A Quality of Experience-aware association algorithm for 5G Heterogeneous Networks

Panagiotis Trakas*, Ferran Adelantado*, Nizar Zorba†, Christos Verikoukis‡
*Open University of Catalonia (UOC), Barcelona, Spain
†Qatar University, Doha, Qatar
‡Centre Tecnologic de Telecomunicacions de Catalunya (CTTC), Barcelona, Spain
Email: ptrakas@uoc.edu, ferranadelantado@uoc.edu, nizarz@qu.edu.qa, cveri@cttc.es

***Abstract*—The rise of Over The Top (OTT) 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 5G networks in 2020. At the same time, the rise of OTT 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 a 5G investment. To this end, we consider a 5G Heterogeneous Network (HetNet) deployment where MNOs use a QoE-based charging scheme. We propose a heuristic, QoE-aware user association algorithm 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 a baseline SINR-based scheme.**

***Index Terms*—Association, Traffic Heterogeneity, QoE, 5G.**

I. Introduction

The exponential increase in mobile data traffic is expected to continue for the next few years. This traffic growth is mainly the result of the emergence of OTT content providers, and the continuous appearance of new applications. These new applications are described by various QoS requirements, which along with the use of multiple devices per user [1] increase the heterogeneity of the traffic demand. As a response to these challenges, the mobile industry has already established the requirements of 5G networks, for their deployment in 2020 [2]. The aforementioned rise of OTT providers has resulted in a growth of the traffic served by the MNOs. Although it may seem contradictory, the described boost of the traffic has increased the OTT profits while, simultaneously, has diminished the MNOs' revenues [3]. This occurs because the usage of the MNO basic services (voice and messaging) has been gradually replaced by their OTT counterparts. Moreover, the MNOs' data service prices have been decreasing over the years, due to the market competition. However, despite these conditions, the MNOs must provide seamless connectivity and high QoE to their users, which is one of the key elements in the design of 5G networks [2].

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 [4]. This means that small degradations in the received QoS can impact significantly the perception of QoE. Yet, QoE is influenced by other factors such as pricing or device characteristics [5]. In this context, it is necessary to design network functionalities adapted to the new requirements, such as QoE-aware Radio Resources Management strategies or new cell association algorithms, and always trying to maximize the profit (to compensate the diminished MNOs' revenues and the increasing deployment investment).

Most works in the user association literature focus solely on the maximization of the served traffic under particular QoS/QoE constraints, without considering the economic effects of their proposal. The user-cell association problem in small cell (SC) networks is solved with the use of matching games with externalities in [6] and [7]. In [6], the proposed algorithm combines the users' context with predictions on QoE, in order to minimize the relative error between the actual and the predicted QoE. In [7], the objective is the maximization of the users' context-aware utility function, which is defined by the service's QoS metrics. The proposed algorithms in these works achieve lower bandwidth utilization [6], and higher user rates [7], when compared to traditional association schemes.

The user association problem in a SC network operating in the 60-GHz band is studied in [8]. The authors' objective is the minimization of the maximum resource utilization in the system. A distributed association algorithm is proposed, based on Lagrangian duality theory. The proposed algorithm is time efficient and converges asymptotically to the optimal values, utilizing substantially less resources compared to conventional association algorithms. The problem of cell association and transmission scheduling in a two-tier network is studied in [9]. The authors propose three approximation algorithms and a scheduling scheme for the sub-optimal solution of the association problem, and the minimization of the delay.

In this paper we study the user association problem aiming to maximize the MNO profit, while offering high QoE to the users. We consider HetNets composed of macrocell base stations (BSs) operating in the sub-6GHz microwave (μWave) band and SCs operating in the mmWave band, dynamic traffic described by numerous QoS/QoE demands, and diverse pricing. In order to address the challenges of traffic heterogeneity we propose a heuristic, QoE-aware user association algorithm to maximize both QoE and MNO profit.

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 association 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 operating in the μWave band and a set of SCs operating in the mmWave band, all of them deployed by a single MNO. We denote this set of BSs, both macrocells and SCs, as $\mathcal{B}$=$\{1, 2, \ldots, N_B\}$, 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 mmWave SC deployment has been extensively addressed in the literature and proposed as a pivotal solution in 5G for two main reasons. First, the bandwidth availability in mmWave bands is higher than in the μWave bands, thereby alleviating the spectrum scarcity problem; second, thanks to the limited interference realized by the use of highly directional antennas, very dense deployments are feasible, thus enhancing the network spectral efficiency [15].

