IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON D2D COMMUNICATIONS: SECURITY ISSUES AND RESOURCE ALLOCATION

Received November 30, 2018, accepted December 20, 2018, date of publication January 30, 2019, date of current version March 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2896008

NOMA-Based Resource Allocation and Mobility Enhancement Framework for IoT in Next Generation Cellular Networks

RAOUF ABOZARIBA[1], (Member, IEEE), MUHAMMAD KAMRAN NAEEM[2], (Member, IEEE), MOHAMMAD PATWARY[3], (Senior Member, IEEE), MIR SEYEDEBRAHIMI[4], PETER BULL[3], (Member, IEEE), AND ADEL ANEIBA[3]

[1]School of Computing and Communications, Lancaster University, Lancaster LA1 4WA, U.K.
[2]Research Innovation and Enterprise, Southampton Solent University, Southampton SO14 0YN, U.K.
[3]Intelligent Systems and Networks, School of Computing and Digital Technologies, Birmingham City University, Birmingham B4 7BD, U.K.
[4]Warwick Manufacturing Group, University of Warwick, Coventry CV4 7AL, U.K.

Corresponding author: Raouf Abozariba (r.abozariba@ieee.org)

The work of R. Abozariba was supported by the Department of Digital, Culture, Media, and Sport (DCMS) and Innovate U.K., under the project 5G Rural Integrated Tested (5GRIT).

ABSTRACT With the unprecedented technological advances witnessed in the last two decades, more devices are connected to the Internet, forming what is called the Internet of Things (IoT). The IoT devices with heterogeneous characteristics and the quality of experience (QoE) requirements may engage in the dynamic spectrum market due to the scarcity of radio resources. We propose a framework to efficiently quantify and supply radio resources to the IoT devices by developing intelligent systems. The primary goal of this paper is to study the characteristics of the next generation of cellular networks with non-orthogonal multiple access (NOMA) to enable connectivity to clustered IoT devices. First, we demonstrate how the distribution and QoE requirements of IoT devices impact the required number of radio resources in real time. Second, we prove that using an extended auction algorithm by implementing a series of complementary functions enhance the radio resource utilization efficiency. The results show a substantial reduction in the number of sub-carriers required when compared with conventional OMA and the intelligent clustering is scalable and adaptable to the cellular environment. Ability to move spectrum usages from one cluster to other clusters after borrowing when a cluster has fewer users or move out of the boundary is another soft feature that contributes to the reported radio resource utilization efficiency. Moreover, the proposed framework provides IoT service providers cost estimation to control their spectrum acquisition to achieve the required quality of service with a guaranteed bit rate (GBR) and non-GBR.

INDEX TERMS 5G networks, clustering, dynamic resource allocation, IoT, NOMA, network slicing.

I. INTRODUCTION

The internet of things (IoT) refers to the interconnection of uniquely-identifiable embedded devices within the internet infrastructure. It is forecast that in the next few years we will witness a deployment of billions more connected devices, enabling new, wide-ranging use cases, including energy and utility monitoring, health-care, autonomous vehicles (AVs) and mission-critical services [1]–[3]. This will generate significant amount of traffic, transmitted over the radio frequency spectrum [4]. For example, it is predicted by 2025, AVs alone will upload over one terabyte (TB) of vehicle and sensor data per month to the cloud.

The associate editor coordinating the review of this manuscript and approving it for publication was Mauro Fadda.

Typically wireless IoT traffic is transmitted over unlicensed spectrum such as the instrument, scientific and medical (ISM) bands. For future sustainability of the IoT technology, however, the question is whether the ISM bands, used by the underlying internet architecture, such as LoRa, Sigfox, Weightless and many other platforms will be flexible enough to stretch to the potential of IoT. For example, low-latency and high-throughput requirements are expected to be necessary to support use cases such as health monitoring and V2X communications. Currently these applications are beyond the capability of current IoT platforms. One solution to provide connectivity and to address the spectrum demands of the IoT devices are the 5G cellular networks, enabling a wide range of data rates with high availability and reliability. Other advantages of 5G cellular networks over traditional

TABLE 1. IoT wireless access technologies.

Parameters \ Technology	LoRa, SigFox & Weightless etc.	eMTC & NB-IoT	5G-NR
Accessing spectrum	Unlicensed (ISM bands)	Licensed	Licensed
Modulation scheme (Uplink)	OMA	OMA	OMA & NOMA
Operating frequency (min-max)	7.8-500 (KHz)	180-1080 (KHz)	Sub-1GHz, 1-6 GHz
Channel bandwidth	15 (KHz)	200 (KHz)	variable (5 - 400 MHz)
Range (max)	50 (km)	25 (km)	2 (km)
Throughput	600 (bps)	1 (Mbps)	>100 Mbps
Transmit power (max)	20 (dBm)	23 (dBm)	30 (dBm)

IoT wireless access technologies include improvement of global coverage, long-term availability and technological advantages with respect to spectral efficiency, latency and data throughput.

Non-orthogonal multiple access (NOMA) has been considered as a promising candidate to increase the connectivity and to improve spectrum utilization in 5G cellular networks. The use of NOMA is not only expected to support the users with good channel conditions but also the users with poor channel conditions, utilizing the same bandwidth resources. Research in this area has demonstrated that NOMA can also support massive connectivity, an important characteristic in the forthcoming 5G networks in order to support the IoT industry [5]. In addition to the capabilities of NOMA to increase spectrum utilization of future cellular communications, network slicing is another enabling technology that is developed from Core Network (CN) to radio link. Radio slices, the physical layer subset of network slicing, are created to fulfill communication requirements and to improve efficiency of radio frequency usage. NOMA and radio slices form the basis in which dynamic spectrum allocation is modeled in this paper.

Radio spectrum is an essential part of the IoT infrastructure. As IoT networks develop towards maturity, diverse bands are perceived to be more attractive to IoT operations. It is expected that the spectrum bands which could be most appropriate for the IoT services would have a wide range of properties, and thus frequencies which serve different types of IoT use cases are increasingly associated with IoT services. Traditionally, most of the existing IoT services rely on unlicensed spectrum to facilitate wireless links between the IoT devices and their associated access point. Unlicensed spectrum will continue to be an important enabler for IoT technology due to its low cost and maturity. However, as the technology evolve to demand more spectrum and the unlicensed spectrum becomes overloaded, an alternative to the unlicensed spectrum is required. A solution which manifest itself as a core part of the future IoT radio architecture is the well-studied dynamic spectrum sharing [6].

In this paper the attention is focused on the 5G candidate frequencies to support IoT spectrum demands through dynamic spectrum sharing. For instance, according to 3GPP, Release 15, FR1 (450-6000 MHz) and FR2 (24250-52600 MHz) are considered in 5G cellular networks [7]. In respect to NOMA, the lower end of the Ultra high frequency (UHF) radio are difficult to control, since the signal in such bands propagate much further and filtering of unwanted RF signals at the receiver end becomes more complex and challenging, as the IoT devices filter out signals with narrow margin of power levels. This in turn increases error probability, undesirable in NOMA. Upper end of the UHF and the millimeter wave-otherwise known as the V-Band-frequencies are considered more viable for NOMA. For example, a comparison between mmWave frequencies and UHF radio in the context of NOMA by Naqvi and Hassan [8] provides useful conclusions. However, with technological advances in software defined radio (SDR) and their effectiveness in filtering, UHF may be a candidate frequency for future NOMA technology as discussed in [9].

IoT devices demand both uplink and downlink connectivity, however, it has been noted that IoT traffic is dominantly uplink and therefore in this paper we have focused only on uplink channel requirements [10]. In addition to overlooking the downlink channel requirements, the control channels, required for periodic update traffic, are also not considered, since the data size in such channels are inherently small. However, with some small changes to the model described in this paper, downlink and control channel requirements can also be accommodated. A summary of IoT communication platforms, including 5G, and their features in terms of capability and operating bandwidth, among other characteristics, are presented in Table 1.

In this paper, a framework is proposed to enhance resource allocation and to manage the mobility of IoT devices using NOMA technique. This is achieved through matching the spectrum demand of IoT devices with the available resources from cellular service providers (CSP), modeling a spectrum trading market, optimizing the spectrum utilization and managing the mobility of IoT devices. The proposed framework offers robust allocation, revenue generation and regulation strategies for both CSP and IoT service providers, contributing to QoS enhancement. The rest of the paper is organized as follows: The literature review is elaborated in section II; the system model is described in section III; the dynamic

spectrum allocation framework is presented in section IV; mobility management model is defined in section V; performance analysis of the proposed frameworks have been presented in section VI along with the comparison of existing frameworks to evaluate the performance of the proposed framework followed by conclusion in section VII.

II. RELATED WORK

5G cellular systems, with its promised reliability, scalability, and efficiency in terms of cost and spectrum utilization, is expected to be a key enabler for IoT technology [11]. Therefore, IoT provision in 5G networks is addressed in the literature to enhance wireless connectivity to IoT devices and to meet the heterogeneous requirements of various IoT use cases. For example, a low-cost and low-complexity operation of IoT communications in 5G networks to support massive connectivity of low-rate and low-power devices has been proposed in [12]. Low latency IoT applications and their requirements in the context of 5G networks are discussed in [13]. Palattella *et al.* [14] have characterized the potential of 5G for the IoT, considering both the technological aspects and their implications on business models and strategies. Hsu *et al.* [15] proposed a scheme that is expected to achieve the uplink data rate for critical tasks in cellular based IoT networks.

