Impact of Limited Backhaul Capacity on User Scheduling in Heterogeneous Networks

Jagadish Ghimire and Catherine Rosenberg

Department of Electrical and Computer Engineering, University of Waterloo, Canada

Email: {jghimire,cath}@uwaterloo.ca

Abstract—In this paper, we investigate the impact of limited backhaul capacity on user scheduling in the context of a heterogeneous network comprising a macro base station overlaid with small-cells. Under a global proportional fairness (PF) criteria, we show that this limited capacity has a fundamental impact on user scheduling. When user association and channel allocation are given, like in the infinite-capacity backhaul case, the global PF user scheduling problem can be decomposed into a set of independent local PF user scheduling problems. However, unlike the case with infinite backhaul where the local PF scheme is equivalent to giving equal time to each user, a local PF scheme with finite backhaul can be of one of three types. We completely characterize these three types and the conditions under which to use them. The results show that a backhaul-aware scheduling scheme is simple to implement and necessary to obtain the best performance.

I. Introduction

Heterogeneous Networks (HetNets) comprise a set of low-power base stations (BSs), called *small cells*, overlaying the existing macro cellular layer [1]. Small cells are connected to the macro infrastructure via wired or wireless backhaul links. This shift from the existing homogeneous structure to a hierarchical heterogeneous architecture is seen as an important innovation required to obtain a 30-fold increase in capacity in LTE cellular networks [2].

The introduction of the new overlapping layer of small cells brings in multiple challenges including *User Scheduling* (US), *Resource Allocation* (RA), *User Association* (UA), and *Transmission Coordination* (TC). If well designed, these processes can improve the system performance significantly. A number of works including [3], [4] and [5] have studied these processes in great detail and have proposed optimal or quasi-optimal schemes. For example, [5] presents an optimal user association and user scheduling scheme whereas [4] presents a jointly optimal user scheduling, user association and resource allocation scheme.

Most of the studies on small cells in the literature suffer from one key limitation. They do not consider the effect of *backhaul's capacity limitation* on system performance, and hence are limited in scope to scenarios with sufficiently large backhaul link capacities. Such an assumption could be justified in the existing homogeneous cellular networks since the cost of backhauling is a small component of the overall cost and hence we can generally expect them to be equipped with sufficiently high backhaul capacity. For small cells however, the picture is very different. The cost of backhauling is a significant part of the total Capital Expenditure (CAPEX) and Operational

Expenditure (OPEX), in some cases exceeding the cost of the small cell BS equipment [6]. It is thus desirable that the backhauling cost for small cells is kept low, which in turn limits the capacity of the installed links. Hence, backhaul links can sometimes be the bottleneck to the performance of small cells. The purpose of this study is to specifically study HetNets where the backhaul links have finite capacities. Our main result is to show that this limitation has a fundamental impact on how user scheduling should be performed.

More precisely, we consider a region covered by one macro cellular area with one macro base station (MBS) with infinite capacity backhaul and a number of small cells connected to the MBS via *wired* backhaul links of limited capacity. We assume that the resource allocation and the user association are given and that each base-station (MBS or small cell) schedules its UEs independently from the other base stations. Our objective is to schedule the users so as to be proportionally fair in a global sense (i.e., over all base stations).

Our main contributions can be summarized as follows.

- Under the assumption of infinite-capacity backhaul links, it was shown in [4] that the global proportional fair (PF) user scheduling problem can be decomposed into a set of decoupled independent local PF user scheduling problems and that the local PF is equivalent to an equal-time scheduling scheme. When the backhaul links have finite capacities, we show that we can still decouple the global problem into independent local PF schemes, but the equivalence between local PF and the equal-time scheduling does not always hold.

- In order to be globally PF in a system with finite backhaul, we show that each small cell j needs to schedule its users in a way which depends on how its backhaul capacity C_j compares to two critical values c_j^* and C_j^* that are specific to the realization that it sees. We give simple closed-form expressions to compute c_j^* and C_j^*. If $C_j \leq c_j^*$ then local PF is equivalent to *equal-throughput* scheduling, while if $C_j \geq C_j^*$ it is equivalent to *equal-time* scheduling. Otherwise, local PF scheduling can be computed easily by solving two non-linear equations with two variables.

- Using numerical results, we quantify the impact of the small cell backhaul capacity on the system performance and the relative difference in performance between the optimal scheduling and the scheme which is optimal when the backhaul capacities are infinite.
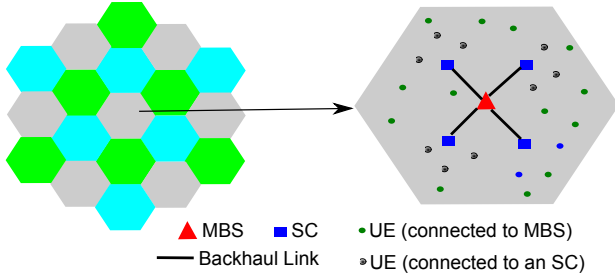
Fig. 1. Our system.

