Performance Evaluation of Scheduling Schemes for D2D Communications

Antonia Maria Masucci, Philippe Olivier, Alain Simonian
Orange Labs, 92320 Châtillon, France
Email: {antoniamaria.masucci, phil.olivier, alain.simonian}@orange.com

***Abstract*—We address the performance of a system where D2D communications add up to standard cellular traffic in future 5G networks. D2D resources are taken from the Uplink spectrum and, in this framework, two scheduling schemes are envisaged: the first one consists in two independent schedulers working on separate subsets of the Uplink channel, respectively dedicated to Uplink cellular communications and to D2D communications; the second one is a joint scheduler making use of the whole Uplink channel. In both schemes, all communications are scheduled, which in particular avoids the intra-cell interference issue. We propose multi-class Processor Sharing queueing models which account for realistic cellular radio conditions and provide traffic performance in terms of the average throughput for the different classes of flows. From the considered numerical scenarios, we conclude that the joint Uplink scheduling scheme provides better performance and allows D2D users to make an efficient use of their higher potential capacity, due to the device proximity. The corresponding throughput is also shown to be significantly higher than that perceived without any dedicated D2D link.**

I. Introduction

Future 5G networks envisage the extension of network functions to the network edge and the integration of devices and private infrastructures into the general network architecture is now a main topic of investigation. In this framework, the promising *Device-to-Device* (D2D) scheme can fulfill this network extension by enabling devices to communicate directly without using the standard cellular network [3], [6]. Allowing direct D2D communications brings some potential advantages, among which are energy saving, increase of spectral efficiency, and improvement of the provided Quality of Service (QoS) in terms of throughput and delay [3].

In the present paper, we address the QoS issue within a use case where an amount of exogenous cellular traffic coexists with a specific D2D communication demand, e.g., related to some proximity-based service, public safety service, etc. Users request explicit demands with a certain probability to be eligible to a D2D communication (i.e., the probability to find the peer device within the radio acceptability area). In the case when D2D communication is not allowed, the demand is alternatively transferred on the cellular system, thus leading to a supplementary amount of traffic, both in the Uplink and in the Downlink.

We develop a performance model where D2D communications are possible only between devices in the same cell; the interference with surrounding cells, however, is explicitly taken into account. The Base Station (BS) is assumed to have a perfect knowledge of device positions, radio conditions and device eligibility to establish a D2D communication. Direct D2D communications are supposed to make use of the cellular Uplink spectrum resources, corresponding to "Inband" communications as standardized in 3GPP LTE-Advanced, Releases 12 and 13 [9] (as opposed to "Outband" where D2D is dedicated a different spectrum from that of a BS).

Besides, 3GPP has defined two D2D resource allocation modes: the scheduled mode, where radio resources are granted by the BS based on user terminal requests; and the autonomous mode, where D2D terminals have random access to the radio resource pool. In this study, we consider the first resource allocation mode acting on the Uplink, which is not essentially different from the conventional scheduling performed by the BS, including both requests for the cellular system as well as D2D communications. This mode is intended to provide a higher QoS guaranty due to its centrally controlled/optimized resource allocation. In particular, a centralized scheduler allocates a succession of orthogonal time-frequency Resource Blocks (RBs) to each on-going communication. This ensures fair sharing between on-going flows within each sub-system, either the cellular Uplink or the direct D2D link, as well as the absence of any intra-cell interference.

Within this mode, we envisage two scheduling schemes on the Uplink:

- Hard Bandwidth Partitioning **(HBP)** implements two distinct Uplink schedulers: this is a specific type of an Overlay system, according to standard terminology [3], [6], as RBs of the Uplink spectrum are segmented into two fixed distinct pools, with a proportion α of radio resources allocated to the User Equipment (UE)-BS communications and the remaining proportion $1-\alpha$ shared among D2D communications;
- Joint Bandwidth Usage **(JBU)** is an Underlay system with a single Uplink scheduler: all Uplink RBs are available to any on-going flow, transferred either on the Uplink sub-system or on the direct D2D link.