In order to maximize the spectral efficiency and to support massive machine-type communications (mMTC) within 5G networks, power-domain NOMA has been considered by the research community as a promising approach. Liu *et al.* [16] have investigated the simultaneous wireless information and power transfer using NOMA within 5G. Recently, several research studies have identified the potential benefits of NOMA in both the downlink and uplink to support IoT services within 5G cellular networks. The work presented in [17] focused on subgrouping in respect to downlink channels, by considering point-to-multipoint and broadcasting in NOMA. Other research such as [18] proposed an edge computing aware NOMA technique which leverages uplink NOMA in reducing users' uplink energy consumption. To overcome the challenge of providing connectivity to a large number of IoT devices, Mostafa *et al.* [19] proposed a power-domain uplink NOMA scheme for narrow band IoT systems. A review study, surveyed recent advancements in NOMA for IoT communications and describes its benefits and challenges, can be found in [20].

NOMA serves multiple users simultaneously using the same sub-carrier, at the cost of increased intra-cell interference. To mitigate intra-cell interference, efficient NOMA design (e.g., user clustering and resource allocation) along with successive interference cancellation (SIC) technique has been considered in the literature to manage large number of devices as discussed in [21]. Clustering technique-grouping users in clusters-provides a way to allocate resources efficiently with minimum intra-cell interference. A NOMA based clustering scheme is proposed by Ali *et al.* [22] and considers the channel gain difference among users to form clusters and optimize their respective power allocations to increase throughput. A NOMA based optimization framework is proposed by Kiani *et al.* [18] that minimizes the energy consumption of the users by optimizing the user clustering and transmit power. An interference aware NOMA framework is presented in [23] that is expected to perform for both intra-cell and inter-cell interference. Tabassum *et al.* [24] proposed a multi-cluster uplink NOMA system and analyze its performance considering SIC, where users are arranged based on the distance of their serving base station.

Several dynamic resource allocation schemes have been presented in the literature [25], [26]. Bandwidth allocation challenges for IoT devices by considering spectrum sharing is discussed in [27]. Spectrum leasing scheme, aimed at providing licensed spectrum to new emerging technologies including IoT applications is investigated in [28]. The authors modeled a monopoly market where femto-cell holders bid for spectrum, owned by mobile network operators, to increase utility.

One challenge faced by IoT service providers (IoT-SPs), when acquiring radio resources is the determination of number of resource blocks required and the efficiency of outcomes. In this context, an auction market approach to maintain the quality of service (QoS) of IoT mobile devices by purchasing bandwidth from the service provider is proposed by [29]. Based on service delay constrain, the aggregated bandwidth requirement is calculated. The authors consider an orthogonal multiple access (OMA) modulation scheme to obtain the spectral efficiency. However, the problem of calculating the bandwidth requirements to support guaranteed bit rate (GBR) in NOMA, has not been addressed.

Predictive techniques to estimate the IoT bandwidth requirements carry a degree of uncertainty between expectations and real-world experiences. The focus of this paper revolves around NOMA enabled devices, with asynchronous data rate requests, served by IoT-SPs. Therefore, this paper quantifies the spectrum required and provide a solution for on-demand based service, which runs in small time windows. Furthermore, it characterizes the QoS demands of the entire IoT devices in a region. Under IoT-SPs strict demands of throughput and transmit power, the focus is on mechanisms to cluster IoT devices based on their location in reference to the base stations of the cellular service providers (CSPs). The main contributions of this paper can be summarized as follows:

1) An algorithm to associate IoT devices with 5G base stations (BSs), and to compute the required number of sub-carrier, providing GBR, is developed with an emphasis on clustering problems. The algorithm involves finding an appropriate combinations of BSs under transmit power and distance restrictions to maximize the average data rate of IoT devices.
2) A second price auction (SPA) algorithm, which matches the requirement of IoT-SPs with CSPs' spectrum availability in a form of radio slices, is provided

to model the spectrum trading market and to optimize the spectrum utilization.

3) Finally, an algorithm is presented to address IoT device mobility, offering a solution to minimize complexity of re-arranging NOMA clusters. More specifically, once spectrum allocation had taken place, the CSP rearranges clusters in response to changes in the cell such as new devices entering the cell, yet without causing any service interruption or quality of service degradation.

III. SYSTEM MODEL

A. NETWORK DESCRIPTION

In this section we describe our model in detail. We consider a scenario consisting of a number of fixed BSs and fixed IoT devices. The horizontal location of the kth IoT device is denoted as $\mathbf{w}_k \in \mathbb{R}^{2\times 1}$, $k \in \{1, 2, \ldots, K\}$, where K is the total number of IoT devices and $\mathbb{R}$ is a set of real numbers. We also consider a set of $\mathcal{B}$ base stations (BSs) arbitrarily located in a given region $\subseteq \mathbb{R}^2$. The coverage area of BS b is denoted as R_b. The locations of IoT devices are modeled as homogeneous Poisson point processes (PPPs) with densities λ. The intensity is sufficiently large to capture the notion of large number of IoT devices predicted in 5G networks. Users generate independent requests of $u_{k,t}$ data rate at time window t when it has pending data to be sent. The data rate requests by the IoT devices are transmitted wirelessly over reliable local communication channels to an IoT-hub (e.g. by sending a SERVICE REQUEST message to the IoT-hub for link establishment) to initiate access. The IoT-hub is used to simplify the interaction between the IoT devices and CSPs. This approach allows overhead communication to be established between IoT-devices and IoT-hub over other unlicensed communication channels using platforms such as LoRa and Lightweight. IoT-hub imposes a set of rules to select between one or more base stations (under strongest cell association). The rules are also based on cost associated with providing IoT devices with sub-carriers. The rules can be periodically updated on the IoT-hub to reflect changes in established agreements between the CSPs. To generalize the system model, the density of IoT devices is considered to be variable over the BSs coverage area and some cells overlap each other by varying λ for each BS, capturing real world scenarios. Furthermore, assume the time is slotted into discrete time slots, during which, the location of IoTs devices are fixed. The analysis provided in this paper can also be extended to a more complex model, taking into account the mobility of IoT devices as will be shown in Section V. An overview of the network model is illustrated in Figure 1.

To set the stage for the system model, we denote with $\mathcal{N} = \{1, 2, \ldots, N\}$ as the set of IoT-hubs in a macro-cell. IoT-hubs do not own spectrum bandwidth by default, which implies that in order for the IoT devices to transmit data to their respective servers, they have to rely on CSPs. Now, let $\mathcal{M}_n$ be the set of IoTs devices which belong to the nth IoT-SP, then we have $\bigcap_{n\in\mathcal{N}} \mathcal{M}_n = \mathcal{M}$. Let $\mathcal{S} = \{1, 2, \ldots, S\}$ be the set of CSPs in the region and $\mathcal{S}_m$ denote the set of CSPs

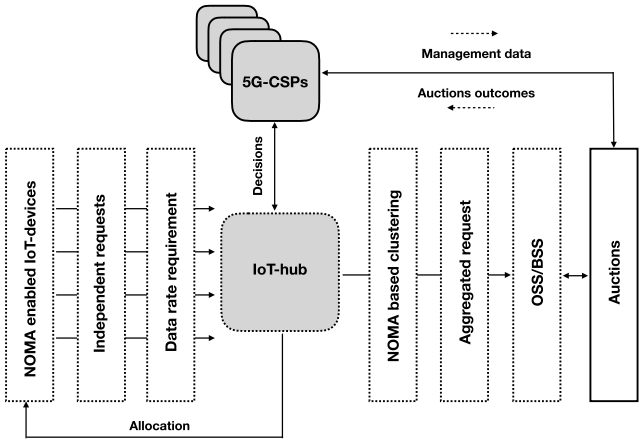


FIGURE 1. Block diagram summarizing the methodological steps of the proposed framework.

who can cover the mth IoTs device, where $m \in \mathcal{M}$, which means that CSP $s \in \mathcal{S}_m$ can provide connectivity to the mth IoT device. The number of CSP which can cover the mth IoT device is represented by

$$|\mathcal{S}_m| = \{s_m : |\mathcal{S}_m| \leq S \geq 0\}, \tag{1}$$

where S is the total number of CSPs and $|\mathcal{S}_m|$ denotes the cardinality of $\mathcal{S}_m$. $|\mathcal{S}_m| = 0$ describes the case where the IoT device k is outside of the coverage area of all CSPs. In this case this IoT device is excluded from further consideration by the IoT-hub. The matrix that defines each IoT-SP devices and their coverage providers is written as

$$\mathcal{M}_n = \{v_{m,s}\}_{M_n\times S}, \text{ where } v_{m,s} \in \{0, 1\}, \quad \forall\, n = 1, 2, \ldots, N. \tag{2}$$

B. FORMULATION OF RESOURCE ALLOCATION PROBLEM

We posit that the BSs are equipped with successive interference cancellers (SIC), which exploit interfering signals structure to mitigate interference. We denote the frequency band of the CSP as W_s Hertz which are divided into N_f orthogonal channels and the bandwidth of each resource block is B. Sub-carrier availability for IoT-SPs in a single time slot is described as $\mathbf{F} = \{f_{i,j}\}_{i\in S, j\in F_s}$ where

$$f_{i,j} = \begin{cases} 1, & \text{if channel is available} \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