The rest of this paper is organized as follows. In Section II, we present the system model. Section III presents the optimization problem as well as the main analysis. Section IV contains numerical results and Section V concludes the paper.

II. SYSTEM MODEL

We consider an OFDM-based HetNet system comprising one macro base station (MBS), X operator-deployed small cells (e.g., pico base stations (PBS)) and N user equipments (UE), all located within a given macro-cell (see Fig. 1). In a given HetNet, let MBS be represented as node 0, $\mathcal{P}$ be the set of small cells and $\mathcal{N}$ be the set of user equipments. Each small cell $j \in \mathcal{P}$ is connected to the MBS via a wired backhaul of capacity C_j. We assume that the capacity of the backhaul infrastructure between the MBS and the backbone is infinite. We focus on the downlink of a given HetNet and assume that all users are active, i.e., there exists a downlink flow from the MBS (source) to each UE (destination). We assume that the users are greedy and hence flows are not throughput constrained. Let MBS have a power budget of P_{MBS} and each of the small cells have a power budget of P_{PBS}. We do not consider a time-domain transmission coordination among the BSs[1] and assume that all the BSs transmit all the time. We assume that the system has been allocated M OFDM subchannels. We consider a channel allocation scheme called Orthogonal Deployment (OD) [1], where K subchannels are allocated to the small cells and the remaining $M - K$ subchannels are allocated to the MBS. There is another type of channel allocation scheme called the Co-channel Deployment (CCD), where all M subchannels are allocated to all BSs. In the absence of explicit inter-BS coordination for interference management, OD is shown to perform much better than CCD [7]. We will thus restrict our study to OD. The results, however, are also true for CCD.

In order to simplify our formulations, we make the following assumptions: **[A1]** A BS transmits on all the subchannels allocated to it; **[A2]** Power allocated to a given BS is equally divided among all the allocated subchannels with P_j representing the per-subchannel power of BS j; **[A3]** Channels are flat, i.e., the channel gains across different subchannels between a BS and a UE are equal. Let, G_{ji} represent the channel gain between BS j and UE i, assumed to be known either by measurements or via a channel model. As discussed in [7], this assumption allows us to reduce a time and frequency

[1]Base station (BS) refers to both the MBS and the small cells.

domain scheduling to just a time domain scheduling, where a BS allocates all of its subchannels to one UE at a given time.

A. Physical interference model and link rates

Since the locations of the MBS and the small cells are fixed, a random realization of N user positions, with a given channel model can be used to generate a random instance of the channel gain matrix $\boldsymbol{G} = (G_{ji})$, which we call a *realization* of the network. This matrix captures the distance dependent path-loss as well as the random shadowing between the transmit-receive pairs. Now, for a given realization $\boldsymbol{G}$, for a given P_{MBS} and P_{PBS}, the SINR between BS j and UE i on each subchannel, represented as γ_{ji}, can be determined as follows. For all $j \in \mathcal{P} \cup \{0\}$ and for all $i \in \mathcal{N}$, we have[2]

$$\gamma_{ji} = \frac{P_j G_{ji}}{N_0 + \mathbf{1}_{\{j \in \mathcal{P}\}} \sum_{j' \in \mathcal{P}: j' \neq j} P_{j'} G_{j'i} + \overline{I}_{out}} \quad (1)$$

where, $P_j = \mathbf{1}_{\{j=0\}} \dfrac{P_{MBS}}{(M-K)} + \mathbf{1}_{\{j \in \mathcal{P}\}} \dfrac{P_{PBS}}{K} \quad (2)$

$\overline{I}_{out}$ is the interference coming from BSs in the nearby HetNets (macro cells), assuming that a reuse factor of 3 is employed among the macro cells and that the nearby HetNets have identical channel allocation scheme and transmit power budget. There is a function $f(.)$ that maps SINR to data rate which defines the maximum supportable rate per subchannel from BS j to UE i. We do not make any a priori assumptions on the exact nature of this rate function. Hence, the maximum supportable rate R_{ji} is given as

$$R_{ji} = \mathbf{1}_{\{j=0\}}(M-K)f(\gamma_{ji}) + \mathbf{1}_{\{j \in \mathcal{P}\}}Kf(\gamma_{ji}) \quad (3)$$

III. OPERATION PROBLEM: SCHEDULING UNDER LIMITED BACKHAUL CAPACITIES

A. Problem Formulation

For a given realization $\boldsymbol{G}$, and given backhaul capacities ($\boldsymbol{C} = (C_j)_{j \in \mathcal{P}}$), we assume that the channel allocation parameter K as well as the rate-function $f(.)$ are given. In this case, the R_{ji}'s can be computed a priori as input parameters to the problem using (1), (2) and (3). Even though our model assumes that the value of K is given, it is important to point out that choosing a good value of K is important (and in general not trivial) [7]. We also assume that the user association is given, with $x_{ji} = 1$ representing the fact that user i is associated to BS j and $x_{ji} = 0$ representing otherwise. We will assume that a user associates to only one BS, i.e., $x_{ji} \in \{0,1\}$ and $\sum_{j \in \{0\} \cup \mathcal{P}} x_{ji} = 1$. Also, we assume that the user association scheme guarantees that each UE has a non-zero rate to its BS:

$$R_{ji} > 0 \qquad \forall i \in \mathcal{N}, \forall j \in \{0\} \cup \mathcal{P} \text{ with } x_{ji} = 1 \quad (4)$$

We intend to schedule the users in a global PF fashion. In [8], it was shown that global PF can be achieved by maximizing the following objective function: $\sum_{i \in \mathcal{N}} \log(\lambda_i)$, where λ_i is the throughput offered to user i. Let us define $0 \leq \alpha_{ji} \leq 1$ as the fraction of time for which BS j transmits to UE i.

