

Citation for published version

Trakas, P. [Panagiotis]. Adelantado, F. [Ferran]. Verikoukis, C. [Christos]. (2019). QoE-Aware Resource Allocation for Profit Maximization Under User Satisfaction Guarantees in HetNets With Differentiated Services. IEEE Systems Journal, 13(3), 2664-2675. doi: 10.1109/JSYST.2018.2876894

DOI

<https://doi.org/10.1109/JSYST.2018.2876894>

Handle

<http://hdl.handle.net/10609/148966>

Document Version

This is the Accepted Manuscript version.

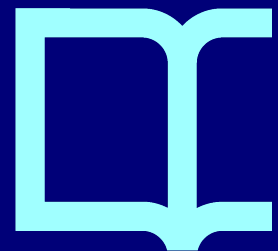
The version published on the UOC's O2 Repository may differ from the final published version.

Copyright

© Duke University Press

Enquiries

If you believe this document infringes copyright, please contact the UOC's O2 Repository administrators: repositori@uoc.edu



QoE-aware resource allocation for profit maximization under user satisfaction guarantees in HetNets with differentiated services

Panagiotis Trakas, *Student Member, IEEE* Ferran Adelantado, *Member, IEEE*, and Christos Verikoukis, *Senior Member, IEEE*

Abstract—The rise of third-party content providers and the introduction of numerous applications has been driving the growth of mobile data traffic in the past few years. The applications’ various Quality of Service (QoS) requirements as well as the use of multiple devices per user have increased the traffic heterogeneity, pressing the telecommunications industry to the deployment of dense Heterogeneous Networks (HetNets). At the same time, the rise of the content providers has also led to the decrease of the Mobile Network Operators’ (MNOs) revenues. Under these circumstances, the MNOs need to guarantee the users’ Quality of Experience (QoE) requirements, while ensuring the sustainability of HetNet investments. To this end, we consider a HetNet deployment where MNOs offer a multitude of services with diverse pricing. We propose a heuristic, greedy, QoE-aware resource allocation algorithm with fairness and overall user satisfaction constraints to maximize the MNO profit, while providing high QoE. Simulation results show that the proposed algorithm can handle traffic heterogeneity by achieving substantial profit and QoE gains, compared to state of the art algorithms.

Index Terms—Resource Allocation, Profit, Traffic Heterogeneity, QoE, Diverse Pricing.

I. INTRODUCTION

The exponential growth of mobile data traffic experienced over the last years is expected to continue in the future. This traffic growth is mainly the result of the surge in demand of multimedia and video content (usually offered by independent content providers) and the explosion in the number of devices and broadband connections. For instance, according to [1], IP traffic is expected to grow from 13 GB per capita in 2016 up to 35 GB per capita by 2021. Moreover, by 2021, the mobile traffic will be heterogeneous from a dual perspective, since there will be a wide range of possible Quality of Service (QoS) requirements (e.g. video on-demand, online gaming and messaging are very different in terms of QoS) and it will be originated/received by diverse devices (e.g. tablets, laptops or smartphones) [2]. In this context, it has been shown that the quality perceived by the users can not be fully captured with QoS metrics, thus making Quality of Experience (QoE) one of the most important Key Performance Indicators (KPI) [3]. Therefore, in the future Mobile Network Operators (MNO) will have to be able to meet not only the envisioned boost of the traffic demand, but also its heterogeneity in terms of QoS/QoE requirements.

The densification of the Radio Access Network (RAN) is one of the main strategies to catch up with the future intense and diverse demand [4]. Mobile industry must invest large amounts of capital in the deployment of dense Heterogeneous Networks (HetNets) to be able to provide seamless connectivity and high QoS/QoE. Therefore, the densification strategy to cope with the increase of the demand has a significant impact on the balance sheet of the different stakeholders, and particularly for MNOs (e.g. increase of the deployment cost). These financial aspects/constraints are exacerbated by the so-called *traffic and revenue paradox/challenge* [5]; specifically, although it may seem contradictory, the described traffic boost has increased the content providers’ profits while, simultaneously, has diminished the MNOs’ revenues. This occurs because the MNO’s basic services (voice and messaging) have been gradually replaced by their third-party counterparts. Moreover, the MNO’s data service prices have been decreasing over the years, due to the market competition.

Therefore, MNOs face a two-fold challenge: meet the QoE requirements and maximize the profit. It has been proven that the relation between QoS and QoE has a non-linear nature [6]. This means that small degradations in the received QoS can significantly impact on the perceived QoE level. Yet, QoE is influenced by other factors such as pricing or device characteristics [7]. In this context, it is necessary to design network functionalities adapted to the new requirements, such as QoE-aware Radio Resource Management strategies, and always trying to maximize the profit (to compensate the diminished MNOs’ revenues and the increasing deployment investment).

The majority of works on Resource Allocation (RA) and scheduling in the context of 4G and 5G networks focus mainly on the provision of high QoS/QoE and other network aspects (e.g. power allocation, fairness etc.), without taking into account the impact of their proposals on the financial aspects of the MNOs (e.g. profit) [8]–[11]. For instance, a joint RA and PA algorithm based on an iterative method for the maximization of proportional fair energy efficiency in energy harvesting-based wireless networks is proposed in [8]. Similarly, a QoE-aware joint RA and PA algorithm with a satisfaction factor that determines the percentage of served users is proposed in [9]. A QoE-based proportional fair scheduling scheme, which aims for the users’ QoE maximization, is proposed in [10]. A concave utility function for QoE provision is used to achieve global optimality through opportunistic gradient scheduling.

P. Trakas and F. Adelantado are with the Open University of Catalonia, 08860 Barcelona, Spain, e-mail: (ptrakas,ferranadelantado@uoc.edu).

C. Verikoukis is with the Telecommunications Technological Centre of Catalonia (CTTC/CERCA), 08860 Castelldefels, Spain, e-mail: (cverik@cttc.es).

The authors in [11] design a hybrid coordinated multipoint transmission scheme for the downlink transmission in cloud-RANs, and propose a stochastic gradient algorithm and a low complexity online learning algorithm for the joint RA and PA of delay-sensitive traffic. In [12], the authors study the bandwidth allocation for video streaming in HetNets. For the minimization of the delay per chunk, they propose a cross layer video streaming algorithm, which offloads users from the small cells to the macrocell. The authors in [13] evaluate the operation of dynamic adaptive streaming over HTTP for three open source players, propose solutions to rate adaptation misbehaviour and suggest guidelines for their improvement.

However, there is a limited number of works on RA or scheduling that consider both network (e.g. QoE or fairness) and financial aspects (e.g. MNO profit). The authors in [14] examine network-assisted mobile streaming, while accounting for the MNO's network and economic factors. They propose a Markov Decision Process (MDP) based adaptation framework, which provides near-optimal profit for the MNO. A downlink packet scheduling scheme for QoS provisioning in wireless networks is proposed in [15]. The scheme's objective is the satisfaction of users with various QoS requirements and priority classes, and the minimization of the network operator's revenue loss. The authors in [16] propose an auction-based RA scheme. The auction is used to decide the users' service levels fairly when a BS's resources do not suffice for the satisfaction of all the users' demands. The RA is formulated as a multi-objective optimization problem, which maximizes the Jain's fairness index, the resource utilization, and the MNO's revenue.

The authors in [17] propose a marginal-based pricing and RA framework for the improvement of the resource utilization and the profit of a wireless video broadcasting MNO as well as the individual user profit. However, the problem formulation as well as the proposed algorithm focus solely on the maximization of the MNO profit with minimum quality guarantees for the users. That is, the authors do not take into account the effect that constraints on fairness and overall user satisfaction may have on the MNO profit, as well as on the user experience.

A method for the simultaneous RA in both licensed and unlicensed bands in the SCs of a HetNet is proposed in [18]. The concurrent RA in both bands for the SC users is designed as an optimization problem. The authors solve the problem twice; first they maximize the users' sum-rate, and then the MNO revenue. Both problems have constraints on the interference on macrocell users, and on the satisfaction of the SC users' minimum rate requirements. However, as the authors consider only usage-based pricing, the revenue maximization problem provides the same results as the sum-rate maximization problem. Moreover, as the considered business model does not take into account the system's operational costs, they cannot study the trade-off between the QoS provided to the users and the MNO revenue (profit).

This paper, which is an extension of our prior work [19], studies the resource allocation problem aiming to maximize the MNO profit, while offering high QoE to the users, under fairness and overall user satisfaction (OS) constraints. We

consider HetNets composed of macrocell and SC base stations (BSs), with dynamic traffic described by numerous QoS/QoE demands. In contrast to the literature, we further consider diverse pricing, that is, various service prices and different pricing schemes. In order to address the challenges of traffic heterogeneity and high network profitability, we propose a QoE-aware resource allocation algorithm that exploits the QoE-awareness and the network's economic aspects (i.e. the MNO profit). Our proposed algorithm consists of three greedy sub-algorithms; a first algorithm that iteratively maximizes the satisfaction of the user with the lowest resource requirements; a second algorithm that adjusts the previous resource allocation by increasing iteratively the BS profit under an OS constraint, and a third algorithm that maximizes the BS profit in the same manner with an additional fairness constraint. The main contributions of the paper are summarized in the following:

- While in [19] we proposed a novel joint resource allocation and pricing algorithm for profit maximization under an overall user satisfaction constraint, in this paper we also introduce fairness among users, extend our analysis on the connection among the individual and overall user satisfaction, fairness, pricing and profit, as well as how they impact each other, through a resources allocation approach.
- We propose a heuristic, greedy, low-complexity, QoE-aware resources allocation algorithm that maximizes the MNO profit while imposing constraints on the minimum overall users satisfaction and on the fairness among users. Our simulation results show that the proposed algorithm outperforms state-of-the-art algorithms.
- We shed light on the trade-off between users' satisfaction, fairness and MNO profit. The paper shows that, given the non-linear relation between QoS and QoE, there is room for profit maximization resources allocation solutions without penalizing the quality perceived by the users.
- We show that the use of diverse pricing schemes (e.g. data-based and time-based pricing) can lead to the prioritization of some services over the others in *pure* profit maximizing algorithms. Moreover, as the QoE is affected by the price level, the manipulation of the service prices by the MNO can affect both the profit and overall satisfaction. In line with this, we provide an accurate set of simulation results that shows the sensitivity of profit and overall satisfaction with respect to pricing schemes and price level, which complete our preliminary analysis in [19].