The subset which contains the available channels for lease is defined as $\overline{\mathbf{F}} = \{f_{i,j} \in \mathbf{F} | f_{i,j} \neq 0\}$. Next, we derive the formulae to determine the number of sub-carriers required by the IoT-SPs, using NOMA technique. It is assumed that each sub-carrier can be allocated to J_i IoT devices, where i represents the index for sub-carriers such as $i = \{1, 2, \ldots, N_f\}$. The signal received by the sth CSP from ith sub-carrier can be represented as y_s^i and defined as:

$$y_s^i = \sqrt{p_1^i \varrho_1}\ g_1^i x_1^i + \sqrt{p_2^i \varrho_2}\ g_2^i x_2^i + \ldots + \sqrt{p_{J_i}^i \varrho_{J_i}}\ g_{J_i}^i x_{J_i}^i + z^i, \tag{4}$$

$$U_j^i(\mathbf{p}) = \mu_1 \log_2\left(1+\frac{\frac{\varrho_1|g_1^i|}{\sigma^2}p_1^i}{\sum_{k=k+1}^{J}\frac{\varrho_k|g_k^i|}{\sigma^2}p_k^i+\sigma^2}\right)+\mu_2\log_2\left(1+\frac{\frac{\varrho_2|g_2^i|}{\sigma^2}p_2^i}{\sum_{k=k+2}^{J}\frac{\varrho_k|g_k^i|}{\sigma^2}p_k^i+\sigma^2}\right)+\ldots$$
$$+\mu_{J-1}\log_2\left(1+\frac{\frac{\varrho_{J-1}|g_{J-1}^i|}{\sigma^2}p_{J-1}^i}{\sum_{k=J-1}^{J}\frac{\varrho_k|g_k^i|}{\sigma^2}p_k^i+\sigma^2}\right)+\mu_J\log_2\left(1+\frac{\frac{\varrho_J|g_J^i|}{\sigma^2}p_J^i}{\sigma^2}\right),\quad \forall j=1,2,\ldots,J. \tag{5}$$

$$U_j^i(\mathbf{p}) = \mu_1 \log_2\left(1+\frac{G_1^i p_1^i}{\sum_{k=k+1}^{J}G_k^i p_k^i+\sigma^2}\right)+\mu_2\log_2\left(1+\frac{G_2^i p_2^i}{\sum_{k=k+2}^{J}G_k^i p_k^i+\sigma^2}\right)+\ldots$$
$$+\mu_{J-1}\log_2\left(1+\frac{G_{J-1}^i p_{J-1}^i}{\sum_{k=J-1}^{J}G_k^i p_k^i+\sigma^2}\right)+\mu_J\log_2\left(1+\frac{G_J^i p_J^i}{\sigma^2}\right),\quad \forall j=1,2,\ldots,J. \tag{6}$$

where the variable x_j^i denotes the symbols transmitted from jth IoT device to sth CSP, $j=\{1,2,\ldots,J\}$. p_j^i is the transmit power of the signal transmitted by jth IoT device on ith sub-carrier. ϱ_j represents the joint effect of path loss and shadowing between the jth IoT device and sth CSP. g_j^i is the small scale fading coefficients for the link between jth IoT device and sth CSP. z^i is the additive white Gaussian noise (AWGN) for the ith sub-carrier. Consider $\mathbf{Y}$ represents the symbols received from all N_f sub-carriers, which is defined as $\mathbf{Y}=\{\mathbf{y}^1,\mathbf{y}^2,\ldots,\mathbf{y}^{N_f}\}$. NOMA is a technique used to realize multiple access by utilizing power domain in a spectrally efficient way and can serve multiple IoT devices in the same sub-carrier. This can be achieved by allocating different power levels to different IoT devices [30]–[33].

Consider that G_j denotes the channel coefficient vector of jth IoT device at sub-carrier i, which includes distance dependent loss, shadowing loss, and instantaneous fading coefficients and assume $|G_1| \le |G_2| \le \ldots \le |G_{J_i}|$ for J_i IoT devices which are expected to transmit data on the ith sub-carrier [34]. The CSP superimposes the IoT devices data by allocating the corresponding power levels, p_j, where it represents the power level for jth IoT device. Moreover, the CSP is expected to allocate more power to the IoT-devices which are experiencing poor channel conditions i.e., $p_1 \ge p_2 \ge \ldots \ge p_{J_i}$ and $p_1^2+p_2^2+\ldots+p_{J_i}^2=1$ if $|G_1| \le |G_2| \le \ldots \le |G_{J_i}|$. The optimal order for decoding is in the order of the increasing channel gain, normalized by the noise and inter-cell interference power. Based on this order, the CSP decodes the signals from any of the J_i IoT devices. The throughput that is expected to be achieved on sub-carrier i denoted as $U_j^i(\mathbf{p})$ and expressed in equation (5), as shown at the top of this page. Where $\mathbf{p} \in \mathbb{R}^{(N_f\cdot J)\times 1}$ represents the transmission power. μ_J is a non-negative constant that represents the priorities of the Jth IoT device in resource allocation i.e. $0 \le \mu_J \le 1$. Consider for jth IoT device $G_j^i=\frac{\varrho_j|g_j^i|}{\sigma^2}$, then the equation in (5) can be written as equation (6), as shown at the top of this page.

The total achievable throughput from N_f sub-carriers can be denoted as $\mathbf{U}$ and defined as

$$\mathbf{U}=\sum_{i=1}^{N_f}\sum_{j=1}^{J_i}U_j^i(\mathbf{p}). \tag{7}$$

Let us assume the required throughput by the nth IoT-SP is U_{req}^n. Based on equation (7), each IoT-SP computes the required number of sub-carriers to achieve a desired QoS in a form of throughput using Algorithm 1. The steps taken by the IoT-SPs to find the combination of CSPs that can accommodate all IoT devices with minimum power or cost are described as follow.

Firstly, the IoT-hub quantifies all the CSPs which could provide access to one or more IoT devices to form the vector $\mathcal{S}' \subseteq \mathcal{S}$ subject to constraints of the form

$$(\mathrm{C}_1): d_{ms} \le d_n',\ d_{ms} \in \mathbb{R}_+ \text{ and} \tag{8}$$
$$(\mathrm{C}_2): p_{ms} \le p_s',\ \forall m=\{1,2,\ldots,M\}, \tag{9}$$

where d_{ms} represents the distance between the mth IoT device and the sth CSP and d_n' is the distance threshold set by the nth IoT-SP. p_{ms}' is the sth maximum allowed power for mth IoT device, imposed by the CSPs and p_n' is the IoT-SP maximum supported power, this is important in 5G networks, supporting a wide range of radio frequencies, operating under various power restrictions to mitigate interference. Next, we find the combinations of CSPs which can accommodate all the IoT devices in the coverage area, enabling the IoT-SPs to identify the CSPs which will be used to provide connectivity. Let $h_{S\times M_n} \in \{0,1\}$, where the value 1 represents the coverage provided by the sth CSP to the $\hat{m} \in \mathcal{M}_n$ IoT device and 0 otherwise. Mathematically the set is given by

$$\begin{cases} l, & \text{if } \prod_{\hat{m}=1}^{M_n} h_{l,\hat{m}} \neq 0 \\ \{l,k\}, \text{where we let } l \neq \hat{m}, & \text{if } \prod_{\hat{m}=1}^{M_n}\sum_{l,k} h \neq 0,\ \forall l,k \in \mathcal{S}' \\ \quad\vdots & \quad\vdots \\ \{l,k,\ldots,S'\} & \text{if } \prod_{\hat{m}=1}^{M_n}\sum_{l,k,\ldots,S'} h \neq 0,\ \forall l,k,\ldots,S' \in \mathcal{S}' \\ \phi, & \text{otherwise,} \end{cases} \tag{10}$$

where $\{l,k,\ldots,S'\} \subseteq \mathcal{S}'$ represents the CSPs that can provide coverage to IoT devices of nth IoT-SP, and ϕ is an

empty set. To derive the optimal allocation strategy, we formulate an optimization problem, which minimizes the cost of radio resources, expressed as

$$(\mathcal{P}): \quad \min \sum_{k=1}^{C} \mathcal{S}'(k) \tag{11}$$

subject to

$$(\mathrm{C}_3): d_{\hat{m}s} = \min\{\mathbf{d}_{\hat{m}s}\}, \ \forall \hat{m} = \{1, 2, \ldots, M_n\}, \quad s = \{1, 2, \ldots, S\} \tag{12}$$

$$(\mathrm{C}_4): p_{\hat{m},n} \leq p'_n, \forall n = \{1, 2, \ldots, N\}. \tag{13}$$

The constraint (C_4) in ($\mathcal{P}$) enforces the IoT-hub to consider spectrum resources with allowable transmit power, as set by the CSP.

Algorithm 1 illustrates the steps to compute the required number of sub-carriers for one IoT-SP in each iteration. It calculates the achievable throughput U^n_{ach}, using Equation (7), considering IoT devices with data to transmit, and in contrast with the required throughput. If the required throughput is not achievable with the available resources, additional sub-carriers are added to fulfill the throughput demand. This is to eliminate packet loss due to interference and fading, requiring retransmission of data and additional time and power which is impractical in real-world deployments. The algorithm starts with measuring the distance between each IoT device and the BSs, updating the vector $\mathcal{M}_n$ (Algorithm 1: Line 1–5). Clustering of IoT devices based on distance, taking into account the maximum number of devices in a cluster, V, is performed to obtain $\mathcal{C}$, using k-means clustering (Algorithm 1: Line 6–12). Based on the number of devices and required QoS, we compute the required number of sub-carrier for each cluster, given in $\mathcal{S}(k)$ (Algorithm 1: Line 14–24). The total number of required sub-carrier is then given by $\mathcal{D}_s = \sum_{k=1}^{C} \mathcal{S}(k)$, where C is the total number of clusters (Algorithm 1: Output).

