Design and Evaluation of a Backhaul-Aware Base Station Assignment Algorithm for OFDMA-Based Cellular Networks

Hiram Galeana-Zapién, *Student Member, IEEE,* and Ramon Ferrús, *Member, IEEE*

***Abstract*—Existing base station (BS) assignment methods in cellular networks are mainly driven by radio criteria since it is assumed that the only limiting resource factor is on the air interface. However, as enhanced air interfaces have been deployed, and mobile data and multimedia traffic increases, a growing concern is that the backhaul of the cellular network can become the bottleneck in certain deployment scenarios. In this paper, we extend the BS assignment problem to cope with possible backhaul congestion situations. A backhaul-aware BS assignment problem is modeled as an optimization problem using a utility-based framework, imposing constraints on both radio and backhaul resources, and mapped into a Multiple-Choice Multidimensional Knapsack Problem (MMKP). A novel heuristic BS assignment algorithm with polynomial time is formulated, evaluated and compared to classical schemes based exclusively on radio conditions. Simulation results demonstrate that the proposed algorithm can provide the same system capacity with less backhaul resources so that, under backhaul bottleneck situations, a better overall network performance is effectively achieved.**

***Index Terms*—BS assignment algorithms, mobile backhaul, OFDMA, radio resource management.**

I. Introduction

WIRELESS access technologies are continuously evolving to provide higher data rates and pave the way for ubiquitous, high speed broadband wireless coverage. Nowadays, the most outstanding radio technologies to meet these requirements are based on orthogonal frequency division multiplexing (OFDM) schemes that have been already adopted for next generation cellular systems such as Long Term Evolution (LTE) and Mobile WiMAX. An OFDM physical layer enables orthogonal frequency division multiple access (OFDMA) that allows for the exploitation of multiuser diversity by managing both time and frequency components in the radio resource allocation process. In an OFDMA-based cellular system, efficient radio resource allocation techniques are crucial to fully exploit OFDMA capabilities [1]. In this context, the base station (BS) assignment problem, that is, the selection of the most appropriate BS to handle radio transmission to/from mobile terminals, constitutes a key component of the overall resource allocation process [2].

So far, existing BS assignment solutions consider that the main capacity bottleneck of cellular networks is always on the air interface. This assumption proved to be valid for voice-dominant cellular networks where, as the aggregate traffic rate at cell sites was relatively low, dimensioning backhaul network[1] capacity to satisfy air interface peak rates was an economically feasible option. Conversely, significantly higher air interface peak rates provided by OFDMA-based systems bring new challenges in how to properly scale the capacity of backhaul networks in a cost-effective manner, especially for the last mile connectivity. Due that backhaul could represent as much as one quarter of the total network costs [4], mobile operators are carefully reviewing their backhaul strategies before making further investments in the transport network infrastructure. In fact, capacity upgrades in the backhaul are expected to be carried out by operators gradually and, while it can be argued that to bring fiber to more cell towers in the backhaul would solve the last mile bandwidth problem, the fact is that there are far too many towers for this to be a near-term strategy. In this context, best practices for efficient backhaul design have been recently issued by NGMN Alliance [5] and there is an increasing number of solutions pushing for the adoption of more cost-effective transmission technologies than those used in most current deployments [3]–[6] along with new resource management functionality specifically tailored to tackle backhaul congestion [7]. As a matter of fact, flow control mechanisms have been already introduced in current mobile networks to partially mitigate traffic peaks in the backhaul at the expenses of an increased delay in some services [8], [9]. Attending to previous considerations, cellular network capacity limitations due to a shortage of backhaul capacity should not be underestimated in some network deployments.

In this paper we propose to take into consideration the amount of backhaul capacity available in each cell site within the BS assignment process of a cellular OFDMA network. Hence, we develop a novel backhaul-aware BS assignment algorithm envisioned as a suitable technique capable to cope, at some extent, with possible backhaul congestion situations.

Manuscript received November 26, 2009; revised April 26, 2010; accepted June 18, 2010. The associate editor coordinating the review of this paper and approving it for publication was I. Habib.

The authors are with the Department of Signal Theory and Communications, Universitat Politècnica de Catalunya (UPC), Jordi Girona 1-3 (Campus Nord), Barcelona, 08034 Spain (e-mail: {hiram.galeana, ferrus}@tsc.upc.edu).

Digital Object Identifier 10.1109/TWC.2010.082110.091735

[1]The *backhaul* or *transport* network refers to the part of the cellular access network that interconnects BSs to the rest of the network infrastructure by means of whatever transmission media (been E1/T1 leased lines over copper wires and microwave radio links the most used transmission technologies [3]).

The main idea behind the proposed algorithm is to distribute traffic among BSs according to a load balancing strategy that considers both radio and backhaul load status. This possibility is shown to constitute a tradeoff between reducing backhaul congestion and using radio resources efficiently since some users can be assigned to BSs not being their "best" radio choice but preventing congestion in other BSs. The proposed algorithm is proven to successfully exploit such a tradeoff, turning ultimately into a better overall network performance.

A. Related Work

The BS assignment problem in cellular systems has deserved significant research efforts. Irrespective of the radio access technology, one of the most common BS assignment approaches is the Minimum Path Loss (MPL) that assigns each user to the BS that provides the highest radio link gain [10]. This approach alone constitutes the core of many BS assignment algorithms used in current 2G/3G cellular systems (where absolute and/or relative received signal level thresholds are used to decide upon the serving BS) and also forms part of more sophisticated approaches aimed to exploit, e.g., multiuser detection and multiple antennas [11]. Another common approach consists of taking into account the Signal to Interference and Noise Ratio (SINR) in the assignment process, which is particularly important when targeting an aggressive reuse of frequencies throughout the network (e.g., single-channel CDMA networks and OFDMA networks with low frequency reuse factors). In this case, there is a mutual dependency between the SINR values and the BS assignment in the downlink[2] (i.e., SINR values depends on the BS allocation and viceversa), that further complicates the resource allocation problem. In addition to channel gain and SINR, different constraints such as maximum transmit power levels or minimum guaranteed bit rates have been also considered under various forms of optimization problems [2], [12], [13]. As to specific works focusing on the BS assignment problem in OFDMA, work in [2] proposes a suboptimal approach based on SINR and constraints on the BS downlink radio capacity, where mutual dependency issue is avoided by performing a greedy BS assignment that sequentially chooses the user with the highest SINR. An iterative BS assignment scheme aimed to balance traffic densities is developed in [14], where the assignment decision is based on the MPL criterion and quality of service (QoS) requirements of users. However, most of the works tackling resource allocation in multi-cell OFDMA [1], [15]–[18], implicitly consider a BS assignment based on a simple MPL criterion, due that these works mainly concentrate on developing algorithmic solutions to the subcarrier and power allocation problems.

B. Our Contributions

The effect of possible backhaul capacity limitations on the BS assignment has not received enough attention so far in the literature. This paper aims to fill this void, and presents three main contributions. First, we formulate a new BS assignment problem in OFDMA-based networks considering both radio and backhaul constraints in the assignment process. The formulated optimization problem is based on utility and resource cost concepts, which have been widely used to develop resource allocation algorithms [19]. The second contribution is the mapping, after some practical considerations, of the BS assignment problem to a Multiple-Choice Multidimensional Knapsack Problem (MMKP), a well-known NP-hard combinatorial optimization problem arisen in many practical and real life problems [20], [21]. Motivated by the need to obtain suboptimal solutions with polynomial time complexity, the third contribution is the derivation of a heuristic backhaul-aware BS assignment algorithm along with its performance comparison respect to classical schemes entirely based on radio conditions.

The rest of the paper is organized as follows. In Section II we describe the system model. Section III presents the formulation of BS assignment problem as an optimization problem, and the mapping to an MMKP. The proposed heuristic BS assignment algorithm is detailed in Section IV, where its computational complexity is also analyzed. Section V provides numerical results of the algorithm's performance. Finally, main conclusions are drawn in Section VI.

II. System Model

We consider a downlink OFDMA-based system with N BSs that cover a geographical area in which there are M active users, as illustrated in Fig. 1. Each user $i \in \{1, \ldots, M\}$ is assumed to have a minimum bit data rate requirement, denoted as $R_i^{\min}$. The overall network uses a total bandwidth BW divided into K OFDM subcarriers so each BS $j \in \{1, \ldots, N\}$ operates a subset of K_j subcarriers attending to a given frequency reuse pattern. Radio and transport resources are assumed to be allocated to each user in a single BS, i.e., no multicell transmission is considered. We assume that each BS is constrained by a limited amount of radio and transport resources. As to radio resources, each BS j is able to allocate simultaneously a maximum of K_j subcarriers and has a maximum transmit power of $P_j^{\max}$. Radio channel gain between BS j and user i is modeled by $\overrightarrow{G}_{i,j} = \{G_{i,j,1}, \ldots, G_{i,j,k}\}$, where $G_{i,j,k}$ denotes the channel gain over subcarrier $k \in \{1, \ldots, K_j\}$. As to transport resources, we assume each BS j has a maximum transport capacity of C_j^{trans} (in bits/sec), which refers to the available bandwidth in the path between BS j and the access gateway[3] (aGW) within the mobile network[4].

