Performance Analysis of Transport and Radio Load Balancing Strategies for BS Assignment in Mobile Access Networks

F. Novillo, H. Galeana, A. Lainz*, R. Ferrús, J. Olmos
Signal Theory and Communications Department
Universitat Politècnica de Catalunya (UPC), Barcelona, Spain.
Email: {fnovillo, hiram.galeana, ferrus, olmos}@tsc.upc.edu, *aferr4197@alu-etsetb.upc.edu

Abstract—In mobile access networks, base station (BS) assignment mechanisms are mainly driven by radio conditions since it is generally assumed that the limiting factor is the air interface. This assumption has been proven to be very reasonable when circuit voice was the dominant service and backhaul capacity provisioning accounting for the peak rate at the BSs was an economically feasible option. However, as more efficient air interface technologies are introduced along with higher data rate services, a growing concern is that the transport part of the mobile network can also represent a bottleneck. In this paper we analyze a BS assignment strategy that simultaneously exploits load balancing in both the radio interface and the transport backhaul links. The BS assignment problem is solved by means of Simulated Annealing heuristics under different deployment scenarios comprising partially-limited backhaul Radio Access Networks (RANs) and the presence of traffic hotspots. Results achieved for the proposed BS assignment strategy are compared to those obtained by common strategies such as minimum path loss (MPL) and load balancing radio (LBR).

Keywords—Backhaul, base station assignment, radio access network.

I. INTRODUCTION

The mobile backhaul network is the infrastructure that interconnects base stations (BS) with network controllers or switching equipments in the core network within a radio access network (RAN). The majority of the 2G/3G backhaul networks are formed by means of point-to-point T1/E1 links, typically employing star/tree topologies, and microwave radio as the most common transmission technology. It is widely recognized that the backhaul represents one of the major contributors to the high cost of building out and running a mobile network since it is nearly one-third of the total network operating cost [1]. During the early stages of 3G systems, operators reused as much as possible existing infrastructure in order to minimize the cost of the transition from 2G to 3G. Nowadays, however, as new air interface technologies are deployed, bandwidth demand in mobile backhaul networks is drastically growing. For instance, the move towards enhanced air interface technologies, such as High Speed Data Packet Access (HSDPA), means that cell sites are likely to increase about seven times their current capacity requirements [1]-[3]. Therefore, operators are currently challenged to look for viable ways to optimize their backhaul networks in order to reduce costs, alleviate technical and operational complexities, and enable the faster rollout of new services. In this sense, a strong effort is being devoted to progress towards innovative packet-based network architectures where IP (Internet Protocol) technologies serve as cornerstone. In this context, commercial solutions from different vendors enable mobile operators to aggregate and compress backhaul traffic in order to make more efficient use of available transport resources in the RAN. In addition, new transmission technologies are positioning as promising proposals for interconnecting RAN network elements in the backhaul. This is the case of multihop mesh networks that use short high bandwidth optical wireless links to interconnect network entities [4]. It is worthy to remark that it is believed that fiber access will take a larger role in many networks in the future. However, deploying fiber to each single cell site in the backhaul will be very costly and challenging, even in dense urban areas where fiber is most prevalent [3].

Taking into account the above context, it should not be precluded the idea that bottlenecks in practical RAN deployments may arise not only at the air interface but also due to resource limitations in the backhaul network [5], [6]. In this paper, we propose a BS assignment strategy for mobile access networks where potential backhaul constraints are considered in the process of assigning the most suitable BS to mobile users. This leads to a new paradigm where transport resources are taken into consideration not only at the network dimensioning stage but are included in an integrated and dynamic resource management framework. In this sense, it will be shown that employing backhaul-aware BS assignment strategies the impact of a limited backhaul capacity in a given BS can be diminished by conveniently re-allocating some users to other BSs with available resources in both air interface and transport network links. In particular, we focus our analysis on the provisioning of high-rate delay-sensitive services where it is mandatory to guarantee a given data rate in both air interface and transport network simultaneously.