Algorithm 1: Pseudocode for NOMA Based IoT Device Clustering and Calculation of the Required Number of sub-Carriers

Input: N : The number of IoT devices,
$\bar{\mathbf{F}}$: available channels,
U^n_{req} : total required throughput,
U^n_{ach} : achieved throughput,
$\hat{\mathcal{R}}_n \in \{\phi\}$.
$\mathcal{M}_n \in \{1, 2, \ldots, M_n\}$,
$\mathcal{R}_n \in \{r_1, r_2, \ldots, r_n\}$.
Output: $\mathcal{D}_s = \sum_{k=1}^{C} \mathcal{S}(k)$ %
total number of required sub-carrier

1 **for** all $i \leftarrow 1 : M_n$ **do**
2 $r_i = \min\{\mathcal{R}_n\}$
3 $\hat{\mathcal{R}}_n \leftarrow r_i$
4 Based on $\hat{\mathcal{R}}_n$, sort all the IoT devices in $\mathcal{M}_n$
5 **return** $\mathcal{M}_n$
6 **for** all $j \leftarrow 1 : V$ **do**
7 $l = j \times C$
8 **if** $l \leq M^n$ **then**
9 $\mathcal{C}_{ji} = \{(C \times (j-1)) + 1, (C \times (j-1)) + 2, \ldots, C \times v\}$ assigning IoT devices to the clusters $\{1, 2, \ldots, C\}$.
10 **else**
11 $\mathcal{C}_{ji} = \{(C \times (j-1)) + 1, (C \times (j-1)) + 2, \ldots, M^n\}$ assigning IoT devices to the clusters $\{1, 2, \ldots, M^n - (C \times (j-1))\}$.
12 **return** $\mathcal{C}$
13 $\mathcal{U}_{req} = \{U_{req}(1), U_{req}(2), \ldots, U_{req}(C)\}$
14 **for** all $k \leftarrow 1 : C$ **do**
15 remaining IoT devices = $\mathcal{C}(k)$
16 **while** *remaining IoT devices* > 0 **do**
17 **while** $\mathcal{U}_{ach}(k) \leq \mathcal{U}_{req}(k)$ **do**
18 calculate $\mathcal{U}_{ach}(k)$ using equation (7) where $N \leq \lambda_N$ and $U \leq \lambda u$.
19 remaining IoT devices = Remaining IoT Devices $-$ 1
20 $\mathcal{U}_{req}(k) = \mathcal{U}_{req}(k) - \mathcal{U}_{ach}(k)$
21 required number of sub-carrier = required number of sub-carrier + 1
22 **if** *remaining IoT devices* > 0 **then**
23 remaining IoT devices = remaining IoT devices + 1
24 **return** $\mathcal{S}(k)$ = required number of sub-carrier for cluster k

IV. DYNAMIC SPECTRUM ALLOCATION

A. RADIO SLICES

Network slicing is considered to be key enabler for 5G's heterogeneous requirements. Network slicing covers all network segments, including core network and the radio access networks (RAN). The core network elements are sliced using network function virtualization (NFV) and software defined networks (SDN) whereas in the RAN, slicing is built on the physical radio resources (e.g., spectrum). In this paper the focus is on the radio segment of the network slices-radio resource slicing-where radio resources are represented in a set of slices. We assume that each CSP broadcast the number of available radio slices, N_{rs}, making them visible to the IoT-SPs. The radio slices are combined by the scheduler according to $\mathbf{C}_{ns}$. The combinations are followed by a linear program to map the optimal combination with the resource requirement, $\mathcal{D}$ (Algorithm 2: Line 2–15).

B. CHARGING SCHEME

Algorithm 1 computes the total number of required sub-carriers, $\mathcal{D}_s$, to meet throughput demand of the IoT devices.

TABLE 2. Key symbols and definitions.

Symbols	Definitions
N	Number of IoT service providers (SP)
M	Number of IoT devices
M_n	Number of IoTs devices belong to the nth IoT-SP
S	The total number of cellular service provides (CSP)
$\mathcal{S}'$	Set of CSPs which can support one or more IoT devices
$\mathcal{S}_m$	Set of CSPs that can cover the mth IoT device
N_f	Number of orthogonal channels
ϱ	Path loss and shadowing
g	Small scale fading coefficient
G	Channel coefficient vector
$\mathbf{F}$	Sub-carrier available for IoT-SPs in a single time slot
$\overline{\mathbf{F}}$	Sub-carrier available for each IoT-SP in a single time slot
$\mathbf{b}$	The account balance of each participant

Using dynamic spectrum sharing, the IoT-SPs attempt to obtain resources from the CSPs based on overlay spectrum sharing technique. The problem of dynamic spectrum sharing (DSS) is addressed by researchers using auctions. The importance of auctions in DSS has generated many methodological papers on auction models and mechanisms as discussed in Section II. Many standard auction models and formats provide reasonable structure to solve the decision making problem of the service providers. In this paper we use the classical second-price auction (SPA) where the winner of the auction, usually pays a price linked to the second-highest bid for the object on lease. As such, the highest bidder wins the auction, but the price is determined by a special hybrid pricing rule, where the winner pays the smaller of either their own bid or the second-highest bid. In the following section we provide rules of entering auctions

C. ENTERING AN AUCTION PROBLEM

CSPs announce information related to auction opening and details on the available resources for lease. The minimum asking price P_m which is set so that bids must exceed the value of P_m, is announced at the beginning of the $\hat{j}$th trade window where $\hat{j} \in \{1, 2, \ldots, \hat{J}\}$ and $\hat{J}$ denotes the total number of trading windows. Based on distributions of IoT-SP valuations, the total expected revenue of the CSP is maximized by a SPA with minimum asking price. The prior distributions of IoT-SP valuations come from past empirical data. The discussion on how much data is necessary and sufficient to guarantee near-optimal expected revenue to the CSPs or IoT-SPs is beyond the scope of this paper.

Next, we present a decision rule used by the IoT-SPs to establish whether entering the auction is required. Suppose $\xi_{\hat{i},\hat{j}}$ denotes the state of a $\hat{i}$th IoT-SP at the beginning of the $\hat{j}$th trading window; $\xi_{\hat{i},\hat{j}} = 0$ or 1 refers to the IoT-SP with data to transfer or not, respectively. Let $\alpha_{\hat{i},\hat{j}} = 0$ refers to the case where the required resources by the $\hat{i}$th IoT-SP (which is computed using Algorithm 1) is smaller than the resources announced by the CSPs at the $\hat{j}$th trade window and $\alpha_{\hat{i},\hat{j}} = 1$ otherwise. Suppose $\varpi_{\hat{i},\hat{j}}$ denote the decision to enter an auction, i.e. $\varpi_{\hat{i},\hat{j}} = 1$ means the ith IoT-SP decides to enter an auction and $\varpi_{\hat{i},\hat{j}} = 0$ represents the decision not entering an auction. The binary rule can be written as

$$\varpi_{\hat{i},\hat{j}} = \begin{cases} 1, & \text{if } \left[\left(1 - \prod_{\hat{j}=1}^{\hat{j}} \xi_{\hat{i},\hat{j}}\right)\left(1 - \prod_{\hat{j}=1}^{\hat{j}} \alpha_{\hat{i},\hat{j}}\right) = 1\right] \\ & \text{and } P(i) > P_m \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

From the above formulation, it can be found that if the IoT-SP valuation of the spectrum resources, $P(i)$, of the sub-carriers being auctioned is less than the minimum asking price, P_m, then not entering the bidding is the optimum decision.

D. RULES OF ENTERING AN AUCTION

An IoT-SP may enter the bidding round once the auction has started within the time duration t_x. Once the time is elapsed, new IoT-SP are not allowed to place bids. Furthermore, the owner of the sub-carriers, the incumbent CSP, my decide to limit the number of bidders to minimize the time spent on deciding the winner of the auction and to avoid crashing or entering infinite loops. This is achieved by setting a threshold value, which represent the maximum number of bidders allowed in one auction. To guarantee efficient usage of spectrum, a minimum number of participant in an auction is not considered.

IoT-SPs who want to place a bid, submit their requests to the auction handler, where requests are registered and all the available RBs from one or more CSPs are found. If a request has not been received earlier in the round, before the time t_x is up, then the request is added giving that the maximum number of bids is not exceeded. IoT-SPs violating the rules of bidding are charged and blocked from subsequent auctions.

The auction in Algorithm 2 (Line 17–31) where the highest bidder wins and pays the second-highest bid incentivize IoT-SPs to place their bids based on their true evaluation of the resources traded in the auction. The case where two or more IoT-SPs submit equal winning bids is resolved by a random selection from the set of winners. Here, the winner pays the full value of the bid, P_b, since in the event of a tie the first-place and second-place bids are equal. Algorithm 2 determines an optimal real-time allocation and pricing of sub-carriers to the winning IoT-SP. SPA is a standard auction, however, the novelty of the algorithm is that the bids of each IoT-SP are based on their exact required number of sub-carriers, which are computed using Algorithm 1.

Once an auction is complete, the winner is allowed to use the auctioned sub-carriers for the specified time and within the area boundaries. This approach simplifies the mechanisms within the spectrum market, allowing supply and demand between CSP and IoT-SPs and facilitating transactions.

Algorithm 2: Matching the Requirement of RBs/Network Slices, Which are Available From the CSPs and Defining Second Price Auction

Input: N' : number of bidders,;
P_m : minimum asking price, ;
N_{rs} : number of radio slices, ;
$\mathbf{b} \in \{b_1, b_2, \ldots, b_{N'}\}$:
account balance for each bidder,;
t_x : time duration of the auction.
Define: P_s : second-best price, P_b : best price,;
$P(i)$: price paid by ith bidder.