978-0-9883045-5-0/18/$31.00 ©2018 ITC
DOI 10.1109/ITC30.2018.00030

The literature on D2D systems is vast, see [3], [6] and references therein. We here focus on closely related works dealing with scheduling or performance analysis. We first mention the proposal of a control policy for inter-device traffic within a classical LTE radio cell (without LTE-A standards) by using both Uplink and Downlink channels, without any direct D2D link [12]; this scheme is intended to be used as a benchmark for further evaluations of actual D2D systems. Regarding the latter, the flow-level performance of LTE-A D2D communications in the Downlink has been analyzed in [5], [8], [10] in the context of cellular traffic offloading combined with content caching. A similar use case as ours, i.e., a specific D2D demand added to standard cellular traffic in the Uplink, has been addressed in [11]; the approach, however, differs from ours in that no scheduling policy is specified for D2D communications, and that D2D traffic is lost, i.e., not transferred to the cellular network, in case the level of intra-cell interference is too high.

The rest of the paper is organized as follows. The generic flow-level traffic and performance model in cellular networks is recalled in Section II. Then the traffic model derived from our use case is provided in Section III. The performance analysis is carried out in Sections IV and V for HBP and JBU schemes, respectively. The cellular radio framework as well as numerical results are presented in Section VI. Finally, Section VII concludes by summarizing our main achievements.

II. Baseline cellular model

Flow-level performance of cellular systems is conveniently handled by means of a multi-class Processor Sharing (PS) queuing model [4] whose basic elements are now recalled. Consider an isolated cell where users are assumed static (we do not account for user mobility in this study). Communication flows, i.e., data transfers, are generated within the whole cell according to a Poisson process with mean arrival rate λ.

The achievable transmission bit rate is known to be highly variable in the cell, reflecting the varying radio conditions encountered by the user. To represent this spatial diversity, the population is segmented into K classes corresponding to geographic areas: in area k, $1 \leqslant k \leqslant K$, flows appear according to a Poisson process with rate λ_k (with $\sum_k \lambda_k = \lambda$) and have i.i.d. data volumes Σ_k to transfer with mean σ_k. The local capacity C_k is defined as the maximum bit rate attainable by a user located in area k when he/she is the unique active user in the cell. The traffic intensity generated in area k is thus $a_k = \lambda_k \sigma_k$ and the corresponding local load is $\varrho_k = \lambda_k \sigma_k / C_k$. The total cell load is then expressed as $\varrho = \sum_k \varrho_k = \sum_k a_k / C_k = a/C$ where $a = \sum_k a_k$ is the total traffic intensity and the equivalent cell capacity C is given by the harmonic mean

$$\frac{1}{C} = \left(\sum_{k=1}^{K} \lambda_k \sigma_k \right)^{-1} \sum_{k=1}^{K} \frac{\lambda_k \sigma_k}{C_k}. \tag{1}$$

Note that, in general, C is not an intrinsic capacity of the system, as it depends on the traffic distribution among classes through the per-class mean arrival rates and data volumes.

The system occupancy can then be described by the K-dimensional vector process $\mathbf{N}(t) = (N_1(t), ..., N_K(t))$, $t \geqslant 0$, where $N_k(t)$ denotes the number of ongoing class-k flows at time t. Following the assumptions made in the Introduction, the resource blocks are equally shared among all active flows on a time slot basis; such a fair sharing policy may be implemented by a Round-Robin (RR) scheduler. Any active flow in the system thus receives a portion $1/N(t)$ of the resource blocks, where $N(t)$ is the total number of concurrent flows at time t, which translates into an instantaneous transmission bit rate of $C_k/N(t)$ for each class-k flow. The process thus evolves as a multi-class PS queue with Poisson arrivals [4], which is known to be stable under the condition $\varrho < 1$.