The amount of network resources required by user i to meet its rate requirement $R_i^{\min}$ depends on the selected BS j. Thus, in order to quantify the resource consumption of user i when assigned to BS j, we define a radio resource cost and a transport resource cost function, denoted as α_{ij} and β_{ij}, respectively. Over such a basis, the BS assignment problem should try to find a feasible assignment (i.e., $R_i^{\min}$ is satisfied for each user i and total radio and transport

[2]This is normally seen as the more restrictive link due to the asymmetric bandwidth demand between downlink and uplink.

[3]The aGW would correspond to the ASN_GW network entity in Mobile WiMAX, or to the Serving Gateway in LTE network architecture.

[4]We have not particularized the analysis of the BS assignment problem to any specific backhaul technology. Hence, the backhaul capacity assumed here could actually correspond to the available bandwidth of the wired/wireless link used for the last mile connection of a given BS.

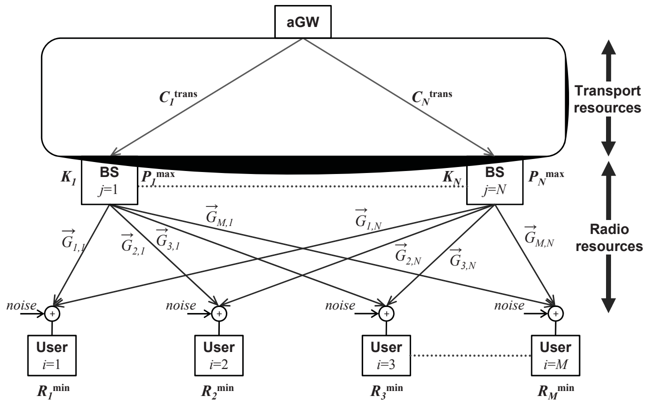

Fig. 1. System model with N BSs serving M users. The arrows connecting users and BSs indicate possible assignment choices, each one having a particular radio channel gain $\vec{G}_{i,j}$. Each BS is assumed to have a limited amount of resources on the air interface (K_j subcarriers, $P_j^{\max}$ power) and transport network (C_j^{trans}) to serve its assigned users.

resources are not exceeded). As well, if several feasible solutions exist (i.e., there are several ways to allocate all the users without exceeding network resources), the "best" one should be chosen. To this end, a utility function is used to quantify the appropriateness of each assignment in terms of bit rate efficiency of the allocated resources. Details of utility and resource cost functions are given in next subsections.

A. Radio Resource Cost

In a cellular OFDMA system, the computation of the SINR achieved at subcarrier k in the receiver of user i served by BS j, is obtained as follows [2]:

$$\text{SINR}_{i,j,k} = \frac{G_{i,j,k}P_{i,j,k}}{I_{i,j,k} + \eta} \tag{1}$$

where $G_{i,j,k}$ is the radio channel gain between BS j and user i over subcarrier k, $P_{i,j,k}$ is the transmit power of BS j on subcarrier k allocated to user i, η is the subcarrier thermal noise, and $I_{i,j,k}$ is the co-channel interference (CCI) power received by user i on that subcarrier. The value of the co-channel interference $I_{i,j,k}$ can be computed as:

$$I_{i,j,k} = \sum_{n=1,n\neq j}^{n=N} G_{i,n,k}P_{m\neq i,n,k} \tag{2}$$

where $P_{m,n,k}$ is the transmit power of interfering BS n, on subcarrier k assigned to other user $m \neq i$. Equation (1) denotes the channel frequency response of user i on subcarrier k, and the achievable transmission rate $r_{i,j,k}$ on this subcarrier of user i assigned to BS j is given by [22]:

$$r_{i,j,k} = \frac{BW}{K}\log_2(1 + \text{SINR}_{i,j,k}) \tag{3}$$

Hence, if all resources of BS j were allocated to user i, the maximum achievable rate would be:

$$R_{ij}^{\max} = \sum_{k=1}^{K_j} r_{i,j,k} \tag{4}$$

Over such a basis, considering that a BS dynamically shares transmission resources among its assigned users by allocating, on average, a given amount of subcarriers to user i, denoted as K_{ij} (being $K_{ij} \leq K_j$) during a given amount of transmission time denoted as ΔT_{ij} (being $T_{ij} \leq T_s$, where T_s is a scheduling reference period), we could relate the achievable rate to the amount of allocated subcarriers and transmission time to meet user's minimum rate requirement by:

$$\frac{K_{ij}}{K_j}\frac{\Delta T_{ij}}{T_s}R_{ij}^{\max} \geq R_i^{\min} \tag{5}$$

From (5), the radio resource cost is defined directly as:

$$\alpha_{ij} \triangleq \frac{R_i^{\min}}{R_{ij}^{\max}} = \frac{K_{ij}}{K_j}\frac{\Delta T_{ij}}{T_s} \leq 1 \tag{6}$$

where $\alpha_{ij} = 1$ would mean that the assignment of user i to BS j requires all available radio resources at the BS to meet its rate requirement. Attending to practical considerations, we consider a limited set of modulation and coding schemes (MCS) that must be used in each subcarrier, thus reducing the output of expressions (3), (4) and (6) to a set of discrete values. Then, we define the aggregate peak rate over the air interface of BS j, denoted as C_j^{air}, as the highest achievable aggregate data rate when using all subcarriers continuously with the highest rate MCS.

B. Transport Resource Cost

The transport resource cost, denoted as β_{ij}, associated with the assignment of user i to BS j is defined as the ratio of the minimum data rate required by user i, $R_i^{\min}$, to the available transport capacity of BS j, denoted as C_j^{trans}, that is:

$$\beta_{ij} = \frac{R_i^{\min}}{C_j^{\text{trans}}} \tag{7}$$

As a matter of clearly relating the transport capacity C_j^{trans}, to the aggregate peak rate of the radio interface C_j^{air}, we define the transport capacity factor ϕ_j as:

$$\phi_j \triangleq \frac{C_j^{\text{trans}}}{C_j^{\text{air}}} \tag{8}$$

Therefore, a transport capacity factor $\phi_j = 1$ would mean that the transport capacity of BS j has been provisioned to support the highest achievable aggregated data rate of the air interface. In this regard, it worth noting that dimensioning the backhaul capacity to satisfy the air interface peak rate may not constitute a resource efficient solution since not all cell connections can always simultaneously exploit the fastest MCS. However, occasionally, user distribution in the cell (e.g., most served users being close to BS and enjoying good radio conditions) may turn into aggregate data rates close to air interface peak rate.

C. Utility Function

Commonly, a utility function is a non-decreasing function of the amount of allocated network resources and its shape (e.g., step, convex, concave or sigmoid are often used) depends on the expected benefit that resource allocation can bring into a given system[5]. Here we formulate the utility function to reflect the bit rate efficiency of the allocated resources to support the data transfer of each user assigned to a given BS. Hence, a utility function denoted as u_{ij} captures the suitability of assigning user i to BS j, so $u_{ij} > u_{il}$ would mean that BS j is more appropriate than BS l to serve user i in terms of bit rate efficiency. As well, $u_{ij} > u_{lj}$ would indicate that is better to assign user i to BS j instead of user l. Over such a basis, the considered bit rate efficiency in the radio interface is directly associated with the spectral efficiency (bits/sec/Hz), while in the transport network it's assumed that all assignments lead to the same bit rate efficiency[6]. Hence, the utility function is defined as:

$$u_{ij} = \frac{1}{K_j}\sum_{k=1}^{K_j}\log_2(1+\mathrm{SINR}_{i,j,k}) \tag{9}$$

As a result, assignments to BSs where users have the highest values of SINR are favored.

III. Problem Setting

The BS assignment problem is formulated as an optimization problem aiming to maximize the total welfare utility, defined as the sum of the utilities of all assignments, subject to resource constraints in BSs. Let $B = \{b_{ij}\}_{MxN}$, be the BS assignment matrix whose entry b_{ij} is equal to one if user i is assigned to BS j and equal to zero otherwise. This problem can be formally written as:

$$\max_{b_{ij}} \quad \sum_{i=1}^{M}\sum_{j=1}^{N} u_{ij}b_{ij} \tag{10}$$

$$\text{s. t.} \quad \sum_{i=1}^{M}\alpha_{ij}b_{ij} \leq 1,\ j=1,\ldots,N \tag{11}$$

$$\sum_{i=1}^{M}\beta_{ij}b_{ij} \leq 1,\ j=1,\ldots,N \tag{12}$$

$$\sum_{j=1}^{N} b_{ij} = 1,\ i=1,\ldots,M \tag{13}$$

$$R_i \geq R_i^{\min},\ i=1,\ldots,M \tag{14}$$

$$b_{ij} \in \{0,1\} \tag{15}$$

The set of constraints considered in (11) and (12) assures that no more radio and transport resources than available are used in each BS. Constraints in (13) indicate that all users need to be assigned to a single BS, while (14) ensures that the expected bit rate of user i, denoted as R_i, meets the minimum data rate requirement of each user. Finally, to avoid splitting or partial assignment of users to BSs, constraint (15) is enforced, which in turn leads to the combinatorial nature of the problem with exponentially growing complexity in the degrees of freedom.