The rest of the paper is organized as follows. Section II discusses the motivation of the proposed approach and presents the generic framework where BS assignment strategies considering backhaul constraints can be applied. Then, in section III, the system model and assumptions considered to assess the performance of transport-aware BS assignment strategies are considered. Section IV presents and analyzes simulation results. Finally, conclusions are drawn in Section V.
strategies are introduced. Afterwards, section IV provides details about the used evaluation methodology based on heuristics and utility functions, and describes how the considered BS assignment strategies are modeled. Section V discusses the simulation results and, finally, section VI draws main concluding remarks.

II. MOTIVATION AND GENERIC FRAMEWORK

In a typical RAN deployment the infrastructure involved in the backhaul network is very complex since there could be hundreds of cell sites that need to be interconnected using different topologies and transmission technologies. Furthermore, a RAN deployment is likely to have multiple cell coverage in some locations of the service area, and therefore some mobile users could have more than one candidate BS to be connected to. In this context, there have been an important number of papers that have studied the problem of BS assignment in RANs, but mainly in terms of air interface resource optimization (e.g. see [7]-[10] for CDMA networks) and not considering potential limitations in the mobile backhauling. Next generation cellular networks will be required to provide high data-rate services to the end-users. This will result in smaller cells, translating to additional cell site requirements to cover a given area which in turn will require more backhaul support. Hence, in this paper we propose to introduce a BS assignment strategy that, besides radio optimization, also takes into account transport resource consumption in the backhaul links. It is worth noting that when addressing BS assignment strategies that consider criteria other than radio, the main challenge is to keep under control the amount of degradation of the radio interface due to not always connecting users to their “best” radio serving BS (e.g. increased path loss and higher interference level) so that the overall performance can be definitively enhanced (e.g. higher number of connections can be served satisfying both radio and transport constraints). In this context, we focus in a homogeneous RAN scenario with a single radio access technology (RAT) and frequency reuse factor of one, which is claimed to be the most critical in terms of using information other than radio metrics to control the BS assignment process. Notice that in scenarios with multiple frequency layers or heterogeneous RAT’s, mobile users can be assigned to a given frequency layer or RAT considering backhaul constraints as well, but the total or partial decoupling of the radio resource pools used in each frequency/RAT could make this decision less critical in terms of radio degradation.

III. SYSTEM MODEL

In view of the arguments outlined in previous sections, the performance analysis of backhaul-aware BS assignment strategies is addressed considering a 3GPP WCDMA (FDD) network. Specifically, we focus on the downlink because it is usually considered the more restrictive link due to the asymmetric bandwidth demand between the downlink and the uplink data services [9]. The network consists of \(N\) BSs that cover a geographical area in which, at a given instant, there are \(M\) active users that have to be allocated to BSs. It is assumed that resources in any BS in the system in the downlink are constrained by two factors: the maximum power limit in the radio interface and the provisioned capacity in the backhaul network. The system state is characterized by a \(M \times N\) matrix, hereafter referred to as \(B = \{b_{ij}\}\), that denotes the BS assignments at a given instant. In particular, \(b_{ij} = 1\), if BS \(j\) is assigned to user \(i\), and \(b_{ij} = 0\), otherwise. A given \(B\) matrix is considered as a feasible solution if radio and transport constraints are fulfilled. Details of these constraints are given in the following.

