On the Need for Coordination Among Base Stations in a Heterogeneous Network

Jagadish Ghimire, Catherine Rosenberg

On the Need for Coordination Among Base Stations in a Heterogeneous Network

Jagadish Ghimire, Catherine Rosenberg
Department of Electrical and Computer Engineering,
University of Waterloo, Canada
Email: {jghimire,cath}@ecemail.uwaterloo.ca

Abstract—**In this paper, we study the impact of different resource allocation schemes, transmission coordination mechanisms among base stations, and user association rules in the context of the downlink of a heterogeneous network comprising of a macro base station overlaid with a pool of low-power pico base stations. We formulate flow-based models for the joint optimization of resource allocation, user association, user scheduling and transmission coordination under a proportional fair throughput objective for two types of resource allocation, i.e, *Full Sharing (FS)* and *Channel Splitting (CS)*, and two types of transmission coordination, i.e., no coordination and ON-OFF coordination. Our formulations allow us to obtain exact solutions for small to medium sized networks. The numerical results show that the performance of FS without transmission coordination among BSs is much worse than CS without coordination. However, FS with our coordination mechanism performs as well as CS with coordination. Interestingly, FS and CS with coordination do not perform much better than CS without coordination. Our study provides a message that we might not need to go for the complexity of transmission coordination among BSs depending upon the resource allocation scheme.**

I. Introduction

Cellular network operators are facing an overwhelming growth in data traffic. One way to deal with this growth is to adopt a heterogeneous network (Hetnet) architecture where low power base stations (BSs) overlay the existing macro infrastructure [1]. Pico BSs (PBSs) are a type of low-power BSs deployed by the operators. In our study, we consider an OFDM based system where a set of PBSs are deployed within the coverage of a macro BS (MBS). We focus on the downlink (D/L). This study is relevant in the context of (but not limited to) 3GPP LTE-Advanced air interface [1]. Overlaying a pool of PBSs over an MBS coverage poses a new set of technological challenges. Interference management is perhaps one of the greatest challenges in this context and is achieved by a mix of several processes such as resource allocation and transmission coordination which will be discussed next.

In the OFDM context, several resource allocation approaches have been proposed including *Channel Splitting (CS)* where PBSs are allocated a pool of subchannels orthogonal to the set of subchannels on which the MBS operates, and *Full Sharing (FS)* where all transmitters (PBSs and MBS) operate in the same set of subchannels, thereby mutually interfering. 3GPP identifies these two approaches as multi-carrier and co-channel deployment, respectively [1]. FS is a relevant approach mainly owing to its simplicity. CS on the other hand is more flexible as the size of the pool of subchannels allocated to the PBSs can be dynamically adjusted if needed. It also results in lower interference. However, the performance of channel splitting largely depends upon the exact way in which the splitting is carried out. The performance of both FS and CS heavily depend upon other network processes (i.e., transmission coordination and user association).

Transmission coordination manages interference by coordinating the BSs, i.e., by scheduling the BSs in time while performing power control. We study a type of transmission coordination, that we call the *ON-OFF* transmission coordination where scheduling is done by allowing a BS at any time to either transmit with the maximum power available or not transmit at all. We compare the performance of such a coordination when performed optimally with the case of *no coordination* where the BSs transmit all the time at full power. Clearly, the optimal ON-OFF coordination of BSs is equivalent to scheduling these BSs in a centralized manner, which is complex but is considered as a viable option in LTE-A [2].

In heterogeneous networks, *user association* is another important network process that affects the performance. Conventionally, in homogeneous networks, a user associates to the BS with the best signal-to-noise-and-interference ratio (SINR). However, in a Hetnet context, such a rule does not work well [3]. *User scheduling* is also an important network process that affects the performance of a wireless network greatly.

The four important network processes introduced so far, namely *resource allocation*, *transmission coordination*, *user association* and *user scheduling* have a complex interplay. For a given system, multiple configurations are possible depending on the resource allocation, transmission coordination, user association and user scheduling options. In our study, we limit the resource allocation options to CS and FS, the transmission coordination options to no coordination and ON-OFF coordination, the user association options to optimal association, and the user scheduling options to a global optimal scheduling based on a global proportional fairness objective function. Hence, we study the following four specific configurations under optimal user association and user scheduling: (1) ***[CS]***: optimal CS without coordination among BSs, (2) ***[CS-O]***: optimal CS with optimal ON-OFF coordination among BSs, (3) ***[FS]***: Full Sharing without coordination among BSs, and (4) ***[FS-O]***: Full Sharing with optimal ON-OFF coordination among BSs. Depending upon the configuration, we formulate a global joint optimal user association and user scheduling problem (***[FS]***), a global joint optimal user association, user scheduling and transmission coordination problem (***[FS-O]***), a global joint optimal user association, user scheduling and resource allocation problem (***[CS]***), and a global joint optimal user association, user scheduling, resource allocation and transmission coordination problem (***[CS-O]***). User scheduling in ***[CS]*** and ***[FS]*** could in fact be performed locally (i.e., each BS scheduling independently from the others) while it has to be done globally along with transmission coordination for ***[CS-O]*** and ***[FS-O]***. A problem formulation based on local scheduling was proposed in [3] and is clearly a simpler approach for solving ***[CS]*** and ***[FS]***. However, problem formulations based on local scheduling do not generalize to ***[CS-O]*** or ***[FS-O]***. By using the concept of "independent sets", we could unify all four configurations into one framework and could model the user scheduling and transmission coordination processes together. This however makes the problem formulations for ***[CS]*** and ***[FS]*** slightly more complex. We accept this complexity for the sake of a unified framework. These problems are formulated using flow-based models for a system "snapshot". By a system "snapshot", we mean that the system parameters like the number of users and the channel gains among the transmitter-receiver pairs are known and are fixed. For such a system snapshot, our optimization models can be used to configure the network processes (i.e, resource allocation, transmission coordination and user association) all at once, optimally. This is thus an offline-static study. For tractability, we assume multipath routing which is equivalent to allowing users to associate with one or more BSs. We later show that these flow-based models with multipath routing provide tight upper bounds on the performance of systems where a user is allowed to associate to only one BS. The upper bounds provided by our model can be used as benchmarks for the performance of practical and online resource allocation, user association and transmission coordination algorithms. Our main contributions can be summarized as follows.