[5]For instance, a step function can be used to model a system where allocating resources below a given threshold has no utility at all, but the maximum utility is just achieved when reaching such a threshold.

[6]The resources needed to transport 1b/s of a user between the aGW and the correspondent BS are considered to be the same for all BSs, noticing here that other assumptions, e.g., based on transport provisioning costs, could be also possible but are out of the scope of this work.

A. Practical Considerations

Problem (10)-(15) is a non-linear combinatorial optimization problem since entries in the assignment matrix B can only take integer values. Notice that utility and radio resource cost functions are non-linear functions that depend on the SINR values, which in turn depend on the BS assignment solution because of CCI (i.e., mutual dependency). So, both utility and radio resource cost function values are coupled with the assignment of users to BSs, making the BS assignment problem very hard to tackle. To overcome this issue, we reformulate the BS assignment problem under the practical consideration of a fully-loaded system [12], where BSs are assumed to transmit at maximum power so that the mutual dependency is avoided. Apart from reducing the complexity of the BS assignment problem, the rationale of considering full-load conditions can be also justified by the fact that it is just under such stressed load conditions where resource management strategies are expected to bring out their potential benefits. We also consider that the maximum transmit power of BS j is shared uniformly, on average[7], over its K_j subcarriers. Hence, BSs are supposed to make use of all available subcarriers in the same way (i.e., there is no a subcarrier more favored than another). Over such a basis, the co-channel interference value observed by user i under full load conditions can be estimated from (2) as:

$$I_{i,j,k}^{\max} = \sum_{n=1,n\neq j}^{n=N_k} G_{i,n,k}\frac{P_n^{\max}}{K_n} \tag{16}$$

where $P_n^{\max}$ and K_n are, respectively, the maximum transmit power and number of used subcarriers in interfering BS n. In this way, the computation of SINR under full load conditions by means of (1) does not depend on the BS assignment, as neither do utility and radio costs values.

B. Problem Mapping

According to previous practical considerations, the BS assignment problem in (10)-(15) can be mapped into a Multiple-Choice Multidimensional Knapsack Problem (MMKP) [23], a variant of the 0-1 knapsack problem. A MMKP considers a set of items, classified in I disjoint groups of J_i items each, and a knapsack (to pack some of them) whose available capacity is modelled by means of S distinct resource constraints represented by $(W_1, W_2, \ldots, W_S)$. Packing item j from group i turns into a benefit (utility) given by u_{ij} at the expenses of using a portion of the knapsack capacity given by $W_{ij} = w_{ij}^1/W_1, w_{ij}^2/W_2, \ldots, w_{ij}^S/W_S$. The objective is to exactly select one item from each group to maximize the aggregated utility subject to knapsack's capacity. The canonical formulation of this problem is as follows:

[7]The average of the transmit power allocated to each subcarrier is assumed to be equal for all subcarriers over the time scale at which the BS assignment algorithm operates.

$$\max_{b_{ij}} \quad \sum_{i=1}^{I}\sum_{j=1}^{J_i} u_{ij} b_{ij} \tag{17}$$

$$\text{s. t.} \quad \sum_{i=1}^{I}\sum_{j=1}^{J_i} \left(\frac{w_{ij}^s}{W_s}\right) b_{ij} \leq 1,\ s=1,\dots,S \tag{18}$$

$$\sum_{j=1}^{J_i} b_{ij} = 1,\ i=1,\dots,I \tag{19}$$

$$b_{ij} \in \{0,1\} \tag{20}$$

The MMKP problem is equivalent to our original optimization problem given by (10)-(15). Hence, the number of groups I corresponds to the number of users M. The set of items J_i within the group i are the set of N BSs where the user can be allocated. The number of limiting resources is $S = 2N$ since there are N BSs, each one having two resource constraints. The amount of resources required for serving user i in BS j (choosing item j from group i) is given by $W_{ij} = (\alpha_{ij}^1, \dots, \alpha_{ij}^s, \dots, \alpha_{ij}^N, \dots, \beta_{ij}^1, \dots, \beta_{ij}^s, \dots, \beta_{ij}^N)$, where α_{ij}^s and β_{ij}^s are described next. Since the allocation of user i only requires resources in the serving BS j, $\alpha_{ij}^s = \alpha_{ij}$ and $\beta_{ij}^s = \beta_{ij}$ if $s = j$, and $\alpha_{ij}^s = 0$ and $\beta_{ij}^s = 0$ otherwise, being α_{ij} and β_{ij} the radio and transport resources modeled by means of (6) and (7), respectively. Hence, inner summation in (18) reduces to $\sum_{j=1}^{J_i}(w_{ij}^s/W_s)b_{ij} \equiv (w_{is}^s/W_s)b_{is} = (w_{ij}^j/W_j)b_{ij} = \alpha_{ij}b_{ij}$ for $s = 1, \dots, N$ and $\sum_{j=1}^{J_i}(w_{ij}^s/W_s)b_{ij} \equiv (w_{is}^s/W_s)b_{is} = (w_{ij}^j/W_j)b_{ij} = \beta_{ij}b_{ij}$ for $s = N+1, \dots, 2N$, resulting in expressions (11) and (12), respectively. Finally, condition (14) is implicitly considered in the computation of resource costs.

IV. BS Assignment Algorithm

Since the BS assignment problem was transformed into an MMKP, any technique available to solve the MMKP can be used. There exist two different types of algorithms to solve the MMKP, namely: exact and heuristic algorithms. Due to its high computational complexity, exact algorithms are not suitable for most real-time decision-making applications [24], so the alternative is to use approximate heuristic approaches with polynomial time complexity. In this work, we develop a heuristic BS assignment algorithm based on [25]. The algorithm of Moser et al. [25] relies on a theorem proven by Everett [26] that makes Lagrange multipliers applicable to discrete optimization problems, such as the MMKP. In this regard, algorithm in [25] has already been considered as a useful tool in some works [27] to solve resource allocation problems in OFDMA wireless networks. Therefore, we have adapted the algorithm of Moser et al. to our specific BS assignment problem and introduced some relevant modifications, discussed later on, to the original algorithm. The main underlying concepts behind the adopted approach and the description of the proposed algorithm are provided in next subsections.

A. Main Concepts

According to [26], the optimal solution $b_{ij}^* \in \{0,1\}$ of the unconstrained maximization problem

$$\max_{b_{ij}} \left(\sum_{i=1}^{M}\sum_{j=1}^{N} u_{ij}b_{ij} - \sum_{j=1}^{N}\lambda_j \sum_{i=1}^{M}\alpha_{ij}b_{ij} - \sum_{j=1}^{N}\mu_j \sum_{i=1}^{M}\beta_{ij}b_{ij} \right) \tag{21}$$

where λ_j and μ_j are non-negative Lagrange multipliers associated with the radio and transport constraint on each BS j, respectively, is also the optimal solution for the constrained optimization problem:

$$\max_{b_{ij}} \quad \sum_{i=1}^{M}\sum_{j=1}^{N} u_{ij}b_{ij} \tag{22}$$

$$\text{s. t.} \quad \sum_{i=1}^{M}\alpha_{ij}b_{ij} \leq \sum_{i=1}^{M}\alpha_{ij}b_{ij}^* \triangleq \pi_j,\ j=1,\dots,N \tag{23}$$

$$\sum_{i=1}^{M}\beta_{ij}b_{ij} \leq \sum_{i=1}^{M}\beta_{ij}b_{ij}^* \triangleq \tau_j,\ j=1,\dots,N \tag{24}$$

that is equivalent to our BS assignment problem except for condition (13) discussed later on. From expression (21) it is easily noted that, if Lagrange multipliers λ_j and μ_j are known, the optimization problem can be easily solved. In fact, rewritten expression (21) as:

$$\max_{b_{ij}} \left\{ \sum_{i=1}^{M}\sum_{j=1}^{N} (u_{ij} - \lambda_j\alpha_{ij} - \mu_j\beta_{ij})b_{ij} \right\} \tag{25}$$

the optimal solution is given by:

$$b_{ij}^* = \begin{cases} 1, & \text{if} \quad w_{ij} = u_{ij} - \lambda_j\alpha_{ij} - \mu_j\beta_{ij} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{26}$$

where we define w_{ij} as the *weighted utility*, a metric that integrates the utility, radio and transport resource costs and associated Lagrange multipliers. It is worthwhile to note that constraint (13) indicating that users need to be assigned to a single BS can be easily taken into account by selecting, among possible assignments choices in (26), the one that provides the maximum weighted utility. Hence, the BS assignment problem can be solved by computing the set of $2N$ Lagrange multipliers. The assignment solution of all users is feasible if the amount of radio and transport resources allocated in each BS, denoted as π_j and τ_j, in (23) and (24), respectively, do not exceed available resources, i.e., $\pi_j \leq 1$ and $\tau_j \leq 1$. Furthermore, the solution is optimal if the following condition is held:

$$\sum_{j=1}^{N}\lambda_j(1-\pi_j) + \sum_{j=1}^{N}\mu_j(1-\tau_j) = 0 \tag{27}$$

The main difficulty in solving the problem is how to efficiently compute the Lagrange multipliers. In this regard, [25] used an approach based upon the concept of graceful degradation of the most valuable choices. First, an initial temporary solution $b_{ij} \in \{0,1\}$ is derived from (26) by considering all Lagrange multipliers equal to zero (i.e., the weighted utility equals to the utility, so that each user is assigned to the "best" BS irrespective of its radio or transport load). Then, Lagrange multipliers associated to BSs that would exceed available resources are iteratively increased in a smart way until a feasible solution, if exists, is found. That is, the increase of Lagrange multiplier associated to a BS cause a reduction in the weighted utility of its served users, so that some of them could be reassigned to other BSs providing higher weighted utility.