The rest of the paper is organized as follows. We present the system model in Section II. Section III describes the MNO's objectives. In Section IV, the profit optimization problem is formulated, and in Section V a QoE-aware profit maximizing RA algorithm is proposed. We validate our algorithm in Section VI, and conclude the paper in Section VII.

II. SYSTEM MODEL

The considered network is composed of a set of macrocells and a set of SCs, all of them deployed by a single MNO. We denote this set of BSs, both macrocells and SCs, as $\mathcal{B} =$

TABLE I: Notation

Notation	Description
\mathcal{B}	Set of BSs
\mathcal{U}	Set of users
\mathcal{U}_i	Set of users in BS i
\mathcal{U}_i^t	Set of time-based charged users in BS i
\mathcal{U}_i^B	Set of data-based charged users in BS i
\mathcal{S}	Set of services
\mathcal{Q}	Set of QoE levels
\mathcal{D}	Set of devices
b_i	BS i 's available bandwidth
π_k	Service Profile (SP) k
s_k, q_k, p_k	Service, QoE class, service price of SP k
θ_k^B	Per data unit charging of service k
θ_k^t	Per time unit charging of service k
r_j	User j 's data rate
r_{kd}	Target data rate for SP k , device d
w_{ij}	Resources allocated to user j by BS i
w_i	$\sum_{j \in \mathcal{U}_i} w_{ij}$
ε_{ij}	Spectral efficiency between user j and BS i
Q_j^{kd}	QoE level of user j with SP k , device d
Q_k^{tg}	Target QoE level for SP k
Q_k^{drop}	Service dropping QoE level for SP k
\hat{Q}_j^{kd}	QoS-based component for QoE mapping
\hat{Q}_p	Price-based component for QoE mapping
$\alpha_{kd}, \gamma_{kd}, \beta_{kd}$	SP-dependent constants for \hat{Q}_j^{kd}
v_k	User-dependent constant for \hat{Q}_p
σ_{ij}	Satisfaction of user $j \in \mathcal{U}_i$
J_i	Jain's fairness index in BS i
OS_i	Overall User Satisfaction (OS) in BS i
OS_i^{max}	Maximum possible OS in BS i
ϕ_i	Relative OS in BS i , OS_i/OS_i^{max}
ϕ^{min}	Relative OS threshold
P_i	BS i profit
R_i	BS i revenue
CB_i	BS i bandwidth utilization cost
c_i, h_i	Cost adjusting factors

$\{1, 2, \dots, N_B\}$, where N_B is the total number of BSs. The bandwidth allocated to each BS $i \in \mathcal{B}$ is hereafter referred to as b_i (in Hz). The notation used henceforth is summarized in Table I.

The MNO serves a set of users $\mathcal{U} = \{1, 2, \dots, N_U\}$, where N_U is the total number of users. It is assumed that users are not served by more than a single BS simultaneously, and therefore we define the set of users served by BS $i \in \mathcal{B}$ as \mathcal{U}_i , where $\mathcal{U} = \cup_{i \in \mathcal{B}} \mathcal{U}_i$ and $\cap_{i \in \mathcal{B}} \mathcal{U}_i = \emptyset$. MNOs have put the focus on the QoS and QoE as the target KPI in the design of networks [3]. Accordingly, in our model each user has a contract with the MNO that specifies a desired QoE for each service, denoted in the sequel as *Service Profile* (SP). If we define the set of services as $\mathcal{S} = \{s : s = 1 \dots S\}$ and the set of QoE classes as $\mathcal{Q} = \{q : q = 1 \dots Q\}$ (Q is assumed to be a discrete and finite set), a generic SP can be defined as $\pi_k = (s_k, q_k, p_k)$, where p_k is the price of the service (in €), $s_k \in \mathcal{S}$ and $q_k \in \mathcal{Q}$. Focusing on p_k , it is worth noting that its definition depends on the service s_k . Thus, some services are charged based on the amount of transmitted/received data and some others are based on the connection time. Let us define the price for a data-based charged service as θ_k^B (in €/MB) and for a time-based charged service as θ_k^t (in €/sec). Moreover, we denote by $\mathcal{U}_i^B, \mathcal{U}_i^t \subseteq \mathcal{U}_i$ the sets of data-based and time-based charged users served by BS i , respectively (i.e. $\mathcal{U}_i^B \cup \mathcal{U}_i^t = \mathcal{U}_i$ and $\mathcal{U}_i^B \cap \mathcal{U}_i^t = \emptyset$). The general expression of p_k for a time period

T can be expressed as

$$p_k = \begin{cases} T\theta_k^t & \text{If user } j \in \mathcal{U}_i^t \\ \frac{T \cdot r}{8} \theta_k^B & \text{If user } j \in \mathcal{U}_i^B \end{cases} \quad [\text{€}], \quad (1)$$

where r (in Mbps) is the user transmission rate¹. As for the perceived QoE, in general any user with a service profile π_k has a target QoE level, Q_k^{tg} , and a minimum QoE level below which the session is dropped, Q_k^{drop} (in the Mean Opinion Score (MOS) scale [6]). We assume that both values are established in the contract between the user (as a customer) and the MNO.

Although the perceived QoE is influenced by multiple factors, as it will be detailed in Section III, we now focus on the impact of the user device. Nowadays, a single user can get connected to the network with different devices (tablet, laptop, smartphone, etc), each one with specific characteristics. These characteristics of the device, such as the screen quality or screen size, are relevant since they may improve or worsen the perceived QoE. For instance, to perceive similar QoE levels, lower image resolution and hence lower transmission bit rate (i.e. lower QoS) is required for a user using a video service in a small-sized screen smartphone than for the same user with a large screen tablet [20]. Therefore, the characteristics of the device must be taken into account to design efficient radio resources management algorithms. We define the set of devices as $\mathcal{D} = \{d : d = 1 \dots D\}$, and the mapping function that links the device-SP pair with the required transmission rate, r_{kd} , as $f : (\pi_k, d) \rightarrow r_{kd}$. According to the definitions, the QoE perceived by a user $j \in \mathcal{U}$ with a SP π_k and using a device $d \in \mathcal{D}$, namely Q_j^{kd} , will be higher than the target QoE Q_k^{tg} if the transmission rate from the serving BS to the user j is higher than r_{kd} . In other words, the target QoE is met at time period t if $r_j(t) = w_{ij}(t)\varepsilon_{ij}(t)b_i \geq r_{kd} = f(\pi_k, d)$, where $r_j(t)$ is the actual transmission rate of user $j \in \mathcal{U}_i$ (in Mbps), $w_{ij}(t) \in [0, 1]$ is the portion of BS $i \in \mathcal{B}$ radio resources allocated to user j , and $\varepsilon_{ij}(t)$ is the spectral efficiency of the link between user j and BS i (in bps/Hz), which can be approximated as $\varepsilon_{ij}(t) = \log_2(1 + SINR_{ij})$, where $SINR_{ij}$ is the Signal to Interference and Noise Ratio received by user j , when served by BS i .

Note that, for a given device d and service s_k , each QoE class is translated into an equivalent QoS level (e.g. different r_{kd} values are required for streaming SD and HD video). This means that the user can choose among Q QoE classes on a contract basis for each service, depending on the personal preferences (e.g. preference for high browsing speed but SD video), and this choice impacts on the minimum QoS requirements. In turn, for a given service s_k and a QoE class q_k , the user can receive the service through a diversity of devices. For each device, the required transmission rate can also differ.

Based on the definitions stated above, it is clear that the satisfaction of users is tightly coupled with the perceived QoE. Specifically, if the satisfaction of user j served by BS i , namely $\sigma_{ij}(t)$, is defined within the interval $[0, 1]$, when $Q_j^{kd}(t) \leq$

¹The user rate in (1) is divided by 8, so that it is expressed in MB/sec, and hence the price p_k in monetary units (i.e. €).

Q_k^{drop} , the session is dropped and the satisfaction is equal to 0. Conversely, when $Q_j^{kd}(t) \geq Q_k^{tg}$, the satisfaction is equal to 1. Thus, according to [7], the satisfaction is defined as

$$\sigma_{ij}(t) = \begin{cases} 0 & \text{if } Q_j^{kd}(t) \leq Q_k^{drop} \\ \frac{Q_j^{kd}(t) - Q_k^{drop}}{Q_k^{tg} - Q_k^{drop}} & \text{if } Q_j^{kd}(t) \in (Q_k^{drop}, Q_k^{tg}) \\ 1 & \text{otherwise.} \end{cases} \quad (2)$$

III. MNO OBJECTIVES

In order to propose a resource allocation scheme based on network and economic functions, we first need to identify and analyse the two MNO's objectives. First, they must offer to the users the QoE agreed in the SP, and guarantee fairness when the system is congested. Second, the network must be managed so as to maximize their economic profit. In the following, the analyses of the QoE, fairness, overall user satisfaction and the profit are detailed.

