A Benchmark for D2D in Cellular Networks: The Importance of Information

Yiğit Özcan, Catherine Rosenberg
University of Waterloo
{yozcan,cath}@uwaterloo.ca

Fabrice Guillemin
Orange Labs, France
fabrice.guillemin@orange.com

***Abstract*—Many new mobile applications create traffic among cellular users. We define *intra-cellular traffic* as the traffic from one cellular user to another user in the same cellular network. This type of traffic introduces new challenges for cellular network operators. Most work in the literature focuses on the possibility to utilize the direct links between those users, if they are close to each other (this is called device-to-device (D2D) communication), to by-pass the base station. However, implementing D2D is not easy, especially because detecting that a traffic is intra-cellular is difficult. In this paper, we assume that we know how to detect if a traffic is intra-cellular or not and focus on designing a type-aware scheduler (i.e., a scheduler which has the information on the type of traffic) in a case where direct communications between users is not enabled. This scheduler can be seen as the benchmark against the case where direct communications are allowed. We show that performance gain can be obtained by jointly scheduling the uplink and downlink with respect to the case where the scheduler is blind to the types. We show for a homogeneous network that when the traffic types are known to a scheduler, a significant performance gain (up to 28%) can be achieved compared to the case where the traffic types are not known. We also analyze heterogeneous networks that consist of macro cells and small cells and show that up to a 36% performance gain can be obtained by performing type-aware user association jointly with user scheduling.**

***Index Terms*—User Scheduling, Uplink, User Association, D2D, Fairness**

I. Introduction

A recent trend in cellular networks is the so called Device-to-Device (D2D) communications that enable cellular users, close to each other, to exchange data directly without using the Base-Stations (BS) as a relay node [1]. Although it has been a hot topic in the last few years, many challenges remain to implement D2D communications. For example, interference management is complicated by D2D communications, detecting that a traffic is intra-cellular (IC) is also hard, and it is not easy to obtain the channel gains between the cellular users and hence to decide if an IC traffic should use D2D communications or be sent via BS(s). In this paper, we focus on IC traffic when D2D communications are not allowed, i.e., the network operates in a classic cellular mode. We show that there is a significant performance gain if the network processes, such as user scheduling and user association, are performed with the knowledge of traffic types even when the direct links between users are not utilized. However, this requires the joint operation of the uplink and downlink.

978-1-5386-3531-5/17/$31.00

We believe that the benchmark against the D2D case (i.e., where direct communications are allowed), which is compared for scheduling or user association, should be type-aware since it is unfair to assume that this information is available for the D2D case and is not available for the case where direct communications are not allowed.

User scheduling [2] is the process of allocating radio resources and power to users at a very short time scale. In a conventional cellular network, uplink and downlink scheduling are performed separately. There are different challenges for uplink scheduling and downlink scheduling. For example, in a realistic multi-cell system, inter-cell interference (ICI) plays a crucial role in system performance and dealing with ICI is not easy especially on the uplink. The power budget on the downlink of a cell comes from a single source (the BS) while it comes from different sources (the user equipments (UE)) on the uplink. Apart from all those challenges, we show that there is an additional challenge that arises if we want to take advantage of the knowledge that a traffic is IC.

We will use the term of flow f_{ij} for a *unidirectional* traffic between cellular users i and j. If i and j are UEs belonging to the same cellular network, we say the flow is IC, if i is a UE and j is a node outside the cellular network, we call it an uplink flow, and if j is a UE and i is a node outside the cellular network, we call it a downlink flow. In a conventional cellular system, an IC flow f_{ij} would use two radio links, one of which is from i to its BS (uplink) and the other one is from the BS of j to j (downlink). Since these two radio links are coupled, we should, if at all possible, couple their scheduling to avoid congestion. Indeed, in a conventional system, if i is very close to its own BS, it might be allocated a very high rate and this could be a problem on the downlink of j especially if j is far from its BS. In practical systems, this problem would translate into buffer overflows. Note that the BSs for i and j might be different or might be the same. To illustrate the gain of a type-aware solution, we will restrict our study to a single macro cell (while taking ICI from the rest of the cells into account) and only call IC flows the flows for which i and j share the same BS. However, we can generalize the concept in a CRAN-based system [3] to the case where i and j are on different BSs if these BSs are connected to the same CRAN, since we could coordinate the scheduling of the uplink of i and the downlink of j within the CRAN.

We show in this paper that the performance can be significantly improved when the uplink and downlink of an IC flow are jointly scheduled (this requires the knowledge of the type of a flow). We call this type of schedulers *type-aware* and we call the schedulers that do not take the type of flows into account and hence schedule the uplink and the downlink independently *type-blind.*

In a cellular network, each UE has typically multiple flows, possibly of different types. Therefore, the issue of fairness among users or flows arises. In this paper, we define the concept of *device fairness* that ensures fairness at the device-level irrespective of the number of flows each user has.