Phase 0: Initialization	01	Set Lagrange multipliers to zero: $\lambda_j \leftarrow 0,\ \mu_j \leftarrow 0$
	02	**for** each user $i=1,\ldots,M$
	03	**for** each BS $j=1,\ldots,n_i$ on the active set of i compute utility and costs
	04	u_{ij} (Eq.9), α_{ij} (Eq.6), β_{ij} (Eq.7)
	05	Find the most valuable BS $n=\operatorname{argmax}_j\{u_{ij}\}$ for each user i and update its assignment accordingly $b_{in} \leftarrow 1$
	06	**for** each BS j compute total radio/transport resource costs
	07	$\pi_j = \sum_{i=1}^{M} \alpha_{ij} b_{ij}, \tau_j = \sum_{i=1}^{M} \beta_{ij} b_{ij}$
	08	Conform total resource cost vector: $\psi=\{\pi_1,\ldots,\pi_N,\tau_1,\ldots,\tau_N\}$
Phase 1: Drop	09	**while** ($\psi_j > 1$ for any j) **do**
	10	Find the BS j^* holding the most offending constraint violation $s = \operatorname{argmax}_j\{\psi_j\}$, where $j^*=s$ if $s=1,\ldots,N$ and $j^*=s-N+1$ otherwise
	11	Compute the increase of the multiplier associated to constraint s (radio/transport) of BS j^*
	12	**for** $\{i \mid b_{ij^*}=1\}$
	13	**for** $\{j=1:n_i\}$
	14	Compute the increase of the weighted utility $\Delta w_{i,j^*\to j} = u_{ij^*} - u_{ij} - \lambda_{j^*}\alpha_{ij^*} + \lambda_j \alpha_{ij} - \mu_{j^*}\beta_{ij^*} + \mu_j \beta_{ij}$
	15	**if** $j^*=s$ **then**
	16	$\Delta\lambda_{i,j^*\to j} = \Delta w_{i,j^*\to j} / \alpha_{ij^*}$
	17	**else**
	18	$\Delta\mu_{i,j^*\to j} = \Delta w_{i,j^*\to j} / \beta_{ij^*}$
	19	Find the user to change its assignment and re-evaluate the corresponding Lagrange multiplier
	20	**if** $j^*=s$ **then**
	21	$I^*J^* = \operatorname{argmin}_{ij}\{\Delta\lambda_{i,j^*\to j}\}$, $I'J' = \operatorname{argmin}_{ij}\{\Delta\lambda_{i,j^*\to j}\}$, $I' \neq I^*$
	22	$\lambda_{j^*} \leftarrow \lambda_{j^*} + (\Delta\lambda_{I^*,j^*\to J^*} + \Delta\lambda_{I',j^*\to J'})/2$
	23	**else**
	24	$I^*J^* = \operatorname{argmin}_{ij}\{\Delta\mu_{i,j^*\to j}\}$, $I'J' = \operatorname{argmin}_{ij}\{\Delta\mu_{i,j^*\to j}\}$, $I' \neq I^*$
	25	$\mu_{j^*} \leftarrow \mu_{j^*} + (\Delta\mu_{I^*,j^*\to J^*} + \Delta\mu_{I',j^*\to J'})/2$
	26	$b_{I^*j^*} \leftarrow 0, b_{I^*J^*} \leftarrow 1$
	27	$\pi_{j^*} \leftarrow \pi_{j^*} - \alpha_{I^*j^*}, \tau_{j^*} \leftarrow \tau_{j^*} - \beta_{I^*j^*}$
	28	$\pi_{J^*} \leftarrow \pi_{J^*} + \alpha_{I^*J^*}, \tau_{j^*} \leftarrow \tau_{j^*} + \beta_{I^*J^*}$
Phase 2: Add	29	**while** more assignments can be exchanged **do**
	30	**for** $\{i=1:M\}$
	31	$j=\operatorname{argmax}_j\{b_{ij}=1\}$
	32	**for** $\{l=1:n_i\}$
	33	$\Delta u_{i,j\to l} \begin{cases} u_{il}-u_{ij} & \text{if} \quad u_{il}-u_{ij}>0, \pi_l+\alpha_{il}\le 1,\ \tau_l+\beta_{il}\le 1 \\ 0 & \text{otherwise} \end{cases}$
	34	Find the best exchangeable assignment: $I''J'' = \operatorname{argmax}_{ij}\{\Delta u_{i,j\to l}\}$ to move user I'' from BS j to BS J''
	35	$b_{I''j} \leftarrow 0, b_{I''J''} \leftarrow 1$
	36	$\pi_j \leftarrow \pi_j - \alpha_{I''j}, \tau_j \leftarrow \tau_j - \beta_{I''j}$
	37	$\pi_{J''} \leftarrow \pi_{J''} + \alpha_{I''J''}, \tau_{J''} \leftarrow \tau_{J''} + \beta_{I''J''}$
Phase 3: Relaxation	38	Each user i with $b_{ij}=0$ is assigned to the BS with max. weighted utility
	39	**for** $\{i \mid b_{ij}=0 \text{ for any } j\}$
	40	**for** $\{j=1:n_i\}$
	41	$w_{ij} = u_{ij} - \lambda_j \alpha_{ij} - \mu_j \beta_{ij}$
	42	$j=\operatorname{argmax}_j\{w_{ij}\}$
	43	$b_{ij} \leftarrow 1$
	44	Final BS assignment configuration $B=\{b_{ij}\}_{M\text{x}N}$

Fig. 2. Heuristic BS Assignment Algorithm

B. Description of the Algorithm

The BS assignment algorithm, shown in Fig. 2, consists of four phases, namely: initialization, drop, add and relaxation. Firstly, Lagrange multipliers are set to zero (line 01), and then resource costs and user utilities are computed (lines 02-04) for each user. In order to reduce the computational complexity, not all BSs are viewed as potential choices. Instead, each user i is assumed to have a candidate set, denoted as n_i, composed by the BSs having the highest channel gain. Then, an initial assignment is obtained by selecting the most valuable BS n for each user (line 05). The total radio and transport costs at each BS j, denoted by π_j and τ_j, respectively, are computed (lines 06-07) and the resource cost vector $\psi = \{\pi_1, \ldots, \pi_N, \tau_1, \ldots, \tau_N\}$ conformed (line 08). If the initial assignment is feasible (i.e., none of the elements of ψ is greater than 1.0), that is an optimal solution. Otherwise, the algorithm continues in the drop phase. Within the drop phase, Lagrange multiplier associated to the most offending constraint violation is repeatedly increased to force user reassignments till a solution not exceeding resource constraints is found. In each iteration of this phase, the BS j^* with the most offending constraint violation s is determined (line 10), where $j^* = s$ if $s = 1, \ldots, N$, and $j^* = s - N + 1$ otherwise. For each user i currently allocated to the BS j^* (line 12) the Lagrange multiplier increase of the most offending constraint s required to move user i from BS j^* to another BS j of its candidate set is computed (lines 12-18). This is done so that the weighted utility of user i at the overloaded BS j^*, w_{ij^*}, is decreased to a value less than or equal to the weighted utility on the candidate BS j, w_{ij}. Thus, if the most offending constraint violation at BS j^* is on transport resources, the increment to Lagrange multiplier μ_{j^*} should be such that:

$$(u_{ij^*} - \lambda_{j^*}\alpha_{ij^*} - (\mu_{j^*} + \Delta\mu_{i,j^*\to j})\beta_{ij^*}) \le (u_{ij} - \lambda_j \alpha_{ij} - \mu_j \beta_{ij}) \tag{28}$$