We consider this system in stationary regime. Regarding data (elastic) traffic, a convenient performance indicator is the per-class average throughput Γ_k defined as the ratio of the mean volume of transferred data to the mean transfer time. By means of Little's formula, this ratio Γ_k eventually reads

$$\Gamma_k = \frac{a_k}{\mathbb{E}(N_k)}. \tag{2}$$

The mean number of class-k flows in the multi-class PS queue is $\mathbb{E}(N_k) = \varrho_k/(1-\varrho)$, thus the performance metrics of this system simply read

$$\Gamma_k = C_k(1-\varrho), \quad \forall\, k \in \{1, ..., K\}, \tag{3}$$

$$\Gamma = C(1-\varrho) \tag{4}$$

for the per-class and overall throughput, respectively.

III. Joint cellular and D2D traffic model

We now apply the generic cellular model of Section II to the case when D2D communication flows add up to the standard cellular traffic. For the cellular Uplink, cellular Downlink and D2D traffic, we respectively consider three types of flows with the following characteristics: flows arrive according to Poisson processes with rates $\lambda_u, \lambda_d, \lambda_D$, and data volumes with means $\sigma_u, \sigma_d, \sigma_D$. The corresponding traffic intensities are thus $a_u = \lambda_u \sigma_u$, $a_d = \lambda_d \sigma_d$ and $a_D = \lambda_D \sigma_D$ within the whole cell.

A. Uplink sub-system

In the Uplink, the cell is partitioned into I areas with associated capacities $(C_{u,i})_{1 \leqslant i \leqslant I}$, while users are randomly distributed in these areas according to probabilities $(p_i)_{1 \leqslant i \leqslant I}$ with $\sum_i p_i = 1$ (each p_i is proportional to the surface of area i if the spatial distribution of users is homogeneous over the cell). In each area i, the request arrival rate of the Uplink traffic is $\lambda_i = p_i \lambda_u$; then the nominal equivalent Uplink capacity C_u^0 (with no supplementary traffic) is given by

$$\frac{1}{C_u^0} = \sum_{i=1}^{I} \frac{p_i}{C_{u,i}} \tag{5}$$

by application of the generic formula (1).

As described in the Introduction, D2D flows are eligible with some probability P for transmission on the direct D2D link; probability P generally depends on the location of the user within the cell (see Section VI). We thus denote by P_i the eligibility probability for a D2D receiver located in area i (by reason of spatial symmetry, the D2D transmitter has the same eligibility). The average eligibility probability is thus

$$P = \sum_{i=1}^{I} p_i P_i. \tag{6}$$

In area i, D2D flows are alternatively handled by the cellular system with probability $1 - P_i$, hence the aggregated traffic intensity in this area is $p_i(a_u + (1-P_i)a_D)$. The overall traffic intensity $a'_u = a_u + (1-P)a_D$ is then offered to the Uplink sub-system, and the associated load reads

$$\varrho_u = \frac{a'_u}{C_u} = \frac{a_u + (1-P)a_D}{C_u},$$

with an equivalent capacity given by

$$\frac{1}{C_u} = \frac{1}{a_u + (1-P)a_D} \sum_{i=1}^{I} \frac{p_i}{C_{u,i}} (a_u + (1-P_i)a_D)$$

by application of (1). Note that capacity C_u explicitly depends on the traffic demands a_u and a_D; it reduces to C_u^0 given by (5), however, in case the eligibility probability does not depend on the user location. Setting $P_u = C_u^0 \sum_{i=1}^{I} P_i p_i / C_{u,i}$, a quantity within $[0,1]$, capacity C_u can be expressed as

$$C_u = C_u^0 \, \frac{a_u + (1-P)a_D}{a_u + (1-P_u)a_D}. \tag{7}$$

B. Downlink sub-system

Downlink analysis is not the main focus of our study, since the envisaged scheduling schemes apply to the Uplink; we need it, however, to evaluate the End-to-End performance of D2D flows (see (24) and results provided in Subsection VI-E).