- We formulate a tractable network flow-based unified framework for a multi-rate heterogeneous network comprising of one macro and several pico base stations. This framework allows us to solve the four different problems introduced above under an assumption of multi-BS association.
- Using numerical results, we provide important engineering insights on the throughput performance of different configurations, under a global proportional fairness (PF) objective function.
  - FS without transmission coordination barely justifies the deployment of PBSs in the first place. However, in the presence of a good interference mitigation via ON-OFF transmission coordination, FS performs significantly better.
  - CS performs very well even in the absence of sophisticated transmission coordination.
  - Multipath routing, as assumed in our model, offers an upper bound on the performance of systems where a user associates to only one BS. We show that the upper bounds provided by our model are tight. This also shows that allowing a user to associate with multiple BSs will not offer significant performance gains.

The paper is organized as follows. In Section II, we present the related work. In Section III, we describe the general system model. Section IV contains the problem formulations. In Section V, we discuss the tightness of the solutions based on our flow-based unified framework with multi-BS association. In Section VI, we present the numerical results and the engineering insights.

II. Related work

Resource allocation, user association and transmission coordination are well-known network processes and have been studied extensively in the heterogeneous network context. However, these network processes are not usually studied jointly. [4] studies the performance of different types of resource allocation schemes under a conventional user association rule where a user is served by the BS offering the highest SINR. However, association of users to BSs following simple user association rules are shown to result in sub-optimal performance [3]. Formulating and solving an optimal *user association* problem on the other hand is not trivial. [5], [6] have studied the optimal user association problem under a network-wide proportionally fair throughput optimization framework. [7] studies a number of approximation algorithms to the user association problem under the total throughput maximization objective. User association problems usually belong to a class of problems identified as generalized assignment problems (GAP). Different aspects of GAP can be found in [8] and the references therein.

[3] and [9] are the closest works to ours where a joint resource allocation, user association and reuse pattern optimization problem in a heterogeneous network comprising of a macro and a number of pico base

stations is formulated. They study the performance of optimal Channel Splitting and Full Sharing under the optimal user association and show that channel splitting performs better than full sharing. The strength of their model is that the authors impose very few assumptions on the system model, and yet could solve for system of moderate sizes. The main problem however is that the model is not applicable to system with transmission coordination. In [10], the authors have studied a joint user association and resource allocation problem in Hetnets. They formulate a "semi-static resource allocation" problem with transmit power, user scheduling and association as problem variables. The resulting problem turns out to be a combinatorial problem of very large complexity (given as $(\mathbb{O}(P^{RN}N^{M}))$ for P power levels, R subchannels, N BSs and M users). Due to the intractability of the problem, they propose heuristic algorithms on user association and resource allocation.

From the modeling point of view, we use the idea of flow-based model and the notion of *independent sets* in our optimization framework. Flow-based optimization models with the notion of independent sets is presented in the context of throughput optimization of wireless mesh networks in [11]. Such an approach allows us to obtain exact solutions for networks of small to medium sizes. For large networks, an efficient computation technique based on *column generation* is presented in [12].

III. General System Description

We consider a system with one macro base station (MBS), X short range pico base stations (PBSs), and N user equipments (UEs). MBS is called node 0. $\mathcal{P}$, $\mathcal{U}$ and $\mathcal{N}$ represent respectively the set of all PBSs, UEs, and all nodes (i.e., the MBS, PBSs and UEs). We assume that there is a downlink flow corresponding to each UE. $\mathcal{F}$ is the set of flows. Any flow $f \in \mathcal{F}$ is characterized by f_d (its destination node). Each flow originates at the MBS and terminates at one of the UEs. The system has a set of M OFDM subchannels on a given frequency band with a per subchannel bandwidth of b.

A. Resource allocation

For a given system, we consider two types of resource allocation.

- Channel Splitting (CS): Under CS, K subchannels are allocated to each PBS and the remaining $M - K$ subchannels are allocated to the MBS. K is a parameter to be configured.
- Full Sharing (FS): Under FS, all M subchannels are allocated to each PBS and the MBS. There is no resource allocation parameter to be configured.