We will explain the type-aware scheduler in an homogeneous context and then focus on heterogeneous networks (HetNets), which consist of macro base stations (MBS) and small cells (SC) [4]. In that case, user association (UA), the process of associating each user to either an MBS or one of the small cells, is critical. We show that when UA is type-aware, system performance can be further improved.
In summary, our contributions can be described as follows:

- We propose an uplink scheduler that offers a better granularity than the one proposed in [5]. This will be critical for the type-aware scheduler.
- We revisit the concept of fairness in the case of multiple flows per UE and propose a single metric that measures (device) fairness and efficiency at the same time.
- We first formulate and study the coupling of the uplink and the downlink schedulers in the type-aware case for an homogeneous system. We compare, for different mixes of flows, the gains in performance with respect to the type-blind case, where the schedulers are decoupled .
- Finally, we consider a HetNet configuration, for which we formulate and study a type-aware UA. We compare the gains in performance with respect to the type-blind UA case for different mixes of flows.

The two main messages of the paper are that 1) important network processes, such as user scheduling and association, should be performed with the knowledge of the type of flows; 2) the uplink and downlink should be jointly scheduled to obtain the best performance when there is IC traffic in the system. This is easy to do when the source and the destination of a flow are associated to the same BS or to different BSs in the same CRAN.

Next, we outline the related work in Section II. We explain the system model in Section III. We define the fairness and efficiency metric in Section IV. We explain the type-blind and type-aware schedulers in Sections V and VI, respectively for a homogeneous system. We compare the performance of the two schedulers in Section VI-B. We examine user association and user scheduling in HetNets with IC flows in Section VII.

II. Related Work

A. D2D communications

D2D communication [1] has emerged as a new paradigm that allows cellular users to communicate directly by bypassing the BS(s). Many works in the literature study resource allocation problem for D2D communication such as [6] and [7]. However, they do not consider a benchmark that takes into account the traffic types and fairness among users.

B. Coupled Uplink and Downlink Scheduling

In a conventional cellular system, the uplink and downlink resources are separated from each other either in the frequency or time domain. The first one is called Frequency Division Duplexing (FDD) and the latter is called Time Division Duplexing (TDD) [8] and the scheduling of the resources are decoupled. There is some work on dynamically allocating the resources between the uplink and downlink such as Dynamic TDD [9], [10]; however, none of them coordinates the scheduling of the uplink and the downlink and they do not consider IC traffic.

C. Fairness with IC Traffic

In a cellular system, where UEs can have multiple flows of possibly multiple types, the notion of fairness has to be revisited carefully. Different fairness criteria for D2D communications are defined in [11]. However, only one type of flow per user is considered. We will define a new fairness metric for multiple types of flows per device.

D. User Scheduling on the Uplink

User scheduling on the downlink has been widely studied in the literature [2], whereas the literature for the uplink is scarce. Most uplink work in the literature considers simplistic scenarios. For example, most of them consider a single cell system while ignoring the ICI. However, ICI plays a crucial role that has a direct impact on the system performance [12], [13]. Since the ICI cannot be known exactly on the uplink (while it can be on the downlink under certain assumptions), it has to be estimated. In a realistic system, some of the resources might be lost due to bad ICI estimation and it is important to consider these losses when quantifying the performance of an uplink scheduler. Most of the uplink work also uses a rate function of the type $\log(1 + SINR)$ [14] instead of a piece-wise constant rate function similar to the one used in LTE, which gives very different results. An uplink scheduler is proposed in [5] that works with a unique ICI estimate for all PRBs in a frame. We will show that this scheduler is not flexible enough for the case of IC traffic and we will propose a modified version that achieves a better granularity.

In summary, there is no work focusing on fairness and how IC flows should be dealt with in a conventional network once their type is detected with a realistic model. However, this would be the fair benchmark for the case where D2D communications are allowed.

III. System Model

We consider a TDD OFDMA system where the time is divided into subframes (of unit duration) and the available frequency band is partitioned into subchannels. A Physical Resource Block (PRB) is the smallest scheduling unit that consists of one subchannel c and one subframe t. A multi-cell environment is considered, where each cell has one MBS equipped with an omni-directional antenna. For the HetNet configuration we consider in Section VII, there are also two

SCs in each cell. We focus on the macro cell in the middle (that we call cell 0) while taking into account the ICI received from the other cells. A UE has a power budget of P_U, an MBS of P_M and a SC of P_S.

There are rM subchannels allocated to the system. A reuse factor r is implemented among the macro cells, which means the group of macro cells that use the same set of subchannels are allocated M subchannels. We assume that all the subchannels have the same characteristics and are flat for a given pair {*i*,*j*} (where i is a UE and j is a BS) during a scheduling frame, which is composed of T subframes.