A. Air Interface Formulation

In the downlink of a WCDMA air interface, the required transmitted power \(P_{ij}\), for user \(i\) being served by BS \(j\) can be expressed as [11]:

\[
P_{ij} = \frac{R}{W} \left( \frac{E_b}{N_0} \right)_{\text{min},i} \left(1-\alpha_i \right) + \sum_{i=1}^{K} L_{ij} - P_k + L_{ij} - p_k \tag{1}\]

where \((E_b/N_0)\), is the minimum bit energy over noise power spectral density requirement, \(P_{ij}\) is the required transmit power devoted to user \(i\) being served by BS \(j\), \(P_k\) is the noise power at the user terminal, \(R\) is the bit rate for the user \(i\), \(W\) is the chip rate, \(P_k\) is the total transmit power of BS \(k\), \(L_{ij}\) is the path loss between BS \(k\) and user \(i\), and \(\alpha_i\) is the orthogonality factor seen by user \(i\) (\(\alpha_i = 1\) means perfect orthogonality). From (1), the required total BS transmission power can be obtained by summing up the power of each served individual user and the radio constraint is formulated as:

\[
P_j = \sum_{i=1}^{M} P_{ij} \leq P_{j,\text{max}} \quad j = 1...N \tag{2}\]

Solving (2) for a fixed BS assignment is a well studied problem so that feasibility conditions and optimal power allocation can be obtained following the algorithm described in [12]. Nevertheless, it is also worth noting that when focusing on the joint power control and BS assignment problem there is not always a Pareto optimal power vector in the downlink as is the case in the uplink.

B. Transport Capacity Formulation

In practical RAN deployments, the transport capacity provisioned for a given BS is normally dimensioned in accordance to the amount of traffic that this BS can serve over the air interface. A commonly used approach to estimate the air interface downlink capacity is based on the computation of the downlink load factor \(n_{DL}\) defined as [13]:

\[
n_{DL} = \frac{K}{W} \frac{R}{W} \frac{E_b}{N_0} \left((1-\alpha_i) + f_{DL,i}\right) \tag{3}\]

where \(K\) is the number of users served by a given BS, \(f_{DL,i}\) is the other-to-own cell received power ratio for the \(i\)-th user at the position where it is located. This means that as the load factor move towards one, the downlink capacity approaches to its maximum pole capacity value. Thus, the maximum
capacity over the air interface $C_{air}$ in bits/s is achieved when the load factor tends to unity. Over such a basis, focusing on one important special case where all mobile users have similar characteristics (i.e. service type, bit rate and $E_b/N_0$ requirements), it is easy to show that the maximum value of $C_{air}$ can be estimated using the following expression:

$$C_{air} = K \cdot R \leq \frac{W}{\sum_{i=1}^{K} E_i / N_{i,\text{min}} \cdot ((1-\alpha) + f_{ia})}$$

(4)

where $\alpha = \frac{1}{K} \sum_{i=1}^{K} \alpha_i$ is the average orthogonality factor in the cell and $f_{ia} = \frac{1}{K} \sum_{i=1}^{K} f_{ia,i}$ is the average ratio of other-to-own cell BS power received by users. In our analysis, the transport capacity $C_{trans}$ of BS is related to the air interface pole capacity $C_{air}$ by means of a multiplicative factor $\beta$ as shown below:

$$C_{trans} = \beta \cdot C_{air}$$

(5)

A value of $\beta = 1$ would mean that the transport capacity has been dimensioned to satisfy the downlink air pole capacity estimated in the planning process. Hence, the transport constraint of a given BS $j$ in the system is expressed as:

$$\sum_{i \in \{P_{i,j} = 1\}} R_i \leq C_{trans}$$

(6)

where $R_i$ is the bit rate required for user $i$.

C. Methodology

The analysis is conducted using the “snapshot” technique [14]. In a given snapshot of the system there are $M$ active users, and the objective is to find a feasible BS assignment so that constraints (2) and (6) can be fulfilled. In order to look for feasible assignments, for each user-BS combination, a utility function is defined to express the degree of fulfillment to the constraints. Then, a system utility $U$ is defined as the summation of the utilities of all assigned users in the system. In our analysis, the considered utility functions are monotonically decreasing and concave functions, although different forms of expressing utilities are possible [15]. In these types of functions, the absolute value of its derivative progressively increases as moving towards a condition of minimum utility. Conversely, these functions exhibit softer variations when they are close to the region of maximum utility. Over such a basis, a simulated annealing-based algorithm targeted to maximize system utility $U$ is used to find a feasible assignment. Repeating this process for a large number of snapshots allow us to determine the average number of users that can be assigned correctly by a particular BS assignment strategy.