A. Network Performance Metrics

Based on the analysis described in [7], the perceived QoE $Q_j^{kd}(t)$ can be divided into two components: the QoS-based component ($\hat{Q}_j^{kd}(t)$) and the price-based component ($Q_p(p_k)$).

$$Q_j^{kd}(t) = \hat{Q}_j^{kd}(t) \cdot Q_p(p_k). \quad (3)$$

The QoS-based component, $\hat{Q}_j^{kd}(t) \in [1, 5]$ (in the MOS scale), shows the effect of QoS level on QoE. In the literature, the QoE is usually modelled to have an exponential interdependency with the QoS, also known as the IQX hypothesis [6]. Using the transmission rate $r_j(t)$ as the reference QoS metric, and according to the IQX hypothesis, we can express $\hat{Q}_j^{kd}(t)$ as

$$\hat{Q}_j^{kd}(t) = \alpha_{k_j d_j} e^{-\beta_{k_j d_j} \Delta r_j(t)} + \gamma_{k_j d_j}, \quad (4)$$

where $\Delta r_j(t) = r_{kd} - r_j(t)$, $\alpha_{k_j d_j}, \gamma_{k_j d_j} > 0$ (both in the MOS scale) and $\beta_{k_j d_j} > 0$ (in sec/bit) are SP-device dependent constants. Regarding the price-based component, it captures how the perception of the quality improves (worsens) as the price falls (rises). As in [7], $Q_p(p_k)$ is modelled as

$$Q_p(p_k) = 1 - v_k p_k, \quad (5)$$

where $v_k > 0$ is an adjusting factor measured in €^{-1} . Particularly, v_k is the factor that determines the sensitivity of the perceived QoE to price variations for a user j and a service s_k . We assume that the value of v_k and hence $Q_p(p_k)$ can be different for each user j , in order to capture the effect of p_k on each user individually. As it can be observed in (5), if the user does not pay for the service (i.e. $p_k = 0$), the price-based component reaches the maximum value, $Q_p(0) = 1$, thereby increasing the perceived QoE in (3). That is, the more a user pays for a service, the higher the expectations on the received quality are.

It should be pointed out that $Q_{k_j}^{tg}$ corresponds to the QoE level a user j wants to perceive, it is constant, and does not depend on the service price. Hence, the MNO's objective is to offer user j a service with $Q_j^{kd} = Q_{k_j}^{tg}$ by providing the required QoS and corresponding price combination.

Fairness: MNOs aim to offer fairness among users both when the available resources suffice to provide them all with the target QoE (i.e. $Q_j^{kd} = Q_k^{tg}$, for all $j \in \mathcal{U}$) and when not all of them can be appropriately served (i.e. $Q_j^{kd} < Q_k^{tg}$ for some users). In this paper, we adopt the well-known Jain's fairness index [21] of the users' satisfaction level $\sigma_{ij}(t)$ as the QoE-fairness metric. Hence, at a specific time period t a BS i 's Jain's index can be expressed as

$$J_i(t) = \frac{\left(\sum_{j \in \mathcal{U}_i} \sigma_{ij}(t)\right)^2}{|\mathcal{U}_i| \sum_{j \in \mathcal{U}_i} \sigma_{ij}^2(t)} \in [0, 1], \quad (6)$$

where $|\mathcal{U}_i|$ denotes the cardinality of \mathcal{U}_i (i.e. the number of users served by BS i).

Overall User Satisfaction: We observe in (6) that $J_i(t)$ depends on the standard deviation of the users' satisfaction in BS i ; the lower the standard deviation, the higher the $J_i(t)$. Hence, a high $J_i(t)$ can be achieved even when both the average and the standard deviation of the users' satisfactions ($\sigma_{ij}(t)$) are low. This means that high fairness does not guarantee high QoE for the users. To this end, we define the Overall User Satisfaction (OS) in a BS i as the sum of of the satisfaction of all the users connected to i , $OS_i(t) = \sum_{j \in \mathcal{U}_i} \sigma_{ij}(t)$. Similarly, the total OS in the system is defined as the aggregate satisfaction of all the users, $OS(t) = \sum_{i \in \mathcal{B}} OS_i(t)$. Thus, the objective of the MNO is to achieve high overall satisfaction (OS_i) and high Jain's index (J_i) values $\forall i \in \mathcal{B}$.

Since the number of users connected to a BS and the satisfaction of each user depends on their spectral efficiency and on the allocation of resources, the maximum overall user satisfaction varies along time. Let us define, for a given time interval t , the maximum achievable Overall User Satisfaction at BS i as $OS_i^{max}(t) = \max_{\forall w_{ij}(t)} \{OS_i(t)\}$. Based on the definition, the Overall User Satisfaction achieved with a specific resource allocation can be expressed as a fraction of the maximum value. Therefore, we define the relative overall user satisfaction $\phi_i(t) = OS_i(t)/OS_i^{max}(t) \in [0, 1]$ as a QoE-aware performance metric. The objective of the MNO is then given by

$$\phi_i(t) \geq \phi^{min}, \forall i \in \mathcal{B}, \quad (7)$$

where ϕ^{min} is a minimum threshold defined by the MNO.

B. MNO Economic Profit

The objective of the MNO is the maximization of the profit while satisfying the QoE required by the users. Specifically, the total profit $P(t)$ of the MNO is the sum of the individual profits of each BS $P_i(t)$, i.e. $P(t) = \sum_{i \in \mathcal{B}} P_i(t)$. In [22], $P_i(t)$ is expressed as the revenue obtained from the traffic served at time t , $R_i(t)$, minus the cost incurred when serving the traffic, which depicts the bandwidth utilization cost, $CB_i(t)$. Therefore,

$$P(t) = \sum_{i \in \mathcal{B}} P_i(t) = \sum_{i \in \mathcal{B}} (R_i(t) - CB_i(t)), [\text{€}]. \quad (8)$$

The revenue of BS i , $R_i(t)$, is usually the price of the services paid by the users in \mathcal{U}_i . That is, $R_i(t) = \sum_{j \in \mathcal{U}_i} R_{ij}(t)$, where

$R_{ij}(t) = p_{k_j}$ is the revenue paid by user j with a SP π_{k_j} , when connected to BS i at time period t for a duration of T seconds. With regard to $CB_i(t)$, it is a convex and increasing exponential function of the total resources used by BS i , $w_i(t) = \sum_{j \in \mathcal{U}_i} w_{ij}(t)$ [22], and for a duration of T seconds it can be written as

$$CB_i(t) = c_i e^{h_i w_i(t) b_i T}, \quad (9)$$

where c_i (in $\text{€}/\text{sec}$) and h_i (in MHz^{-1}) are adjusting factors that capture the differences in the operational cost of the different BSs (e.g. macrocells and SCs have different transmit power, maintenance cost, site rent, etc). Substituting (1) and (9) into (8), and denoting the SP of a generic user j as π_{k_j} , the profit of BS i at time period t with a duration of T seconds when $Q_j^{k_j d_j}(t) \in (Q_{k_j}^{drop}, Q_{k_j}^{tg}]$ is given by

$$\begin{aligned} P_i(t) &= \sum_{j \in \mathcal{U}_i} p_{k_j} - c_i e^{h_i \sum_{j \in \mathcal{U}_i} w_{ij}(t) b_i} T \\ &= T \left(\sum_{j \in \mathcal{U}_i^B} \frac{\varepsilon_{ij}(t) w_{ij}(t) b_i}{8} \theta_{k_j}^B \mathbb{1}(\sigma_{ij}(t) > 0) \right. \\ &\quad \left. + \sum_{j \in \mathcal{U}_i^t} \theta_{k_j}^t \mathbb{1}(\sigma_{ij}(t) > 0) - c_i e^{h_i \sum_{j \in \mathcal{U}_i} w_{ij}(t) b_i} \right), \end{aligned} \quad (10)$$

where $\mathbb{1}(\cdot)$ is the binary indicator function, which is equal to 1 if the condition is true and 0 otherwise. We use the binary indicator function in order to emphasize that a BS i will not receive revenue if the allocated resources to user j , $w_{ij}(t)$, do not suffice for a satisfactory service with $\sigma_{ij}(t) > 0$.

It can be seen in (10) that the profit is influenced by multifarious factors, such as the perceived QoE (which in turn depends on multiple factors), the cost, the radio resources usage, etc. The maximization of the profit involves all these factors.