B. Power allocation

MBS has a total transmit power of P_{MBS} and each PBS has a total transmit power of P_{PBS}. The power allocation is carried out by assigning equal power to all of the allocated subchannels. Under Channel Splitting (CS), for a given channel split parameter K, the power per subchannel for PBS j ($P_j = P$) and MBS (P_0) is then given as,

$$P_0 = \frac{P_{MBS}}{M-K}, \qquad P_j = P = \frac{P_{PBS}}{K}, \quad \forall j \in \mathcal{P} \quad (1)$$

Under Full sharing (FS), power per subchannel is simply given as,

$$P_0 = \frac{P_{MBS}}{M}, \qquad P_j = P = \frac{P_{PBS}}{M} \qquad \forall j \in \mathcal{P} \quad (2)$$

C. Links

We consider three types of links: (a) *MBS-PBS* links which are the wired backhaul links from MBS to PBSs, (b) *MBS-UE* links which are the direct (wireless) links from MBS to UEs, and (c) *PBS-UE* links which are the direct (wireless) links from PBSs to UEs. The wired MBS-PBS backhaul links are the logical representation of the backhaul network. We represent the backhaul link from MBS to PBS j as L_j with the link capacity C_j. The set of these wired MBS-PBS links is represented as $L = \{L_j : j \in \mathcal{P}\}$. Moreover, we define a *wireless link* l as a tuple (j, i, r_m) for $j \in \{0\} \cup \mathcal{P}$, $i \in \mathcal{U}$ where j and i are the origin node and destination node of the link respectively. r_m is the associated link-rate. $o(l)$, $d(l)$ and R_l are used to represent respectively, the origin node, destination node and the associated link-rate of link l. A wireless link l is *feasible* if the signal-to-noise ratio (SNR) received by $d(l)$ from $o(l)$ is greater than the threshold SNR required for supporting the rate R_l as defined in the Modulation and Coding Scheme (MCS). We take an MCS with a set of supported data rates $\mathcal{R} = \{r_1, r_2, \cdots, r_D\}$ and the corresponding SNR threshold set $\boldsymbol{\beta} = \{\beta(r_1), \beta(r_2), \cdots, \beta(r_D)\}$. Under this MCS, we can define the set of feasible wireless PBS-UE links ($\overline{\mathcal{L}}$) and feasible wireless MBS-UE links ($\overline{\mathcal{L}}_0$) as follows.

$$\overline{\mathcal{L}} = \{(j, i, r_m) : j \in \mathcal{P}, i \in \mathcal{U}, r_m \in \mathcal{R}, \frac{P_j G_{j,i}}{N_0} \geq \beta(r_m)\} \quad (3)$$

$$\overline{\mathcal{L}}_0 = \{(0, i, r_m) : i \in \mathcal{U}, r_m \in \mathcal{R}, \frac{P_0 G_{0,i}}{N_0} \geq \beta(r_m)\} \quad (4)$$

where $G_{m,n}$ represents the channel gain between node $m \in \{\mathcal{P} \cup \{0\}\}$ and $n \in \mathcal{U}$, assumed to be known either by measurements or a channel model and N_0 is the additive white Gaussian noise power. We make no restricting assumptions on the channel gains except that we assume that the channel gains are flat and thus are the same over all subchannels, for a given transmitter-receiver pair. We note here that the links are logical and thus we can have multiple links for a given transmitter-receiver pair (these logical links differ in terms of the rate being supported). By definition, all wired links L_j are feasible. $\mathcal{L} = \overline{\mathcal{L}} \cup \overline{\mathcal{L}}_0 \cup L$ represents the set of all feasible links (including the wired and the wireless links). We do not allow links between any two PBSs or between any two UEs.

D. Independent sets, transmission coordination and user scheduling

We first introduce the notion of "independent sets" before we describe how different transmission coordination mechanisms can be modeled using this notion. An independent set (ISet) is defined as a subset of feasible links which can be activated simultaneously without harmful interference at any of the destination nodes (i.e., all destination nodes can decode their received signals). The definition of an ISet should take into account the "half-duplex" communication constraint of a wireless node, in the sense that a wireless node cannot transmit and receive simultaneously on the same subchannel. Wired links can always be included in every ISet without causing infeasibility of other (wired or wireless) links. The idea of independent set has been used extensively in the area of scheduled wireless mesh networks [11]. Next, we define the two types of transmission coordination scheme that we consider in our study and describe how these coordination mechanisms can be modeled, together with user scheduling using the notion of independent sets.

1) No coordination: Under no coordination, we activate all BSs all the time with their maximum power. We assume that each user has at least one feasible link to a BS when all BSs are transmitting simultaneously both under the ***[CS]*** and ***[FS]*** configuration. An independent set has then a specific simple form. It contains all the wired links in L and one wireless link l from each BS, i.e., an ISet contains L (all MBS-PBS links), 1 MBS-UE link, and X PBS-UE links (i.e., one PBS-UE link from each PBS). Moreover, we require that no two wireless links share the same destination node in an ISet. More formally, we define the set of ISets for ***[CS]*** (represented as $\mathcal{I}^{CS}$) as follows.

$$\begin{aligned}
&\mathcal{I}^{CS} = \{L \cup s \cup \tilde{l} : s \subset \overline{\mathcal{L}}, |s| = X, \tilde{l} \in \overline{\mathcal{L}}_0, \\
&\forall (l_i, l_j \in s, l_i \neq l_j), \\
&\quad d(l_i) \neq d(l_j), o(l_i) \neq o(l_j), d(l_i) \neq d(\tilde{l}) \ \& \\
&\forall l_i, \\
&\quad \frac{P_{o(l_i)} G_{o(l_i),d(l_i)}}{N_0 + \sum_{\substack{l_j \in s, \\ l_j \neq l_i}} P_{o(l_j)} G_{o(l_j),d(l_i)}} \geq \beta(R_{l_i})\} \qquad (5)
\end{aligned}$$

The last constraint in (5) is the feasibility constraint of a link that checks whether the received signal-to-noise-and-interference (SINR) ratio for a given link exceeds the minimum threshold required for supporting the link-rate. As interference does not affect the MBS-UE links, we check the feasibility of the PBS-UE links only.