for $k = 1 \leftarrow N_{rs}$ **do**
  $\mathbf{C}_{ns} = \frac{\mathbf{n}!}{k!(\mathbf{n}-k)!}$, $\mathbf{n} = \{1, 2, \ldots, N_{rs}\}$;
  define $\mathbf{C_x}$ from $\mathbf{s}$ with respect to the combinations in $\mathbf{C}_{ns}$;
  $\mathbf{c}_i = \sum_{j}^{J} \mathbf{C}_x^{ij}$, $i = 1, 2, \ldots,$ number of rows of $\mathbf{C}_{ns}$;
  $\mathbf{v}_i = \mathbf{c}_i - \mathcal{D}_s$;
  if $\sum \mathbf{v}_i \geq 0$ *(% all values in* $\mathbf{v}_i$ *are +ive)* **then**
    $\hat{\mathcal{D}}_s = \min\{\mathbf{v}_i\} + \mathcal{D}_s$;
    else if $\sum \mathbf{v}_i < 0$ *(% all values in* $\mathbf{v}_i$ *are -ive)* **then**
      $\hat{\mathcal{D}}_s = \max\{\mathbf{v}_i\} + \mathcal{D}_s$;
    else if *(%*$\mathbf{v}_i$ *consists of +ive and -ive values)* **then**
      for *l=1:J* **do**
        if $\mathbf{v}_i(l) > 0$ **then**
          $\hat{\mathbf{v}}_i(l) = \mathbf{v}_i(l)$
        $\hat{\mathcal{D}}_s = \min\{\hat{\mathbf{v}}_i\} + \mathcal{D}_s$;
  $\mathcal{D} = \mathcal{D}_s + \hat{\mathcal{D}}_s$
next we model the auction mechanism according to the second best price auction with incentives and privacy ;
$b_i = b_i - \mathbf{\Gamma}$;
while $t_x \neq 0$ **do**
  for $i = 1 \leftarrow N'$ **do**
    $P_b = P_m$;
    if $b_i \geq b_m$ *(% check account balance)* **then**
      if $(P(i) > P_b)$ and $(P(i) \geq P_m)$ **then**
        $P_s = P_b$;
        $P_b = P(i)$;
        winner $= i$
      else if $P(i) > P_s$ **then**
        $P_s = P(i)$
  if $\omega = 1$ *(% check privacy)* **then**
    $b_i = b_i - (P_s + P_p)$;
    else
      $b_i = b_i - P_s$
calculate $\mathbf{\Gamma}$ using equation (15) and (16) ;
return;

E. ACCOUNT BALANCE AND MODELING OF INCENTIVES

To limit IoT-SPs from breaching the rules or abandoning the auction, the CSPs hold a monetary account of each IoT-SP participant. We denote the account balance of each IoT-SP as $\mathbf{b} = \{b_1, b_2, \ldots, b_N\}$. Also, IoT-SPs whose account value are below a certain threshold (e.g., $b_i < b_m$), their bids are blocked from entering the auction. The IoT-SPs may withdraw money from their respective accounts at anytime but not during an auction they entered. Modeling the account balance in the auction is important, since it captures and removes the IoT-SPs which are considered high risk. This process has advantages over, for example, making a payment on ad-hoc basis using bank transactions where it can be time consuming because every transaction must be approved through a complicated process before IoT devices are able to transmit data over the radio spectrum. Advanced transactions between the IoT-SP to the CSP could be processed off-line to expedite payments between them, and provide cover against aborting IoT-SPs. In addition, the system offers the advantage of requiring a single transaction for one sum, used for multiple purchases, depending on the size of the account balance. Similar systems exists, such as prepaid cards.

Moreover, we present an incentive scheme for rewarding (positive incentive) or penalizing (negative incentive) IoT-SPs for cooperating with CSPs (such as giving access to available leased spectrum in peak hours) or violating operational rules (such as out-of-band spectrum and interference violation [35]) respectively. We assume that the CSP defines a set of rules, prior to auction commencement, rules can be defined as $\{z_{\tilde{i}} | \tilde{i} = 1, 2, \ldots, \Upsilon\}$. If the CSP has identified a particular cooperation or violation for a rule $z_{\tilde{i}}$, it rewards or penalizes based on the magnitude of the perceived cooperation or violation. Cooperation or violation of a particular rule by the IoT-SP will result in activation of rewards or penalties $\gamma_{\tilde{i}} \in (-1, 0, 1)$, which are credited to IoT-SPs by the CSP or payable by the IoT-SPs to the CSP at the end of the lease period. $\gamma_{\tilde{i}}$ can be defined by the rule

$$\gamma_{\tilde{i}} = \begin{cases} 1, & \text{when IoT-SP cooperate} \\ -1, & \text{when rules are violated} \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

where $\tilde{j} \in \{1, 2, \ldots, N'\}$ and N' is the number of CSPs participating in auction. The reward of cooperation and cost of violating the rules are defined as the set $\{\zeta_{\tilde{i}}^{+} | \tilde{i} = 1, 2, \ldots, \upsilon^{+}\}$ and $\{\zeta_{\tilde{i}}^{-} | \tilde{i} = 1, 2, \ldots, \upsilon^{-}\}$ respectively. Hence, when the CSP detects a cooperation or violation, it measures the total incentives, $\mathbf{\Gamma}$, and assigns these incentives to the IoT-SP. The total incentives can be calculated as

$$\mathbf{\Gamma} = \sum_{\tilde{i}=1}^{\Upsilon} (\zeta_{\tilde{i}}^{+} \gamma_{\tilde{i}} + \zeta_{\tilde{i}}^{-} \gamma_{\tilde{i}}). \tag{16}$$

As the incentives added is proportional to the number of rules cooperated or violated by the IoT-SP, the total incentives for an IoT-SP will be proportional to number of rules

cooperated or violated. The incentives assigned to a particular IoT-SP are credited or deducted from their respective account as discussed in the previous section. However, if the total negative incentive is higher than the positive incentive plus the balance of the IoT-SP (i.e. an inequality $\mathbf{\Gamma}^- > (b_i + \mathbf{\Gamma}^+)$), the IoT-SP is not allowed to enter subsequent auctions unless the full negative incentives are cleared.

In this formalization the total payment, $(P_s + \mathbf{\Gamma})$, is paid to the CSP, depending on whether a violation has occurred. After each time window, the balance of all involved IoT-SPs are updated in two steps, subtracting incentives $\mathbf{\Gamma}$ from the balance b_i and storing the updated balance in association with the IoT-SP. The new balance can be computed as $[b_i = b_i - \mathbf{\Gamma}]$. b_i is the open inventory of the ith IoT-SP which is the monetary balance of the account that the ith IoT-SP has on hand at the beginning of a trade window or prior to entering an auction.

There is a possibility that CSPs may violate the rules by applying excessive charges that do not reflect the true cost of spectrum utility and/or by favoring particular IoT-SPs over others unfairly. In infrastructure-based networks, this can be halted by trusted monitoring governmental organization or by assigning a third-party to observe the process and to guarantee fair charging, preventing abuse of service provisioning.

F. PRIVACY

Prior to placing bids, the only information available to the IoT-SPs is details of sub-carriers being offered for lease along with the lease time and the maximum number of bidders allowed. The latter can provide an indication to the IoT-SPs on the probability of winning an auction to determine whether to enter an auction. In addition, the proposed scheme achieves a degree of privacy by limiting access to information on the IoT-SPs involved. Although it is pivotal to inform auction participants of the outcome, privacy can still be protected by keeping the participants and the winners anonymous. The IoT-SPs are informed of the auction outcome, but without information linking to the identity of the winner IoT-SP. This can be achieved by assigning a new random identity code to each IoT-SP to keep them from being identified.

While we do not explicitly model communication, sending and receiving control data including bidding, updating account balance and all other related communications between the IoT-SPs and CSPs can be made using wired communications via internet, which can provide higher stability and security to the system. Even though, the proposed dynamic scheme requires careful implementation, especially for large-scale networks, supporting massive machine-type communication (mMTC), it remains practical on the grounds of computational and cost efficiency.

V. ALLOCATION AND MOBILITY MANAGEMENT

The CSPs allocate channels to IoT devices and assign a QoS Flow ID (QFI) to each IoT-device. The QFI and session establishment remains fixed throughout the lifetime of the trade window, assuming connection stays active and without failure or detachment from the BS. Thereafter, for every event in the cell, such as when IoT devices join or leave a cell, a set of procedures should be executed. Events in the network may or may not induce the need for deploying additional resources. Resources are only required when the number of active users in the cell increases or decrease. When the number of users entering the cell equals the number of users leaving at a given time, then the existing resources will only be redistributed according to the new distribution of the IoT devices in the cell. And in the case where the modulation scheme NOMA is considered, the BS rearranges the clusters to ensure devices are supported with the resources sufficient to provide their required data rate. For some events, however, it is not necessary that all clusters need to be rearranged to accommodate the changes in the cell. For example, if an IoT device enters the cell in a location which can be added to a cluster, without significantly degrading the data rate to the existing IoT devices in the same cluster, then adjustments are only required in that cluster, leaving other clusters unchanged.

Due to the large number of cells in 5G networks in contrast with earlier cellular generations, to improve mobility and handover, secondary cells (SCells) can be configured to form together with the primary cell (PCell) a set of serving cells. The configured set of serving cells for an IoT device therefore should always consist of one PCell and one or more SCells [7]. As such, when an IoT device enters the cell from a neighboring one in which the IoT device is pre-configured to a new cell as its SCell, the IoT device should have been considered in previous allocation and does not have an impact on the clustering.

Events such as connection failure and users moving from inactive (disconnected) state to active (connected) (or from active/connected to inactive/disconnected state) should also be treated as part of the mobility. Modification procedures to other IoT devices in the associated cluster will also take place when an event of this nature occurs. Such events need to be handled autonomously and with minimum interruption to the network resource allocation. Algorithm 3 is developed to accommodate IoT device mobility challenges in NOMA.

We denote by λ_q the maximum data rate that can be handled by a cluster. And if this threshold is exceeded then the BS could deploy additional channels to that cluster following the procedure in Algorithm 3.

VI. RESULTS AND ANALYSIS

This section demonstrates the performance analysis of the proposed NOMA based resource allocation framework which is expected to facilitate the IoT-SP to calculate the required number of resource blocks that can meet their throughput demand and seamless acquisition of resource blocks from CSPs through proposed second-price auction. Moreover, we investigate the performance of the proposed solution and outline the merits of our framework compared to related work from the literature.

Algorithm 3: Mobility Management

Input:

$x_{[enter]}$: The number of IoT devices entering from neighboring cell.

$x_{[leave]}$: The number of IoT devices leaving the cell.

$x_{[active]}$: The number of IoT devices becoming active.

$x_{[inactive]}$: The number of IoT devices becoming inactive.

t_d: The time duration, t_i: The time instant within t_d, λ_q: The QoS threshold.