A. The Flows

In our system, a user i may have up to four types of flows:

- U/L: Uplink flow to the Internet
- D/L: Downlink flow from the Internet
- IC$_o$: From i to another device in the same macro cell
- IC$_d$: From another device in the same macro cell to i

Hence, we categorize a flow from i to a device outside cell 0 as an uplink flow and a flow from a device outside cell 0 to i as a downlink flow. We assume that each user has, at most, one flow of the first three types (it can have multiple flows of the fourth type). Let $y(i)$ be the destination of the IC$_o$ flow of user i (if any) and let $z(i)$ be the source of the IC$_d$ flow of user i (if any). We define the throughput for each flow type as λ_i^{UL}, λ_i^{DL}, $\lambda_i^{IC_o}$, and $\lambda_i^{IC_d}$ for the U/L, D/L, IC$_o$ and IC$_d$ flows of user i, respectively. An IC flow contains two hops: the hop between the source (either i or $z(i)$) and the BS and the hop between the BS and destination (either i or $y(i)$).

We assume that the binary matrix X, whose dimension is 4xN, showing whether a user has a given type of flow or not, is given. More specifically:

- $X(1, i)$ = 1 if device i has a U/L flow
- $X(2, i)$ = 1 if device i has a D/L flow
- $X(3, i)$ = 1 if device i has a IC$_o$ flow
- $X(4, i)$ = 1 if device i has IC$_d$ flow(s)

B. SINR and Power

If transmitter i is given PRB (c, t) to transmit to receiver j with power p, the SINR on that PRB at j is:

$$\gamma_j(c,t) = \frac{p \times G_{i,j}^c}{\mu + I_j(c,t)}, \tag{1}$$

where $G_{i,j}^c$ is the channel gain between i and j on c, μ is the thermal noise on that channel, and $I_j(c, t)$ is the ICI at j due to co-channel transmission on c at t. We assume that there is a rate function, $f(.)$ that maps the SINR on each subchannel to a corresponding data rate for a given block error rate. It is constant per part. It selects the best modulation and coding scheme (MCS) for a given SINR.

On the downlink, we assume that a BS allocates all its power equally to its allocated subchannels and we know the interferers which are the other BSs transmitting on the same channels as the MBS in cell 0. Then, if we assume that the channel gains between i and all these BSs are known exactly, we can compute exactly the ICI seen at each user device and hence the SINR.

On the uplink calculating the ICI on the uplink (i.e., seen by the BS) is not as easy as calculating it for the downlink since we do not know how the power is used and who the interferers in the other cells are (it depends on the schedule in the cells). Hence, we have to estimate the ICI and take decoding errors, due to a bad estimation of the ICI, into account.

The final parameter that we introduce is $0 < \beta < 1$, which determines how much of the scheduling frame time is allocated to the downlink (βT) and the uplink ($(1 - \beta)T$). Selecting β might not be so straightforward when IC traffic is present.

IV. A Single Metric for Fairness and Efficiency

Since each user might have a different number of flows, it is very important to decide how to define fairness among those users. A network operator can offer fairness to flows irrespective of the devices on which they are or it can offer fairness to devices without considering the number of flows each device has. We focus on a *device fairness*, where each device is treated as a single entity rather than considering each flow separately. This is because offering flow fairness might cause unfairness among the users with different numbers of flows by assigning higher weights to the users with higher numbers of flows. To this end, we define a utility function for each user that is the geometric mean (GM) of the throughput of each flow of types 1, 2, and 3. The reason why we do not include type 4 flows, i.e., IC$_d$ flows, is that we do not want to double count flows. This will become clearer when we define the objective function across all users in the cell. Let $F(i)$ be the set of flows of types 1, 2, and 3 of user i where $1 \le \|F(i)\| \le 3$. Then, the utility φ_i of user i is defined as:

$$\varphi_i = \sqrt[|F(i)|]{\prod_{j \in F(i)} \lambda_i^j}, \tag{2}$$

where $F(i) = \{j \in \{1, 2, 3\} \mid X(j, i) = 1\}$. Note that if we use an arithmetic mean instead of a geometric mean in Eq. (2), we might assign zero resource to some IC flows.

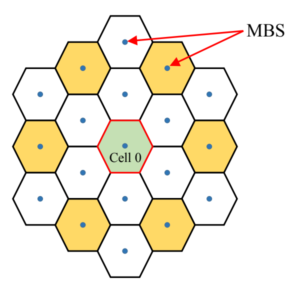


Fig. 1: Homogeneous network configuration with a reuse factor (r) of 3. The cells interfering with cell 0 are highlighted in yellow.

V. Type-Blind Scheduling

In this section, we explain how user scheduling is performed in a conventional homogeneous cellular network. We consider the homogeneous network shown in Figure 1[1]. Typically, user

[1]The hexagonal shape of the coverage areas is only to be taken symbolically. It does not represent the exact geometrical shape of a coverage area.

scheduling is performed separately for the uplink and the downlink and the scheduler does not know if a flow is IC or not, i.e., it is *type-blind* and we explain its operation in this subsection. It allocates resources on the downlink for a time βT and then on the uplink for the remaining frame time. The downlink scheduler that we describe is state-of-the-art (SoA) while the uplink scheduler is new since the SoA is not adequate as will be discussed.