The MNO serves a set of users $\mathcal{U} = \{1, 2, \ldots, N_U\}$, 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 QoE as the target Key Performance Indicator (KPI) in the design of 5G networks. 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 $\mathcal{S} = \{s : s = 1 \ldots S\}$ and the set of QoE classes as $\mathcal{Q} = \{q : q = 1 \ldots Q\}$ ($\mathcal{Q}$ is assumed to be a discrete and finite set), a generic service profile 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). The general expression of p_k for a time period T can be expressed as

$$p_k = \begin{cases} T\theta_k^t & \text{if } s_k \text{ is time-based charged} \\ \frac{T\cdot r}{8}\theta_k^B & \text{otherwise,} \end{cases} \tag{1}$$

where r (in Mbps) is the transmission rate of the user. 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 MOS scale). 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 is required for a user using a video service in a small-size screen smartphone than for the same user with a large screen tablet [10]. Therefore, we define the set of devices as $\mathcal{D} = \{d : d = 1 \ldots D\}$, and the mapping function that links the device-SP pair with the required transmission rate, r_{kd}, as $f : (\pi_k, d) \to r_{kd}$. 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).

Note that the QoE classes differentiate the perceived quality by offering different maximum rate values for the same service (e.g. SD and HD video). That is, the user can opt between Q quality levels for every service and create a *user profile*, since each service may be of different importance to the user (e.g. preference for high browsing speed, but SD video).

Based on the definitions stated above, it is clear that the satisfaction of users is tightly coupled with the perceived QoE. Specifically, if satisfaction of user j served by BS i, namely $\sigma_{ij}(t)$ is defined within the interval [0,1], when $Q_j^{kd}(t) = Q_k^{drop}$, 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 [5], the satisfaction can be defined as

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

III. MNO Objectives

MNOs have a two-fold objective. First, they must offer the users the QoE agreed in the SP. Second, the network must be managed so as to maximize their economic profit. In the following, the analyses of the QoE and the profit are detailed.

A. User QoE

Based on the analysis described in [5], the perceived QoE $Q_j^{kd}(t)$ can be divided into two components: the QoS-based component ($\widehat{Q}_j^{kd}(t)$) and the price-based component ($Q_p(p_k)$).

$$Q_j^{kd}(t) = \widehat{Q}_j^{kd}(t) \cdot Q_p(p_k) \tag{3}$$

The QoS-based component, $\widehat{Q}_j^{kd}(t) \in [1, 5]$ (in the MOS scale), shows the effect of QoS level on QoE. In the literature, the connection between QoE and QoS is usually modelled according to the IQX hypothesis [4], which defines it as an exponential relationship. Using the transmission rate $r_j(t)$

as the reference QoS metric, and according to the IQX hypothesis, we can express $\widehat{Q}_j^{kd}(t)$ as

$$\widehat{Q}_j^{kd}(t) = \alpha_{kd} e^{-\beta_{kd} \Delta r_j(t)} + \gamma_{kd}, \tag{4}$$

where $\Delta r_j(t) = r_{kd} - r_j(t)$, and $\alpha_{kd} > 0$, $\gamma_{kd} > 0$ (both in the MOS scale), $\beta_{kd} > 0$ (in sec/bit) are SP-device dependent constants. Regarding the price-based component, it captures how the perception of the quality improves (or worsens) as the price falls (or rises). As in [5], $Q_p(p_k)$ is modelled as

$$Q_p(p_k) = 1 - v_k p_k, \tag{5}$$

where $v_k > 0$ is an adjusting factor measured in €$^{-1}$. 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 stated in (3). That is, the more a user pays for a service, the higher her expectations on the received quality are.