IV. BASE STATION ASSIGNMENT STRATEGIES

The considered BS assignment strategies are expressed in terms of utility functions. In particular, the analyzed strategies are referred to as Load Balancing Radio (LBR) and Joint Radio and Transport (JRT). The former is a strategy that is based exclusively on radio aspects, while JRT is the strategy we propose to account for potential backhaul limitations.

A. Load Balancing Radio

The considered LBR strategy aims to distribute some terminals among BSs in order to balance the transmitted power of BSs, whenever propagation losses between terminals and candidate BSs do not exceed a given margin $\Delta$ above the minimum path loss. To that end, the utility function of the LBR strategy is formulated as:

$$u_i = \left(1 - \frac{\min(L_i - L_{i,bs}, \Delta)}{\Delta}\right) \left(1 - \frac{\min(P_i, P_{trans})}{P_{trans}}\right) \left(1 - \frac{\min(R_i, C_{trans})}{C_{trans}}\right)$$

(7)

where $L_{i,bs}$ is the minimum path loss attenuation between user $i$ and its best server, and $L_{i,j}$ is the attenuation that this user would have if it is served by BS $j$. Thus, as observed in expression (7), when the total power of BS $j$ is increasing towards its maximum power limit $P_{trans}$ the utility to connect user $i$ to BS $j$ decreases. Similarly, the utility is decreased by assigning a user to a BS with a path loss approaching $\Delta$ above the minimum.

B. Joint Radio and Transport

The JRT approach extends the LBR strategy to include transport restrictions in the utility function. Thus, the utility function of this strategy can be written as follows:

$$u_i = \left(1 - \frac{\min(L_i - L_{i,bs}, \Delta)}{\Delta}\right) \left(1 - \frac{\min(P_i, P_{trans})}{P_{trans}}\right) \left(1 - \frac{\min(R_i, C_{trans})}{C_{trans}}\right)$$

(8)

where $R_i$ is the aggregated rate of all users being served by BS $j$ and $C_{trans}$ is the associated transport capacity of the BS $j$. The rightmost term in (8) takes into account the transport occupancy of BS $j$. Thus, if a user is assigned to a BS with high transport/power utilization, the resulting utility will be lower. Although some degree of coupling exists between the BS transmitted power and the served aggregate rate, different situations can arise where one constraint could become more restrictive than the other. For instance, for the same aggregate traffic load, different power levels will be required depending on how far from the BS users are located.

C. Algorithm Description

The method to find a feasible BS assignment under a given snapshot is realized using the simulated annealing (SA) technique [16]. The algorithm aims to maximize the system utility $U$ while it searches for a feasible BS assignment. It begins the search from an initial BS assignment $B$ and an initial temperature value $T$, as illustrated in Fig. 1. The initial assignment $B$ is obtained using a minimum path loss (MPL) criterion, while the method for generating the new solution $B'$ is based on an estimation of the utility increment that users would have if they are reallocated to a new BS. For each temperature, the inner loop is performed until a feasible
BS solution (i.e. stopping criterion) is found or the maximum number of iterations is met. Users with highest utility increment estimation are potentially considered for generating a new assignment. The utility increment of a user is computed as the difference between the utility obtained when it is assigned to a given BS under the current allocation and the estimated utility if this user is moved to a different BS. We have seen that using this approach to generate a new solution $B'$ is a good tradeoff between avoiding local minima and reducing the overall number of iterations, in comparison to generate $B'$ in a completely random fashion.