Without loss of generality, the cell is partitioned into the same $J = I$ areas (the distinct notation I and J is kept in the following, however, to avoid confusion between the two link directions) with associated capacities $(C_{d,j})_{1 \leqslant j \leqslant J}$, while users in the cell are randomly distributed in these zones according to probabilities $(q_j)_{1 \leqslant j \leqslant J}$. In a way similar to the Uplink, the Downlink load is $\varrho_d = (a_d + (1-P)a_D)/C_d$, with the nominal and equivalent capacities, C_d^0 and C_d, given by

$$\frac{1}{C_d^0} = \sum_{j=1}^{J} \frac{q_j}{C_{d,j}}, \tag{8}$$

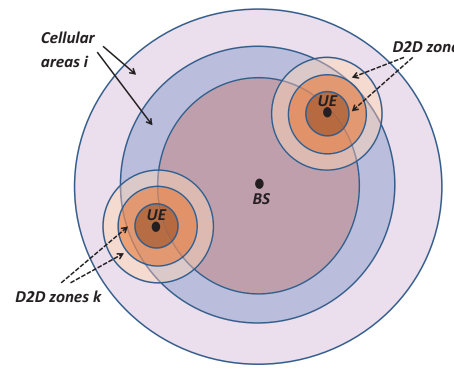

Fig. 1. D2D capacity zones $k \in \{1, ..., K\}$ within areas $i \in \{1, ..., I\}$.

$$C_d = C_d^0 \, \frac{a_d + (1-P)a_D}{a_d + (1-P_d)a_D}, \tag{9}$$

where we set $P_d = C_d^0 \sum_{j=1}^{J} P_j q_j / C_{d,j} \in [0,1]$.

C. D2D sub-system

Now consider a given D2D receiver in area i within the cell, with a D2D communication eligible to actual transfer through the direct D2D link. The spatially varying capacity around this device is represented by K zones [10], depending on i, within the region of eligibility. This leads us to define $I \times K$ zones (or classes) for D2D traffic; the probability for any D2D transmitter to pertain to given zone k and area i is then denoted by $r_{i,k}$ with $\sum_{1 \leqslant k \leqslant K} r_{i,k} = 1$. Note that by doing so, a zone k corresponds to several geographical zones centered on any other potential D2D receiver, see Fig. 1.

From the above definitions, the intensity of traffic actually generated in the D2D link from users in any given class (i,k) is thus $a_{i,k} = P_i p_i r_{i,k} \lambda_D \sigma_D = P_i p_i r_{i,k} a_D$. The capacity corresponding to the D2D traffic class (i,k) is denoted by $C_{i,k}$. On the D2D link, the total offered traffic intensity is $a'_D = \sum_{1 \leqslant i \leqslant I, 1 \leqslant k \leqslant K} a_{i,k} = P a_D$ and the generated load therefore equals $\varrho_D = a'_D / C_D = P a_D / C_D$ with an equivalent capacity given by

$$\frac{1}{C_D} = \frac{1}{P} \sum_{1 \leqslant i \leqslant I, 1 \leqslant k \leqslant K} \frac{P_i p_i r_{i,k}}{C_{i,k}}. \tag{10}$$

IV. Performance of HBP scheme

The HBP scheme partitions the pool of radio resources in the Uplink. Denote by α the proportion of Uplink RBs allocated to the UE-BS communications; the remaining proportion $1 - \alpha$ is shared by the D2D communications. This principle entails that there is no interference between those two parts of the Uplink spectrum, and that all capacities (maximum physical bit rates) $(C_{u,i})_{1 \leqslant i \leqslant I}$ related to the Uplink, and also C_u^0, may be simply multiplied by α to evaluate the performance on the Uplink. Likewise, all capacities $(C_{i,k})_{1 \leqslant i \leqslant I, 1 \leqslant k \leqslant K}$ related to the D2D link, and also C_D, have to be multiplied by $1-\alpha$ to evaluate the performance on the D2D sub-system.