A. Downlink Scheduler

We consider a simple downlink scheduler [15], where a BS allocates equal power to all its available subchannels and serves one user at a time on all the subchannels. Then, the users are time scheduled. With these assumptions, we can compute the exact ICI (and the SINR) seen at each user and hence, knowing the rate function $f(.)$, we can pick the best MCS, i.e., the one that yields the maximum achievable rate for each user. For a given realization ω, where a realization corresponds to the random deployment of the users within cell 0 and their corresponding channel gains, let r_i be the rate user i sees (over all the channels) when it is scheduled and U_0 be the set of users associated to BS_0. Then, assuming the full buffer case, the following problem maximizes the proportional fair objective function:

$$\mathbb{P}_{\mathbb{DL}}(\omega): \quad \max_{\alpha_i \geq 0, \lambda_i} \sum_{i \in U_0} \log(\lambda_i) \tag{3}$$
$$s.t. \quad \lambda_i = \alpha_i r_i \ , \ \forall i \in U_0 \tag{4}$$
$$\sum_{i \in U_0} \alpha_i \leq \beta T \tag{5}$$

where λ_i is the throughput user i sees and α_i is the fraction of time user i is scheduled. Note that $\mathbb{P}_{\mathbb{DL}}(\omega)$ is a convex problem and it was previously shown that the optimal scheduler allocates equal time to each user [15] if there is no additional constraint, i.e., $\lambda_i = \frac{\beta T}{|U_0|} r_i$.

B. Uplink Scheduler

As discussed above, scheduling on the uplink is more challenging. First, the ICI cannot be known exactly since the transmitters and the power they allocate on each subchannel in the neighboring cells are unknown. Furthermore, power allocation is not as simple as downlink since there are multiple possible transmitters in a cell.

The scheduler proposed in [5] allocates $m_i \geq 1$ subchannels to user i for the duration of a frame, where m_i is an integer. The power budget of user i is divided equally between these m_i channels. The scheduler uses the same ICI estimate $\hat{I}$ on all subchannels. This scheduler is not flexible enough for the type-aware scheduler because it might be necessary to allocate a user less than T PRBs (i.e., one subchannel during the whole frame). Hence, we propose a scheduler that can be seen as a more flexible version of the scheduler proposed in [5]. We continue to use the same ICI estimate $\hat{I}$ on all subchannels.

We assume the subchannels are organized into blocks of different sizes and that a UE can only be allocated one block at a time for transmission. If a UE is allocated a block of $1 \leq k \leq M$ subchannels at a given time, we assume that its power budget is shared equally among the k subchannels. Let the number of blocks of size k be t_k. Our uplink scheduler computes for every frame the values of t_k (since the realization can change from one frame to another) and allocates a block of size k to user i for a fraction of time θ_i^k.

Let $R_i^k(\hat{I})$ be the rate seen by user i on a subchannel block of size k when the ICI estimate is $\hat{I}$. This can be computed by first computing the SINR with equation (1) with $\hat{I}$ as the ICI estimate and then mapping this SINR to a data rate using the rate function $f(.)$ and k. Specifically, for a realization ω, given $\hat{I}$, $R_i^k(\hat{I})$ and U_0, the uplink scheduler solves the following problem $\mathbb{P}_{\mathbb{UL}}(\omega, \hat{\mathbb{I}})$:

$$\mathbb{P}_{\mathbb{UL}}(\omega, \hat{\mathbb{I}}): \quad \max_{(\theta_i^k),(t_k),(\lambda_i((\hat{I}))} \sum_{i \in U_0} \log(\lambda_i(\hat{I})) \tag{6}$$
$$s.t. \quad \lambda_i(\hat{I}) = \sum_{k=1}^{M} \theta_i^k R_i^k(\hat{I}) \ , \ \forall i \in U_0 \tag{7}$$
$$\sum_{k=1}^{M} \theta_i^k \leq (1-\beta)T, \ \forall i \in U_0 \tag{8}$$
$$\sum_{i \in U_0} \theta_i^k \leq t_k(1-\beta)T, \ \forall k \in \{1..M\} \tag{9}$$
$$\sum_{k=1}^{M} k t_k \leq M \tag{10}$$
$$t_k \in \mathbb{Z}^+, \ \theta_i^k \geq 0, \ \forall k \in \{1..M\}, \ \forall i \in U_0 \tag{11}$$

The throughput $\lambda_i(\hat{I})$ of user i is defined as the sum of rates it sees on each block (constraint (7)). Constraint (8) ensures that the total time a user is scheduled cannot exceed the uplink frame duration. Constraint (9) ensures that the total time users are scheduled on blocks of size k cannot exceed $t_k(1-\beta)T$. Constraint (10) enforces that the total number of subchannels allocated to the blocks cannot exceed the total number of subchannels M.

A crucial part of this scheduler is the computation of the $R_i^k(\hat{I})$'s. If we use an optimistic ICI estimate (i.e., a small value for $\hat{I}$), we might see many losses since the real ICI might be much higher. We define the *goodput* seen by a user as the effective rate this user sees after taking into account PRB losses[2]. For a low value of $\hat{I}$, we will show that the GM estimated by solving $\mathbb{P}_{\mathbb{UL}}(\omega, \hat{\mathbb{I}})$ is very different from the goodput GM. We illustrate this numerically next.