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). Similarly to the transmission rate proportional fairness described in [11], resource allocation algorithms based on QoE are intended to guarantee fairness among users in terms of QoE. Thus, for the set of users served by BS i, QoE fairness is achieved if $\sigma_{ij}(t) = \sigma_{in}(t)$ for any $j, n \in \mathcal{U}_i$ with service profiles $\pi_{k_j} = (s_{k_j}, q_{k_j}, p_{k_j})$ and $\pi_{k_n} = (s_{k_n}, q_{k_n}, p_{k_n})$, and devices $d_j, d_n \in \mathcal{D}$, respectively. When b_i is not enough to offer $\sigma_{ij}(t) = 1, \forall j \in \mathcal{U}_i$, all $\sigma_{ij}(t)$ are decreased (by reducing $w_{ij}(t)$) until $\sigma_{ij}(t) = \sigma_{in}(t)$ $\forall j, n \in \mathcal{U}_i$. If $\sigma_{ij}(t) = \sigma_{in}(t)$ is only true for the trivial solution (i.e. $\sigma_{ij}(t) = 0$), users with $\sigma_{ij}(t) = 0$ are dropped (i.e. $w_{ij}(t) = 0$).

B. MNO Profit

The objective of the MNO is the maximization of the profit while satisfying the QoE required by the users. Specifically, the total profit $P(t)$ 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 [12], $P_i(t)$ is expressed as the revenue obtained from the traffic served at time t, $R_i(t)$, minus the cost incurred when serving the traffic. In turn, the cost can be decoupled into the bandwidth utilization cost, $CB_i(t)$, and the fixed cost, $CF_i(t)$. Therefore,

$$P(t) = \sum_{i \in \mathcal{B}} P_i(t) = \sum_{i \in \mathcal{B}} (R_i(t) - CB_i(t) - CF_i), \text{ [€]}, \tag{6}$$

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 $R_{ij}(t)$ is the revenue paid by user j when connected to BS i at time period t. In [5], the authors propose a QoE-based charging policy where the price (and revenue) is reduced when satisfaction is below 1. In other words, the price of the service is reduced when $\sigma_{ij}(t)$<1. Thus, based on [5], for a user j served by BS i and with a SP π_{k_j}, the BS i revenue is given by

$$R_{ij}(t) = \sigma_{ij}(t) \cdot p_{k_j} \tag{7}$$

With regard to $CB_i(t)$, it is a convex and increasing exponential function of the total resources used by BS i, $w_i(t) = \sum_{j \in \mathcal{U}_i} w_{ij}(t)$ [12], and it can be expressed as

$$CB_i(t) = c_i e^{h_i w_i(t) b_i}, \tag{8}$$

where c_i (in €) 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 (7) and (8) into (6), and denoting the SP of a generic user j as π_{k_j}, the profit of BS i at time period t when $Q_j^{k_j d_j}(t) \in (Q_{k_j}^{drop}, Q_{k_j}^{tg}]$ is given by

$$P_i(t) = \sum_{j \in \mathcal{U}_i} \left[\frac{\left(\alpha_{k_j d_j} e^{-\beta_{k_j d_j} \Delta r_j(t)} + \gamma_{k_j d_j}\right)(1 - v_{k_j} p_{k_j})}{Q_{k_j}^{tg} - Q_{k_j}^{drop}} - \frac{Q_{k_j}^{drop}}{Q_{k_j}^{tg} - Q_{k_j}^{drop}} \right] p_{k_j} - c_i e^{h_i \sum_{j \in \mathcal{U}_i} w_{ij}(t) b_i} - CF_i \tag{9}$$

It can be seen in (9) that the profit is impacted by multifarious factors, such as the perceived QoE (which in turn depends on multiple factors), the cost, the usage of the radio resources, etc. In the subsequent Sections the optimization of the MNO profit is stated and a low complexity algorithm is proposed, also taking into account the users' QoE.

IV. Profit Optimization

As explained in the previous Section, the MNO aims to maximize the profit $P(t)$ 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. Let us define the association of user j to BS i at time period t as $x_{ij}(t)$, where $x_{ij}(t) = 1$ if user j is served by BS i and $x_{ij}(t) = 0$ otherwise. In order to capture the impact of dynamic traffic demand and channel conditions, the profit is maximized for a period of N_S subframes. The association problem is formulated as

$$\max \sum_{t=1}^{N_S} P(t) = \sum_{t=1}^{N_S} \sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{U}} x_{ij}(t) \sigma_{ij}(t) p_{k_j} - \sum_{t=1}^{N_S} \sum_{i \in \mathcal{B}} c_i e^{h_i b_i \sum_{j \in \mathcal{U}} x_{ij}(t) w_{ij}(t)} - \sum_{i \in \mathcal{B}} CF_i, \tag{10a}$$

$$s.t. \quad \sum_{i \in \mathcal{B}} x_{ij} \leq 1, \ \forall i \in \mathcal{B}, \ \forall j \in \mathcal{U}, \tag{10b}$$

$$w_i \in [0, 1], \ \forall i \in \mathcal{B}, \tag{10c}$$

$$\sigma_{ij} = \sigma_{in}, \ \forall i \in \mathcal{B}, \ \forall j, n \in \mathcal{U}_i, \tag{10d}$$