The set of ISets for ***[FS]*** (represented as $\mathcal{I}^{FS}$) can also be defined similarly, as follows.

$$\begin{aligned}
&\mathcal{I}^{FS} = \{L \cup s \cup \tilde{l} : s \subset \overline{\mathcal{L}}, |s| = X, \tilde{l} \in \overline{\mathcal{L}}_0, \\
&\forall (l_i, l_j \in s, l_i \neq l_j), \\
&\quad d(l_i) \neq d(l_j), o(l_i) \neq o(l_j), d(l_i) \neq d(\tilde{l}) \ \& \\
&\forall l_i \in s \cup \{\tilde{l}\}, \\
&\quad \frac{P_{o(l_i)} G_{o(l_i),d(l_i)}}{N_0 + \sum_{\substack{l_j \in s \cup \{\tilde{l}\}, \\ l_j \neq l_i}} P_{o(l_j)} G_{o(l_j),d(l_i)}} \geq \beta(R_{l_i})\}, \qquad (6)
\end{aligned}$$

We note here that there is now a need to take into account the interference created by the MBS on the PBS-UE links as well as the interference created by the PBSs on the MBS-UE links.

2) ON-OFF transmission coordination: Under ON-OFF transmission coordination, we activate different groups of BSs for a different fractions of time. An optimal ON-OFF coordination determines the right proportion of times for which these groups of BSs have to be activated. Under CS resource allocation with ON-OFF coordination (i.e., ***[CS-O]*** configuration), MBS operates on an orthogonal set of subchannels to the PBSs. It can be activated all the time without affecting the other links (i.e., the PBS-UE links). Hence, under ***[CS-O]***, the MBS transmits all the time and a group of PBSs $g \subseteq \mathcal{P}$ is activated for a proportion of time α_g. Under the FS resource allocation with ON-OFF coordination (i.e., ***[FS-O]***), MBS interferes with the PBS transmissions (and vice-versa) and thus at any time a group of BSs $g \subseteq \{0\} \cup \mathcal{P}$ is activated for a proportion of time α_g. Clearly, such a transmission coordination provides a greater degree of freedom in terms of interference mitigation as compared to the systems without coordination and hence ***[CS-O]*** and ***[FS-O]*** perform better than their respective "no coordination" counterparts. An optimal ON-OFF coordination can be equivalently translated into an optimal ISet activation schedule. The set of ISets for ***[CS-O]***, represented as $\mathcal{I}_O^{CS}$, can be defined as follows.

$$\begin{aligned}
&\mathcal{I}_O^{CS} = \{L \cup s \cup \tilde{l} : s \subset \overline{\mathcal{L}}, \tilde{l} \in \overline{\mathcal{L}}_0, \\
&\forall (l_i, l_j \in s, l_i \neq l_j), \\
&\quad d(l_i) \neq d(l_j), o(l_i) \neq o(l_j), d(l_i) \neq d(\tilde{l}), \ \& \\
&\forall l_i, \\
&\quad \frac{P_{o(l_i)} G_{o(l_i),d(l_i)}}{N_0 + \sum_{\substack{l_j \in s, \\ l_j \neq l_i}} P_{o(l_j)} G_{o(l_j),d(l_i)}} \geq \beta(R_{l_i})\} \qquad (7)
\end{aligned}$$

Compared to (5), we now do not restrict the set of PBS-UE links in each ISet to have a cardinality of X.

Similarly, the set of ISets for ***[FS-O]***, represented as

$\mathcal{I}_O^{FS}$, can be defined as follows.

$$\mathcal{I}_O^{FS} = \{L \cup s : s \subset \overline{\mathcal{L}} \cup \overline{\mathcal{L}}_0,$$
$$\forall (l_i, l_j \in s, l_i \neq l_j),$$
$$d(l_i) \neq d(l_j), o(l_i) \neq o(l_j), d(l_i) \neq o(l_j) \ \&$$
$$\forall l_i \in s,$$
$$\frac{P_{o(l_i)} G_{o(l_i),d(l_i)}}{N_0 + \sum_{\substack{l_j \in s, \\ l_j \neq l_i}} P_{o(l_j)} G_{o(l_j),d(l_i)}} \geq \beta(R_{l_i})\}, \quad (8)$$

We note here that, under ***[CS-O]***, the wireless MBS-UE links $\tilde{l} \in \overline{\mathcal{L}}_0$ are included in each ISet, implying that the MBS transmits all the time. However, under ***[FS-O]***, the MBS is also included in the set of coordinating BSs and can be potentially turned off if required.