[2]Indicator function $\mathbf{1}_{\{A\}} = 1$ if A is true, 0 otherwise.

Then, we want to find the optimal values of the α_{ji}'s so as to maximize our objective function, i.e., we want to solve the following problem.

$$[\mathbf{P}] \quad \max_{(\lambda_i),(\alpha_{ji})} \sum_{i \in \mathcal{N}} \log(\lambda_i)$$

$$\text{subject to: } \lambda_i = \sum_{j \in \mathcal{P} \cup \{0\}} R_{ji}\alpha_{ji}, \forall i \in \mathcal{N} \tag{5}$$

$$\sum_{i \in \mathcal{N}} R_{ji}\alpha_{ji} \leq C_j, \quad \forall j \in \mathcal{P} \tag{6}$$

$$\sum_{i \in \mathcal{N}} \alpha_{ji} \leq 1, \qquad \forall j \in \mathcal{P} \cup \{0\} \tag{7}$$

$$\alpha_{ji} \leq x_{ji}, \qquad \forall i \in \mathcal{N}, \forall j \in \mathcal{P} \cup \{0\} \tag{8}$$

$$\alpha_{ji} \geq 0, \qquad \forall i \in \mathcal{N}, \forall j \in \mathcal{P} \cup \{0\} \tag{9}$$

(5) relates user schedules to throughputs, (6) is the constraint due to finite backhaul capacities, and (7) represents the scheduling constraints at each BS. (8) is the user-association constraint. We call $[\mathbf{P}]$ the *global* PF scheduling problem. It is well-known that the proportional fairness scheme maximizes the *Geometric Mean* (GM) throughput $\sqrt[|\mathcal{N}|]{\prod_{i \in \mathcal{N}} \lambda_i}$, which we take as our key performance metric.

Let us define A_j as the set of UEs associated to BS $j \in \mathcal{P} \cup \{0\}$, i.e., $A_j = \{i \in \mathcal{N} : x_{ji} = 1\}$. In [4], Fooladivanda and Rosenberg have shown that the following properties hold for $[\mathbf{P}]$ if the backhaul capacities are unlimited (i.e., if (6) is removed).

- *Decomposability*: The global PF problem $[\mathbf{P}]$ can be decoupled into a set of $X + 1$ independent local PF problems.
- *Equal-time equivalence*: A local PF scheduling at BS j is equivalent to an *equal-time* scheduling where each user $i \in A_j$ is allocated $\frac{1}{|A_j|}$ fraction of time.

We want to see whether this remains valid under limited backhaul capacities, and if not, characterize the nature of the optimal scheduling under limited backhaul.

B. Decomposition

First we will define a set of $X + 1$ independent local PF optimization problems (one per BS). For the MBS, we define:

$$[\mathbf{P^0_{Local}}] \quad \max_{(\alpha_{0i})_{i \in A_0}} \sum_{i \in A_0} \log(R_{0i}\alpha_{0i}) \text{ s. t. } \sum_{i \in A_0} \alpha_{0i} \leq 1;$$
$$\alpha_{0i} \geq 0, \quad \forall i \in A_0$$

For small cell $j \in \mathcal{P}$, we define:

$$[\mathbf{P^j_{Local}}(C_j)] \max_{(\alpha_{ji})_{i \in A_j}} \sum_{i \in A_j} \log(R_{ji}\alpha_{ji}) \text{ s. t.}$$

$$\sum_{i \in A_j} R_{ji}\alpha_{ji} \leq C_j \qquad (\mu_j) \tag{10}$$

$$\sum_{i \in A_j} \alpha_{ji} \leq 1 \qquad (\zeta_j) \tag{11}$$

$$\alpha_{ji} \geq 0, \quad \forall i \in A_j \qquad (l_{ji}) \tag{12}$$

where $\mu_j \geq 0$, $\zeta_j \geq 0$ and $l_{ji} \geq 0$ are the dual-variables, associated with the backhaul limitation constraint, the scheduling constraint, and the non-negativity constraint respectively. The log function in the objective is defined only for strictly positive values of α_{ji}.

The following proposition establishes that it is indeed possible to decompose the global problem $[\mathbf{P}]$ into the above stated *local* problems and hence *decomposability* holds even with backhaul capacity limitations.

Proposition 1: For a given $\mathbf{G}$, a given user association satisfying (4) and hence a given set of users A_j associated to each BS j, $[\mathbf{P}]$ can be decomposed into $[\mathbf{P^0_{Local}}]$ and $[\mathbf{P^j_{Local}}(C_j)]$ for $j \in \mathcal{P}$ and hence optimal schedules to these local problems solve the global problem $[\mathbf{P}]$ to optimality.