TABLE I: SIMULATION PARAMETERS

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base station</td>
<td>384 Kbps</td>
</tr>
<tr>
<td>Bit rate, $R_i$</td>
<td>5.2 dB</td>
</tr>
<tr>
<td>$E_b/N_0$ target</td>
<td>1152 Kbps</td>
</tr>
<tr>
<td>Pole capacity, $C_{av}$</td>
<td>1 Km</td>
</tr>
<tr>
<td>Propagation model</td>
<td>L(dB)=128.1+37.6log[d(km)]+S(dB)</td>
</tr>
<tr>
<td>Shadowing standard deviation, $S$</td>
<td>10 dB</td>
</tr>
<tr>
<td>Chip rate, $W$</td>
<td>3.84 MHz/s</td>
</tr>
<tr>
<td>BS max. transmitted power</td>
<td>43 dBm</td>
</tr>
<tr>
<td>Noise power, $P_N$</td>
<td>-101.15 dBm</td>
</tr>
<tr>
<td>User Equipment</td>
<td></td>
</tr>
<tr>
<td>Bit rate</td>
<td>128 Kbps</td>
</tr>
<tr>
<td>$E_b/N_0$ target</td>
<td>3.2 dB</td>
</tr>
<tr>
<td>Chip rate, $W$</td>
<td>3.84 MHz/s</td>
</tr>
<tr>
<td>Max. mobile transmission power</td>
<td>21 dBm</td>
</tr>
<tr>
<td>Noise power, $P_N$</td>
<td>-100.15 dBm</td>
</tr>
</tbody>
</table>

A. Homogeneous transport capacity and uniform users distribution scenario.

In this scenario, we compare the different BS assignment strategies when all BSs have been provisioned with the same transport capacity (i.e. $\beta=\beta_i$ for $i=1...N$) and users are uniformly distributed in the overall service area. In particular, the comparison is addressed in terms of the number of active users that can be effectively allocated to BSs so that both radio and transport constraints are fulfilled with a given probability (referred to as satisfaction level). Results are shown in Fig.2 for $\beta=1, 1.5$ and 2, where it can be seen that the proposed JRT strategy can effectively accommodate more active users with higher satisfaction levels using much less provisioned transport capacity in BSs than LRB and MPL strategies. In particular, we can see that JRT with a transport capacity equal to the air pole capacity ($\beta=1$) can support a higher number of active users than MPL and JRT strategies even when BSs have been provisioned with an additional $50\%$ transport capacity ($\beta=1.5$). Moreover, we can see that benefits of balancing the usage of the transport capacity using the JRT strategy are kept even for transport capacity as high as twofold the considered air pole capacity ($\beta=2$) where radio limitations become the dominant factor. Notice that, at that point, load balancing in the radio by using LBR also outperforms a classical MPL strategy. On the other hand, the benefits achieved by the JRT strategy come at the expense of increased power consumption due to the possibility to assign a user to a BS other than the “best” one from a radio perspective. This power increase is shown in Fig. 3 for the downlink and uplink cases. In the uplink case where this increase can be more critical, it is shown that in the case of $\beta=1.5$ where a capacity gain around $55\%$ can be achieved by JRT over LBR and MPL (i.e. 47 active users can be allocated by JRT with a $95\%$ satisfaction level in front of 30 users with the other strategies) uplink power increase is kept below $1.5$ dB.
B. Non homogeneous transport capacity and uniform users distribution scenario.

The analysis has been extended to scenarios with non-homogeneous backhaul capacity, that is, BSs with different transport capacities. These scenarios are likely to be quite usual since normally mobile operators initiate the rollout of 3G infrastructure focusing on coverage as the main requirement and consequently cell sites can be provisioned with relatively low transport capacity. In a subsequent phase the focus shifts to the transport capacity, thus capacity can be added gradually to cell sites attending to traffic demand increase. Non-homogeneous transport conditions are modeled by considering two different transport capacity values for BSs in the scenario. In this way, the parameter \( N_{BS} \) is used to indicate a given number of BSs that are provisioned with different transport capacity than the rest of BSs in the service area. Specifically two different scenarios have been considered. First, in Scenario B1 most of the BSs are provisioned with enough transport capacity (\( \beta = 2.5 \)) except \( N_{BS} \) cells which have lower transport capacity (\( \beta \) is varied between 1 and 2.5). On the other hand, in Scenario B2 the transport capacity of most of the BSs is low (\( \beta = 1.5 \)) whereas \( N_{BS} \) cells are provisioned with higher capacity (\( \beta \) is varied from 1 to 2.5).