In the optimization problem, users cannot be connected to more than a single BS (10b), the maximum bandwidth allocated by BS i is b_i, that is $\sum_{j \in \mathcal{U}_i} w_{ij}(t) = w_i(t) \leq 1$ (10c), and QoE fairness must be guaranteed (10d). As this maximization problem cannot be solved in polynomial time (i.e. it is NP-hard), in the following section a heuristic, QoE-aware association algorithm is proposed.

V. QoE-Aware Association Algorithm

The proposed low complexity algorithm, $O(n^2)$, which is presented in Algorithm 1, takes as input the user's SP-device pair (π_{k_j}, d_j), as well as the BSs' state in the previous subframe (i.e. $w_{ij}(t-1), R_{ij}(t-1),\ \forall j \in \mathcal{U}_i, \forall i \in \mathcal{B}$), to maximize the MNO profit through the user association. At the beginning of the subframe t, the algorithm creates the set $\mathcal{A}(t)$ with all the users that, being served by a BS, had a satisfaction below 1 at time $t-1$, i.e. $\sigma_{ij}(t-1) < 1$, and the users that consumed a lot of resources due to a Non-Line-Of-Sight (NLOS) connection (line 1). If the set is not empty, all users with $\sigma_{ij}(t-1) = 1$ are associated to the same BS they were associated with at time $t-1$ (line 4). For the association of the rest of the users, the algorithm makes use of the estimates of $w_i(t)$, $P_i(t)$ and $R_{ij}(t)$, denoted as $\overline{w_i}(t)$, $\overline{P_i}(t)$ and $\overline{R_{ij}}(t)$.

Users in $\mathcal{A}(t)$ are selected randomly, one by one, in order to avoid the prioritization of users during the association procedure, and hence guarantee fairness . Each user's contribution to the estimated BS profit, where the user was connected to at time $t-1$ is subtracted (lines 8-11). The algorithm estimates the resources needed/available in each BS, $\widehat{w_{ij}}(t)$, the expected revenue, $\widehat{R_{ij}}(t)$, and the expected satisfaction, $\widehat{\sigma_{ij}}(t)$ (line 13). Then, for all BSs, if there are BSs that provide satisfaction equal to 1, the user will be associated to the BS that maximizes the MNO profit while $\widehat{\sigma_{ij}}(t) = 1$ (lines 18-21). If there are not BSs that could provide $\widehat{\sigma_{ij}}(t) = 1$, but $\widehat{\sigma_{ij}}(t) > 0$, the user will be associated with the BS that maximizes the MNO profit with $\widehat{\sigma_{ij}}(t) > 0$ (lines 22-25). The rest of users are not associated to any BS, since there are not enough resources in the neighbouring BSs or the channel between the user and the BSs is in outage (line 27). The procedure is repeated for all users in $\mathcal{A}(t)$.

Feasibility: The feasibility of the algorithm depends on the availability of $w_{ij}(t-1)$ and $R_{ij}(t-1)$, the users' SINR, and the information needed to calculate $\widehat{\sigma_{ij}}(t)$ and $\widehat{R_{ij}}(t)$. In LTE-A, the SINR is calculated by the device and sent to the BS over the PUCCH or PUSCH [13]. The necessary information for $\widehat{\sigma_{ij}}(t)$ and $\widehat{R_{ij}}(t)$ can be obtained in real time with a module such as the Policy and Charging Control (PCC) in LTE-A, which can control the QoS on a per service data flow, apply different charging models, as well as control usage monitoring to make dynamic policy decisions [14].