Proof: For UE $i \in A_j$ for $j \in \{0\} \cup \mathcal{P}$, we can combine (5) and (8) to obtain $\lambda_i = \alpha_{(j_i)i}R_{(j_i)i}$ where (j_i) is the BS to which i is associated, i.e. $x_{(j_i)i} = 1$. Also, if a solution $\boldsymbol{\alpha} = (\alpha_{ji}, \forall i \in \mathcal{N}, \forall j \in \{0\} \cup \mathcal{P})$ is feasible, then (8) requires that we have $\alpha_{ji'} = 0$ for all $i' \neq A_j$. This allows us to remove all constraints that couple any two $i, i' \in \mathcal{N}$ such that $(j_i) \neq (j_{i'})$. The objective can be rewritten in the separable form: $\sum_{j \in \{0\} \cup \mathcal{P}} \sum_{i \in A_j} \log(\alpha_{ji}R_{ji})$. For $i, i' \in A_0$, the only coupling constraint is (7) and for $i, i' \in A_j$ (where $j \in \mathcal{P}$), there are two coupling constraints (6) and (7). Thus we achieve the stated decomposition. ∎

$[\mathbf{P^0_{Local}}]$ is a simple unweighted proportional fair scheduling problem that is known to have a solution $\alpha_{0i} = \frac{1}{|A_0|}$ as long as $R_{0i} > 0$ for all $i \in A_0$ (see (See [4], [9])). In other words, the MBS schedules on *equal-time* basis. Note that if the backhaul capacity of the MBS was not infinite, Proposition 1 would not be true in general as there would be a strong coupling constraint due to that capacity since all traffic is routed through the MBS.

Next, we focus on one of the small cells j and characterize the optimal scheduling of $[\mathbf{P^j_{Local}}(C_j)]$.

C. Analysis of $[\mathbf{P^j_{Local}}(C_j)]$

We first show an important property of the problem $[\mathbf{P^j_{Local}}(C_j)]$ as follows.

Proposition 2: If $C_j > 0$, there exists a unique optimal solution to $[\mathbf{P^j_{Local}}(C_j)]$ with $\alpha_{ji} > 0$ for all $i \in A_j$.

Proof: By the definition of log function, the optimal solution, if it exists, needs to satisfy $\alpha_{ji} > 0$ for all $i \in A_j$. Note that, (4) makes sure that a UE i is associated to BS j only when $R_{ji} > 0$. Also, if $C_j > 0$, we can easily show that there exists some $\alpha_{ji} > 0$ for all $i \in A_j$, such that (10) and (11) are feasible. In other words, the set of feasible solutions with $\alpha_{ji} > 0$ for $i \in A_j$ is non-empty and hence an optimal solution with $\alpha_{ji} > 0$ for $i \in A_j$ always exists if $C_j > 0$. Since the problem involves the maximization of a strictly concave function over a convex set, there is a unique optimal solution [10]. ∎

The Lagrangian function of the problem can be defined as follows.

$$L(\boldsymbol{\alpha}_j; \mu_j, \zeta_j, \boldsymbol{l}_j) = -\sum_{i \in A_j} \log(\alpha_{ji}) + \mu_j \left(\sum_{i \in A_j} R_{ji}\alpha_{ji} - C_j \right)$$
$$+ \zeta_j \left(\sum_{i \in A_j} \alpha_{ji} - 1 \right) - \sum_{i \in A_j} l_{ji}\alpha_{ji} \tag{13}$$

where $\boldsymbol{\alpha}_j$ and $\boldsymbol{l}_j$ are respectively the vectors comprising of all α_{ji} and all l_{ji} for $i \in A_j$. The dual-problem that solves the local problem $[\mathbf{P^j_{Local}}(C_j)]$ is given as follows.

$$\max_{\mu_j \geq 0, \zeta_j \geq 0, \boldsymbol{l}_j \geq 0} \quad \min_{\boldsymbol{\alpha}_j \geq 0} L(\boldsymbol{\alpha}_j; \mu_j, \zeta_j, \boldsymbol{l}_j)$$

We can obtain the Karush-Kuhn-Tucker (KKT) conditions [10], necessary for optimality, as follows.

$$\frac{\partial L}{\partial \alpha_{ji}} = 0 \implies \alpha_{ji} = \frac{1}{\mu_j R_{ji} + \zeta_j - l_{ji}} \quad \forall i \in A_j \quad (14)$$

$$\zeta_j \left(\sum_{i \in A_j} \alpha_{ji} - 1 \right) = 0 \quad (15)$$

$$\mu_j \left(\sum_{i \in A_j} R_{ji} \alpha_{ji} - C_j \right) = 0 \quad (16)$$

$$l_{ji} \alpha_{ji} = 0, \quad \forall i \in A_j \quad (17)$$

$$\mu_j \geq 0; \quad \zeta_j \geq 0; \quad \boldsymbol{l}_j \geq 0; \quad (10); \quad (11); \quad (12);$$