So, the increment $\Delta\mu_{i,j^*\to j}$ to the transport Lagrange multiplier can be computed as:

$$\Delta\mu_{i,j^*\to j} \ge \frac{u_{ij^*} - u_{ij} - \lambda_{j^*}\alpha_{ij^*} + \lambda_j \alpha_{ij} - \mu_{j^*}\beta_{ij^*} + \mu_j \beta_{ij}}{\beta_{ij^*}} \tag{29}$$

Similarly, the increment $\Delta\lambda_{i,j^*\to j}$ to the Lagrange multiplier of the radio constraint can be computed as:

$$\Delta\lambda_{i,j^*\to j} \ge \frac{u_{ij^*} - u_{ij} - \lambda_{j^*}\alpha_{ij^*} + \lambda_j \alpha_{ij} - \mu_{j^*}\beta_{ij^*} + \mu_j \beta_{ij}}{\alpha_{ij^*}} \tag{30}$$

where numerator in (29) and (30) is the increase of the weighted utility of user i, denoted as $\Delta w_{i,j^*\to j}$. For each BS j in the candidate set n_i of users currently allocated to BS j^*, the increase of the corresponding Lagrange multiplier is computed in lines 12-18. Then, as suggested in [25], the user I^* and candidate BS J^* causing the least increase of the corresponding multiplier is chosen for exchange (lines 19-25) as this choice minimizes the gap between the optimal solution characterized by (26) and the new assignment solution obtained at this point. However, if the multiplier increase is just computed as the equality as done in [25], important convergence problems arise since users tend to have the same weighted utility towards multiple BSs. To avoid this problem, we compute the increment to be added to the corresponding multiplier as the average between the least increase, corresponding to user I^* and BS J^*, and the second least increase obtained with user I' and BS J'. This choice guarantees that only one user is reassigned at each iteration and the next BS assignment solution is stable (equal weighted utilities due to the update of the multipliers are avoided). Furthermore, as the BS assignment problem could have no feasible solution (not enough resources to allocate all the users), condition (13) is relaxed at this point by allowing that a user i may not have allocated resources in any BS (i.e., resource costs and utility equal to zero). The reassignment of the selected user is performed (line 26), and radio and transport resource costs updated accordingly (lines 27-28). The process is repeated until a solution not exceeding resource constraints is determined, yet there may be some users not served by any BS. Solution arisen from the drop phase may not be the most efficient BS assignment configuration in terms of resource utilization as some BSs could still have available resources. Then, the solution is improved in the add phase by applying the following procedure. For each user i it is verified whether, amongst the BSs in its candidate set, there is an assignment option BS l that provides a higher utility ($u_{il} > u_{ij}$) than current assignment at BS j. The utility increment, denoted as $\Delta u_{i,j\to l}$, is computed in lines 30-33. Among all user assignments satisfying $\Delta u_{i,j\to l} > 0$, as well as radio and transport constraints of BS l, the user I'' causing the largest increase in the utility is selected for reassignment (line 34). The exchange is done in line 35 and costs associated with radio and transport constraints are updated in lines 36-37. This process is repeated until no more reassignments are possible. If the achieved solution after the add phase is a feasible solution (all users have been allocated and resources are not exceeded), the algorithm ends, otherwise the algorithm continues in the relaxation phase.

When a feasible solution cannot be found, users without allocated resources after the add phase would have to be dropped or not served temporarily (e.g., in case of a joint scheduling and BS allocation problem) in order to guarantee the minimum data rate requirements to the rest of served users. Alternatively, these users can be served at the expenses of allowing some degree of service degradation. This second approach is the one used in this work since it allows a fair comparison of the proposed algorithm with other strategies in Section V in terms of service degradation. Hence, a relaxation phase is considered after the add phase where users without allocated resources are finally allocated to the BS with the highest weighted utility w_{ij} among those of its candidate set. In any case, notice that, as a full load condition has been assumed for the computation of radio resource costs, the resulting BS assignment after the relaxation phase may not necessarily lead to service degradation when real load conditions are accounted. Hence, the output of the presented algorithm is always a complete BS assignment and its feasibility and level of service degradation caused by exceeding resource constraints is numerically assessed in Section V by considering accurate load and interference level estimations.

C. Complexity Analysis

The algorithm's complexity is determined in this section based on the analysis given in [25]. The initialization phase has a complexity of $O(2N+3Mn_i)$. In line 09, the `while` loop could be executed up to $O(Mn_i)$ times, as in each iteration one user can be changed from BS j^* to BS J^*. Inner loop (line 12) could perform up to n_i iterations for each user assigned to BS j^*, thus its complexity is $O(Mn_i)$. The increase of multipliers (lines 13-18) results in a complexity of $O(n_i)$. The complexity of lines 20-25 and lines 26-28 is $O(2Mn_i)$ and $O(3)$, respectively. Thus, in the worst case the complexity order of the drop phase is $O(M^2(n_i)^3+2M^2(n_i)^2+3Mn_i)$. In the add phase, the complexity of line 34 and lines 35-37 is $O(Mn_i)$ and $O(3)$, respectively. At line 30, for each user we have at most n_i BSs resulting in a complexity of $O(Mn_i)$. The complexity of line 32 is $O(n_i)$, while the `while` loop (line 29) is executed at most Mn_i times. Thus, the complexity of add phase is $O(M^2(n_i)^3+M^2(n_i)^2+Mn_i)$, while the complexity of third phase is $O(Mn_i)$. Therefore, the complexity order of the algorithm is given by $O(M^2(n_i)^3)$.

V. Performance Evaluation

In this section, we study the performance of the proposed BS assignment algorithm. The considered scenario is composed by 19 hexagonal cells (one central cell and its two concentric tiers). We consider three frequency channels with 20 MHz bandwidth and a frequency reuse pattern of 3. The maximum transmit power of each BS j is set to 47 dBm. The transport capacity of each BS j is expressed in terms of the transport capacity factor ϕ_j, assumed to be the same for all BSs so that herein we drop index j. The size of the candidate BS set is limited to seven. Users are uniformly distributed over the entire service area and all are assumed to have the same downlink data bit rate requirement $R^{\min}$. A maximum radio cost $\alpha_{ij}^{\max}$ is used to prevent that a user may consume an excessive share of overall BS radio resources to meet its requirement. So, the expected data bit rate for user i, denoted as R_i, is always limited by $R_i \leq \min(R^{\min}, R_{ij}^{\max}\alpha_{ij}^{\max})$. In the presented analysis, it's considered that the BS decision-making process is able to follow channel variations due to propagation path loss and slow shadowing changes. Hence, minimum user bit rate requirements and resource costs considered in the algorithm would represent average values taken over the time scale dictated by long-term channel variations (i.e., few hundreds of milliseconds). Under such an approach, the mean channel gain in each subcarrier k from BS j to user i, referred to as $G_{i,j,k}$, is the same for all subcarriers. Consequently, the computation of $\mathrm{SINR}_{i,j,k}$ according to (1) leads also to the same value for all subcarriers, namely $\mathrm{SINR}_{i,j}$, since the interference levels are assumed to be uniformly distributed over the entire bandwidth, as argued in section III-A and captured by (16). So, upon the average $\mathrm{SINR}_{i,j}$, the MCS, and, consequently the corresponding achievable rate at the air interface, are taken from the look-up table provided in Table I. Notice that the approach adopted in this work does not preclude the applicability of the proposed algorithm in a problem also tackling fast fading fluctuations in the channel gain, e.g., a

TABLE I
MCS THRESHOLDS AND PHY DATA RATES.

#	Modulation	Coding	$\mathrm{SINR}_{\min}$ [dB]	PHY data rate [Mbps]
1	BPSK	1/2	3.4	6.99
2	QPSK	1/2	6.4	13.99
3	QPSK	3/4	8.2	20.99
4	16 QAM	1/2	13.4	27.98
5	16 QAM	3/4	15.2	41.98
6	64 QAM	2/3	19.7	55.97
7	64 QAM	3/4	21.4	62.97

TABLE II
OFDMA SYSTEM PARAMETERS.

Parameter	Value
Total number of cells, N	19
Max. BS transmit power, $P_j^{\max}$	47 dBm
Transmit antenna gain	18.7 dBi
Cell radius	1060 m
Antenna pattern	Omnidirectional
Operating frequency	2500 MHz
Reuse factor	3
Number of channels	3
Channel bandwidth	20 MHz
Number of data subcarriers, K_j	1440
OFDM symbol duration	102.9 μs
Path loss model	COST-231 Hata
BS height	32 m
Mobile terminal height	1.5 m
Shadowing standard deviation	8 dB
Shadowing correlation	50%
Shadow fade margin	13.2 dB
Thermal noise	-174 dBm/Hz
Receiver noise figure	7 dB
User rate requirements, $R^{\min}$	600, 1200, 1800 2400 Kbps
Maximum radio cost, $\alpha_{ij}^{\max}$	0.2

joint scheduling and BS assignment problem[8]. As well, it is worth noting that bit rate values provided in the Table I could also account for any performance gain associated with the usage of mechanisms exploiting (subcarrier) frequency selectivity as well as multi-user diversity that would operate at shorter time scales than that considered for the BS assignment process. Propagation losses are computed using the COST-231 Hata model [28] with parameters as given in Table II. Lognormal shadowing is accounted with standard deviation equal to 8 dB and 50% spatial correlation. The radius of the cell has been chosen so that a signal to noise ratio $\mathrm{SNR}_{\mathrm{req}}$=3.4 dB is assured at the cell border with a probability of 95%, considering typical sample link budgets for mobile broadband systems [28]. All system parameters are presented in Table II.