Algorithm 1: QoE-Aware user Association Algorithm

1 Create $\mathcal{A}(t)$ as the set of users with $\sigma_{ij}(t-1) < 1$ or a NLOS channel with serving BS.
2 **if** $\mathcal{A}(t) \neq \emptyset$ **then**
3   **for** $j \in \mathcal{U} \setminus \mathcal{A}(t)$ **do**
4     $x_{ij}(t) = x_{ij}(t-1),\ \forall i \in \mathcal{B}$
5   **end**
6   Initialize $\overline{w_i}(t) = w_i(t-1)$; $\overline{P_i}(t) = P_i(t-1)$; $\overline{R_{ij}}(t) = R_{ij}(t-1)$; $\forall i \in \mathcal{B}, \forall j \in \mathcal{U}_i$
7   **while** $\mathcal{A}(t) \neq \emptyset$ **do**
8   Select user j randomly from $\mathcal{A}(t)$
9     $m = \arg\max_{i \in \mathcal{B}} x_{ij}(t-1)$
10     $\overline{P_m}(t) = \overline{P_m}(t) - \overline{R_{mj}}(t) + {} + c_m e^{h_m \overline{w_m}(t) b_m} \left(1 - e^{-h_m w_{mj}(t-1) b_m}\right)$
11     $\overline{w_m}(t) = \overline{w_m}(t) - w_{mj}(t-1)$
12   **for** $i \in \mathcal{B}$ **do**
13       $\widehat{w_{ij}}(t) = \min\left(\frac{r_{kd}}{\varepsilon_{ij}(t) b_i}, 1 - \overline{w_i}(t)\right)$
14   Calculate $\widehat{\sigma_{ij}}(t), \widehat{R_{ij}}(t)$ according to (2),(7) for $\widehat{w_{ij}}(t)$
15       $\widehat{P_i}(t) = \overline{P_i}(t) + \widehat{R_{ij}}(t) - {} - c_i e^{h_i \overline{w_i}(t) b_i} (e^{h_i \widehat{w_{ij}}(t) b_i} - 1)$
16       $\widehat{w_i}(t) = \overline{w_i}(t) + \widehat{w_{ij}}(t)$
17   **end**
18   **if** $\exists i \in \mathcal{B}$ *such that* $\widehat{\sigma_{ij}}(t) = 1$ **then**
19       $m = \arg\max_{i \in \mathcal{B}: \widehat{\sigma_{ij}}(t) = 1} \{\widehat{P_i}(t) + \sum_{v \in \mathcal{B} \setminus \{i\}} \overline{P_v}(t)\}$
20       $x_{mj}(t) = 1$; $x_{vj}(t) = 0,\ \forall v \in \mathcal{B} \setminus m$
21       $\overline{w_m}(t) = \widehat{w_i}(t)$; $\overline{R_{mj}}(t) = \widehat{R_{mj}}(t)$; $\overline{P_m}(t) = \widehat{P_m}(t)$
22   **else if** $\exists i \in \mathcal{B}$ *such that* $\widehat{\sigma_{ij}}(t) \in (0,1)$ **then**
23       $m = \arg\max_{i \in \mathcal{B}: \widehat{\sigma_{ij}}(t) \in (0,1)} \{\widehat{P_i}(t) + \sum_{v \in \mathcal{B} \setminus \{i\}} \overline{P_v}(t)\}$
24       $x_{mj}(t) = 1$; $x_{vj}(t) = 0,\ \forall v \in \mathcal{B} \setminus m$
25       $\overline{w_m}(t) = \widehat{w_i}(t)$; $\overline{R_{mj}}(t) = \widehat{R_{mj}}(t)$; $\overline{P_m}(t) = \widehat{P_m}(t)$
26   **else**
27       $x_{vj}(t) = 0,\ \forall v \in \mathcal{B}$
28   **end**
29     $\mathcal{A}(t) = \mathcal{A}(t) \setminus \{j\}$
30   **end**
31 **end**

VI. Performance Evaluation

The scenario used for the performance evaluation consists of a cluster with 4 mmWave SCs deployed in the coverage area of a macrocell eNB sector. The cluster is square 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 mmWave channel is modelled as a three-states channel [15], with LOS, NLOS and outage states. Although high directivity of antennas compensates partially the path loss, the probability of LOS communications falls rapidly as the distance between transmitter and receiver increases. In the scenario we define R_{sc} as the distance at which the probability of having a LOS communication is 0.55. According to [15], $\mathbb{P}_{LOS}(R_{sc}) = 0.55$ holds for R_{sc}=40m in the 28GHz band. In the sequel the Inter-Site Distance (ISD) between SCs, R_{ISD}, is expressed as a multiple of R_{sc}, i.e. R_{ISD}=nR_{sc}, with n∈$\mathbb{N}$.