Now, after defining precisely the set of ISets corresponding to each configuration, we are ready to model the transmission coordination as well as user scheduling together using the idea of ISet activation schedule. Let α_s be the fraction of time ISet s is scheduled. Then, for a given set of ISets $\mathcal{I}$, the problem of user scheduling (in ***[CS]*** and ***[FS]***) and the problem of user scheduling and ON-OFF transmission coordination (in ***[CS-O]*** and ***[FS-O]***) is equivalent to finding the optimal values of $(\alpha_s)_{s \in \mathcal{I}}$ to maximize the global PF objective function.

Next, we formulate four flow-based optimization models corresponding to the four configurations.

IV. Problem Formulation

A. Routing variable under multipath routing

We first present how user association is included in our framework via the notion of multipath routing which is a key element of our optimization models. Recall that typically a user would associate to exactly one BS. Such a *single-association* would translate into a single-path routing which is very hard to solve since the problem becomes an Integer problem (IP). For tractability, we make the a priori unrealistic assumption that a user can associate to multiple BSs. We call the resulting solutions as *multi-association*. Such assumption yields a much more tractable model. We expect that our optimization models based on multipath routing will provide upper-bounds on the performance of the configurations under *single-association*. It is however unclear a priori if such upper-bounds are tight. We will later show that it is indeed the case. Let x_l^f be the amount of flow f routed through link l. Under a multipath routing assumption, a flow can take a direct MBS-UE path, and/or via one or more of the PBSs in two hops. x_l^f, which models the user association aspect of the system is also called the routing variable in the optimization problems to be presented next.

B. Optimization models for CS

Under CS (both ***[CS]*** and ***[CS-O]***), the set of available subchannels is split such that PBS-UE links are operating on K subchannels whereas MBS-UE links are operating over the remaining $M-K$ subchannels. Let $\mathcal{I}$ be the set of ISets. Clearly, $\mathcal{I} = \mathcal{I}^{CS}$ for ***[CS]*** and $\mathcal{I} = \mathcal{I}_O^{CS}$ for ***[CS-O]***. Given $\mathcal{I}$, P_{MBS}, P_{PBS}, the set of flows $\mathcal{F}$ and the MCS in a network comprising of one MBS, X PBSs and N UEs with known channel gains ($G_{j,i}$) and M subchannels each with a bandwidth of b, the problem is to find the optimal setting of the parameters α_s (the fraction of time each ISet is scheduled), x_l^f (the routing parameter) and K (the resource allocation parameter) such that $\sum_{f \in \mathcal{F}} \log(\lambda_f)$ is maximized where λ_f represents the throughput allocated to flow f. Such an allocation is called a global proportional fair allocation. Proportional fair allocation maximizes the geometric mean throughput offered to the network flows (or equivalently to the users in our case). Proportional fairness (PF) is generally accepted as a solution that exhibits a good trade-off between efficiency and fairness. The optimization model $\tilde{\mathbf{P}}_{\mathbf{CS}}$ is given as follows.

$\tilde{\mathbf{P}}_{\mathbf{CS}}$:

$$\max_{(\alpha_s),(x_l^f),K,(\lambda_f)} \quad \sum_{f \in \mathcal{F}} \log(\lambda_f) \quad (9)$$

Subject to:

$$\sum_{\substack{l \in \mathcal{L}: \\ o(l)=n}} x_l^f - \sum_{\substack{l \in \mathcal{L}: \\ d(l)=n}} x_l^f = \begin{cases} \lambda_f & n = f_s, \\ -\lambda_f, & n = f_d, \\ 0, & \text{o.w.}, \end{cases} \quad \substack{\forall n \in \mathcal{N}, \\ \forall f \in \mathcal{F}} \quad (10)$$

$$\sum_{f \in \mathcal{F}} x_l^f \leq \begin{cases} b(M-K)\sum_{s \in \mathcal{I}} R_l \alpha_s \mathbf{1}_{\{l \in s\}}, & \forall l \in \overline{\mathcal{L}}_0 \\ bK \sum_{s \in \mathcal{I}} R_l \alpha_s \mathbf{1}_{\{l \in s\}}, & \forall l \in \overline{\mathcal{L}} \\ C_{o(l)}, & \forall l \in L \end{cases} \quad (11)$$

$$\sum_{s \in \mathcal{I}} \alpha_s \leq 1 \quad (12)$$

$$1 \leq K \leq M \quad (13)$$

$$\textit{Eqs. } (1), (3), (4), \quad (14)$$

$$x_l^f, \alpha_s, \lambda_f \geq 0, \qquad \forall f \in \mathcal{F}, \forall l \in \mathcal{L}, \forall s \in \mathcal{I} \quad (15)$$

(10) is the flow conservation constraint. It guarantees that the amount of flow f entering an intermediate node is equal to the amount exiting from it. Also, the amount of flow exiting at the source node and the amount of flow entering at the destination node is the total throughput offered to the flow. (11) is the link capacity constraint. It guarantees that the sum of the amount of all flows at a link cannot exceed the scheduled link capacity. The scheduled capacity of a wireless link is given by the product of the number of subchannels allocated to it, the per subchannel bandwidth and the link-rate multiplied by the fraction of times it is scheduled. The scheduled capacity of a wired link is simply the capacity of the link since it is scheduled all the time. (12) is the scheduling constraint. In the above problem, by taking $\mathcal{I} = \mathcal{I}^{CS}$,

we can obtain the optimization model for ***[CS]***. Also, by taking $\mathcal{I} = \mathcal{I}_O^{CS}$, we can obtain the optimization model for ***[CS-O]***. In other words, the optimization model for ***[CS]*** and ***[CS-O]*** are similar, except the definition of the set of ISets, which is different.