We consider the 19 cell system shown in Figure 1 and focus on cell 0. We only consider the six other cells that use the same set of subchannels as cell 0. P_U is set to 24 dBm. The number of subchannels M used by cell 0 is 33. We use the following distance based path loss formula: $128.1 + 37.6 \times \log_{10}(d/1000)$ [8]. The antenna gains are 15 dBi for the BS and 0 dBi for the UEs. Penetration loss is set to 20 dB. We obtain the

[2]Note that on the downlink, the scheduler computes the goodput directly since we assume that it has the exact ICI value.

channel gains by further applying a log-normal shadowing of 8 dB standard deviation. We use the piece-wise constant rate function given in Table III of [15]. The number of users in U_0 is set to 10. We assume there are the same number of users in each of the six cells.

Since we focus on proportional fairness (PF), efficiency and fairness can be measured using a single metric, the geometric mean (GM) of the user throughputs [16] defined as $\Gamma(\omega) = (\Pi_{i=1}^{N} \lambda_i(\hat{I}))^{1/N}$, where N is the number of users and $\lambda_i(\hat{I})$ is the throughput of user i when the ICI estimate is $\hat{I}$.

We consider a snapshot scenario in which we create a global realization made of $N = 10$ users per cell. We schedule each cell locally using the same ICI estimate $\hat{I}$ and obtain the estimated GM for cell 0 as the value of the local objective function. The resultant schedule is mapped to the PRBs for each cell. We can then compute the goodput GM since we now have the real ICI values (once we know the scheduling in each cell, we know the exact ICI). We perform this simulation for multiple time slots with different PRB allocation and take the time average. The decoding rule is as follows for a given PRB: If the real SINR is higher than the threshold of the MCS, the user gets the rate of that MCS from the PRB. Otherwise, the PRB cannot be decoded and we consider it lost. We repeat this computation for 100 realizations and take the average GM goodput for cell 0. The results are given in Figure 2.

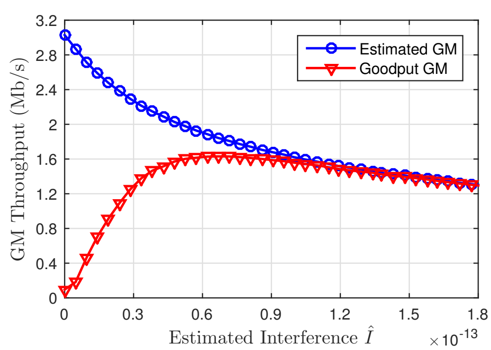


Fig. 2: Comparison of goodput and estimated GMs as a function of $\hat{I}$ for the uplink scheduler

We can see that the throughput computed with $\mathbb{P}_{\mathrm{UL}}(\omega, \hat{\mathbb{I}})$ is significantly lower than the real goodput for low values of $\hat{I}$. However, the estimated throughput and the goodput overlaps after some point. In the following, we will select the lowest value of $\hat{I}$ that yields a difference of less than 0.5% between the two curves.

VI. Type-Aware Scheduling

A. Formulation

Recall that an IC flow (unidirectional by definition) uses simultaneously the uplink and the downlink. If its uplink and downlink scheduling are not coupled as in the type-blind case, it is possible that the flow receives a higher goodput on the uplink than on the downlink and this would create a buffer overflow at the BS. To avoid overflows and wastage, we have to constrain the goodput seen by an IC flow on the uplink to be equal to the goodput seen on the downlink. This couples the scheduling on the uplink and the downlink and makes the computations of the schedules more difficult. We will show next how to do it and then what can be gained in terms of performance by doing it. We formulate the problem for the optimal type-aware scheduler. Our aim is to be proportionally fair in the utilities (defined in Eq. (2)) of the users. Since we focus on device fairness, each flow of each user is not treated as a separate entity, but we consider a user as a single entity in our scheduling problem. To avoid double-counting, we do not include IC$_d$ flows in the computation of the utility of a UE and hence we do not explicitly take them into account in the problem but each IC$_o$ flow for i is an IC$_d$ flow for another UE. We consider only the first 3 rows of the matrix X.

Specifically, for a user i with an IC$_o$ flow, i.e., an intra-cellular flow originated in i, the IC$_o$ throughput λ_i^{IC} is defined as the minimum of two different equations, one corresponding to the throughput on the uplink hop and the other to the throughput on the downlink hop. Both throughputs must be equal to each other to avoid wastage or overflow. In a sense, we do rate matching to avoid possible overflow at the downlink buffers, which we do not model. The throughput of the uplink (resp. downlink) flow of UE i is denoted as λ_i^{UL} (resp. λ_i^{DL}). If there is no flow of this type, the throughput is zero.