The number of active users with a satisfaction level of 95% supported in Scenario B1 for \( N_{BS} = 5 \) cells is shown if Fig. 4. Notice that the value of the \( \beta \) factor represented on the x-axis refers only to the transport capacity of the set of \( N_{BS} \) cells, whereas the rest of cells have a fixed \( \beta \) factor equal to 2.5. Under this situation, when focusing on small \( \beta \) values, it is shown how the proposed JRT strategy can significantly alleviate the transport limitations of such cells by making use of spare transport capacity available in surrounding cells. As a result, a capacity increase around 50% can be achieved in scenarios where there are few cells with insufficient transport capacity by effectively using capacity available in neighboring cells. On the other side, as \( \beta \) increases for this set of \( N_{BS} \) cells, transport capacity is becoming available in a more homogeneous way in the overall scenario and, the LBR and JRT strategies tend to converge as the backhaul capacity increases.

Fig. 4 also depicts some results for Scenario B2 for \( N_{BS} = 5 \) cells and for the three considered strategies. In this case, the situation is that almost all cells in the scenario have transport limitations (i.e. \( \beta \) equal to 1.5) except \( N_{BS} \) cells. It is observed how JRT can always accommodate a higher number of users satisfying power and transport constraints than a LBR scheme because transport resources remain as the main limitation aspect even when few cells have been upgraded up to \( \beta = 2.5 \).

C. Homogeneous transport capacity and non uniform users distribution scenario.

Here we analyze the effect of the presence of traffic hot-spots in the service area. A hot-spot region could appear due to the existence of a shopping mall or large office which might be 100–200 m across in a macrocell whose cell radius is a couple of kilometres. In particular, circular hot-spots are considered in two possible positions (see Fig. 5), called hot-spot in corner and hot-spot in center. The distribution of users in the system is characterized by means of a non-uniformity factor \( \mu_{D} \) that represents the percentage of users that located in the hot-spot region, while the rest of the users are uniformly distributed.
distributed [9]. Under this analysis, hot-spots are randomly located within the service area of some cells (a cell can have one or zero hot-pots) and the parameter $N_{BS}$ is used to indicate the number of BSs with a hot-spot region. Results have been obtained for $N_{BS}=5$ cells, $\mu_0=0.8$ and for the two possible hot-spot locations within the cell. As shown in Fig. 6, irrespective of the location of the hot-spot, we can conclude that JRT strategy is again able to allocate more active users for $\beta$ values below 2 due to the load balancing of radio and transport resources. It is also worth noting that, when radio interference is the limiting factor ($\beta = 2.5$), the presence of the hot-spot in the center leads to a slight increase in the number of served users for all three strategies because of less power consumption.

VI. CONCLUSION

In this paper we have proposed and analyzed a BS assignment strategy named JRT that incorporates simultaneously both radio and transport network capacity constraints. The motivation of such a kind of novel approach is based on the growing concern about the possibility that the transport part of the mobile network can also constitute a bottleneck in certain deployment scenarios. This situation is a reality in current 3G networks with the deployment of high speed channels and is expected to persist in next generation networks. The performance of JRT has been assessed for WCDMA networks under different deployment scenarios and compared to classical approaches such as Minimum Path Loss (MPL) allocation and Load Balancing Radio (LBR). Obtained results demonstrate that JRT is able to allocate more users to BSs so that both radio and transport constraints can be satisfied under different scenarios comprising partially-limited backhaul RANs and the presence of traffic hotspots.

REFERENCES