A. Evaluated BS Assignment Algorithms

We evaluate three BS assignment schemes: 1) *Algorithm A*, is based on the common MPL scheme, which under full radio load conditions would also be equivalent to an algorithm that assigns a user to the BS that provides the highest SINR; 2) *Algorithm B* constitutes a load balancing scheme exclusively based on radio load information, implemented by only considering radio resource constraints in the

[8]This alternative approach is out of the scope of the current work that mainly tries to expose the benefits (or needs) to incorporate both radio and transport information in the ordinary BS assignment problem.

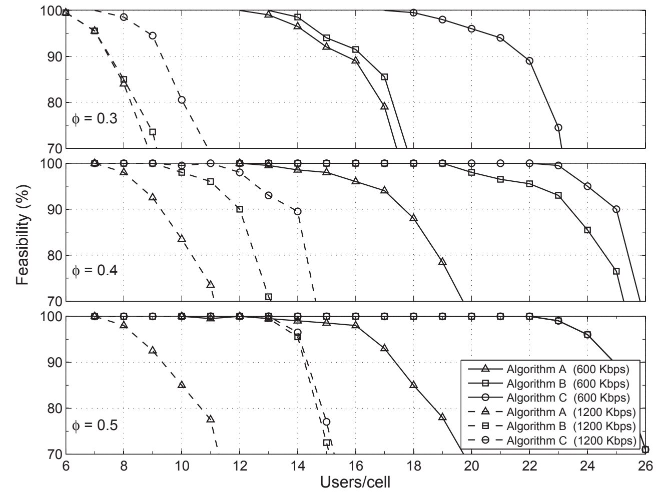

Fig. 3. Percentage of feasible solutions found by each BS assignment algorithm under different transport capacity factors $\phi = \{0.3, 0.4, 0.5\}$ and data rate requirements $R^{\min}$={600, 1200 Kbps}.

heuristic algorithm shown in Fig. 2; and 3) *Algorithm C* considers both radio and transport load of BSs to determine the BS assignment solution, and is realized by means of the proposed heuristic algorithm. Over such a basis, for a given snapshot of the system (i.e., random distribution of M users in the service area), a BS assignment solution is computed with each algorithm assuming full load conditions. Then, for each obtained solution, performance metrics discussed in next section are computed considering a more accurate estimation of load and interference levels than the one provided assuming full load conditions. This can be achieved because once the BS assignment has been fixed it is possible to compute the power levels by means of a recursive approach such as the one proposed in [29]. This step is needed to allow a fair comparison of the different schemes. For each analyzed case, results are obtained over 10000 different snapshots.

B. Numerical Results

Fig. 3 presents the percentage of feasible solutions (i.e., all users assigned without service degradation) that each algorithm is able to achieve attending to the mean number of users per cell and considering different minimum rate requirements and transport capacity factors. As shown in Fig. 3, *Algorithm A*'s performance is always quite poor when compared to load aware schemes. On the other side, *Algorithm C* clearly achieves the highest number of feasible solutions by exploiting both radio and transport load balancing. Notice that, only for transport capacity factors equal to or higher than half the value of the radio peak rate ($\phi \geq 0.5$), *Algorithm B* converges to *Algorithm C* for the considered user bit rates.

Over such a basis, Fig. 4 provides the maximum number of users per cell supported by each algorithm when targeting a percentage of feasible solutions equal to 90% (i.e., a BS solution satisfying all user rate requirements and BS resource constraints is found in the 90% of the snapshots). Results are obtained for transport capacities $0.3 \leq \phi \leq 0.6$ and minimum rate requirements $R^{\min}$={600, 1200, 1800, 2400 Kbps}. Notice that minimum rate requirements are between

1% and 4% when normalized to the BS peak rate[9]. As shown in Fig. 4, the relative number of users that can be successfully allocated by algorithms *B* and *C* in front of *Algorithm A* is very noticeable for any transport capacity, specifically under high data rate requirements. For instance, for $R^{\min}$=2400 Kbps and transport capacity factor $\phi = 0.5$, see Fig. 4(b), algorithms *B* and *C* provide capacity gains of 75% and 100%, respectively, over *Algorithm A* (i.e., 4, 7 and 8 users per cell achieved by algorithms *A*, *B*, and *C*, respectively). Under the same transport capacity but a lower data rate requirement $R^{\min}$=1200 Kbps, algorithms *B* and *C* both can both support 14 users per cell, in front of 9 users per cell supported *Algorithm A*, which turns into a capacity gain of 56% over *Algorithm A*. Then, under the more limited transport capacity conditions (i.e., $\phi < 0.5$) and the higher data rate requirements, the more is the capacity gain achieved by *Algorithm C*, or, equivalently, the less is the transport capacity needed to support the same number of users in the system. Fig. 4(b) shows, for instance, that in order to support 8 users/cell with a data rate requirement of 1800 Kbps (i.e., a total aggregated rate of 14.4 Mbps), *Algorithm B* requires around 28 Mbps of backhaul capacity to meet the considered network availability. On the other hand, the level of backhaul resources needed in this case by *Algorithm C* is around 22 Mbps, which turns into a capacity reduction of about 28% in respect to *Algorithm B*.

When a feasible BS assignment solution cannot be found, users assigned to overloaded BSs will suffer a service degradation since the provided data rate would be lower than the minimum expected. The extent of such degradation is quantified here by considering that a BS exceeding its radio and/or transport resources proportionally reduces the bit rate allocated to each served user. Fig. 5 illustrates the cumulative distribution of the allocated data rate for a transport capacity factor $\phi = 0.3$, a distribution of 12 users per cell and a data rate requirement $R^{\min}$=1200 Kbps. It is shown that *Algorithm A* exhibits the highest degradation, so the requested minimum data rate is only guaranteed to a 74% of the total users. Conversely, the degradation is less pronounced for algorithms *B* and *C*, which can provide the minimum rate requirement to around 80% and 90% of users, respectively. In this context, in Table III we extend previous results by considering different transport capacity factors $\phi = \{0.3, 0.4, 0.5\}$, data rate requirements $R^{\min}$={1200, 2400 Kbps}, and traffic load conditions (e.g., mean aggregated rates of 14.4, 19.2 and 24.0 Mbps). Each table cell provides the percentage of users receiving the minimum requested data rate (upper row), and the percentage of users receiving at least 90% of the requested data rate (bottom row). It is shown that for a mean aggregated rate of 19.2 Mbps at BSs, with $R^{\min}$=1200 Kbps and $\phi = 0.4$, *Algorithm C* guarantees that 89.6% of users receives the minimum requested rate, whereas algorithms *B* and *A* lead to 83.2% and 68.2% of fully satisfied users, respectively. At higher rate requirements, but same mean aggregated rate per BS (i.e., 19.2 Mbps) and transport capacity factor, *Algorithm C* even achieves better performance, where 88.2% of users are fully satisfied in front of 76.3% and 68% for algorithms *B* and *A*, respectively. This is because, under the same aggregated rate, for a higher minimum rate requirement, less users can be allocated in the overall system and, the less the number of users supported per BS, the more important becomes the need to account for (radio and transport) load balancing schemes to properly distribute traffic among neighbouring BSs. Notice that focusing on the percentage of users receiving at least 90% of the requested rate, similar trends are obtained but differences are less noticeable between algorithms *C* and *B* (2% and 3% for previous considered cases), but still quite significant compared to *Algorithm A* (above 10%).

[9]In fact, the analysis provided in this paper would be valid for different BS peak rate whenever the normalized transport capacity (ϕ) and normalized minimum rate requirement are kept.