Given this partitioning pattern, the D2D and cellular Uplink sub-systems evolve independently. The same does not hold for the Downlink and Uplink sub-systems: in fact, the presence of the same additional D2D traffic on these two links makes them interdependent. The analysis of the system accounting for this dependence is out of scope of the present work.

We develop instead a simpler model based on a "Store-and-Forward" worst case: the transfer of flows is supposed to be first completed in the Uplink direction and then resumed in the Downlink. By this assumption, we are allowed to handle the whole system as a set of three decoupled PS queues.

A. *Performance analysis*

1) Performance of D2D sub-system: This system is a PS queue with $I \times K$ classes of D2D flows. Applying general formulas (3) and (4), we obtain $\Gamma_{i,k}^{(\alpha)} = (1-\alpha)C_{i,k}(1-\varrho_D^{(\alpha)})$ for the throughput of any class (i,k), under the stability condition $\varrho_D^{(\alpha)} < 1$, and

$$\Gamma_D^{(\alpha)} = (1-\alpha)C_D(1-\varrho_D^{(\alpha)}) = (1-\alpha)C_D - Pa_D \quad (11)$$

for the overall throughput, where the D2D capacity C_D is given by (10) and the effective, α-dependent, load is

$$\varrho_D^{(\alpha)} = \frac{\varrho_D}{1-\alpha} = \frac{P\,a_D}{(1-\alpha)\,C_D}. \quad (12)$$

2) Performance of Uplink sub-system: Consider a PS queue with $2I$ classes with I classes for Uplink traffic and I classes for D2D traffic. For any $i \in \{1,...,I\}$, class (i,u) refers to Uplink traffic flows while class (i,D) refers to D2D flows, both of them being generated in area i. These two classes have traffic intensities $p_i a_u$ and $p_i(1-P_i)a_D$, respectively, and share the same capacity $\alpha C_{u,i}$. We then obtain, under the stability condition $\varrho_u^{(\alpha)} < 1$,

$$\Gamma_{i,u}^{(\alpha)} = \Gamma_{i,D}^{(\alpha)} = \alpha C_{u,i}(1-\varrho_u^{(\alpha)}) \quad (13)$$

for the throughputs of classes (i,u) and (i,D), and

$$\Gamma_u^{(\alpha)} = \alpha C_u(1-\varrho_u^{(\alpha)}) = \alpha C_u - a_u - (1-P)a_D \quad (14)$$

for the overall throughput on the Uplink, where the effective, α-dependent, load $\varrho_u^{(\alpha)}$ on the Uplink is

$$\varrho_u^{(\alpha)} = \frac{\varrho_u}{\alpha} = \frac{a_u + (1-P_u)a_D}{\alpha\,C_u^0}. \quad (15)$$

The throughput given by (14) reflects the performance perceived by any active flow on the Uplink sub-system and may be called a "system" performance measure. We introduce the notion of "user-perceived" performance by measuring the throughput perceived by the different traffic types on the Uplink; this is first given by (13) as regards the individual traffic classes. We further define the overall throughputs $\Gamma_{u,u}^{(\alpha)}$ and $\Gamma_{u,D}^{(\alpha)}$ perceived by the Uplink and D2D flows, respectively. Both are evaluated by a successive use of identity (2) for the average per-class (or per aggregate of classes) throughput:

$$\begin{aligned}\Gamma_{u,u}^{(\alpha)} &= \frac{a_u}{\mathbb{E}(N_{u,u})} = \frac{a_u}{\sum_i \mathbb{E}(N_{i,u})} = \frac{a_u}{\sum_i p_i a_u / \Gamma_{i,u}^{(\alpha)}} \\ &= \alpha C_u^0(1-\varrho_u^{(\alpha)}) = \alpha C_u^0 - a_u - (1-P_u)a_D \quad (16)\end{aligned}$$

where we have used (13) and introduced some self-explanatory notation regarding the number of concurrent flows. Similarly, the throughput of D2D flows on the Uplink is derived as

$$\Gamma_{u,D}^{(\alpha)} = \frac{1-P}{1-P_u}\left(\alpha C_u^0 - a_u - (1-P_u)a_D\right). \quad (17)$$

Note that the three overall throughputs given by (14), (16) and (17) are simply related by

$$\Gamma_u^{(\alpha)} = \frac{C_u}{C_u^0}\,\Gamma_{u,u}^{(\alpha)} = \frac{1-P_u}{1-P}\,\frac{C_u}{C_u^0}\,\Gamma_{u,D}^{(\alpha)}$$

after using (7), and are thus identical when all the P_i are equal.