Users are uniformly distributed within the cluster with a speed of 5 km/h and the SP of each user is selected with equal probability among the SPs defined in Table I. As it can be observed in Table I, three services are considered, each one with two QoE classes $\mathcal{Q}$ = {Basic, Premium}:

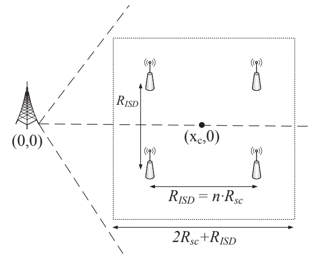

Fig. 1: Simulation scenario topology

TABLE I: Service Profiles' parameters

Service	**QoE class**	$\{\mathbf{r_{k1}}, \mathbf{r_{k2}}, \mathbf{r_{k3}}\}$ (Mbps)	$\theta_\mathbf{k}^\mathbf{t}$ **or** $\theta_\mathbf{k}^\mathbf{B}$
Service 1 (Data Based)	Basic	80	$3 \cdot 10^{-4}$€/MB
	Premium	100	$4 \cdot 10^{-4}$€/MB
Service 2 (Time Based)	Basic	$\{50, 65, 80\}$	3.5€/h
	Premium	$\{60, 75, 90\}$	5€/h
Service 3 (Time Based)	Basic	$\{70, 85, 90\}$	3.5€/h
	Premium	$\{80, 100, 120\}$	5€/h

TABLE II: BS parameters

Parameter	**Macrocell**	**Small cell**
c_i (€)	$5 \cdot 10^{-6}$	$5 \cdot 10^{-6}$
h_i (MHz^{-1})	$39 \cdot 10^{-3}$	$16 \cdot 10^{-3}$
b_i (MHz)	200	500
CF_i (€) [16]	$2.41 \cdot 10^{-4}$	$14 \cdot 10^{-4}$
Transmission Power (dBm)	43	37

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 I. Note that for each SP, r_{kd} is the transmission rate required to perceive a QoE equal to 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. Moreover, v_k is selected so as to have $Q_p(p_k) = 0.9$ in (5), and

$$\begin{cases} \alpha_{kd} = \frac{Q_k^{tg}}{Q_p(p_k)} - \gamma_{kd} & \text{(11a)} \\ \beta_{kd} = -\frac{1}{\Delta r_j^{drop}} \ln\left(\frac{Q_k^{drop} - \gamma_{kd} Q_p(p_k)}{Q_k^{tg} - \gamma_{kd} Q_p(p_k)}\right), & \text{(11b)} \end{cases}$$

where $\Delta r_j^{drop} = r_{kd} - r_{kd}^{drop}$, and $\gamma_{kd} = 1$, for all π_k, $d \in \mathcal{D}$ and $(Q_k^{tg}, Q_k^{drop}) = (3.5, 2.5)$ for Basic QoE class of all services 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 II. For the bandwidth allocation in the two tiers, we adopted the 5G configuration proposed by a leading telecommunications vendor [17]. The 3GPP LTE-A channel model used for macrocells is described in [13] and for SCs (in the 28GHz band) in [15]. Antenna gains are set to 0 dB and, due to high antenna directivity in the mmWave band, SCs communications are assumed to be noise-limited. In conventional cellular networks, the cell selection is based on schemes that connect the users to the BS with the strongest signal [18]. Hence, in the following we compare the proposed

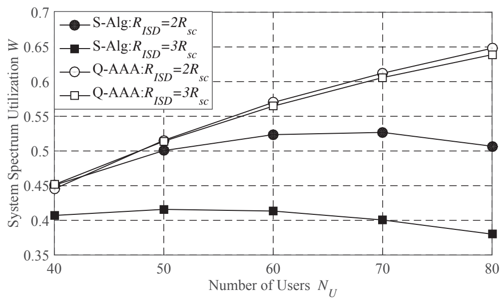

Fig. 2: Bandwidth utilization (W)

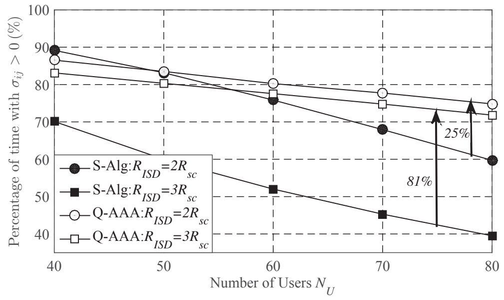

Fig. 3: Percentage of time with a satisfaction above 0 ($\sigma_{ij} > 0$)

QoE-aware Association Algorithm (denoted as Q-AAA) with a SINR-based cell selection algorithm (referred to as SINR-Alg)[1]. It should be noted that the resource allocation in both algorithms satisfies the condition for QoE fairness (i.e. $\sigma_{ij}(t) = \sigma_{in}(t)$ for any $j, n \in \mathcal{U}_i$).