We use the uplink and downlink schedulers defined in the previous section. However, we need to extend the notation since each user might have different types of flows. Previously, we used α_i for the fraction of time user i is served on the downlink. Now, we define α_i^{DL} and α_i^{IC} for the fraction of time user i is served on the downlink for its D/L flow and IC$_o$ flow, respectively. Similarly, we extend the notation of θ_i^k to $\theta_i^{k,UL}$ and $\theta_i^{k,IC}$ as the fraction of time user i uses subchannel block k on the uplink for its U/L and IC$_o$ flows, respectively.

For a realization ω, we formulate $\mathbb{P}_{\mathrm{OPT}}(\omega, \hat{\mathbb{I}})$ (see the box in the next page), given $\{(F(i)),\ X,\ (y(i)),\ \beta,\ \hat{I},\ (R_i^k(\hat{I})),\ (r_i)\}$. The variables are $\{(\lambda_i^{IC}),\ (\lambda_i^{DL}),\ (\lambda_i^{UL}),\ (\alpha_i^{DL}),\ (\alpha_i^{IC}),\ (\theta_i^{k,IC}),\ (\theta_i^{k,UL}),\ (\varphi_i), \text{ and } (t_k)\}$.

The throughput of the U/L flow of user i is defined by constraint (14) and of the D/L flow by constraint (16). The throughput of the IC$_o$ flow of user i is defined by (15) and (17). Constraints (22), (23), (24), and (25) ensure that a device does not get resources if it does not have a flow that uses that type of resources. Constraints (18) and (21) ensure that the time users are scheduled on a subchannel cannot exceed the total uplink and downlink subframe time, respectively. Since the objective function is the GM of user utilities and each utility is the GM of flow throughputs, we guarantee that none of the flows gets a zero rate.

$\mathbb{P}_{\mathrm{OPT}}(\omega, \hat{\mathbb{I}})$ is an NP-hard problem since it is a mixed integer program. For reasonable size problems, it can be solved by commercial solvers such as Bonmin [17]. Its computational complexity is high but this is not a problem for our offline study, which is focused on showing how much we can gain by jointly scheduling the uplink and downlink when the types of the flows are known.

B. Numerical Results

In this section, we compare the performance of the type-aware scheduler with the type-blind scheduler. We consider the same cellular system with the same parameters as described in Section V. We set P_M to 46 dBm. The destination node $y(i)$ of the IC$_o$ flow of each user i is selected randomly (and hence a device can receive multiple IC flows). We select $\hat{I}$ so that losses are negligible as discussed previously. Our performance metric is the GM of the φ_i of the users. We consider two scenarios:

- **Scenario 1:** 10 users in cell 0 with all three types of flows, i.e., $X(1,i) = X(2,i) = X(3,i) = 1$,
- **Scenario 2:** D users with an IC$_o$ flow, i.e. $X(3,i) = 1$, and $\{10 - D\}$ users with U/L and D/L flows and no IC$_o$ flow, i.e., $X(1,i) = X(2,i) = 1$.

$$\mathbb{P}_{\text{OPT}}(\omega, \hat{\mathbb{I}})) : \quad \max \sum_{i \in U_0} \log(\varphi_i) \tag{12}$$

$$s.t. \quad \varphi_i = \sqrt[|F(i)|]{\prod_{j \in F(i)} \lambda_i^j} \ , \ \forall i \in U_0, \tag{13}$$

$$\lambda_i^{UL} = \sum_{k \in \{1...M\}} \theta_i^{k,UL} R_i^k(\hat{I}) \ , \ \forall i \in U_0 \tag{14}$$

$$\lambda_i^{IC} = \sum_{k \in \{1...M\}} \theta_i^{k,IC} R_i^k(\hat{I}) \ , \ \forall i \in U_0 \tag{15}$$

$$\lambda_i^{DL} = \alpha_i^{DL} r_i \ , \ \forall i \in U_0 \tag{16}$$

$$\lambda_i^{IC} = \alpha_i^{IC} r_{y(i)} \ , \ \forall i \in U_0 \tag{17}$$

$$\sum_{i \in U_0} (\theta_i^{k,UL} + \theta_i^{k,IC}) \le t_k(1-\beta)T \ , \ \forall k \in \{1...M\} \tag{18}$$

$$\sum_{k \in \{1...M\}} (\theta_i^{k,UL} + \theta_i^{k,IC}) \le (1-\beta)T \ , \ \forall i \in U_0 \tag{19}$$

$$\sum_{k=1}^{M} k t_k \le M \tag{20}$$

$$\sum_{i \in U_0} (\alpha_i^{DL} + \alpha_i^{IC}) \le \beta T \tag{21}$$

$$\theta_i^{k,UL} \le X(1,i) \ , \ \forall i \in U_0 \ , \ \forall k \in \{1...M\} \tag{22}$$

$$\theta_i^{k,IC} \le X(3,i) \ , \ \forall i \in U_0 \ , \ \forall k \in \{1...M\} \tag{23}$$

$$\alpha_i^{DL} \le X(2,i) \ , \ \forall i \in U_0 \tag{24}$$

$$\alpha_i^{IC} \le X(3,i) \ , \ \forall i \in U_0 \tag{25}$$

$$\theta_i^{k,UL} \ge 0, \ \theta_i^{k,IC} \ge 0 \ , \ \forall i \in U_0, \forall k \in \{1...M\} \tag{26}$$

$$\alpha_i^{DL} \ge 0, \ \alpha_i^{IC} \ge 0 \ , \ \forall i \in U_0, \tag{27}$$

$$t_k \in \mathbb{Z}^+ \ , \ \forall k \in \{1...M\} \tag{28}$$

In the type-blind case, the flows within a UE using the uplink (resp. the downlink) are aggregated and offered a goodput which is independent on the composition of the aggregation.