3) Performance on the Uplink spectrum: As another performance metric of interest, introduce the average throughput $\Gamma_{u,S}^{(\alpha)}$ perceived by all flows using the Uplink spectrum. This overall throughput will be used, in particular, to analyze the impact of the bandwidth partitioning parameter α. Following the same methodology as above, it is evaluated by

$$\Gamma_{u,S}^{(\alpha)} = \frac{a'_u + a'_D}{\mathbb{E}(N_{u,u}) + \mathbb{E}(N_{u,D}) + \mathbb{E}(N_D)},$$

$$\text{so that}\quad \frac{1}{\Gamma_{u,S}^{(\alpha)}} = \frac{1}{a_u + a_D}\left(\frac{\varrho_u}{\alpha-\varrho_u} + \frac{\varrho_D}{1-\alpha-\varrho_D}\right) \quad (18)$$

where we successively used (2), (11), (14), (12), (15) and (7).

We can now look for an optimal choice for the partitioning parameter α by maximizing the overall Uplink throughput $\Gamma_{u,S}^{(\alpha)}$. Differentiating the r.h.s of (18) shows that $\Gamma_{u,S}^{(\alpha)}$ has a unique maximum at $\alpha = \alpha^\# \in]\varrho_u, 1-\varrho_D[$, where

$$\alpha^\# = \frac{\sqrt{\varrho_u}}{\sqrt{\varrho_u}+\sqrt{\varrho_D}}\left(1-\varrho_D+\sqrt{\varrho_u\varrho_D}\right). \quad (19)$$

Observe that this optimum depends on both traffic loads, ϱ_u and ϱ_D, so that in practical implementations, the long-term averages of traffic intensities should be dynamically monitored in order to properly estimate the optimal partitioning parameter.

The overall and "per-type of traffic" optimal throughputs are obtained by setting $\alpha = \alpha^\#$ in (18), (16), (17) and (11). Denoting the total load by $\varrho = \varrho_u + \varrho_D$ and the overall capacity of the Uplink spectrum by $C_{u,S} = (a_u + a_D)/\varrho$, we get

$$\begin{cases} \Gamma_{u,S}^\# = C_{u,S}(1-\varrho)\,\dfrac{\varrho}{\left(\sqrt{\varrho_u}+\sqrt{\varrho_D}\right)^2}, \\ \Gamma_{u,u}^\# = C_u^0(1-\varrho)\,\dfrac{\sqrt{\varrho_u}}{\sqrt{\varrho_u}+\sqrt{\varrho_D}}, \\ \Gamma_{u,D}^\# = C_u^0(1-\varrho)\,\dfrac{1-P}{1-P_u}\,\dfrac{\sqrt{\varrho_u}}{\sqrt{\varrho_u}+\sqrt{\varrho_D}}, \\ \Gamma_D^\# = C_D(1-\varrho)\,\dfrac{\sqrt{\varrho_D}}{\sqrt{\varrho_u}+\sqrt{\varrho_D}}. \end{cases} \quad (20)$$