Define: $\Im$: measure of QoS

1 **while** $t_i \leq t_d$ **do**

2 $\quad \epsilon \leftarrow \frac{\left(\sum x_{[enter]} + \sum x_{[active]}\right)}{\left(\sum x_{[leave]} + \sum x_{[inactive]}\right)}$

3 $\quad$ **if** $\epsilon > 1$ **then**

4 $\quad\quad$ **if** $\Im \leq \lambda_q$ **then**

5 $\quad\quad\quad$ Compute $\mathcal{S}(k)$ from Algorithm 1, step: 14 to 24

6 $\quad\quad\quad$ **if** $D_s > D$ **then**

7 $\quad\quad\quad\quad$ get RBs using Algorithm 2

8 $\quad\quad\quad$ **else**

9 $\quad\quad\quad\quad$ assign additional required RBs

10 $\quad\quad$ **else**

11 $\quad\quad\quad$ no additional RBs are required

12 $\quad$ **else if** $\epsilon \leq 1$ **then**

13 $\quad\quad$ no additional RBs are required

14 **return**

TABLE 3. Simulation parameters and their values.

Parameter	Value
Max Transmit Power	30 dBm
Bandwidth (B)	1 Hz
Path Loss and Shadowing (ϱ)	4
Small Scale Fading Coefficient (g)	1
σ^2	0.5 dBm/Hz

A. SCENARIO

Consider a cellular network covered by 5 BSs, where each BS belongs to multiple CSPs. The devices which belong to N IoT-SPs seek to transmit data to their designated servers/cloud. However, the devices need to have access to the BSs to enable uplink data transmission. To transmit the data to BSs, the IoT provider pays the CSPs in a form of monetary according to the amount of spectrum required to transmit the data and the time this spectrum is occupied by the IoT devices as discussed in Section IV-E.

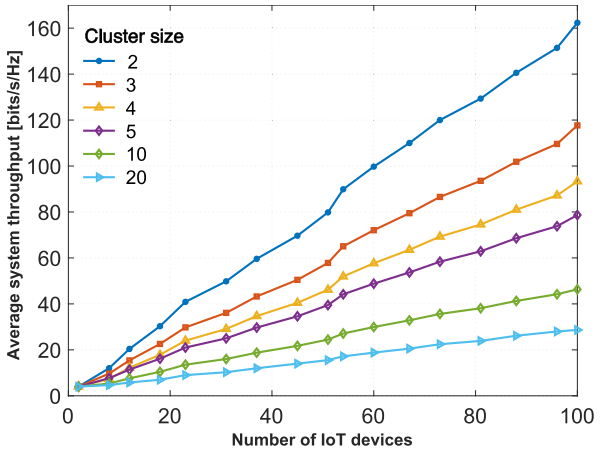


FIGURE 2. Average cell throughput with varied cluster size using NOMA for maximum transmit power = 27 dBm.

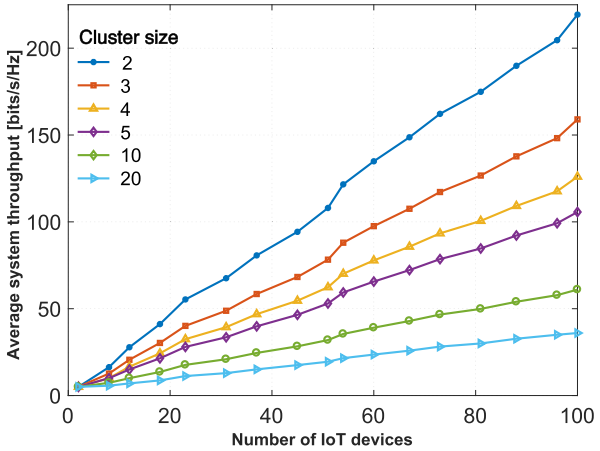


FIGURE 3. Average cell throughput with varied cluster size using NOMA for maximum transmit power = 30 dBm.

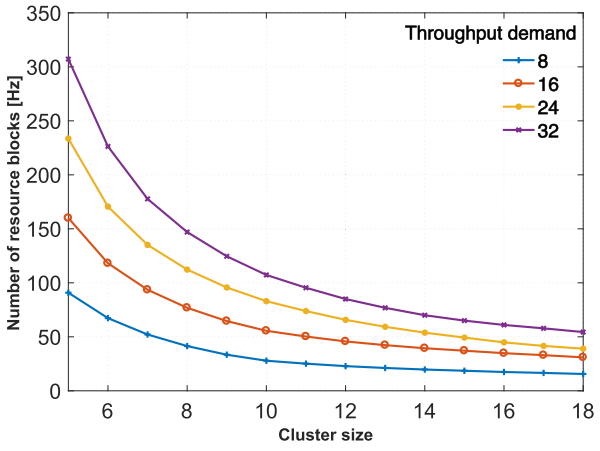


FIGURE 4. Required number of resource blocks with varied cluster sizes to guarantee throughput demand [bits/s/Hz] using NOMA for maximum transmit power = 27 dBm.

B. PERFORMANCE ANALYSIS

To demonstrate the effectiveness of the proposed framework, an IoT network model is simulated. The proposed framework is analyzed in terms of average system throughput, required

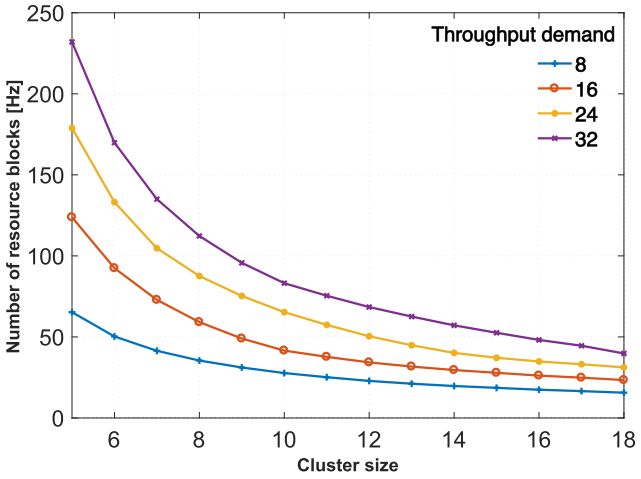


FIGURE 5. Required number of resource blocks with varied cluster sizes to guarantee throughput demand [bits/s/Hz] using NOMA for maximum transmit power = 30 dBm.

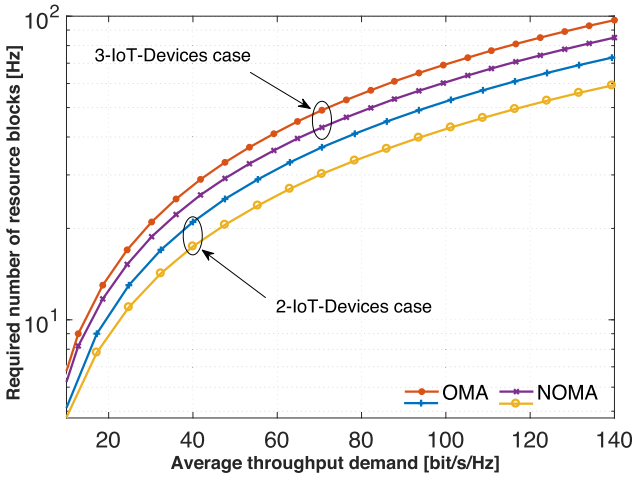


FIGURE 6. The number of resource blocks required for NOMA and OMA to guarantee throughput demand.

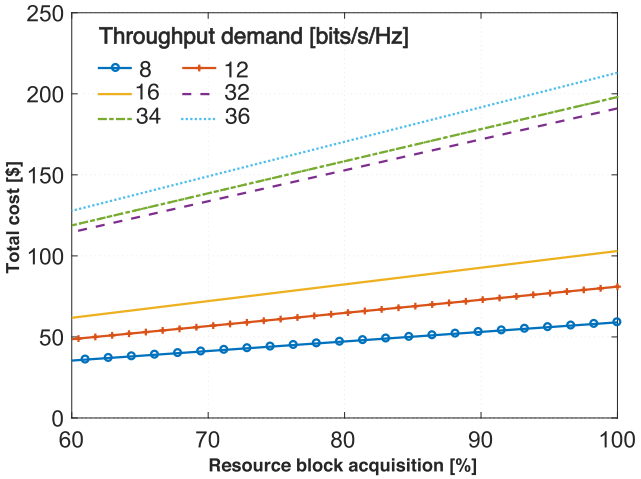


FIGURE 7. Cost estimate to acquire required resource blocks to achieve throughput demand.

number of resource blocks and incurred cost of resource allocations. Note that the average system throughput is drawn from the theoretical maximum achievable throughput of each IoT device, averaged over the entire network.

1) IMPACT OF CLUSTERING ON IOT DEVICES DATA RATE DEMAND AND SUB-CARRIER REQUIREMENTS

The performance of the proposed framework is analyzed with average system throughput and different cluster sizes. The average system throughput is calculated using equation (6), where it is assumed that one resource block is assigned to each cluster. It is observed that reducing cluster sizes using NOMA gains significant average system throughput as shown in Fig. 2 and 3, but it will require additional resource blocks to achieve the targeted throughput demand as shown in Fig. 4 and 5. It can be found from Fig. 2 that providing services to 100 devices with 10 devices in each cluster, the maximum achievable average throughput is 46.33 bits/s/Hz. However reducing the cluster size to 5, 4, 3 and 2 will increase the average achievable throughput by 69%, 100%, 154% and 250% but with additional resource block of 100%, 150%, 230% and 500% respectively. Similarly, it is shown in Fig. 3 that with maximum allowable transmit power of 30dBm and providing services to 100 devices with cluster size 10, the maximum achievable average throughput is 61 bits/s/Hz. Reducing the cluster size to 5, 4, 3 and 2 will increase the average achievable throughput by 73%, 106%, 160% and 259% but with additional resource block of 100%, 150%, 230% and 500% respectively.