$\tilde{\mathbf{P}}_{\mathbf{CS}}$ is a non-convex problem. The non-convexity of the problem is related to the following:

- Power splitting model: the SINR of a link is a non-linear function of K. The general lack of convexity of the rate functions on non-linear SINR functions makes the feasible region non-convex.
- Non-convexity of constraint (11).

However, by fixing K in $\tilde{\mathbf{P}}_{\mathbf{CS}}$, we can define $\mathbf{P}_{\mathbf{CS}}(K)$ as a parametrized problem which is a convex optimization problem and thus can be solved efficiently to obtain a global maximum for small to medium sized problems. The solution to the original problem can then be derived as

$$\mathbf{P}_{\mathbf{CS}}(K^*) \text{ where } K^* = \underset{K\in\{0,1,\cdots,M\}}{\arg\max}\ \mathbf{P}_{\mathbf{CS}}(K) \quad (16)$$

C. Optimization models for FS

Unlike the Channel Splitting based resource allocation, Full Sharing does not require any configuration of a resource allocation parameter. All M subchannels are allocated to all the BSs. Moreover, the capacity constraints for the wireless links need to be redefined. The capacity constraint is obtained by replacing $M-K$ by M for the MBS-UE links and K by M for the PBS-UE links, as shown in (19). We formulate the optimization problem for FS with the set of ISets $\mathcal{I}$, as follows.

$\tilde{\mathbf{P}}_{\mathbf{FS}}$:

$$\max_{(\alpha_s),(x_l^f),(\lambda_f)} \quad \sum_{f\in\mathcal{F}} \log(\lambda_f) \quad (17)$$

Subject to:

$$\sum_{\substack{l\in\mathcal{L}:\\ o(l)=n}} x_l^f - \sum_{\substack{l\in\mathcal{L}:\\ d(l)=n}} x_l^f = \begin{cases} \lambda_f & n=f_s,\\ -\lambda_f, & n=f_d,\\ 0, & \text{o.w.,}\end{cases} \ \ \substack{\forall n\in\mathcal{N},\\ \forall f\in\mathcal{F}} \quad (18)$$

$$\sum_{f\in\mathcal{F}} x_l^f \le \begin{cases} bM\sum_{s\in\mathcal{I}} R_l\alpha_s\mathbf{1}_{\{l\in s\}}, & \forall l\in\overline{\mathcal{L}}_0\cup\overline{\mathcal{L}}\\ C_{o(l)}, & \forall l\in L\end{cases} \quad (19)$$

$$\sum_{s\in\mathcal{I}}\alpha_s \le 1 \quad (20)$$

$$Eqs.\ (2),(3),(4), \quad (21)$$

$$x_l^f,\alpha_s,\lambda_f \ge 0, \qquad \forall f\in\mathcal{F},\forall l\in\mathcal{L},\forall s\in\mathcal{I} \quad (22)$$

where $\mathcal{I} = \mathcal{I}^{FS}$ results in the optimization model for ***[FS]*** and $\mathcal{I} = \mathcal{I}_O^{FS}$ results in the optimization model for ***[FS-O]***. In other words, the optimization model for ***[FS]*** and ***[FS-O]*** are similar, except for the definition of the set of ISets, which is different.

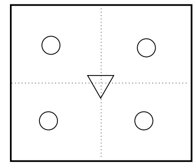


Fig. 1. 4 Pico BSs are placed in a grid layout on a macro coverage of a 500 *m.* × 500 *m.* square

V. Tightness of the upper-bounds based on multipath routing

So far, our models consider multipath routing and thus a UE can associate to more than one BS. However, in cellular networks, conventionally, a UE associates to only *one* BS. In the earlier section, we called such an association as *single-association* and mentioned that the optimization problem under single-association restriction results in an integer non-linear program which has a much larger complexity than its multi-association counterpart. Clearly, the optimization models based on multipath routing provide upper bounds to systems with single-association constraints. In order to show that the upper-bounds based on multipath routing are tight, we present a simple user association rule called the *pico-cell first* (PCF) rule with a parameter β. This PCF(β) rule was first introduced in [3]. Under the PCF(β) rule, a user associates to the PBS that provides the best SINR as long as it is greater than or equal to β, even if the MBS is offering strictly greater SINR. This simple user association rule can be translated into the routing variables (x_l^f). For example, if user i associates to BS j, the corresponding flow routing variable x_l^f (where flow f corresponds to user i and thus $f_d = i$) will be 0 for all wireless links l that do not belong to BS j (i.e., $o(l) \ne j$). We can compute the optimal settings of the other parameters under a given user association by solving the four problems given above with fixed (x_l^f). This will provide a feasible solution to the optimal single-association problem. By numerically showing that these feasible solutions are performing close to the upper-bounds provided by the multipath routing, we will show that our flow models provide good upper-bounds.