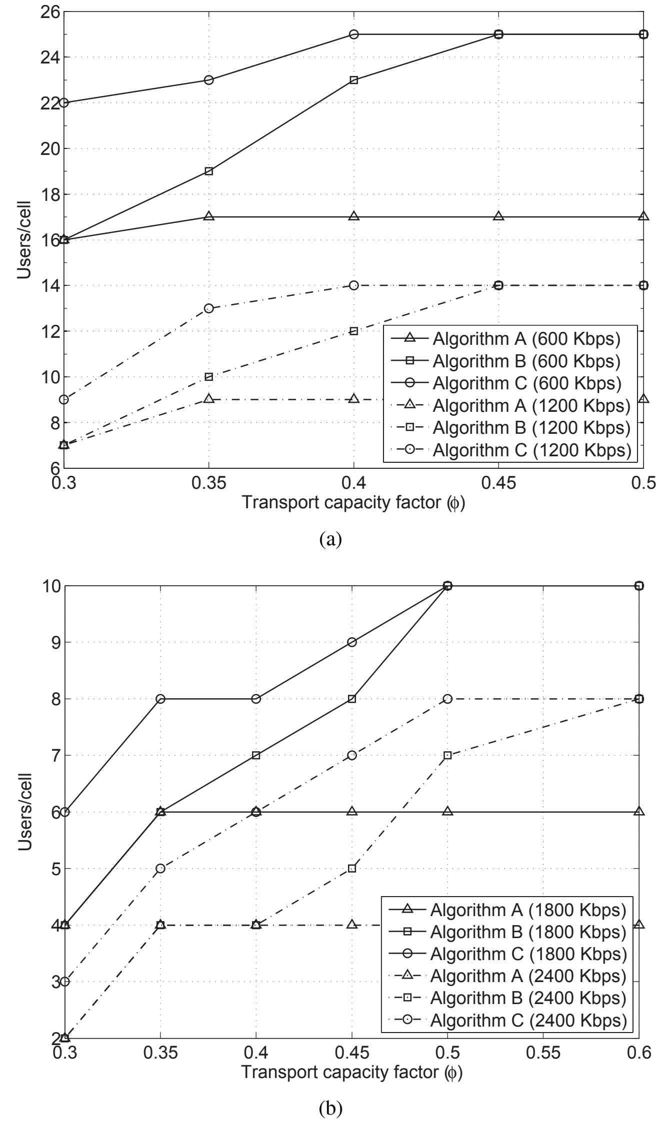

Fig. 4. Users/cell supported by each algorithm so that a BS assignment solution can be found in 90% of the snapshots, considering different data rate requirements (a) $R^{\min}$={600, 1200 Kbps}; and (b) $R^{\min}$={1800, 2400 Kbps}.

So far, it has been demonstrated the performance gain that can effectively be attained by *Algorithm C* with respect to algorithms *A* and *B*. We now examine in more detail how each algorithm impacts on the underlying radio and transport resource consumption. We consider a distribution of 8 users per cell, a transport capacity factor in the range $0.3 \leq \phi \leq 0.6$ and a minimum data rate requirement $R^{\min}$=2400 Kbps. Fig.

TABLE III
PERCENTAGE OF USER SATISFACTION UNDER DIFFERENT MEAN OF USERS PER CELL AND TRANSPORT CAPACITY CONDITIONS

User rate (Kbps)	Users/cell	Mean aggregated rate (Mbps)	Algorithm A			Algorithm B			Algorithm C		
			ϕ=0.3	ϕ=0.4	ϕ=0.5	ϕ=0.3	ϕ=0.4	ϕ=0.5	ϕ=0.3	ϕ=0.4	ϕ=0.5
1200	12	14.4	74.6% 87.4%	88.9% 96.5%	89.9% 96.5%	78.1% 91.3%	96.1% 98.9%	97.0% 98.9%	92.7% 98.2%	96.7% 98.9%	97.0% 98.9%
	16	19.2	33.8% 53.7%	68.2% 81.5%	74.0% 83.5%	34.5% 58.8%	83.2% 94.6%	91.9% 96.2%	38.1% 75.3%	89.6% 95.6%	92.3% 96.2%
	20	24.0	3.3% 17.4%	34.2% 52.3%	42.8% 57.9%	4.6% 17.9%	36.6% 65.2%	66.3% 75.3%	10.0% 31.0%	42.6% 69.8%	66.5% 75.6%
2400	6	14.4	63.2% 78.2%	87.3% 89.6%	90.8% 92.3%	67.5% 79.0%	91.3% 94.2%	95.8% 95.8%	78.3% 93.0%	94.5% 95.4%	95.8% 95.8%
	8	19.2	32.8% 52.6%	68.0% 80.1%	78.7% 84.0%	36.1% 52.6%	76.3% 88.4%	92.0% 92.4%	50.4% 84.0%	88.2% 91.4%	92.0% 92.4%
	10	24.0	10.4% 21.5%	42.2% 55.1%	57.0% 68.0%	11.4% 24.6%	45.9% 62.6%	79.4% 86.9%	14.15% 53.1%	59.2% 78.6%	80.7% 87.7%

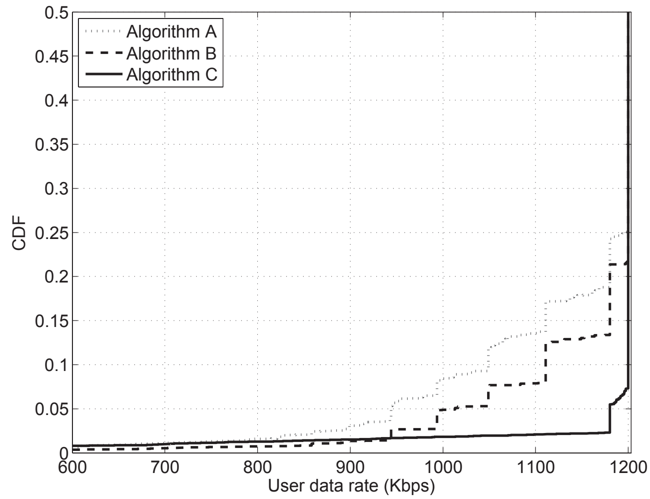

Fig. 5. Cumulative distribution function (CDF) of allocated data rate for a distribution of 12 users/cell, data rate requirement $R^{\min}$=1200 Kbps, and transport capacity factor $\phi = 0.3$.

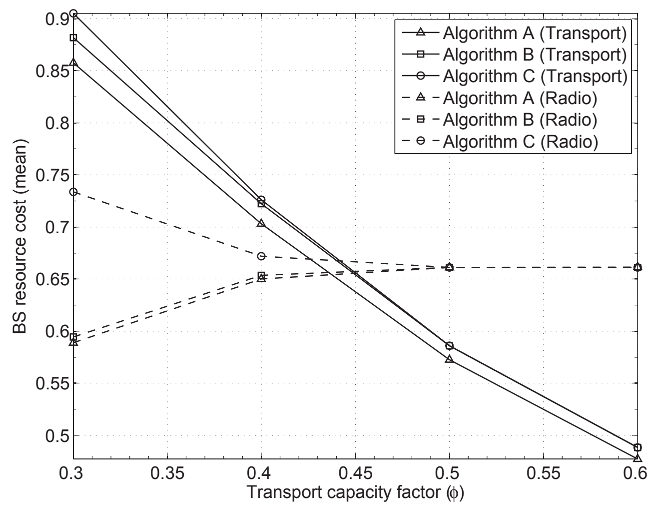

Fig. 6. Mean of BS radio and transport resource costs for data rate requirement $R^{\min}$=2400 Kbps under different transport capacity conditions and a distribution of 8 users/cell.

6 shows the mean value of BS radio and transport resource costs incurred by each algorithm over all obtained snapshots. We observe that, as expected, the performance gain achieved by *Algorithm C* is realized at the expenses of a higher usage of BS radio resources. Specifically, BS mean radio resource costs with *Algorithm C* are over 23% higher than with *Algorithm B* in the most restricted transport condition (i.e., $\phi = 0.3$). However, it's worthwhile to note that for $\phi = 0.4$ where, according to Fig. 4, *Algorithm C* allows to accommodate up to 6 users/cell in front of 4 users/cell achieved by algorithms *A* and *B*, the mean radio resource cost of *Algorithm C* is only 4% higher than the other strategies (i.e., *Algorithm C* uses 0.675 while the others 0.65). So, *Algorithm C* is able to use this, otherwise unused, radio resources to wisely steer traffic and avoid transport limitations. As a result, a slightly higher transport resource utilization is obtained with *Algorithm C* as it leads to lower data rate degradation. Finally, performance gains achieved by *Algorithm C* are also tightly coupled with its capability to distribute traffic load in a smooth manner among BSs. This fact can be noticed in Fig. 7 that presents the coefficient of variation of radio/transport resource costs (defined as the ratio of standard deviation of radio/transport costs to the mean of radio/transport costs in all BSs). It can be seen that coefficients of variation of *Algorithm C* are always lower than those obtained by the other two algorithms. Lastly, based on the provided complexity analysis, it can be determined that the add and drop phase represents around 47% and 52%, respectively, of the total complexity of *Algorithm C*, while remaining corresponds to phases 0 and 3. In analyzed scenarios the add phase provides a performance enhancement of around 8% with respect to the drop phase. Furthermore, the number of iterations required for the algorithm to converge is less than 800 iterations in the 95% of the computed snapshots.