(14) are the first-order necessary conditions for optimality. (15), (16) and (17) are the so-called complementary-slackness conditions. Since the primal problem involves maximization of a concave function over a convex set, any tuple of primal and dual variables $((\alpha_{ji})_{i \in A_j}, \mu_j, \zeta_j, (l_{ji})_{i \in A_j})$ that satisfies all of the KKT conditions is optimal [10]. Also, using Proposition 2, such a solution is unique. Moreover, since the optimal solution is known to satisfy $\alpha_{ji} > 0$, we have $l_{ji} = 0$ for all $i \in A_j$ from (17). We thus do not mention the dual variables $\boldsymbol{l}_j$ in the rest of the analysis, as they should always be zero for optimality. We can also rewrite the first order condition (14) as

$$\alpha_{ji} = \frac{1}{\mu_j R_{ji} + \zeta_j} \quad \forall i \in A_j \quad (18)$$

It is also useful to observe that the optimal dual variables obey one of the *three* conditions: $(\mu_j = 0, \zeta_j > 0)$, $(\mu_j > 0, \zeta_j = 0)$, and $(\mu_j > 0, \zeta_j > 0)$. This is because, (18) imposes $\mu_j R_{ji} + \zeta_j \neq 0$. Hence, $(\mu_j = 0, \zeta_j = 0)$ is not possible. Next, we define

$$c_j^* \triangleq \frac{|A_j|}{\sum_{i \in A_j} \frac{1}{R_{ji}}}; \quad C_j^* \triangleq \frac{1}{|A_j|} \sum_{i \in A_j} R_{ji} \quad (19)$$

We will make use of the following lemmas to establish our main result.

Lemma 1: (a) If $C_j \geq C_j^*$, then $(\alpha_{ji} = \frac{1}{|A_j|}, \forall i \in A_j)$ is the unique optimal solution to $[\mathbf{P^j_{Local}}(C_j)]$.
(b) If $C_j < C_j^*$, then $(\alpha_{ji} = \frac{1}{|A_j|}, \forall i \in A_j)$ is not feasible.
Proof: It is easy to verify that $\alpha_{ji} = \frac{1}{|A_j|}$ for all $i \in A_j$, $\mu_j = 0$ and $\zeta_j = |A_j|$ satisfy all KKT conditions if $C_j \geq C_j^*$. It is thus an optimal solution consistent with the backhaul capacity value $C_j \geq C_j^*$. Proposition 2 implies that this is in fact the only optimal solution.
If $C_j < C_j^*$, then substituting $\alpha_{ji} = \frac{1}{|A_j|}$ for all $i \in A_j$ in $\sum_{i \in A_j} \alpha_{ji} R_{ji} \leq C_j$ results in a contradiction. $\blacksquare$
Lemma 2: (a) If $C_j \leq c_j^*$, then $(\alpha_{ji} = \frac{C_j}{|A_j| R_{ji}}, \forall i \in A_j)$ is the unique optimal solution to $[\mathbf{P^j_{Local}}(C_j)]$.
(b) If $C_j > c_j^*$, then $(\alpha_{ji} = \frac{C_j}{|A_j| R_{ji}}, \forall i \in A_j)$ is not feasible.

Proof: We can easily verify that $\alpha_{ji} = \frac{C_j}{|A_j| R_{ji}}$ for all $i \in A_j$, $\mu_j = \frac{|A_j|}{C_j}$ and $\zeta_j = 0$ satisfy all KKT conditions if $C_j \leq c_j^*$. It is thus an optimal solution consistent with the backhaul capacity value $C_j \leq c_j^*$. Proposition 2 implies that this is also the only optimal solution.
If $C_j > c_j^*$, then substituting $\alpha_{ji} = \frac{C_j}{|A_j| R_{ji}}$ for all $i \in A_j$ in $\sum_{i \in A_j} \alpha_{ji} \leq 1$ results in a contradiction. $\blacksquare$
Lemma 3: If $c_j^* < C_j < C_j^*$, the optimal dual solution is obtained by solving the following equations for $\mu_j > 0$ and $\zeta_j > 0$.

$$\sum_{i \in A_j} \frac{R_{ji}}{\mu_j R_{ji} + \zeta_j} = C_j \quad (20)$$

$$\sum_{i \in A_j} \frac{1}{\mu_j R_{ji} + \zeta_j} = 1 \quad (21)$$

The primal solution can then be obtained as $\alpha_{ji} = (\mu_j R_{ji} + \zeta_j)^{-1}$ for all $i \in A_j$.
Proof: We will first show that the optimal dual variables have to satisfy $\mu_j > 0$ and $\zeta_j > 0$.
First, we assume that there exists a dual optimal solution such that $\mu_j = 0$. $\mu_j = 0$ implies $\zeta_j > 0$, and hence

$$\alpha_{ji} = \frac{1}{\zeta_j} \text{ and } \sum_{i \in A_j} \alpha_{ji} = 1 \implies \alpha_{ji} = \frac{1}{|A_j|}, \forall i \in A_j \quad (22)$$