Fig. 4 and Fig. 5 presents the effect of different cluster sizes on the required number of resource blocks to guarantee average system throughput. It is observed that increasing cluster size reduces the number of sub-carrier requirements, however, power management at such scale is difficult and could result in higher bit error rate (BER). It is found that to serve high number of IoT devices, it is not always cost effective to have smaller cluster size. As shown in Fig. 4, to achieve 8 bits/s/Hz of average throughput with cluster size 5, 10, 12 and 14 will require 91, 28, 22 and 20 resource blocks. Similarly it is observed in Fig. 5, to achieve 8 bits/s/Hz of average throughput with cluster size 5, 10, 12 and 14 will require 66, 26, 23 and 20 resource blocks. However, as discussed earlier, higher number of devices in one cluster will result in higher BER and complex power management which will result in poor QoS. Hence, selection of cluster size is important in NOMA to maintain QoS and cost, this will be addressed in the future work.

2) NOMA VS OMA

A comparison between the efficiency of NOMA and OMA in respect to sub-carrier demand is provided in Fig. 6. The required number of resource blocks for OMA are calculated through simulations based on the theoretical equations for average system throughput presented in [36]. It is observed from simulation results that at lower throughput demands the difference between the two techniques is small, however, as throughput demand grows, NOMA requires less

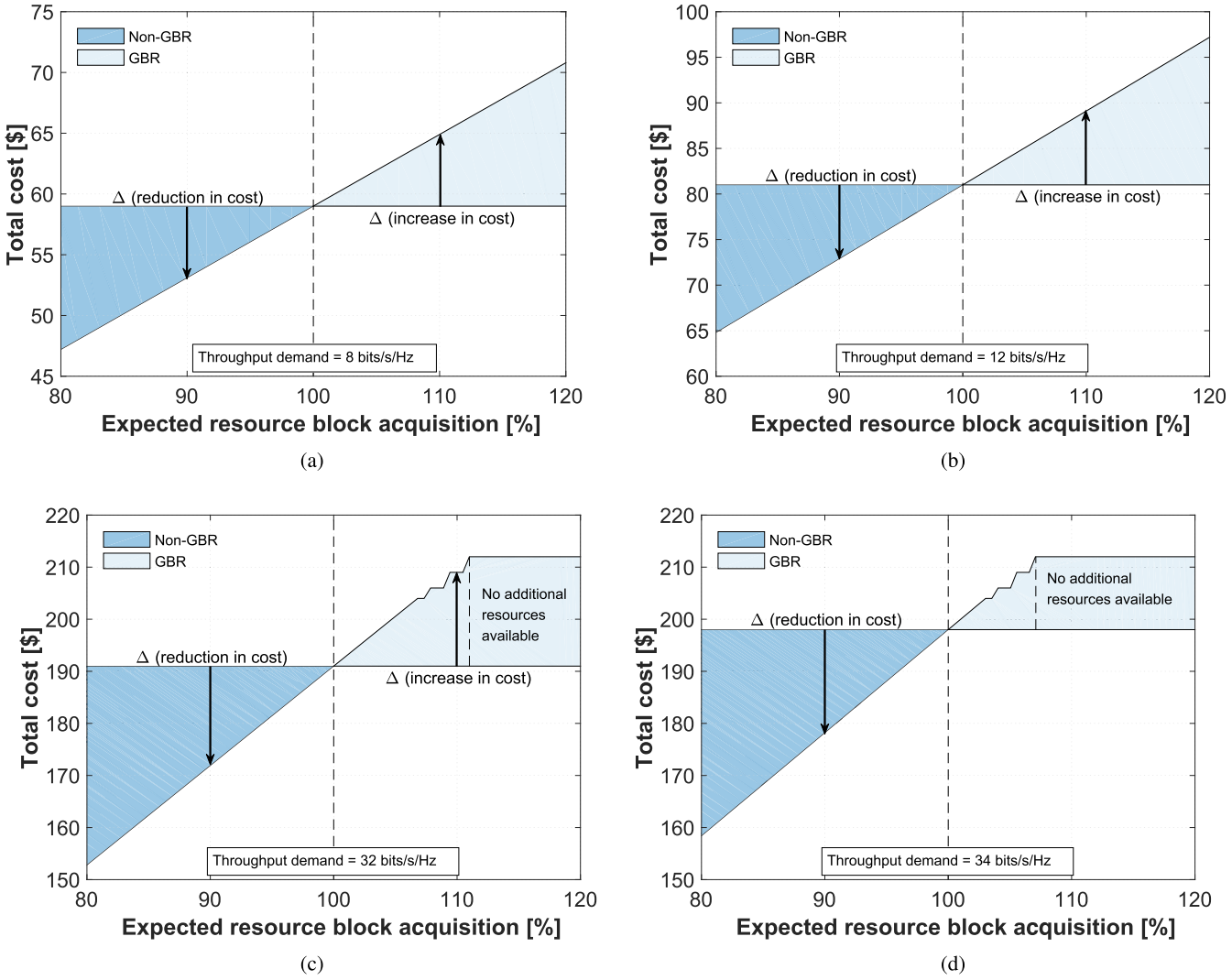


FIGURE 8. Cost estimation for expected resource block acquisition to achieve throughput demand (a) 8 bits/s/Hz, (b) 12 bits/s/Hz, (c) 32 bits/s/Hz and (d) 34 bits/s/Hz with guaranteed bit rate (GBR) and non-guaranteed bit rate (Non-GBR).

sub-carriers to deliver the same QoS. The simulation results also demonstrate that to achieve average system throughput of 20 bits/s/Hz, OMA requires 22% and 27% more resource blocks than NOMA requires for cluster size 2 and 3 respectively. Moreover, for achievable average system throughput of 40 bits/s/Hz, OMA requires 23.5% and 16% additional resource blocks for cluster size 2 and 3 respectively. Therefore, NOMA performs better than OMA technique as it can achieve more system throughput by exploiting the power domain for multi-user multiplexing and utilizing SIC to harness inter-user interference. The comparison between OMA and NOMA in terms of channel requirement represent the sole purpose of our investigation. The framework proposed in this paper is the first of its kind to address this challenge using NOMA in real-time. Previous works on this subject focused on the problem from different perspectives using OMA only. It is worth noting that several NOMA techniques could provide results with varied resource blocks requirement, however, the framework presented in this paper is generic and could be applied to other NOMA techniques with minor changes to the algorithms and throughput calculations.

3) SPECTRUM ACQUISITION

In this subsection, the results are presented to estimate the cost of acquiring resource blocks for different application requirements such as Guaranteed Bit Rate (GBR) and Non-GBR. Fig. 7 presents the cost estimation for the acquisition of 60 to 100% of required resource blocks to meet the targeted throughput demand. Where, the price of each resource block is assumed to be one unit of price.

The Fig. 8 illustrates how IoT service providers may control their spectrum acquisition. For example, for applications where the users required GBR, the IoT-SP may choose to acquire additional resource blocks to facilitate the users with QoS requirements. Similarly, an IoT-SP may acquire less than the required number of resource block for Non-GBR users. The 100% mark on the X-Axis of Fig. 8 means the

IoT-SP gets all the channels it theoretically required to fulfill the throughput demand. 80% mark means 20% less resource blocks provided for its IoT devices, however in some applications, this can be tolerated and the IoT-SP may save money. We provide a quantification of all these values to improve IoT-SP decisions on channel acquisition strategies. As discussed earlier, the throughput demand may increase due to GBR demand or providing services in an active user area where QoS is also very important and IoT-SP may acquire additional resource blocks to guarantee the QoS. The results in Fig. 8 provides a cost estimation to help the IoT-SP to take decision of resource block acquisition. In Fig. 8c and 8d, the area with ''no additional resources available'' presents a scenario where no additional resource blocks are available from CSPs. The analysis provides useful details to enable advance planning to resource acquisition, established using account balance and application requirements. Therefore, it will reduce the delaying factors in auction procedures such as low account balance and bidding for less/more than the required resource blocks. Therefore, the overall speedup in this process will enhance the user experience-user experience refers to better quality of experience (QoE).

VII. CONCLUDING REMARKS

This paper has provided methodological contributions to compute radio resource demand for IoT devices within 5G networks by formulating a new algorithm which takes into account the characteristics of NOMA such as throughput calculation and k-means clustering. This approach is based on recent advancements made in realizing the potential of 5G in support of IoT technology. The paper also include the algorithm for efficient sub-carriers trading between the IoT-SP and CSP, based on Second Price Auction (SPA). The algorithm also provide methods to model penalties and incentives as well as account balance, important for maintaining the integrity of the spectrum market design in future settings. The effectiveness of the proposed algorithm and its role to facilitate spectrum borrowing strategies have also been quantified. In addition, to address IoT device mobility, such as devices deployed to enable autonomous vehicles, the algorithm offers a solution for minimizing complexity of re-arranging NOMA clusters to respond to the dynamics of the cells. For next generation cellular networks' context, where network slicing will play crucial role to have access to radio resources – the proposed framework offers robust allocation, revenue generation and regulation strategies for both CSP and IoT-SP. There are other issues that can be solved in the context of IoT device mobility within NOMA clusters. One issue that would come to light is the effect of mobility in intra-cell on re-establishing links to keep the communication rate within a given QoS.

REFERENCES

[1] F. Chiti, D. Di Giacomo, R. Fantacci, L. Pierucci, and C. Carlini, ''Optimized narrow-band M2M systems for massive cellular IoT communications,'' in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.

[2] J. Britt and S. Zimmerman, ''System and method for flow control in an Internet of Things (IoT) system,'' U.S. Patent App. 14/967 870, Jun. 15, 2017.

[3] I. Farris *et al.*, ''Federations of connected things for delay-sensitive IoT services in 5G environments,'' in *Proc. Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[4] S. Buzzi, I. Chih-Lin, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, ''A survey of energy-efficient techniques for 5G networks and challenges ahead,'' *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 697–709, Apr. 2016.

