D2D communication underlying cellular 5G and beyond.

Sajal K. Das (Fellow, IEEE) is currently a professor of computer science and Daniel St. Clair Endowed chair at the Missouri University of Science and Technology, Rolla, Missouri, where he was the chair of Computer Science Department during 2013-2017. His research interests include wireless sensor networks, mobile and pervasive computing, cyber-physical systems and IoTs, smart environments, cloud computing, cyber security, biological and social network. He has published more than 700 papers in high quality journals and refereed conference proceedings. He holds 5 US patents and coauthored 4 books. He is a recipient of ten best paper awards, IEEE Computer Society's Technical Achievement Award for pioneering contributions to sensor networks, and University of Missouri System President's Award for Sustained Career Excellence. He is the founding editor-in-chief of the *Elsevier's Pervasive and Mobile Computing*, and associate editor of the *IEEE Transactions on Dependable and Secure Computing*, the *IEEE Transactions on Mobile Computing*, and the *ACM Transactions on Sensor Networks*.