IV. PROFIT OPTIMIZATION

As explained in the previous section, the MNO aims to maximize his profit in each BS i $P_i(t)$, while satisfying the required QoE of all users. However, when not all users can be served with the required QoE due to network congestion, the MNO must ensure fairness among them and the highest possible QoE level. Based on (10), the BS i 's profit maximization problem at time t is formulated as

$$\begin{aligned} \max_{w_{ij}, j \in \mathcal{U}_i} P_i(t) &= T \left(\sum_{j \in \mathcal{U}_i^B} \frac{\varepsilon_{ij}(t) w_{ij}(t) b_i}{8} \theta_{k_j}^B \mathbb{1}(\sigma_{ij}(t) > 0) \right. \\ &\quad \left. + \sum_{j \in \mathcal{U}_i^t} \theta_{k_j}^t \mathbb{1}(\sigma_{ij}(t) > 0) - c_i e^{h_i \sum_{j \in \mathcal{U}_i} w_{ij}(t) b_i} \right), \end{aligned} \quad (11)$$

$$w_i \in [0, 1], \quad \forall i \in \mathcal{B}, \quad (11a)$$

$$J_i(t) \geq J^{min}, \quad \forall i \in \mathcal{B}, \quad (11b)$$

$$\phi_i(t) \geq \phi^{min}. \quad (11c)$$

In the optimization problem, the maximum bandwidth allocated by BS i is b_i , that is $\sum_{j \in \mathcal{U}_i} w_{ij}(t) = w_i(t) \leq 1$

(11a), and QoE fairness must be guaranteed for a minimum Jain's index value, J^{min} (11b). Finally, the relative overall user satisfaction must be higher than the minimum threshold ϕ^{min} (11c). Since we use the binary indicator function in $P_i(t)$, the optimization problem in (11) is a Mixed-Integer Non-linear Programming (MINLP) problem, whose computational complexity is NP-hard [23]. As this maximization problem cannot be solved in polynomial time due to its complexity, in the following section we propose a heuristic, QoE-aware resources allocation algorithm.

V. PROFIT MAXIMIZING RESOURCE ALLOCATION

As it can be observed in (11), the resources allocation for profit optimization is constrained by minimum fairness and overall user satisfaction values. In this section we analyse the interaction between satisfaction, fairness and profit, and propose a greedy, heuristic, QoE-aware resources allocation algorithm for profit maximization.

Based on (11), the MNO profit can be maximized by reducing the cost and/or increasing the revenue. However, it is noteworthy that for a given association of users to BS i , \mathcal{U}_i , the bandwidth utilization cost depends on the total amount of resources allocated to users, regardless of how they are distributed among the users. Therefore, the utilization cost ($CB_i(t)$) is fixed for a given number of total resources ($w_i(t)$).

In turn, (1) shows that the revenue presents a differentiated behaviour for time-based charged services and data-based charged services. Whereas the revenue generated by a user with a time-based charged service remains constant as long as the connection is not dropped, the revenue generated by users with a data-based charged service increases with the amount of transferred data. In other words, after providing time-based charged users with the minimum amount of resources required to guarantee that the perceived QoE is above the minimum QoE level, the revenue can be increased by allocating the rest of resources to data-based charged users. Note that in terms of QoE the aforementioned resources allocation strategy is translated into low satisfaction of time-based charged users (their $\sigma_{ij}(t)$ is low but above 0) and higher satisfaction of data-based charged users. These differences between the two types of services lead to resources allocation unfairness (i.e. low Jain's index values, $J_i(t)$) and could result in low relative overall user satisfaction ($\phi_i(t)$). This is the reason why (11b) and (11c) impose minimum fairness and minimum relative overall user satisfaction levels, respectively, and this is also the reason why the heuristic algorithm presented in the sequel takes both aspects into account. This analysis shows that there is a trade-off between the MNO profit and the network performance, as the optimization of one of the two objectives comes at the expense of the other.

In order to overcome the complexity of the maximization problem stated in (11), we propose a low complexity resource allocation algorithm, $O(n^2)$, presented in Algorithm 1, that maximizes the MNO profit for a minimum fairness and OS level (i.e. J^{min} and ϕ^{min}). This greedy, Profit Maximizing resource allocation algorithm (referred to as PM) takes as input the users associated to a BS at period t , and each user's SP-device pair (p_k, d) .

Algorithm 1: Profit Maximizing RA Algorithm (PM)

```

1 Set  $\mathcal{U}'_i = \mathcal{U}_i$ ,  $w_i(t) = 0$  and compute  $w_{ij}(t) \leq 1$  to
   maximize  $\sigma_{ij}(t)$  for each  $j \in \mathcal{U}_i$ 
2 while  $\mathcal{U}'_i \neq \emptyset$  do
3   Find user  $j \in \mathcal{U}'_i$  with  $\min(w_{ij}(t))$ 
4   if  $w_{ij}(t) \leq 1 - w_i(t)$  then
5      $w_i(t) = w_i(t) + w_{ij}(t)$ 
6   else if  $\sigma_{ij}(t) \geq \sigma_{ij}^{min}$  for  $w_{ij}(t) = 1 - w_i(t)$  then
7      $w_{ij}(t) = 1 - w_i(t)$  and  $w_i(t) = 1$ 
8   else
9      $w_{ij}(t) = 0$ 
10  end
11   $\mathcal{U}'_i \leftarrow \mathcal{U}'_i - \{j\}$ 
12 end
13 Set  $OS_i^{max}(t) = OS_i(t)$ ,  $\phi_i(t) = 1$  and  $\sigma_{step} = \sigma_{step}^{max}$ 
14 while  $\phi_i(t) \geq \phi_i^{min}$  and  $\sigma_{step} \geq \sigma_{step}^{min}$  do
15   forall  $j \in \mathcal{U}_i$  do
16      $\sigma_{ij}^-(t) = \max(\sigma_{ij}(t) - \sigma_{step}, 0)$ 
17      $\sigma_{ij}^+(t) = \min(\sigma_{ij}(t) + \sigma_{step}, 1)$ 
18     Calculate the total profit  $P_i(t)$  and  $\phi_i(t)$  with
19      $\sigma_{ij}(t) = \sigma_{ij}^-(t)$  and  $\sigma_{ij}(t) = \sigma_{ij}^+(t)$ 
20     Store the maximum profit  $P_i(t)$  s.t.
21      $\phi_i(t) \geq \phi_i^{min}$  and  $w_i(t) \leq 1$  in  $P'_{ij}(t)$  and the
22     corresponding  $\phi_i(t)$  and  $\sigma_{ij}(t)$  in  $\phi'_{ij}(t)$  and
23      $\sigma'_{ij}(t)$ 
24   end
25    $j^* = \arg \max_j (P'_{ij}(t))$ 
26   if  $P'_{ij^*}(t) > P_i(t)$  then
27      $P_i(t) = P'_{ij^*}(t)$ ,  $\phi_i(t) = \phi'_{ij^*}(t)$ ,
28      $\sigma_{ij^*}(t) = \sigma'_{ij^*}(t)$  and the corresponding
29      $w_{ij^*}(t)$  values are updated
30   else
31     Reduce  $\sigma_{step}$ 
32   end
33 end
34 Calculate Jain's index  $J_i(t)$  and set  $\sigma_{step} = \sigma_{step}^{max}$ 
35 Repeat steps 14-27 while  $J_i(t) < J_i^{min}$ 
36 Select the maximum  $P'_{ij}(t)$  in 21 that reduces
    $|J_i(t) - J_i^{min}|$ 

```

PM is divided into three parts. In the first part, PM determines the resource allocation that maximizes BS i 's Overall Satisfaction $OS_i(t)$ (steps 1-13). Subsequently, using as input the resource allocation obtained from the first part of the algorithm, PM determines the resource allocation that maximizes $P_i(t)$, while satisfying $\phi_i(t) \geq \phi_i^{min}$ (steps 14-27). Finally, the profit maximizing resources allocation determined in the previous step is iteratively modified until the minimum overall user satisfaction and the minimum fairness constraints are satisfied (steps 28-30).

The first part of Algorithm 1 (steps 1-13) is a greedy algorithm that aims to determine the resources allocation (i.e. the value of $w_{ij}(t) \forall j \in \mathcal{U}_i$) that maximizes the OS_i . The algorithm initially computes the resources required to meet the maximum user satisfaction for each user. By substituting

(2) in (3), $w_{ij}(t)$ can be expressed as

$$w_{ij}(t) = \frac{1}{\varepsilon_{ij}(t)b_i\beta_{k_j d_j}} \left[r_{kd}\beta_{k_j d_j} + \log \left(\frac{\sigma_{ij}(t)(Q_{k_j}^{tg} - Q_{k_j}^{drop}) + Q_{k_j}^{drop}}{\alpha_{k_j d_j} Q_p(p_{k_j})} - \frac{\gamma_{k_j d_j}}{\alpha_{k_j d_j}} \right) \right] \quad (12)$$

Subsequently, the algorithm assigns resources iteratively to the user $j \in \mathcal{U}_i$ with the least resource requirements. Particularly, each user is allocated the resources calculated previously until the total available resources are depleted. In case of resources depletion, the resources allocated to the remaining users are reduced as long as the user satisfaction is above the minimum acceptable satisfaction value σ_{ij}^{min} or are set to 0 otherwise. With this, the overall user satisfaction is maximized, but fairness is not taken into account.

The second part of PM (steps 14-27) is a greedy profit maximization algorithm on the basis of the resources allocation resulted from the first part of the algorithm (steps 1-13). Specifically, for each user the user satisfaction is decreased (step 16) and increased (step 17) with σ_{step} (the resources allocation $w_{ij}(t)$ is calculated from (12)), and then the profit is calculated for both satisfaction values. Only $\sigma_{ij}(t)$ values that increase the profit are considered as feasible results. This procedure is repeated iteratively for each user and for different values of σ_{step} as long as the relative overall user satisfaction is above the minimum threshold. The resources allocation is updated with the distribution of resources that provides the maximum profit for a relative user satisfaction level above ϕ_i^{min} , that is, the optimal solution in a single iteration. Therefore, the resulting resources allocation of the second part of the PM algorithm converges to a local optimal solution of the profit maximization problem for a given minimum satisfaction, which at times can be a global maximum [24].