VI. Numerical Results

We take a square deployment area of length 500 m with a centrally placed MBS. $X = 4$ PBS are deployed on a 2×2 grid as shown in Figure 1. We consider $N = 75$ users , deployed uniform randomly in the area. We use the path-loss model recommended by 3GPP [2]. The path loss $PL_{j,i}$ for the transmitter-receiver pair (j,i) separated by a distance d_{ji} (m.) is given as $PL_{0,i} = 20 - A.G. + 128.1 + 37.6\log_{10}\left(\frac{d_{0i}}{1000}\right)(dB), d_{0i} \ge 35m$ for MBS links and $PL_{j,i} = 20 - A.G._p + 140.7 + 36.7\log_{10}\left(\frac{d_{ji}}{1000}\right)(dB), d \ge 10m$ for PBS links (i.e., $j \in \mathcal{P}$). $A.G.$ represents the antenna gain of

TABLE I
AVAILABLE RATES AND THE CORRESPONDING SNR THRESHOLDS

Threshold SNR (dB)	-6.5	-4	-2.6	-1	1	3	6.6	10	11.4	11.8	13	13.8	15.6	16.8	17.6
Efficiency (bits/symbol)	0.15	0.23	0.38	0.60	0.88	1.18	1.48	1.91	2.41	2.73	3.32	3.9	4.52	5.12	5.55

the macro downlink and $A.G._p$ is the antenna gain of the pico downlink. We take $A.G. = 15dB$ and $A.G._p = 5dB$. On top of this path-loss model, we further apply log-normal shadowing with zero mean and standard deviation of 8 dB to obtain random path-loss $\overline{PL}$. i.e., $\overline{PL}_{j,i} = PL_{j,i} + \mathbf{N}(0, 8)$ where $\mathbf{N}(a, b)$ is a normal random variable with mean a and standard deviation b. Then, the channel gains can be obtained as $G_{j,i} = 10^{-\frac{\overline{PL}_{j,i}}{10}}$. We take $P_{MBS} = 46dBm$, $N_0 = -112.4245dBm$, and $M = 100$ subchannels each with $b = 180KHz$ bandwidth. We assume that the system employs an adaptive modulation and coding scheme with 15 discrete rates. The rates (expressed in terms of bits/symbol efficiencies, r_j) and the corresponding threshold SNRs are listed in Table I. The rate R achieved when employing a scheme of efficiency r_j is given as $r_j \frac{N_{sc}N_{sym}}{T_{fr}}$ where N_{sc} and N_{sym} represent respectively the number of sub-carriers per subchannel and the number of symbols per sub-frame. T_{fr} is the duration of a sub-frame. For a given system, the factor $\frac{N_{sc}N_{sym}}{T_{fr}}$ is a fixed quantity and thus we normalize the rates to this quantity. We assume that the wired MBS-PBS backhaul links are not the bottleneck and hence are considered to be of infinite capacity. Our model however can also be used for finite capacity backhaul links. We study 100 random realizations of the network, where each realization is obtained by generating 75 users uniformly randomly in the deployment area. The numerical results are obtained by solving convex optimization problems to global maximum for each realization using the commercial solver, Minos [13]. Since PF maximizes the geometric mean (GM) of the throughput of the users, i.e., $(\prod_{f \in \mathcal{F}} \lambda_f^*)^{(1/N)}$, the GM throughput is taken as the performance metric with which we compare different configurations. Fig. 2 shows the GM throughput for one typical realization for different configurations as a function of P_{PBS}. The horizontal line represents the GM throughput under the case without PBS deployment. This MBS-only case is considered as the *base-case* for calculating the throughput gains. The throughput gain under a given configuration Y for a particular realization i is given as

$$\mathcal{G}_Y(i) = 100 \times \frac{\lambda_Y^{GM}(i) - \lambda_0^{GM}(i)}{\lambda_0^{GM}(i)} \quad (23)$$

where $\lambda_Y^{GM}(i)$ is the GM throughput of realization i under configuration Y. Fig. 3 shows the average gain in GM throughput obtained by the deployment of PBSs over the base-case. The average is calculated as the sample average of the gain over all 100 realizations.

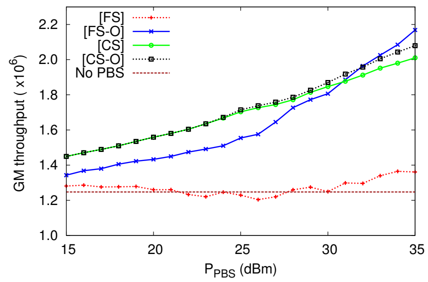


Fig. 2. GM Throughput, a particular realization

These results show the following.

- In absence of transmission coordination, FS performs very poorly as compared to the optimal CS scheme. By comparing with the base-case, we show that not only is ***[FS]*** a magnitude worse than ***[CS]***, it does not offer any significant throughput gains over the base-case. It would not make much sense to deploy PBSs under ***[FS]*** despite its simplicity.
- On the other hand, FS with optimal ON-OFF transmission coordination (***[FS-O]***) performs very well, often exceeding the performance of CS with ON-OFF coordination (***[CS-O]***) at high powers. This can be a motivating result for the argument in favor of FS. Such a magnitude of improvement is not experienced under CS with the introduction of coordination.
- *Transmission coordination versus Channel Splitting:* Under CS, we can operate the network without transmission coordination which is much simpler and yet sacrifice very little in terms of network throughput, whereas under FS, transmission coordination is essential for satisfactory throughput performance. This is a major result in favor of CS over FS.
- We also see that the GM throughput of ***[FS-O]***, ***[CS]*** and ***[CS-O]*** increases with P_{PBS}. However, this is not true under ***[FS]***. Under CS, this can be attributed to the fact that increasing P_{PBS} does not affect the SINR of MBS-UE links, and in the same time results in better SINR for each PBS-UE links. On the other hand, under FS, increasing P_{PBS} causes more interference to MBS-UE links and thus the behavior is not predictable. However, with an added degree of freedom in terms of ON-OFF transmission coordination, ***[FS-O]*** benefits from the increase in P_{PBS}.