We know from Lemma 1(b) that this is an infeasible solution since $C_j < C_j^*$. Thus, we require $\mu_j > 0$.
Similarly, we assume that there exists a dual optimal solution such that $\zeta_j = 0$. $\zeta_j = 0$ implies $\mu_j > 0$, and hence

$$\alpha_j = \frac{1}{\mu_j R_{ji}} \text{ and } \sum_{i \in A_j} \alpha_{ji} R_{ji} = C \implies \mu_j = \frac{|A_j|}{C_j} \quad (23)$$

$$\implies \alpha_j = \frac{C}{|A_j| R_{ji}}, \forall i \in A_j \quad (24)$$

We know from Lemma 2(b) that this is an infeasible solution since $C_j > c_j^*$. Thus, we require $\zeta_j > 0$.
Thus, the optimal solution has to satisfy $\mu_j > 0$ and $\zeta_j > 0$. In such case, (15) and (16) mandate that the primal constraints (10) and (11) are satisfied with equality, i.e.,

$$\sum_{i \in A_j} R_{ji} \alpha_{ji} = C_j \text{ and } \sum_{i \in A_j} \alpha_{ji} = 1 \quad (25)$$

Substituting the value of α_{ji} from (18), we get the required equations (20) and (21). A strictly positive solution of (μ_j, ζ_j) to (20) and (21) should exist as we know from Proposition 2 that the problem always has a unique solution with $\alpha_{ji} > 0, \mu_j \geq 0, \zeta_j \geq 0$, given $C_j > 0$. $\blacksquare$
These results characterize the local scheduling solutions at the small cells that yield the global optimal solution depending on the values of the C_j's.

D. Interpretation

In Fig. 2, the solid curve represents the typical shape of the plot of the geometric mean throughput ($\left(\sqrt[|A_j|]{\prod_{i \in A_j} \lambda_i} \right)$) as a function of the backhaul capacity for one of the small cells $j \in \mathcal{P}$ when the local PF scheduling (which yields the global
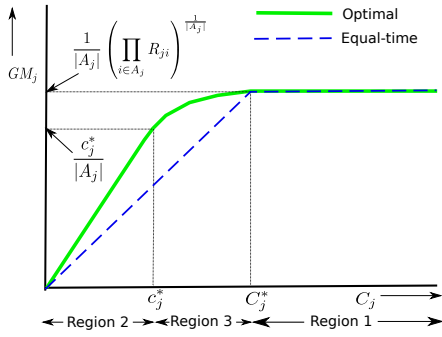
Fig. 2. The three regions and the two critical values of backhaul capacity

PF) is performed. This figure shows the three regions as a function of the two critical values of the backhaul capacity.

1) As we mentioned earlier, it is known that if C_j is infinitely large, *local PF* is equivalent to *equal-time* scheduling with $\alpha_{ji} = \frac{1}{|A_j|}$ for all $i \in A_j$. Such an assumption of infinite backhaul capacity is so ubiquitous in the literature of cellular networks that *equal-time* scheduling has often been seen as being synonymous to *proportional fairness*. Our result (Lemma 1) shows that this equivalence is true only when the backhaul capacity is greater than or equal to C_j^* (Region 1 in Fig. 2). However if $C_j < C_j^*$, *equal-time* scheduling with $\alpha_{ji} = \frac{1}{|A_j|}$ for all $i \in A_j$ is not feasible. In this case, if we still impose an *equal-time* scheduling among users (i.e., $\alpha_{ji} = \alpha_{ji'}, \quad \forall i, i' \in A_j$), we will in general be under-utilizing the time resource (i.e., $\sum_{i \in A_j} \alpha_{ji} < 1$) and the performance will be sub-optimal. Let us define a new scheduling scheme that adds the following *equal-time scheduling* constraint on $[\mathbf{P}^{\mathbf{j}}_{\mathbf{Local}}(C_j)]$:

$$\alpha_{ji} = \alpha_{ji'}, \quad \forall i, i' \in A_j$$

We call it the *local equal-time scheduling*. We can show that it imposes $\alpha_{ji} = \min\{\frac{1}{|A_j|}, \frac{C_j}{\sum_{i \in A_j} R_{ji}}\}$ for all $i \in A_j$, and the corresponding GM throughput $\sqrt[|A_j|]{\prod_{i \in A_j} \lambda_i}$ is given by $\frac{C_j}{\sum_{i \in A_j} R_{ji}} \tilde{R}_j$ if $C_j < C_j^*$ and by $\frac{1}{|A_j|} \tilde{R}_j$ otherwise, where $\tilde{R}_j = \sqrt[|A_j|]{\prod_{i \in A_j} R_{ji}}$. In Section IV, we will compare the performance between the optimal local PF scheduling and the *local equal-time scheduling* as a function of backhaul capacity. Such a comparison between optimal scheduling and *local equal-time* scheme is intended to show how much performance degradation can be expected if equal-time scheme (which would otherwise be optimal if the backhaul capacity was infinite) is employed.