[5] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, ''A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends,'' *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[6] M. Nitti, M. Murroni, M. Fadda, and L. Atzori, ''Exploiting social Internet of Things features in cognitive radio,'' *IEEE Access*, vol. 4, pp. 9204–9212, 2016.

[7] *NR and NG-RAN Overall Description; Stage 2 (Release 15)*, document 3GPP TS 38.300 v15.0.0, 2017.

[8] S. A. R. Naqvi and S. A. Hassan, ''Combining NOMA and mmWave technology for cellular communication,'' in *Proc. IEEE 84th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2016, pp. 1–5.

[9] P. Angueira, M. Fadda, J. Morgade, M. Murroni, and V. Popescu, ''Field measurements for practical unlicensed communication in the UHF band,'' *Telecommun. Syst.*, vol. 61, no. 3, pp. 443–449, 2016.

[10] X. Wang, M.-J. Sheng, Y.-Y. Lou, Y.-Y. Shih, and M. Chiang, ''Internet of Things session management over LTE—Balancing signal load, power, and delay,'' *IEEE Internet Things J.*, vol. 3, no. 3, pp. 339–353, Jun. 2016.

[11] J. G. Andrews *et al.*, ''What will 5G be?'' *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[12] A. Ijaz *et al.*, ''Enabling massive IoT in 5G and beyond systems: PHY radio frame design considerations,'' *IEEE Access*, vol. 4, pp. 3322–3339, 2016.

[13] P. Schulz *et al.*, ''Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture,'' *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 70–78, Feb. 2017.

[14] M. R. Palattella *et al.*, ''Internet of Things in the 5G era: Enablers, architecture, and business models,'' *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 510–527, Mar. 2016.

[15] S.-H. Hsu, C.-H. Lin, C.-Y. Wang, and W.-T. Chen, ''Minimizing upload latency for critical tasks in cellular-based IoT networks using multiple relays,'' in *Proc. Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.

[16] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, ''Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer,'' *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[17] J. Montalban *et al.*, ''Multimedia multicast services in 5G networks: Subgrouping and non-orthogonal multiple access techniques,'' *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 91–95, Mar. 2018.

[18] A. Kiani and N. Ansari, ''Edge computing aware NOMA for 5G networks,'' *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1299–1306, Apr. 2018.

[19] A. E. Mostafa, Y. Zhou, and V. W. S. Wong, ''Connectivity maximization for narrowband IoT systems with NOMA,'' in *Proc. Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[20] M. Shirvanimoghaddam *et al.*, ''Massive non-orthogonal multiple access for cellular IoT: Potentials and limitations,'' *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 55–61, Sep. 2017.

[21] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, ''Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges,'' *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[22] M. S. Ali, H. Tabassum, and E. Hossain, ''Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems,'' *IEEE Access*, vol. 4, pp. 6325–6343, 2016.

[23] K. S. Ali, H. Elsawy, A. Chaaban, and M.-S. Alouini, ''Non-orthogonal multiple access for large-scale 5G networks: Interference aware design,'' *IEEE Access*, vol. 5, pp. 21204–21216, 2017.

[24] H. Tabassum, E. Hossain, and M. J. Hossain, ''Modeling and analysis of uplink non-orthogonal multiple access in large-scale cellular networks using poisson cluster processes,'' *IEEE Trans. Commun.*, vol. 65, no. 8, pp. 3555–3570, Aug. 2017.

[25] M. Asaduzzaman, R. Abozariba, and M. Patwary, ''Dynamic spectrum sharing optimization and post-optimization analysis with multiple operators in cellular networks,'' *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1589–1603, Mar. 2018.

[26] M. K. Naeem, M. Patwary, and M. Abdel-Maguid, "Universal and dynamic clustering scheme for energy constrained cooperative wireless sensor networks," *IEEE Access*, vol. 5, pp. 12318–12337, 2017.
[27] N. C. Luong, P. Wang, D. Niyato, Y. Wen, and Z. Han, "Resource management in cloud networking using economic analysis and pricing models: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 954–1001, 2nd Quart., 2017.
[28] X. Chen, L. Xing, T. Qiu, and Z. Li, "An auction-based spectrum leasing mechanism for mobile macro-femtocell networks of IoT," *Sensors*, vol. 17, no. 2, p. 380, 2017.
[29] S. Misra, S. Das, M. Khatua, and M. S. Obaidat, "QoS-guaranteed bandwidth shifting and redistribution in mobile cloud environment," *IEEE Trans. Cloud Comput.*, vol. 2, no. 2, pp. 181–193, Apr. 2014.
[30] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.
[31] J. Choi, "Non-orthogonal multiple access in downlink coordinated two-point systems," *IEEE Commun. Lett.*, vol. 18, no. 2, pp. 313–316, Feb. 2014.
[32] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, Oct. 2015.
[33] J. Choi, "Power allocation for max-sum rate and max-min rate proportional fairness in NOMA," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2055–2058, Oct. 2016.
[34] N. Otao, Y. Kishiyama, and K. Higuchi, "Performance of non-orthogonal access with SIC in cellular downlink using proportional fair-based resource allocation," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2012, pp. 476–480.
[35] W.-Y. Lee and I. F. Akyildiz, "Joint spectrum and power allocation for inter-cell spectrum sharing in cognitive radio networks," in *Proc. 3rd IEEE Symp. New Frontiers Dyn. Spectr. Access Netw.*, Oct. 2008, pp. 1–12.
[36] A. Benjebbour, K. Saito, A. Li, Y. Kishiyama, and T. Nakamura, "Non-orthogonal multiple access (noma): Concept and design," in *Signal Processing for 5G: Algorithms and Implementations*. 2016, pp. 143–168.

RAOUF ABOZARIBA (S'16–M'17) received the Ph.D. degree from Staffordshire University, in 2017. He is currently a Senior Research Associate with the School of Computing and Communications, Lancaster University, U.K. His current research focuses on 5G networks and the IoT communications, with an emphasis on dynamic spectrum sharing and allocation.

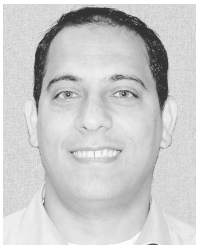
MUHAMMAD KAMRAN NAEEM (M'17) received the B.Eng. degree (Hons.) in electrical engineering from Air University, Pakistan, in 2008, and the M.Sc. and Ph.D. degrees in telecommunication engineering from Staffordshire University, Stafford, U.K., in 2011 and 2017, respectively.

He was a Researcher and a Lecturer with Birmingham City University, from 2017 to 2018. He is currently a Postdoctoral Researcher with Southampton Solent University. His research interests include the Internet of Things, 5G networks, wireless communications, channel estimation and equalization techniques, and multi antenna systems.

MOHAMMAD PATWARY (SM'11) received the B.Eng. degree (Hons.) in electrical and electronic engineering from the Chittagong University of Engineering and Technology, Bangladesh, in 1998, and the Ph.D. degree in telecommunication engineering from The University of New South Wales, Sydney, Australia, in 2005. He was with the General Electric Company of Bangladesh, from 1998 to 2000, and Southern Poro Communications, Sydney, from 2001 to 2002, as a Research and Development Engineer. He was a Lecturer with The University of New South Wales, from 2005 to 2006, and then a Senior Lecturer with Staffordshire University, U.K., from 2006 to 2010. He was then a Full Professor of wireless systems and digital productivity and the Chair of the Centre of Excellence on Digital Productivity with Connected Services, Staffordshire University, until 2016. He is currently a Full Professor of telecommunication networks and digital productivity and the Head of Research with the Centre for Intelligent Systems and Networks, School of Computing and Digital Technology, Birmingham City University. He is also a Principal Data Architect for a large-scale 5G testbed in U.K. to accelerate digital productivity and develop urban connected community. His current research interests include sensing and processing for intelligent systems, wireless communication systems design and optimization, signal processing and energy-efficient systems, future generation of cellular network architecture, and business modeling for data economy.

MIR SEYEDEBRAHIMI received the Ph.D. degree in wireless and mobile communications from Aston University, U.K., in 2015. He is currently a Senior Fellow in embedded systems and networking with WMG, The University of Warwick, U.K. He is also leading various teaching modules in the area of networking, computing, digital transmission, and corresponding discrete mathematics. He also has previous industrial experiences within Telecommunication Sector as a Senior Network Management Designer and as an Instructor. His main research interests include software-defined and embedded systems employed in cyber physical systems, the Internet of Things, and wireless networking in general.

PETER BULL (M'–) was a Research Engineer on a collaborative project with BAE Systems, working on an industrial sub-contract for BAE investigating the application of a safety critical Ethernet concept to an OpenFlow-based switch. He is currently a Lecturer with the School of Computing and Digital Technology, Birmingham City University. His current research interests are in the area of software-defined networking and related issues, such as network performance and security.



ADEL ANEIBA received the Ph.D. degree in the field of mobile computing and distributed systems from Staffordshire University, in 2008. He has worked as a Senior ICT Consultant for international organizations, for 10 years, including UNESCO and several governmental organizations for many years, and has participated in managing mega ICT projects mainly on data center designing and development, and reengineering business processes. He is currently an Associate Professor in the area of computer networks and the Internet of Things (IoT) with Birmingham City University.He is also the BCU-IBM Academic Coordinator, who is maintaining the relationship on various academic activities, including R&D around the IoT and BlockChain. He is supervising several Ph.D. students on the research topics, such as the IoT, SDN, Resources allocations, and optimization in 5G and Blockchain applications in smart cities domain. His current research interests include the IoT, computer networks, evaluation and optimization, and blockchain. He is co-leading several IoT research projects in partnership with West Midlands Combined Authority. Recently, he was awarded the Innovative UK-Nesta Project on using drones in future cities. In addition, he participates in the Digital Catapult Innovation Programme on waste management and air-quality challenges that are facing local authorities.