Fig. 2 shows the expected total utilization of the spectrum, which is defined as $W = \mathbb{E}\left[W(t) = \frac{\sum_{i\in\mathcal{B}} w_i(t) b_i}{\sum_{i\in\mathcal{B}} b_i}\right]$. As it can be observed, the proposed algorithm (Q-AAA) presents higher bandwidth utilization than the cell association algorithm based on the SINR (SINR-Alg) for both $R_{ISD} = 2R_{sc}$ and $R_{ISD} = 3R_{sc}$. As expected, SINR-Alg should always present the best spectrum efficiency, and consequently, the lowest bandwidth utilization, since users tend to use the most efficient Modulation and Coding Scheme (MCS). Interesting enough, a different behaviour can be noticed between the two algorithms regarding N_U and R_{ISD}. As it can be seen, for Q-AAA utilization increases with N_U, and is comparable regarding R_{ISD}. As R_{ISD} rises, the average distance between BSs and users grows as well. Thus, the probability of serving users with NLOS links raises, increasing W. The broad difference in W between the two R_{ISD} values observed for SINR-Alg can be explained with the help of Fig. 3.

Fig. 3 shows the percentage of time during which the users had a satisfaction above 0, i.e. $\sigma_{ij}>0$. For both algorithms, the users perceive $\sigma_{ij}>0$ for less time as N_U and R_{ISD} increase. As the system's bandwidth demands increase with both N_U and R_{ISD}, the system can accommodate a lower percentage of users. We further observe that Q-AAA achieves gains up

[1]In the SINR-Alg, users are served by the BS with the highest SINR.

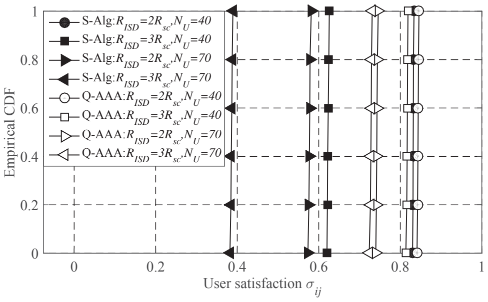

Fig. 4: CDF of the user satisfaction

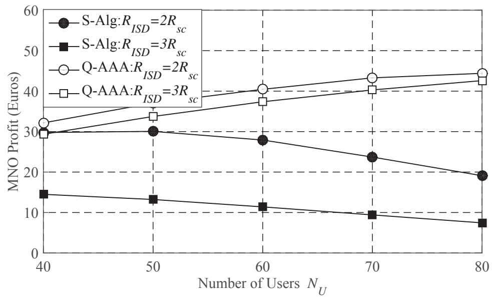

Fig. 5: Total MNO profit (P)

to 25% and 81% for R_{ISD}={2,3}R_{sc} respectively. This can be explained by the fact that Q-AAA prioritizes users that can obtain an appropriate QoE over users that cannot, whereas SINR-Alg connects users without considering the available BS resources. Thus, SINR-Alg congests the BSs more frequently, dropping users to satisfy the QoE fairness condition.

Fig. 4 presents the empirical CDFs of σ_{ij}, for $N_U = \{40, 70\}$ and $R_{ISD} = \{2, 3\}R_{sc}$. As it can be observed, σ_{ij} diminishes as N_U and R_{ISD} increase and the system gets congested more frequently. Hence, the QoE is degraded more regularly, while the service time with $\sigma_{ij} > 0$ decreases as well (see Fig. 3). It can be deduced from Fig. 2-4 that Q-AAA results in less frequent congestion of the BSs, and provides higher QoE for longer time periods. The above explains why P increases with Q-AAA, whereas it decreases for the SINR-Alg not only with R_{ISD}, but also with N_U, as shown in Fig. 5. This happens because as Q-AAA offers higher user satisfaction for longer time period, the MNO generates larger revenue, compensating the higher bandwidth utilization.