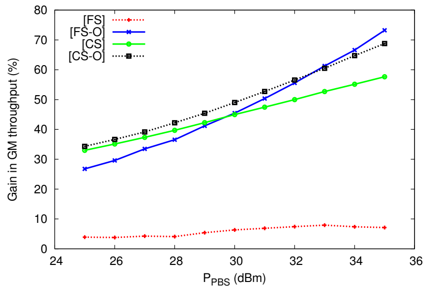
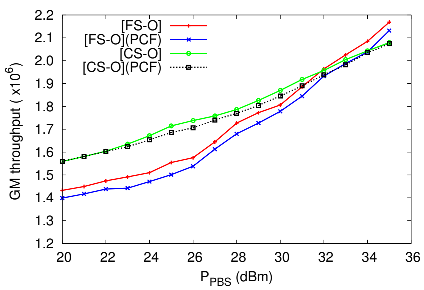
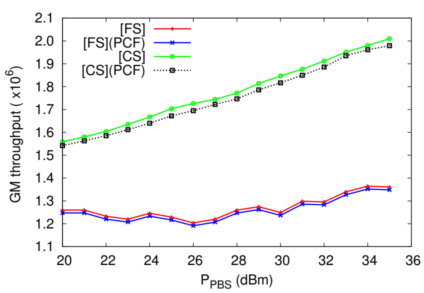


Fig. 3. Average gain in GM Throughput



(a) ***[FS-O]*** and ***[CS-O]***



(b) ***[FS]*** and ***[CS]***

Fig. 4. GM Throughput for Optimal and PCF association.

So far, the results are based on the multipath routing framework. Next, we present the results obtained under the PCF(β) rule for each configuration and show the results along-side the results obtained with multipath routing. For each configuration and each realization, we select the value of β that provides the best performance in terms of the GM throughput. In Figures 4(a) and 4(b), we plot the GM throughput of the optimally tuned PCF(β) along with the GM throughput under the optimal multi-association. These results show that a well tuned PCF results in performance very close to the optimal solution based on multi-association (a difference of less than 3%). A solution based on PCF is a feasible solution to the single-association. This closeness validates the tightness of the solutions obtained by our flow model under multipath routing.

VII. Conclusion

In this paper, we presented a flow-based unified optimization framework for the joint optimization of resource allocation, user scheduling and user association under the optimal ON-OFF transmission coordination of base stations. More complex coordination where we play with different power-levels are also possible but it would make the problem much more difficult. Under a global proportional fairness objective, we solved four configurations in terms of resource allocation and transmission coordination under optimal user scheduling and multi-path routing. We showed that under Full Sharing based resource allocation without transmission coordination, the deployment of PBSs over a macro coverage is not expected to offer much performance benefits. However, under the optimal ON-OFF transmission coordination, we showed that Full Sharing based systems can reach or exceed the performance of systems with optimal Channel Splitting. Despite this, on a more practical note, we showed that optimal Channel Splitting performs well even with no transmission coordination and thus can still be the best practical solution compared to Full Sharing which requires a fine-grained ON-OFF coordination. We also showed that our multipath routing framework provides a tight upper bound to systems with optimal single-association.

References

[1] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *Wireless Communications, IEEE*, vol. 18, no. 3, pp. 10 –21, june 2011.

[2] 3GPP-TSG-RAN-WG1, "Evolved universal terrestrial radio access (EUTRA)," *3GPP, Tech. Rep. TR 36.814*, 2010.

[3] D. Fooladivanda, A. A. Daoud, and C. Rosenberg, "Joint channel allocation and user association for heterogeneous wireless cellular networks," in *Proc. IEEE PIMRC*, Sept. 2011.

[4] W. Lei, W. Hai, Y. Yinghui, and Z. Fei, "Heterogeneous network in LTE-advanced system," in *Proc. IEEE ICCS*, Nov. 2010.

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

[7] D. Amzallag and D. Raz, "Resource allocation algorithms for the next generation cellular networks," pp. 99–129, 2010.

[8] D. Shmoys and É. Tardos, "An approximation algorithm for the generalized assignment problem," *Mathematical Programming*, vol. 62, no. 1, pp. 461–474, 1993.

[9] D. Fooladivanda and C. Rosenberg, "Joint resource allocation and user association for heterogeneous wireless cellular networks," *submitted to IEEE trans. on Wireless Comm.*, Jan. 2012.

[10] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "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.

[11] A. Karnik, A. Iyer, and C. Rosenberg, "Throughput-optimal configuration of fixed wireless networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 5, pp. 1161–1174, 2008.

[12] J. Luo, C. Rosenberg, and A. Girard, "Engineering wireless mesh networks: Joint scheduling, routing, power control, and rate adaptation," *IEEE/ACM Trans. Netw.*, vol. 18, no. 5, pp. 1387–1400, 2010.

[13] "Minos 5.5," [Online], Website, http://www.sbsi-sol-optimize.com/asp/solproductminos.htm.