The last part of the PM algorithm (steps 28-30) introduces the fairness. Thus, starting from the resources allocation obtained in the second part of the algorithm, PM executes the same greedy process run in the second part (steps 14-27) but this time only solutions that converge to J_i^{min} are selected (i.e. with declining $|J_i(t) - J_i^{min}|$ values). This last part, and hence PM, is terminated when there are no alternative resource allocations that satisfy the relative overall satisfaction threshold and the convergence of $J_i(t)$ to J_i^{min} .

It should be noted that PM consists of three greedy iterative algorithms, which in each of their iterations make the optimal decision for a subproblem (e.g. maximize the OS_i or P_i by allocating resources to a single user at a time). However, it is probable that it converges to locally optimal solutions instead of the global optimum [24]. Thus, PM may not always perform optimally, which we will examine in Section VI.

Feasibility: The feasibility of our system model and Algorithm 1 require the monitoring of the user rate $r_j(t)$ for calculating the user satisfaction $\sigma_{ij}(t)$, as well as the calculation of the revenue $R_{ij}(t)$ for the diverse pricing we examine. In LTE-A, real-time monitoring of the User Equipment (UE) application layer data throughput performance is used for measuring the provided QoS [25]. Similarly in LTE-A, $R_{ij}(t)$ can be obtained in real time with a module such as the Policy

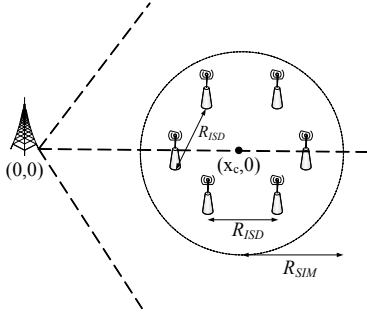


Fig. 1: Simulation scenario topology

and Charging Control (PCC), which can control the QoS, apply different charging models, and control usage monitoring to make dynamic policy decisions [26].

VI. PERFORMANCE EVALUATION

A. Scenario description and parameters

The scenario used for the performance evaluation consists of a cluster with 6 SCs deployed in the coverage area of a macrocell sector. The cluster is circular shaped and centred at location $c = (x_c, 0)$, as shown in the layout depicted in Fig. 1. Along simulations, x_c is randomly selected according to a uniform distribution with $x_c \in [100, 190]m$. The Inter-Site Distance (ISD) between two SCs equals $R_{ISD} = 50m$.

Users are uniformly distributed within a radius of $R_{SIM} = 75m$ from c , and the SP of each user is selected with equal probability among the SPs defined in Table II. As it can be observed in Table II, three services are considered, each one with two QoE classes $\mathcal{Q} = \{\text{Basic, Premium}\}$: Service 1 is a data-based charged service, and Services 2 and 3 are time-based charged services. Likewise, 3 different devices are considered, and the corresponding transmission rates associated to each SP, r_{kd} , are also included in Table II. Note that for each SP, r_{kd} is the transmission rate required to perceive the service's target QoE level, Q_k^{tg} . In the simulations, the transmission rate that results in a perceived QoE equal to Q_k^{drop} is set to $r_{kd}^{drop} = 0.7r_{kd}$ for all SPs. Therefore, when a user is served with a data rate below or equal to $0.7r_{kd}$, the connection is dropped. Moreover, v_k is selected randomly so as to have a price-based QoE component $Q_p(p_k) \in [0.8, 0.9]$ in (5), and

$$\left\{ \begin{array}{l} \alpha_{kd} = \frac{Q_k^{tg}}{Q_p(p_k)} - \gamma_{kd}, \\ \beta_{kd} = -\frac{1}{\Delta r_j^{drop}} \ln \left(\frac{Q_k^{drop} - \gamma_{kd} Q_p(p_k)}{Q_k^{tg} - \gamma_{kd} Q_p(p_k)} \right), \end{array} \right. \quad (13a, 13b)$$

where $\Delta r_j^{drop} = r_{kd} - r_{kd}^{drop}$, and $\gamma_{kd} = 1$, for all SPs π_k , devices $d \in \mathcal{D}$ and target and drop QoE levels $(Q_k^{tg}, Q_k^{drop}) = (3.5, 2.5)$ for Basic QoE class and $(Q_k^{tg}, Q_k^{drop}) = (4.5, 3.5)$ for Premium QoE class of all services.

Parameters used for the BSs, both eNBs and SCs, are listed in Table III. For the carrier bandwidth allocated to each tier, we adopted 3GPP LTE-A's channel models described in [27], and the antenna gains are set to 0 dB. For the cell selection, we associate the users to the BS with the highest SINR, as it is common practice in mobile networks [28].

TABLE II: Service Profiles' parameters

Service	QoE class	$\{r_{k1}, r_{k2}, r_{k3}\}$ (Mbps)	θ_k^t or θ_k^B
Service 1 (Data Based)	Basic	5.5	1.5€/GB
	Premium	7	2€/GB
Service 2 (Time Based)	Basic	{3.5, 4, 5}	4€/h
	Premium	{4, 4.5, 5.5}	7€/h
Service 3 (Time Based)	Basic	{4.5, 5.5, 6}	4€/h
	Premium	{5, 6, 7}	7€/h

TABLE III: BS parameters

Parameter	Macrocell	Small cell
c_i (€/sec)	$5 \cdot 10^{-5}$	$5 \cdot 10^{-5}$
h_i (MHz ⁻¹)	0.28	0.275
b_i (MHz)	20	20
Transmission Power (dBm)	43	30

For PM, the values for the change in user satisfaction are $\sigma_{step} = \{0.01, 0.05\}$. Moreover, the minimum acceptable satisfaction level for all algorithms is $\sigma_{ij}^{min} = 0.01$ (note that with null user satisfaction, i.e. $\sigma_{ij}(t) = 0$, the session is dropped). The results shown in the following subsections were acquired through Monte-Carlo simulations, where each simulation iteration examines a single network instance of a $T = 1 \text{ sec}$ duration. It should be noted that we assume perfect channel estimation, and hence we do not inspect the network and economic impact of imperfect channel estimations or rate fluctuations during consecutive instances.

B. Impact of fairness and Satisfaction constraints

Initially, we present how the fairness and relative overall satisfaction objectives (i.e. J^{min} and ϕ^{min}) affect the user satisfaction, the network performance and the MNO profit, when we apply PM on a system with $N_U = 80$ users. We also present the results generated by the optimal solution of the problem in (11) (henceforth labelled as Opt).

In Fig. 2 and 3 we show the expected user satisfaction of the data-based and time-based charged users, denoted by $E[\sigma_{ij}^B]$ and $E[\sigma_{ij}^t]$, respectively. Unserved users are not considered in the calculation of $E[\sigma_{ij}^t]$ and $E[\sigma_{ij}^B]$, since when users are not served the satisfaction of the user is $\sigma_{ij}(t) = 0$. We observe that for low to medium relative overall user satisfaction levels ϕ^{min} and no fairness constraint ($J^{min} = 0$) there is a substantial difference between the user satisfaction of data-based and time-based charged users (i.e. 521% and 69% higher mean satisfaction of data-based charged users ($E[\sigma_{ij}^B]$) than of time-based charged users ($E[\sigma_{ij}^t]$), when relative overall satisfaction is $\phi^{min} = 0$ and $\phi^{min} = 0.6$ respectively). As explained in section V, a BS i can maximize its profit by serving time-based charged users with the minimum acceptable satisfaction (thus with the minimum allocation of resources $w_{ij}(t)$). At the same time, the data-based charged users require to be served with high rates (and satisfaction) in order for BS i to gain high revenue. However, as the fairness requirement increases ($J^{min} \uparrow$), the difference between the mean satisfaction of data-based and time-based charged users ($E[\sigma_{ij}^B]$ and $E[\sigma_{ij}^t]$) decreases. In order to increase fairness ($J_i(t)$), the standard deviation of the satisfaction among the users in BS i must be decreased (as explained in Section V).

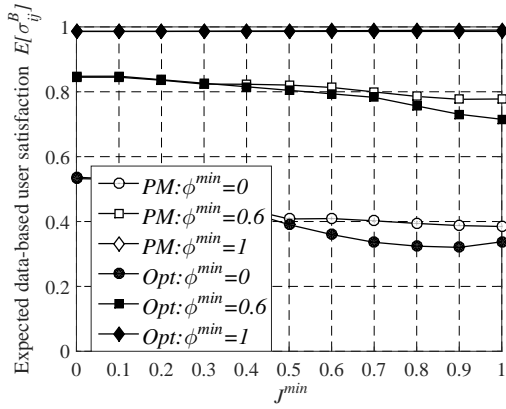


Fig. 2: Expected user satisfaction for data-based users ($E[\sigma_{ij}^B]$) versus target minimum Jain's index (J^{min}) for $N_U = 80$ users

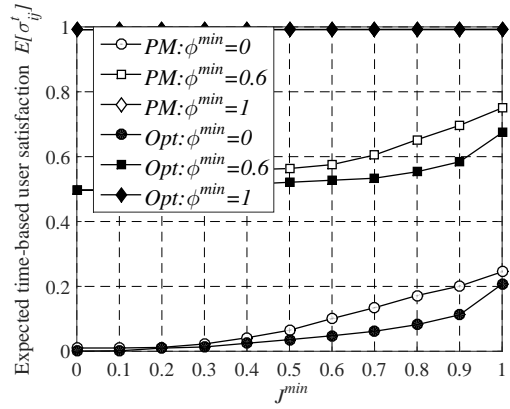


Fig. 3: Expected user satisfaction for time-based users ($E[\sigma_{ij}^t]$) versus target minimum Jain's index (J^{min}) for $N_U = 80$ users