2) When $C_j \leq c_j^*$ (Region 2 in Fig. 2), we have $\alpha_{ji} R_{ji} = \frac{C_j}{|A_j|}$ for all $i \in A_j$ (from Lemma 2(a)). This is a region where users in a given small cell are offered equal throughput. Thus for $C_j \leq c_j^*$, a local *equal-throughput* scheduling is equivalent to the local PF scheduling. Interestingly in the case with infinite backhaul, *max-min* scheduling is equivalent to an equal-throughput scheduling scheme.

3) For $c_j^* < C_j < C_j^*$ (Region 3 in Fig. 2), neither *local equal-throughput* nor *local equal-time* scheduling is optimal. The optimal solution to local PF can however be simply

obtained by solving the set of two non-linear equations (20) and (21) in two variables (μ_j, ζ_j).

In the next section, we present numerical results that quantify the impact of the small cell backhaul capacity on the system performance and the loss in GM throughput due to performing *local equal-time* scheduling as oppose to the local optimal scheduling as a function of the backhaul capacity, in a realistic HetNet setting.

IV. NUMERICAL RESULTS

We consider a hexagonal HetNet deployment area with each side equal to $500/\sqrt{3}$ m. This corresponds to a macro cell for a scenario with an inter-site distance of $500m$. In addition to a centrally placed MBS, we have $X = 4$ symmetrically placed small cells ($j = 1, 2, 3, 4$) at a distance of $178m$ from the MBS. We take $P_{MBS} = 46dBm$ and $P_{PBS} = 30dBm$. Outside interference is calculated by considering 18 identical macro cells around the given macro cell and a reuse factor of 3. We consider scenarios where the small cells are identical, i.e. they all have the same $C_j = C$. Out of $M = 100$ subchannels, $K = 50$ subchannels are used by each small cell and the remaining $M - K = 50$ subchannels are allocated to the MBS. We assume that there are $N = 75$ users uniformly distributed in this area. We use a distance-based path-loss model recommended by 3GPP [11], shown in Table II in [7], in order to calculate the path-loss (in decibel) from j to i. The channel-gains G_{ji} are obtained by further applying a log-normal shadowing of zero mean and $8dB$ standard deviation. A random realization corresponds to a realization of channel-gains for a random instance of uniformly deployed user positions and randomly generated shadowing coefficients. The rate function $f(.)$ is taken as the 15-rate MCS available in LTE, as shown in Table III of [7]. The table shows the efficiency e_l in terms of bits per symbol for a given threshold SNR. $f(\gamma)$ can then be calculated as $e_l \frac{n_{sc} n_{sym}}{T}$ if γ is between l^{th} and $(l+1)^{th}$ SNR threshold. $n_{sc} = 12$, $n_{sym} = 14$ and $T = 1ms$ are respectively the number of subcarriers in one subchannel, the number of OFDM symbols in one subframe and the duration of a subframe. We take $N_0 = -112.45dBm$ as the noise power per subchannel.

User association (UA) is known to play a crucial role in the performance of HetNets. The conventional approach of associating to the BS providing the highest SINR is known to perform poorly as it tends to overload the MBS. We consider two promising user association schemes introduced in the literature.

1) **Small cell First (SCF)**(δ) [4]: This scheme has a tunable parameter δ. For each i, we set $x_{j^*i} = 1$ for $j^* = \arg\max_{j \in \mathcal{P}} \gamma_{ji} \mathbf{1}_{\{\gamma_{ji} \geq \delta\}}$. In other words, we associate to the best small cell as long as the SINR from the small cell is greater than or equal to δ. If $\gamma_{ji} < \delta$ for all $j \in \mathcal{P}$, UE i associates to the MBS and thus we set $x_{0i} = 1$. δ is a parameter that can be chosen to reflect the relative SINR-bias in favor of small cells over the MBS.

2) **Range Extension (RE)** [12]: This scheme associates a UE to the BS with the highest channel-gain. For each i, we set
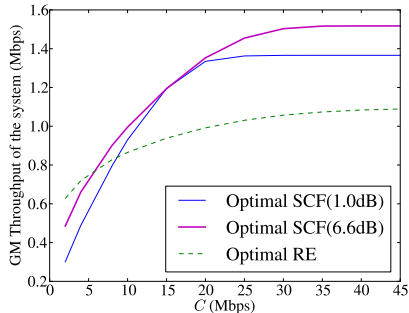
Fig. 3. GM throughput of the system as a function of backhaul capacity.

Fig. 4. The average loss in geometric mean throughput per small cell

$x_{j^*i} = 1$ only for $j^* = \arg\max_{j\in\mathcal{P}\cup\{0\}} G_{ji}$.

For a given realization ω, we can compute the global PF GM throughput for a given UA and a given K defined as $\sqrt[|\mathcal{N}|]{\prod_{i\in\mathcal{N}} \lambda_i}$. Fig. 3 shows how the GM throughput averaged over 100 realizations evolves as a a function of C (the common small cell backhaul capacity) for K=50 and 3 UAs (RE, SCF with $\delta = 1.0dB$ and SCF with $\delta = 6.6dB$). Clearly the impact of a limited backhaul capacity on the system performance is significant. As expected, we see that the GM throughput increases with C and reaches a plateau when C goes beyond C^*. Fig. 3 also confirms that SCF can perform much better than RE when C is not the bottleneck (this was shown in [4]).