Our proposed algorithm, Q-AAA, manages to offer higher QoE to the users and profit to the MNO compared to the reference algorithm, because it bases its decisions on both the technological (i.e. QoS/QoE requirements) and economic (i.e. pricing and profit) context of the network. This scheme can guarantee the sustainability of a network, providing the incentives for adoption in future 5G deployments.

VII. CONCLUSIONS

In this paper, we studied the user association problem in a single MNO's 5G HetNet described by traffic heterogeneity. Our objective was the maximization of the MNO's profit, while providing high QoE to the users. Thus, we proposed a heuristic user association algorithm, which bases its decisions on the capability of the BSs to offer an acceptable QoE level to the user, while at the same time maximizing the MNO profit. We evaluated the performance of the proposed algorithm by comparing it with a SINR-based algorithm. The simulation results show the adaptability of the proposed algorithm to traffic heterogeneity by achieving substantially higher profit and QoE, in contrast to the SINR-based one.

ACKNOWLEDGEMENT

This work has been funded by the MITN Project CROSS-FIRE (PITN-GA-2012-317126), Project CellFive (TEC2014-60130-P), and Project SGR (2014-SGR-1551).

REFERENCES

[1] Appier Research Report: *"Cross-Screen User Behavior Insights, Asia 1H 2015"*, Aug 2015
[2] 5G-PPP, *"The 5G Infrastructure Public Private Partnership: the next generation of communication networks and services,"*
[3] Mobile Squared, *"Annual Market Review 2014 and PRS market outlook 2015,"* Jul. 2015
[4] Fiedler, M.; Hossfeld, T.; Tran-Gia, P., *"A generic quantitative relationship between quality of experience and quality of service,"* in Network, IEEE , vol.24, no.2, pp.36-41, March-April 2010
[5] Reichl, P.; Maille, P.; Zwickl, P.; Sackl, A., *"On the fixpoint problem of QoE-based charging,"* 6th International Conference on Performance Evaluation Methodologies and Tools, pp.235-242, Oct. 2012
[6] Pantisano, F.; Bennis, M.; Saad, W.; Valentin, S.; Debbah, M.; Zappone, A., *"Proactive user association in wireless small cell networks via collaborative filtering,"* in Signals, Systems and Computers, Asilomar Conference on , pp.1601-1605, 3-6 Nov. 2013
[7] F. Pantisano, M. Bennis, W. Saad, S. Valentin and M. Debbah, *"Matching with externalities for context-aware user-cell association in small cell networks,"* IEEE Global Communications Conference (GLOBECOM), Atlanta, GA, 2013, pp. 4483-4488.
[8] G. Athanasiou, P. C. Weeraddana, C. Fischione and L. Tassiulas, *"Optimizing Client Association for Load Balancing and Fairness in Millimeter-Wave Wireless Networks,"* in IEEE/ACM Transactions on Networking, vol. 23, no. 3, pp. 836-850, June 2015.
[9] H. Zhou, S. Mao and P. Agrawal, *"Approximation Algorithms for Cell Association and Scheduling in Femtocell Networks,"* in IEEE Trans. Emerging Topics in Computing, vol. 3, no. 3, pp. 432-443, Sept. 2015.
[10] 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)
[11] Kibria, M.G.; Lin Shan, *"Resource Allocation Optimization for Users with Different Levels of Service in Multicarrier Systems,"* in Signal Processing Letters, IEEE , vol.22, no.11, pp.1869-1873, Nov. 2015
[12] Iosifidis, G.; Lin Gao; Jianwei Huang; Tassiulas, 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
[13] 3GPP TR 36.814 v9.0.0, *"Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects,"* Mar. 2010
[14] 3GPP TR 23.203 v14.0.0, *"Policy and charging control architecture,"* Jun. 2016
[15] M. R. Akdeniz et al., *"Millimeter Wave Channel Modeling and Cellular Capacity Evaluation,"* in IEEE Journal on Selected Areas in Communications, vol. 32, no. 6, pp. 1164-1179, June 2014.
[16] Real Wireless, *"The Business Case for Urban Small Cells"*, Report for Small Cell Forum, Feb 2014.
[17] *"Ten key rules of 5G deployment,"* Nokia Networks, White Paper 2015
[18] J. G. Andrews, *"Seven ways that HetNets are a cellular paradigm shift,"* in IEEE Commun. Mag., vol. 51, no. 3, pp. 136-144, March 2013.