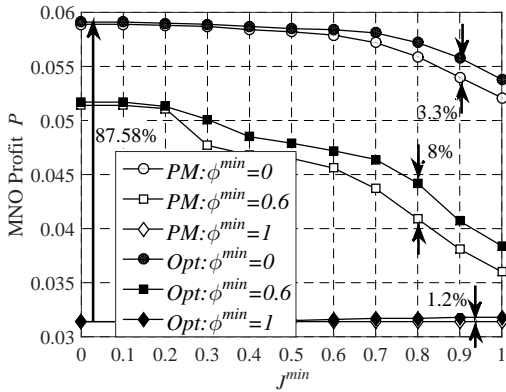


Fig. 4: Profit (P) versus target minimum Jain's index (J^{min}) for $N_U = 80$ users

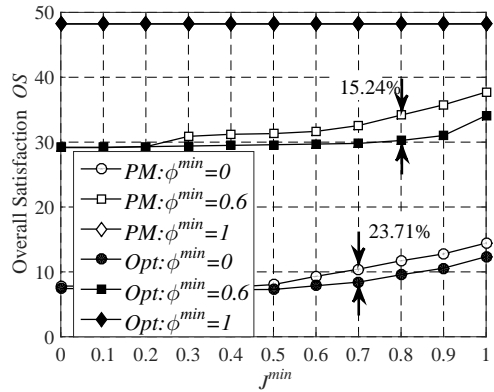


Fig. 5: Overall satisfaction (OS) versus target minimum Jain's index (J^{min}) for $N_U = 80$ users

This can only be achieved by increasing the resources allocated to time-based charged users at the expense of reducing the resources allocated to data-based charged users. Finally, we observe that for high relative overall satisfaction constraints (i.e. $\phi^{min} = 1$, or in other words, the overall user satisfaction must be always the maximum one, OS_i^{max}), the PM algorithm skips the second and the third parts of Algorithm 1, since the overall user satisfaction achieved in the first part of PM (step 13) can not be reduced. Therefore, users are either served with maximum satisfaction ($\sigma_{ij}(t) = 1$) or dropped². Regarding the optimal results, we observe that they show the same trend as PM, and small differences for medium to high J^{min} values.

TABLE IV: Percentage of Served Users for $N_U = 80$

PM	Data-Charged Users		Time-Charged Users	
ϕ^{min} / J^{min}	0	1	0	1
0	52.69%	54.13%	67.12%	68.74%
1	52.68%	52.68%	65.05%	65.05%

Table IV shows the percentage of users served according to

²As resources are allocated until their depletion, the last user served could receive resources that result in $0 < \sigma_{ij}(t) < 1$ (steps 6-7).

their service's pricing scheme, for target minimum Jain's index and relative overall users satisfaction (J^{min} and ϕ^{min}) equal to 0 and 1 (i.e. minimum and maximum levels) for $N_U = 80$ users, according to the PM results³. Note that J^{min} is the objective, but the actual value of the Jain's index $J_i(t)$ could not reach J^{min} if there are users with very low SINR levels. We observe that for a given number of users N_U , fairness and relative overall satisfaction objectives (J^{min} and ϕ^{min}) have little or no effect on the percentage of served users. However, we observe that more time-based than data-based charged users are served in all cases (i.e. [23, 27]% more time-based charged users). This occurs because in general time-based charged services have lower requirements in terms of transmission rate than data-based charged services. Therefore, although more time-based charged users are served (see Table IV), they are served with lower satisfaction, for low and medium relative overall satisfaction objectives (see Fig. 2 and 3).

Fig. 4 and Fig. 5 depict the MNO profit P and the total

³The corresponding optimal results are approximately the same with absolute differences below 1%, and hence they are omitted.

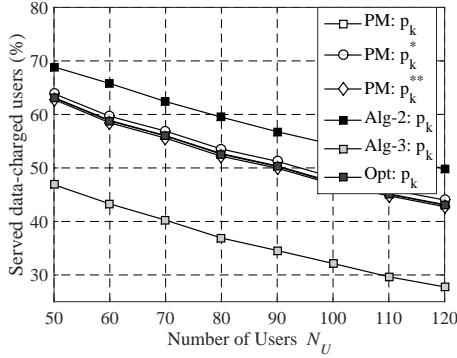


Fig. 6: Percentage of served Data-based charged users

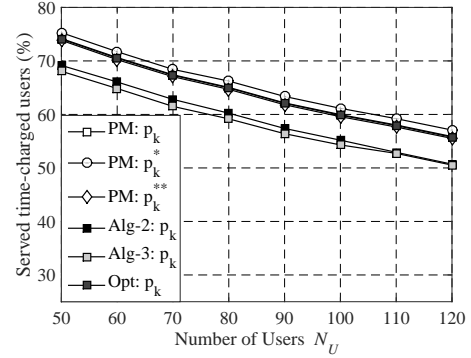


Fig. 7: Percentage of served Time-based charged users

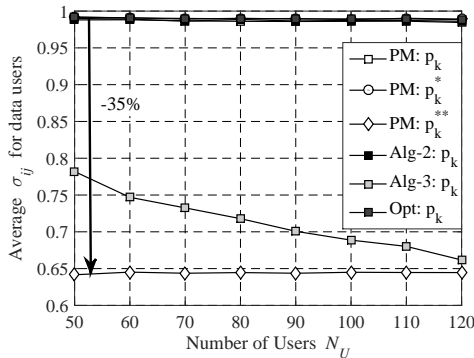


Fig. 8: Expected satisfaction for data-based charged users.

overall satisfaction OS for $N_U = 80$ users, respectively. In these two figures, we can observe the trade-off among the MNO profit P , the Overall Satisfaction (OS), and the objective minimum fairness (J^{min}). Particularly, profit is maximized when there are no satisfaction and fairness constraints, which also leads to the lowest Overall Satisfaction. This occurs because the lack of fairness and satisfaction objectives turns the PM algorithm into a *pure* profit maximization solution. Conversely, when the objective relative overall satisfaction (ϕ^{min}) is high, the profit is substantially lower than the maximum (i.e. [65, 87]% profit gain for $\phi^{min} = 0$ over $\phi^{min} = 1$). This decrease of the profit is caused by the fact that the proposed algorithm initially maximizes the satisfaction; then, it modifies the solution to increase the profit as long as the minimum relative satisfaction objective (ϕ^{min}) is satisfied. Therefore, the higher the value of ϕ^{min} , the closer to a *pure* satisfaction maximization algorithm PM is. Similar to Fig. 2 and 3, the optimal results follow the same trend as the PM results, with differences appearing for medium to high J^{min} values. Regarding the MNO profit, we see that PM performs close to the optimum in most cases. The highest differences are observed for the low and medium OS requirements (i.e. 3.3% and 8% difference for $\phi^{min} = 0$ and $\phi^{min} = 0.6$ respectively), whereas there are minuscule differences when the OS must be maximized (i.e. $\phi^{min} = 1$).

From the results in Fig. 2-5, it can be concluded that when

there are two pricing schemes (data-based and time-based charging), the profit can be maximized only when data-based charged users are prioritized over the time-based charged users in terms of average user satisfaction (which leads to low fairness). At the same time, serving more time-based charged users allows the MNO to gain revenue with a low cost. Finally, the trade-off between Profit and Overall Satisfaction shows that the increase of one of them implies the decrease of the other, as specified in the analysis in Section V.

C. Comparison with SoA algorithms and impact of pricing

In the following (i.e. Fig. 6-10 and Table V), we compare PM with two algorithms referred to as Alg-2 [9] and Alg-3 [17], and the optimal solution of (11) (i.e. Opt).

Alg-2 [9] is an iterative resources allocation algorithm that maximizes the user QoE. In each iteration, Alg-2 allocates enough resources to satisfy a single user, starting from the user with the highest spectral efficiency towards the user with the lowest spectral efficiency.

Alg-3 is an iterative resources allocation algorithm proposed in [17]. In the scenario proposed in [17], a MNO offers wireless video broadcasting services, and aims to maximize his profit. Particularly, the MNO broadcasts a set of video contents (e.g. tv channels) to different users groups. Each user has her own utility function, which is defined as the perceived QoE minus the charge for the service. In order to maximize the profit, Alg-3 increases the rate of a single content until either all contents are served with their “*ideal rate*”⁴ or the available bandwidth is fully utilized. The content whose rate will be increased is the one that maximizes the marginal MNO profit, provided that it is non-zero.

In order to compare Alg-3 with PM, we assume that each user demands a content, which is unique to herself. The “*ideal rate*” of a broadcast content described in Alg-3 corresponds to the user’s SP-device pair required transmission rate r_{kd} , in our system model. Therefore, Alg-3 has to broadcast N_U unique contents, whose rate requirements are defined by each user’s SP-device pair. Hence, the broadcasting resources allocation problem is transformed into the typical resources allocation

⁴In [17], “*ideal rate*” is defined as the rate that satisfies perfectly all the users that share a particular content.

problem, where each user's rate (i.e. QoE) is decided separately.