We now compare the scenario in which the optimal local scheduling schemes are performed in each small cell with the scenario in which each small cell uses *equal time scheduling*. For a given realization ω, we can obtain the geometric mean throughput for small cell j, defined as $\sqrt[|A_j|]{\prod_{i\in A_j} \lambda_i}$, for the *Global optimal (G)* scheduling, and the case where *local equal-time (ET)* scheduling is performed in each small cell irrespective of the value of C_j. Let $\Lambda_j^G(\omega)$, and $\Lambda_j^{ET}(\omega)$ be these GM throughputs. The percentage-loss in geometric mean throughput for the realization ω due to ET scheduling is calculated as $100 \times \frac{\Lambda_j^G(\omega)-\Lambda_j^{ET}(\omega)}{\Lambda_j^G(\omega)}$. Note that, the loss is zero if $C_j^* \leq C_j$, where C_j^* varies across different realizations. So, for a given realization, C_j^* characterizes the capacity above which the small cell backhaul is not a bottleneck.

In Fig. 4, we plot the average percentage-loss in geometric mean throughput at one of the small cells due to the equal-time scheme over a set of 100 random realizations for the SCF user association scheme with two choices of δ, and for the Range Extension scheme. As evident from the plots for SCF($\delta = 1.0dB$), we could lose up to about 20% performance if equal-time scheme is employed when the backhaul capacity is very low. The results clearly show the importance of a backhaul-aware scheduling. For SCF($\delta = 6.6dB$) and Range Extension, we see a much smaller performance loss. This shows that the impact of backhaul limitation varies for different UA schemes.

V. CONCLUSION

In this paper, we investigated the impact of limited backhaul capacity on the way optimal scheduling should be performed in a HetNet comprising of a macro base station with a number of overlaid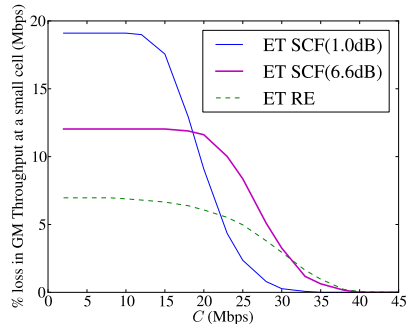 small cells. We showed that the problem can be separated into per-BS independent local PF problems. However, unlike the case with infinite backhaul, where the local PF scheme is equivalent to an equal time scheduling, a local PF scheme with finite backhaul can be of one of three types: *equal-throughput*, *equal-time*, and neither of the two. We derive closed-form expressions for the two critical values of backhaul capacity that can be used to determine the type of the best local scheduling. With numerical results, we quantified the impact of limited backhaul capacity on the system performance and the performance gap between the optimal scheme and a scheme where each user is scheduled for an equal amount of time. These results show that the loss in performance with a sub-optimal scheme can be significant. In conclusion, we have shown that a backhaul limitation-aware optimal scheme can not only be necessary but also easy to implement.

REFERENCES

[1] A. Damnjanovic *et al.*, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun. Mag.*, vol. 18, no. 3, pp. 10 –21, June 2011.

[2] Huwawei, "The second phase of LTE-Advanced," http://www.huawei.com/ilink/en/download/HW_259010.

[3] R. Madan *et al.*, "Cell association and interference coordination in heterogeneous LTE-a cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1479 –1489, December 2010.

[4] D. Fooladivanda and C. Rosenberg, "Joint resource allocation and user association for heterogeneous wireless cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 248–257, 2013.

[5] T. Bu, L. Li, and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," in *Proc. IEEE INFOCOM*, April 2006, pp. 1 –12.

[6] Patrick Donegan, "Small Cell Backhaul: What, Why and How? (white paper)," [Online], Website, July 2012, http://www.tellabs.com/resources/papers/tlab_smallcellbackhaul_wp.pdf.

[7] J. Ghimire and C. Rosenberg, "Resource allocation, transmission coordination and user association in heterogeneous networks: A flow-based unified approach," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1340–1351, 2013.

[8] F. Kelly, "Charging and rate control for elastic traffic," *European transactions on Telecommunications*, vol. 8, no. 1, pp. 33–37, 1997.

[9] L. Li, M. Pal, and Y. Yang, "Proportional fairness in multi-rate wireless LANs," in *Proc. IEEE INFOCOM 2008*. IEEE, 2008, pp. 1004–1012.

[10] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge Univ Pr, 2004.

[11] 3GPP-TSG-RAN-WG1, "Evolved universal terrestrial radio access (EU-TRA)," *3GPP, Tech. Rep. TR 36.814*, 2010.

[12] A. Khandekar, N. Bhushan, J. Tingfang, and V. Vanghi, "LTE-Advanced: Heterogeneous networks," in *Wireless Conference (EW), 2010 European*. IEEE, 2010, pp. 978–982.