We start with the first scenario where all users have all types of traffic. It is seen in Figure 3a that there is a significant performance difference between the two schedulers. Note that both schedulers achieve their peak performance when β is 0.5, which means equal time for the uplink and downlink. In that case, there is a 14% gain for the type-aware scheduler.

Next, we consider the second scenario. It is important to note that the two schedulers perform exactly the same if there is no IC traffic in the network. Figure 3b shows the performance of the two schedulers when D is 2 and 10. The difference in GM utility is 9% for $D = 2$, whereas it reaches 28% for $D = 10$. To examine this further, we plot Figure 3c which shows the performance of the two schedulers as D increases when $\beta = 0.5$. The gain obtained with the type-aware scheduler increases with the amount of IC traffic.

By avoiding wastage and overflows, the type-aware scheduler can do much better that the type-blind one. Hence, the researchers who study the performance gain of D2D communications should use the type-aware scheduler as their benchmark since this is what can be achieved with a well designed scheduler when direct communications is not enabled. There are two main reasons for this difference. First, the type-blind scheduler does not limit the throughput of the uplink hop of an IC flow if the downlink hop has worse channel characteristics. This avoids some of the data to be transmitted on the second hop. Furthermore, since the type-blind scheduler does not know the flow types, it cannot share the resources in an efficient way between the flows.

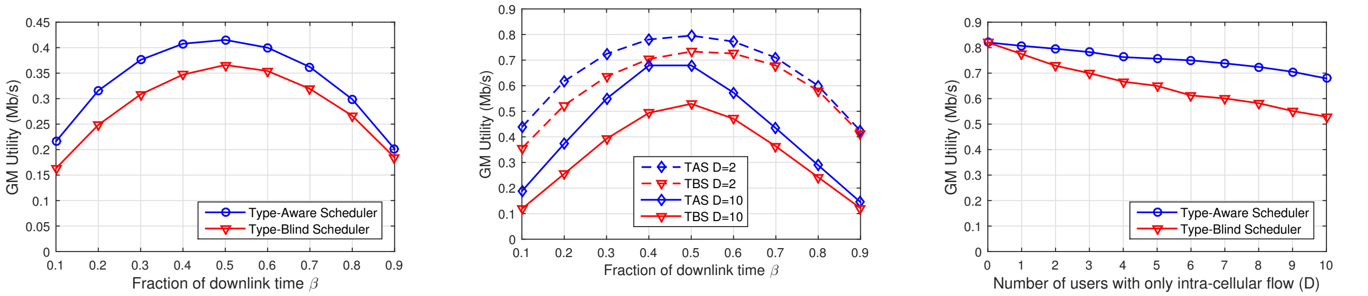
VII. Joint User Association and User Scheduling in Heterogeneous Networks

We now examine the effects of type knowledge on the UA process of an HetNet. We consider the cellular network shown in Figure 1 with two small cells (SCs) added to each cell at a distance of 230 meters left and right of the MBS. We continue to focus on cell 0 but now the users in U_0 have the choice to associate with the MBS or one of the two SCs. For this case, an IC flow is defined between two users in U_0, irrespective of their association.

We consider an orthogonal deployment, where c subchannels are allocated to the small cells and $M - c$ to the MBS. For downlink scheduling, we assume each BS (MBS and SC) allocates equal power to its subchannels and serve one user at a time. For the uplink, we adapt $\mathbb{P}_{\text{UL}}(\omega, \hat{\mathbb{I}})$ to the HetNet case. For the ICI estimation, we consider two estimates, one for the MBS and one for the SCs. These two estimates are independent of each other due to the orthogonal deployment. We performed simulations to obtain a goodput vs. ICI estimates graph similar to Figure 2 and we selected estimates for the MBS and SCs to avoid losses. The results of these simulations are not given here due to lack of space.

UA is a critical process that associates each user to a single BS. Furthermore, the best performance can be obtained only when it is jointly performed with user scheduling [15]. Here, we compare the performance of joint type-aware UA/scheduling and type-blind UA. For the type-blind UA, we use the optimal UA for the downlink, which can be found by solving the integer program described in [15]. Once the UA is given, the user scheduling can be performed independently at each BS for the type-blind scheme as explained in Section VI.