It should be noted that we do not concur with pure profit maximizing policies that result in low quality service provision, as presented previously. To that end, the results provided in the following for PM and Opt are obtained with strict fairness and satisfaction objectives, that is $(J^{min}, \phi^{min}) = (1, 1)$, which maximize the OS. In order to study the price variations on PM, two additional price values have been defined with respect to the price stated in Table II: $p_k^* = 0.5p_k$ and $p_k^{**} = 1.5p_k$. However, initially we only compare PM, Alg-2, Alg-3 and the optimal results with price equal to p_k .

Comparison of PM, Alg-2, Alg-3 and Opt: Fig. 6 and 7 show the percentage of served data-based and time-based charged users respectively, whereas Fig. 8 and Table V show the expected satisfaction for data-based and time-based charged users ($E[\sigma_{ij}^B]$ and $E[\sigma_{ij}^t]$ respectively⁵). Initially, only the price p_k is considered. As expected, we observe in Fig. 6 and 7 that the percentage of served users decreases as the total number of users N_U is increased, for both data-based and time-based users. This fact is caused by the network congestion and, consequently, by the lack of resources to serve all the users.

With Alg-2, both the percentage of served users (Fig. 6 and 7) and the expected user satisfaction for data-based charged users ($E[\sigma_{ij}^B]$ in Fig. 8) and time-based charged users (first row of Table V) is almost the same. This occurs because resources allocation algorithms that are exclusively aimed to QoE maximization (like Alg-2) do not prioritize services based on their pricing scheme. Conversely, with Alg-3 substantially more time-based charged users are served ([45, 81]% more time-based than data-based charged users are served when comparing Fig. 6 and Fig. 7), but with significantly lower expected user satisfaction (in line with PM results when $\phi^{min} = 0$ - see Fig. 2 and 3). This is explained by the fact that Alg-3 is aimed to maximize the profit. If the BS allocates to time-based users the minimum amount of resources to avoid their connection being dropped: i) resources are better distributed among users; ii) dropping is reduced; and iii) the number of served time-based users and the profit are increased. Similarly to Alg-3, it can be observed in Fig. 6 and Fig. 7 that PM always serves a larger percentage of time-based charged users, and this difference increases with the total number of users N_U ([17, 29]% more time than data-based charged users are served). This can be explained by the fact that the time-based charged users have lower rate requirements on average than the data-based charged users due to the use of different devices, as it can be seen in Table II. Therefore, time-based charged users with low rate requirements are prioritized over data-based charged users in order to achieve the high user satisfaction performance observed in Table V.

Summing up, given a particular relative overall satisfaction constraint (ϕ^{min}) and two charging schemes (i.e. data and time-based charging), the profit will be maximized by serving as many time-based charged users as possible. Additionally, if the objective of the MNO is the maximization of the profit

TABLE V: Expected satisfaction for time-based charged users ($E[\sigma_{ij}^t]$)

Time-Charged Users				
Price level	PM	Alg-2	Alg-3	Opt
p_k	0.9909	0.9901	0.01	0.9929
p_k^*	0.9903	-	-	-
p_k^{**}	0.6427	-	-	-

without any fairness and satisfaction constraint (i.e. PM with $(J^{min}, \phi^{min}) = (0, 0)$ and Alg-3), the data-based charged users will have higher expected satisfaction than the time-based charged users.

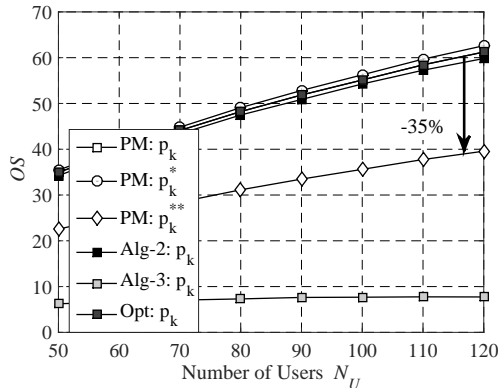
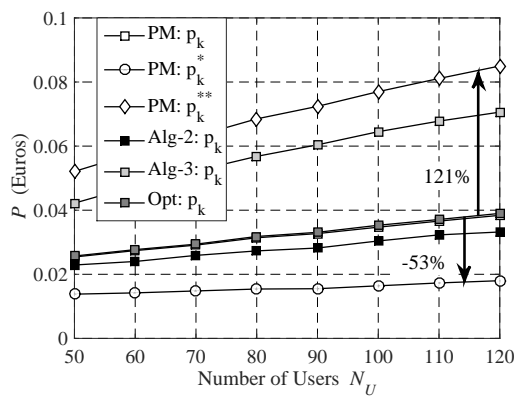
Fig. 9 shows the comparison of the system's overall user satisfaction OS . As expected, Alg-2 and PM offer the highest OS , whereas for Alg-3 OS is kept low and slightly increasing with the number of users N_U . As mentioned earlier, Alg-2 sorts the users according to their spectral efficiency, and then allocates the resources until they are exhausted. This means that Alg-2 will first serve the users with the highest spectral efficiency regardless of their service's requirements. Conversely, PM finds iteratively the user with least resource requirements $w_{ij}(t)$ and serves her with the maximum satisfaction. In Fig. 9, the difference between PM and Alg-2 in terms of Overall Satisfaction (OS) is within the range [1.14, 2.36]%. Alg-2 performs well when there is a single service with a single rate requirement. However, in a scenario with heterogeneous traffic as well as diverse pricing, a more elaborate algorithm such as PM is required in order to serve the users with even higher satisfaction, while gaining large MNO profit.

Fig. 10 presents the comparison of the MNO profit for the PM, Alg-2, Alg-3 and Opt. As expected, we observe that Alg-3 gains the highest profit, which increases with the number of users N_U . Due to the use of strict OS and fairness constraints, PM achieves lower profit than Alg-3. Nevertheless, it outperforms notably Alg-2 (i.e. [10.92, 15.96]%), even though they share an almost equal OS performance. Regarding the results from the optimal solution of (11), we observe that in all of the Fig. 6-10 and Table V PM and Opt have the same performance. A small difference is visible only in Fig. 10, where we can see that Opt generates more profit than PM by a small margin (i.e. [1.18, 1.73]% profit gain for Opt over PM). This gain in profit corresponds to an even smaller loss in OS performance (i.e. < 1% smaller OS than PM).

In light of the results presented above, PM is a profit maximizing resource allocation algorithm, which guarantees similar QoE performance results to the ones achieved with Alg-2 (a QoE maximizing algorithm), thanks to the satisfaction and fairness constraints (in Fig. 6-10 $J^{min} = \phi^{min} = 1$). However, the good performance in terms of QoE is not translated in a decrease of the profit. Specifically, PM achieves higher profit than Alg-2 for all scenarios. Finally, when strict satisfaction and fairness constraints are applied, PM provides approximately optimal results for the solution of problem (11).

Impact of price level: Results of Fig. 6-10 have been so far analysed with the price level p_k stated in Table II. Now, we examine in the same figures how the price levels affect the PM algorithm in terms of users satisfaction, the percentage of served users, the MNO profit and the Overall Satisfaction.

⁵The results in Table V are included as a table because they are constant for $N_U \in [50, 120]$ users.

Fig. 9: Overall Satisfaction (OS)Fig. 10: MNO profit P

In order to do this, the price values p_k , $p_k^* = 0.75p_k$ and $p_k^{**} = 1.5p_k$ are considered.

Taking the above into consideration as well as the corresponding results, it can be observed that the decrease of the price ($p_k \rightarrow p_k^*$) results in an increase of the percentage of served users (see Fig. 6 and 7) and the mean users' satisfaction (see Fig. 8 and Table V). Note that the reduction of the price improves the perception of the user thanks to the price-based component of the QoE (as shown in expression (5)). The opposite behaviour is observed when the price is increased ($p_k \rightarrow p_k^{**}$). Consequently, as we observe in Fig. 9 and 10 the higher MNO profit gained by the higher price p_k^{**} (a gain in the range of 105-121%), comes at the expense of lower expected user satisfaction $E[\sigma_{ij}^B]$, $E[\sigma_{ij}^t]$ (with a decrease around 35%), and lower Overall Satisfaction OS (around 35%). Conversely, when p_k is decreased to $p_k^* = 0.75p_k$, there is a slight increase in the percentage of served data-based and time-based charged users (in the range of 1.1-2% for data-based charged users and 1.4-2.1% for time-based charged users), and in the Overall Satisfaction (1.4 – 2.2% gain). However, the improvement in the percentage of served users and in the overall satisfaction is achieved at the cost of a significant profit loss (around 45 – 53% drop of the profit). Thus, a decrease of the price improves the perception of the users (i.e. higher satisfaction) while it reduces the MNO profit, and vice versa.

Summary of results:

- The trade-off between Profit and Overall Satisfaction shows that the increase of one of them implies the decrease of the other.
- Given a particular minimum relative overall satisfaction (ϕ^{min}) and two charging schemes (i.e. data and time-based charging), profit (P) will be maximized by serving as many time-based charged users as possible. Additionally, if the MNO maximizes the profit without any fairness and satisfaction constraints, the data-based charged users will have higher expected satisfaction than the time-based charged users.
- PM guarantees similar QoE performance results to the ones achieved with Alg-2 (a QoE maximizing algorithm), thanks to the satisfaction and fairness constraints, while achieving higher profit than Alg-2 for all scenarios.
- When fairness and relative overall satisfaction are forced to be maximum ($J^{min} = \phi^{min} = 1$), PM provides approximately optimal results for the solution of problem (11).
- Decreasing the price improves the users' perception while it reduces the MNO profit, and vice versa.