VI. CONCLUDING REMARKS

In this paper, a BS assignment algorithm for OFDMA-based systems has been proposed. Unlike most of the existing approaches, the proposed algorithm accounts for potential backhaul network constraints in the BS decision making process. We have shown that, in scenarios with limited transport capacity (i.e., where the transport capacity is less than half of the peak rate in the radio interface), the proposed algorithm brings out significant gains with respect to algorithms that are completely based on radio criteria in terms of feasible BS assignment solutions found, and percentage of users meeting their minimum bit rate requirement. We claim that the pro-

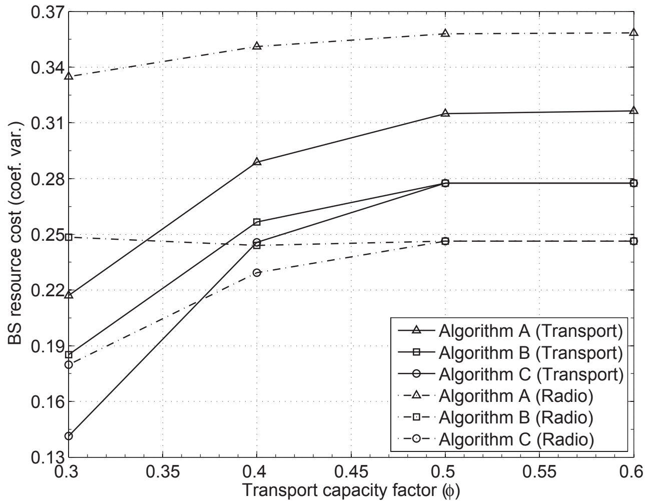

Fig. 7. Coefficient of variation of BS radio and transport resource costs for data rate requirement $R^{\min}$=2400 Kbps under different transport capacity conditions and a distribution of 8 users/cell.

posed algorithm can be used to alleviate potential transport capacity restrictions in cellular system deployments. As future work, we aim to assess the performance of the proposed *backhaul-aware* BS assignment algorithm under partially limited backhaul scenarios (i.e., not homogeneous transport capacity limitations) and under different spatial distribution of users over the service area (i.e., hot-spots). As well, motivated by the adoption of this kind of strategies within flat network architectures (e.g., LTE), a distributed implementation of the proposed algorithm is envisaged based upon a dynamic pricing approach aligned to that proposed in [12].

ACKNOWLEDGMENTS

This work has been supported by the Spanish Research Council (CICYT) and FEDER funds under COGNOS (ref. TEC2007-60985) and ODISSEY grants (ref. TEC2006-09109). The support from the Mexican National Council of Science and Technology (CONACYT) under grant 187469 is hereby acknowledged. Authors are very thankful to the reviewers for their constructive comments.

REFERENCES

[1] G. Li and H. Liu, "Downlink radio resource allocation for multi-cell OFDMA system," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3451-3459, Dec. 2006.

[2] S. Pietrzyk and G. Janssen, "Radio resource allocation for cellular networks based on OFDMA with QoS guarantees," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM'04)*, vol. 4, Nov. 2004, pp. 2694-2699.

[3] S. Chia, M. Gasparroni, and P. Brick, "The next challenge for cellular networks: backhaul," *IEEE Microw. Mag.*, vol. 10, no. 5, pp. 54-66, Aug. 2009.

[4] A. Bolle and H. Herbertsson, "Backhaul must make room for HSDPA," *Wireless Europe*, vol. 43, pp. 17-18, Mar. 2006.

[5] A. Cauvin, *Next Generation Mobile Networks Optimised Backhaul Requirements*. NGMN Alliance, Aug. 2007.

[6] K. Ayagari and J. Tang, *Backhaul Solutions for 3G Networks*. Alcatel-Lucent White Paper, 2007.

[7] J. Lakkakorpi and A. Sayenko, "Backhaul as a bottleneck in IEEE 802.16e networks," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM'08)*, Nov. 2008, pp. 1-6.

[8] S. Nadas, S. Racz, Z. Nagy, and S. Molnar, "Providing congestion control in the IUB transport network for HSDPA," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM'07)*, Nov. 2007, pp. 5293-5297.

[9] 3GPP, IUB/IUR congestion control (Release 7), TR 25.902 V7.1.0, Mar. 2007.

[10] C. Saraydar, N. Mandayam, and D. Goodman, "Pricing and power control in a multicell wireless data network," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 1883-1892, Oct. 2001.

[11] V. Lau, "On the macroscopic optimization of multicell wireless systems with multiuser detection and multiple antennas–uplink analysis," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1388-1393, July 2005.

[12] J.-W. Lee, R. Mazumdar, and N. Shroff, "Joint resource allocation and base-station assignment for the downlink in CDMA networks," *IEEE/ACM Trans. Networking*, vol. 14, no. 1, pp. 1-14, Feb. 2006.

[13] D. Amzallag, R. Bar-Yehuda, D. Raz, and G. Scalosub, "Cell selection in 4g cellular networks," in *Proc Int'l IEEE Conf. Comput. Commun. (INFOCOM'08)*, Apr. 2008, pp. 700-708.

[14] Y. J. Zhang and K. Letaief, "Multiuser adaptive subcarrier-and-bit allocation with adaptive cell selection for ofdm systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1566-1575, Sep. 2004.

[15] I. Stiakogiannakis, D. Zarbouti, G. Tsoulos, and D. Kaklamani, "Subcarrier allocation algorithms for multicellular OFDMA networks without channel state information," in *Proc. Int'l Symp. Wireless Pervasive Comput. (ISWPC'08)*, May 2008, pp. 73-77.

[16] C. Koutsimanis and G. Fodor, "A dynamic resource allocation scheme for guaranteed bit rate services in OFDMA networks," in *Proc. IEEE International Conf. Commun. (ICC'08)*, May 2008, pp. 2524-2530.

[17] C. Wengerter, J. Ohlhorst, and A. von Elbwart, "Fairness and throughput analysis for generalized proportional fair frequency scheduling in OFDMA," in *Proc. IEEE Veh. Technol. Conf. (VTC'05 Spring)*, vol. 3, May 2005, pp. 1903-1907.

[18] L. Jorguseski, T. Le, E. Fledderus, and R. Prasad, "Downlink resource allocation for evolved UTRAN and WiMAX cellular systems," in *Proc. Int'l Symp. Personal Indoor Mobile Radio Commun. (PIMRC'08)*, Sep. 2008, pp. 1-5.

[19] W.-H. Kuo and W. Liao, "Utility-based radio resource allocation for QoS traffic in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2714-2722, July 2008.

[20] K. Navaie and H. Yanikomeroglu, "Optimal downlink resource allocation for non-real time traffic in cellular CDMA/TDMA networks," *IEEE Commun. Lett.*, vol. 10, no. 4, pp. 278-280, Apr. 2006.

[21] G. Chen, S. Khan, K. F. Li, and E. Manning, "Building an adaptive multimedia system using the utility model," in *Proc. Int'l Workshop Parallel Distributed Realtime Syst.*, 1999.

[22] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 171-178, Feb. 2003.

[23] H. Kellerer, U. Pferschy, and D. Pisinger, *Knapsack Problems*. Springer-Verlag, 2004.

[24] M. Michrafy, M. Hifi, and A. Sbihi, "Heuristic algorithms for the multiple-choice multidimensional knapsack problem," *J. Operational Research Society*, vol. 55, no. 12, pp. 1323-1332, Dec. 2004.

[25] M. Moser, D. P. Jokanovic, and N. Shiratori, "An algorithm for the multidimesional multiple-choice knapsack problem," *IEICE Trans. Fundam. Electron., Commun.*, vol. E80-2, no. 3, Mar. 1997.

[26] H. Everett III, "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," *Operations Research*, vol. 11, pp. 399-417, 1963.

[27] M. Katoozian, K. Navaie, and H. Yanikomeroglu, "Optimal utility-based resource allocation for OFDM networks with multiple types of traffic," in *Proc. IEEE Veh. Technol. Conf. (VTC'08 Spring)*, May 2008, pp. 2223-2227.

[28] K. Ramadas and R. Jain, *WiMAX System Evaluation Methodology (Tech. Report)*. Wimax Forum, 2007.

[29] R. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341-1347, Sep. 1995.

Hiram Galeana-Zapién received the Electronics and Telecommunications Engineering degree from the Universidad Autónoma de Nuevo León, Monterrey, México, in June 2001. From February 2001 to September 2004 he worked at Avantel, a Mexican fixed-line and internet carrier. In October 2004, he joined the Radio Communication Research Group in the Department of Signal Theory and Communications at the Universitat Politècnica de Catalunya, where he is currently working toward the PhD degree in communications. His current research interests are focused on the field of mobile radio communication systems, with a special emphasis on radio resource management solutions. He is a student member of the IEEE.

Ramon Ferrús received the Telecommunications Engineering (BS plus MS) and PhD degrees from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 1996 and 2000, respectively. He is currently an Associate Professor with the Department of Signal Theory and Communications, UPC. His actual research interest is focused on architectures, QoS, mobility and radio and spectrum resource management in the context of heterogeneous wireless communications systems. He has participated in several research projects within the 6th and 7th Framework Programmes of the European Commission, as well as in research and technology transfer projects for private companies. He is co-author of 50+ papers published in international journals, magazines and conference proceedings.