For the type-aware case, we perform the UA and scheduling jointly while coupling the uplink and downlink. We assume that the BSs of a macro-cell are coordinated using a CRAN



(a) Scenario 1: GM utility vs β

(b) Scenario 2: GM utility vs β for D=2 and D=10

(c) Scenario 2: GM utility vs D with β=0.5

Fig. 3: GM comparison for the type-aware and type-blind schedulers for different scenarios in the homogeneous case

[3]. The problem formulation is not given in this paper for brevity but essentially, we need to introduce binary variables to indicate to which BS a user is associated, and then solve $\mathbb{P}_{\mathbb{OPT}}(\omega, \hat{\mathbb{I}})$ with the additional UA and HetNet constraints.

We consider Scenario 2, which was explained in the previous section with $\beta = 0.5$. We use the system parameters described in [15] and set P_S to 30 dBm. The performance difference of the type-aware UA and type-blind UA is given in Figure 4 as a function of c, the number of subchannels allocated to the small cells.

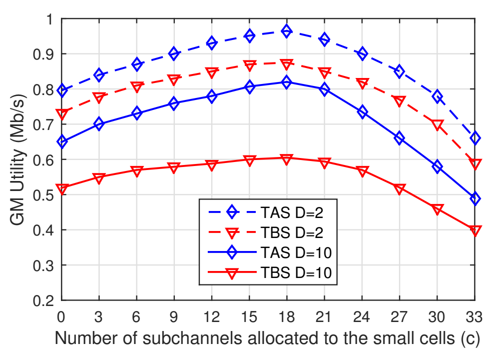


Fig. 4: GM comparison for the type-aware scheme (TAS) and type-blind scheme (TBS) as a function of c in the HetNet scenario

We consider two cases where D, the number of users with IC flows, is 2 and 10. It is obvious that when the UA and scheduling is performed with the knowledge of the type of traffic, the performance is much better. Furthermore, the difference increases as D increases. The maximum performance is achieved for both schemes when 18 subchannels are allocated to the SCs and in that case, the GM difference is 11% and 36% for $D = 2$ and $D = 10$, respectively.

VIII. Conclusion

We analyze cellular networks with intra-cellular (IC) traffic and show that the operation of the network can be improved significantly if the traffic types are known. Essentially, user scheduling should be performed jointly on the uplink and downlink for IC traffic to avoid resource losses caused by bottlenecks. We also claim that a fair benchmark to use to evaluate the performance gains of D2D direct communications should be type-aware since any solution involving D2D direct communications would require the knowledge of the type of the traffic. We also show that user association in an Hetnet should also be type-aware.

References

[1] G. Fodor *et al.*, "Design aspects of network assisted device-to-device communications," *IEEE Comm. Mag.*, vol. 50, no. 3, pp. 170–177, 2012.

[2] F. Capozzi *et al.*, "Downlink Packet Scheduling in LTE Cellular Networks: Key Design Issues and a Survey," *IEEE Communications Surveys Tutorials*, vol. 15, no. 2, pp. 678–700, Second 2013.

[3] A. Checko *et al.*, "Cloud RAN for Mobile Networks - A Technology Overview," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 405–426, Firstquarter 2015.

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

[5] A. Kumar and C. Rosenberg, "Energy and Throughput Trade-Offs in Cellular Networks Using Base Station Switching," *IEEE Transactions on Mobile Computing*, vol. 15, no. 2, Feb 2016.

[6] Q. Ye *et al.*, "Distributed resource allocation in device-to-device enhanced cellular networks," *IEEE Trans. on Communications*, vol. 63, no. 2, pp. 441–454, Feb 2015.

[7] C. H. Yu *et al.*, "Resource sharing optimization for device-to-device communication underlaying cellular networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2752–2763, August 2011.

[8] 3GPP TS 36.211 v10.2.0, 3rd Generation Partnership Project, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (EUTRA); Physical Channels and Modulation (Release 10)."

[9] Z. Shen *et al.*, "Dynamic uplink-downlink configuration and interference management in TD-LTE," *IEEE Comm. Mag.*, vol. 50, pp. 51–59, 2012.

[10] S. Kim and J. W. Lee, "Joint resource allocation for uplink and downlink in wireless networks: A case study with user-level utility functions," in *IEEE VTC Spring 2009*, April 2009, pp. 1–5.

[11] P. C. Nguyen and B. D. Rao, "Fair Scheduling Policies Exploiting Multiuser Diversity in Cellular Systems With Device-to-Device Communications," *IEEE Trans. on Wireless Communications*, vol. 14, no. 9, pp. 4757–4771, Sept 2015.

[12] D. C. Dimitrova *et al.*, "Scheduler-dependent Inter-cell Interference and Its Impact on LTE Uplink Performance at Flow Level," ser. WWIC'11. Berlin, Heidelberg: Springer-Verlag, 2011.

[13] A. Racz, N. Reider, and G. Fodor, "On the Impact of Inter-Cell Interference in LTE," in *IEEE GLOBECOM*, 2008, pp. 1–6.

[14] C. Shi, M. Honig, S. Nagaraj, and P. Fleming, "Uplink distributed power and receiver optimization across multiple cells," in *IEEE WCNC*, 2012.

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

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

[17] "Bonmin (basic open-source mixed integer programming)," http://projects.coin-or.org/Bonmin.