VII. CONCLUSIONS

In this paper, we study the resource allocation problem in a HetNet characterized by traffic heterogeneity and diverse pricing (i.e. two different schemes and various service prices). The objective of this paper is the maximization of the MNO's profit while providing high QoE to the users. With this objective, we propose a low complexity heuristic, greedy resource allocation algorithm, namely PM, which maximizes the profit of the MNO as long as fairness and overall users satisfaction (a QoE performance metric) levels are kept above specific thresholds. We highlight the overall satisfaction-fairness-profit trade-off and show that the proposed resources allocation algorithm achieves similar results in terms of users' satisfaction when compared to QoE maximization algorithms (e.g. Alg-2), while outperforming them in terms of profit gains. Moreover, we show that our algorithm approximates the optimal solution of the profit maximizing problem. With the analysis of two pricing schemes (one for data-based charged users and another for time-based charged users), we further show that *pure* profit maximizing algorithms prioritize the satisfaction of data-based charged users over time-based charged users, and we show that PM with strict fairness and satisfaction constraints smooths this effect. Finally, we shed light on how the changes in price levels can improve or worsen the user experience, and the corresponding effects of these changes on the network performance and the MNO profit.

ACKNOWLEDGMENT

This work was supported by the Spanish Ministry of Economy and the FEDER regional development fund under SINERGIA project (TEC2015-71303-R), AGAUR (2017 SGR 60) and AGAUR (2017 SGR 891), and the research projects SPOT5G (TEC2017-87456-P), SEMIoTICS (780315).

REFERENCES

- [1] Cisco, San Jose, CA, USA white paper: *The Zettabyte Era: Trends and Analysis*, June 2017
- [2] Appier, *Cross-Screen User Behavior Insights, Asia 1H 2015*, Taipei, Taiwan, Research Report, Aug 2015
- [3] Ofcom (Jan 2013), *Measuring mobile voice and data quality of experience*.
- [4] A. Gotsis, S. Stefanatos and A. Alexiou, "UltraDense Networks: The New Wireless Frontier for Enabling 5G Access," in IEEE Vehicular Technology Magazine, vol. 11, no. 2, pp. 71-78, June 2016
- [5] Mobile Squared, *Annual Market Review 2014 and PRS market outlook 2015*, Bracknell, UK, Jul. 2015
- [6] R. Imran, M. Odeh, N. Zorba and C. Verikoukis, "Quality of Experience for Spatial Cognitive Systems within Multiple Antenna Scenarios," in IEEE Transactions on Wireless Communications, vol. 12, no. 8, pp. 4153-4161, August 2013.
- [7] Reichl, P.; Maille, P.; Zwickl, P.; Sackl, A., "On the fixpoint problem of QoE-based charging," in Performance Evaluation Methodologies and Tools (VALUETOOLS), 2012 6th International Conference on , pp.235-242, 9-12 Oct. 2012
- [8] B. C. Chung, K. Lee and D. H. Cho, "Proportional Fair Energy-Efficient Resource Allocation in Energy-Harvesting-Based Wireless Networks," in IEEE Systems Journal, vol. PP, no. 99, pp. 1-11, 2017.
- [9] V. F. Monteiro, D. A. Sousa, T. F. Maciel, F. R. M. Lima, E. B. Rodrigues and F. R. P. Cavalcanti, "Radio resource allocation framework for quality of experience optimization in wireless networks," in IEEE Network, vol. 29, no. 6, pp. 33-39, Nov.-Dec. 2015.
- [10] Y. H. Cho, H. Kim, S. H. Lee and H. S. Lee, "A QoE-Aware Proportional Fair Resource Allocation for Multi-Cell OFDMA Networks," in IEEE Communications Letters, vol. 19, no. 1, pp. 82-85, Jan. 2015.
- [11] J. Li, M. Peng, A. Cheng, Y. Yu and C. Wang, "Resource Allocation Optimization for Delay-Sensitive Traffic in Fronthaul Constrained Cloud Radio Access Networks," in IEEE Systems Journal, vol. 11, no. 4, pp. 2267-2278, Dec. 2017.
- [12] S. Colonnese, F. Cuomo, L. Chiaraviglio, V. Salvatore, T. Melodia and I. Rubin, "CLEVER: a Cooperative and Cross-Layer Approach to Video Streaming in HetNets," in IEEE Transactions on Mobile Computing.
- [13] I. Ayad, Y. Im, E. Keller, and S. Ha, "A practical evaluation of rate adaptation algorithms in http-based adaptive streaming," *Computer Networks*, vol. 133, pp. 90-103, 2018.
- [14] Z. Yan, C. Westphal, X. Wang and C. W. Chen, "Service provisioning and profit maximization in network-assisted adaptive HTTP streaming," 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, 2015, pp. 2786-2790.
- [15] B. Al-Manthari, H. Hassanein, N. A. Ali and N. Nasser, "Fair Class-Based Downlink Scheduling with Revenue Considerations in Next Generation Broadband Wireless Access Systems," in IEEE Transactions on Mobile Computing, vol. 8, no. 6, pp. 721-734, June 2009
- [16] T. Taleb, N. Nasser and M. P. Anastasopoulos, "An Auction-Based Pareto-Optimal Strategy for Dynamic and Fair Allotment of Resources in Wireless Mobile Networks," in IEEE Transactions on Vehicular Technology, vol. 60, no. 9, pp. 4587-4597, Nov. 2011
- [17] W. Ji, B. W. Chen, Y. Chen and S. Y. Kung, "Profit Improvement in Wireless Video Broadcasting System: A Marginal Principle Approach," in IEEE Transactions on Mobile Computing, vol. 14, no. 8, pp. 1659-1671, Aug. 1 2015.
- [18] A. R. Elsharif, W. P. Chen, A. Ito and Z. Ding, "Resource Allocation and Inter-Cell Interference Management for Dual-Access Small Cells," in IEEE Journal on Selected Areas in Communications, vol. 33, no. 6, pp. 1082-1096, June 2015
- [19] P. Trakas, F. Adelantado, N. Zorba and C. Verikoukis, "A QoE-Aware Joint Resource Allocation and Dynamic Pricing Algorithm for Heterogeneous Networks," GLOBECOM 2017 - 2017 IEEE Global Communications Conference, Singapore, 2017, pp. 1-6.
- [20] Abdul Rehman ; Kai Zeng and Zhou Wang, "Display device-adapted video quality-of-experience assessment," Proc. SPIE 9394, Human Vision and Electronic Imaging XX, 939406 (March 17, 2015)
- [21] R Jain, DM Chiu, WR Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer system". (DigitalEquipment Corporation, Hudson, 1984)
- [22] Iosifidis, G.; Lin Gao; Jianwei Huang; Tassioulas, L., "A Double-Auction Mechanism for Mobile Data-Offloading Markets," in Networking, IEEE/ACM Transactions on , vol.23, no.5, pp.1634-1647, Oct. 2015
- [23] Lee, J., Leyffer, S.; "Mixed Integer Nonlinear Programming," Springer, New York (2012)
- [24] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. 2009. "Introduction to Algorithms," Third Edition (3rd ed.). The MIT Press.
- [25] *User Equipment (UE) application layer data throughput performance*, document 3GPP TR 37.901 v15.0.0, Mar. 2018
- [26] *Policy and charging control architecture*, document 3GPP TR 23.203 v15.2.0, Mar. 2018
- [27] *Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects*, document 3GPP TR 36.814 v9.0.0, Mar. 2010
- [28] J. G. Andrews, "Seven ways that HetNets are a cellular paradigm shift," in IEEE Communications Magazine, vol. 51, no. 3, pp. 136-144, March 2013.



Panagiotis Trakas (S'11) received the M.Eng. degree in electrical and computer engineering from the Aristotle University of Thessaloniki in 2012. He is currently pursuing the Ph.D degree at the Signal Theory and Communications Department (TSC) of Universitat Politècnica de Catalunya BarcelonaTech (UPC). His research has been concerned with outsourced traffic offloading and HetNet Radio Resource Management in LTE-A and 5G environments.



Ferran Adelantado (M'07) received his engineering degree in telecommunications (2001) and his Ph.D. degree in telecommunications (2007) from Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, and his B.Sc. in business sciences (2012) from Universitat Oberta de Catalunya (UOC). Currently, he is an associate professor at UOC and a researcher at the Wireless Networks Research Group (WINE). His research interests are wireless networks, particularly 5G, LPWAN, and IoT technologies.



Christos Verikoukis (SM'07) received the Ph.D. degree from the Technical University of Catalonia, Barcelona, Spain, in 2000, in the area of wireless communications.

He is currently a Fellow Researcher with Telecommunications Technological Centre of Catalonia (CTTC/CERCA), Spain, and an Adjunct Professor with University of Barcelona. He has authored more than 122 journal papers and more than 180 conference papers. He has co-authored three books, 14 chapters, and three patents. He has participated in more than 30 competitive projects and has served as the Principal Investigator of national projects. He has supervised 15 Ph.D. students and five postdoctoral researchers.

Dr. Verikoukis is the Chair of the IEEE Communications Society's Communications Systems Integration and Modeling Technical Committee. He is the Associate Editor-in-Chief of the IEEE Network Letters. He received the Best Paper Award at the 2011 IEEE International Communications Conference, the IEEE GLOBECOM 2014 and 2015, the 2016 European Conference on Networks and Communications, and the EURASIP 2013 Best Paper Award of the Journal on Advances in Signal Processing. He has also received the Best Demo Award in IEEE CAMAD 2018.