4) Performance of Downlink sub-system: On the Downlink, the capacity C_d is given by (9). As for the Uplink, we derive $\Gamma_{j,d} = \Gamma_{j,D} = C_{d,j}(1-\varrho_d)$ for the throughputs of classes (j,d) and (j,D), and

$$\Gamma_d = C_d(1-\varrho_d) = C_d - a_d - (1-P)a_D \tag{21}$$

for the overall throughput on the Downlink, under the stability condition $\varrho_d < 1$. The per-traffic type throughputs are then

$$\Gamma_{d,d} = C_d^0 - a_d - (1-P_d)a_D, \tag{22}$$

$$\Gamma_{d,D} = \frac{1-P}{1-P_d}\,(C_d^0 - a_d - (1-P_d)a_D). \tag{23}$$

The throughputs given by (21), (22) and (23) verify

$$\Gamma_d = \frac{C_d}{C_d^0}\,\Gamma_{d,d} = \frac{1-P_d}{1-P}\,\frac{C_d}{C_d^0}\,\Gamma_{d,D}.$$

5) Performance of D2D on the cellular path: We are now able to evaluate the End-to-End (E2E) throughput $\Gamma_{E,D}^{(\alpha)}$ for the fraction of D2D flows which cross the Uplink and Downlink sub-systems. According to the above "Store-and-Forward" assumption, the average E2E delay $1/\Gamma_{E,D}^{(\alpha)}$ is the sum $1/\Gamma_{u,D}^{(\alpha)} + 1/\Gamma_{d,D}$ of the average delays in each sub-system, hence

$$\frac{1}{\Gamma_{E,D}^{(\alpha)}} = \frac{1}{1-P}\left[\frac{1-P_u}{C_u^0(\alpha-\varrho_u)} + \frac{1-P_d}{C_d^0(1-\varrho_d)}\right] \tag{24}$$

after (17) and (23).

B. Stability condition

Let us specify the stability condition of the overall system under which all previous calculations are valid. The condition for an admissible D2D traffic to be fed in addition to the cellular traffic stems from the stability conditions of the D2D, Uplink and Downlink sub-systems of Subsections IV-A1, IV-A2 and IV-A4 above. This consequently provides the overall stability condition $a_D < C^{(\alpha)}$, where

$$C^{(\alpha)} = \mathrm{Min}\left(\frac{(1-\alpha)C_D}{P}, \frac{\alpha\, C_u^0 - a_u}{1-P_u}, \frac{C_d^0 - a_d}{1-P_d}\right). \tag{25}$$

Another approach to optimize the bandwidth partitioning scheme now aims at maximizing this stability region. Obviously, parameter α is not involved in the last argument (related to Downlink) of the minimum (25) and we can ignore the latter for clarity[1]. In that case, the optimal value is given by

$$\begin{aligned}\alpha^* &\triangleq \underset{\varrho_u<\alpha<1-\varrho_D}{\mathrm{Argmax}}\left[\mathrm{Min}\left(\frac{(1-\alpha)C_D}{P}, \frac{\alpha\, C_u^0 - a_u}{1-P_u}\right)\right] \\ &= \frac{Pa_u + (1-P_u)C_D}{PC_u^0 + (1-P_u)C_D}.\end{aligned} \tag{26}$$

[1]If one explicitly accounts for the Downlink stability, it can be shown that the optimal partitioning factor either equals α^* given by (26) or lies in an interval including α^*, so that we may always choose α^* as the optimal value.

The corresponding optimal capacity C^*, given by

$$C^* = \frac{C_D(C_u^0 - a_u)}{PC_u^0 + (1-P_u)C_D}, \tag{27}$$

must be understood as the supremum of D2D intensity that can be fed into the system while preserving its stability. Here, the optimal values α^* and C^*, by construction, only depend on the Uplink traffic intensity, contrary to $\alpha^{\#}$ as noted previously.

In the particular case when all D2D eligibility probabilities (P_i) are constant and equal to P, as noted in Subsection IV-A2, we have $\Gamma_u^{(\alpha)} = \Gamma_{u,u}^{(\alpha)} = \Gamma_{u,D}^{(\alpha)}$ for any value of α. Moreover, when bandwidth partitioning is optimized as above, the respective throughputs Γ_u^* and Γ_D^* of Uplink and D2D queuing systems become proportional to the fractions $1-P$ and P of D2D traffic actually handled by the two sub-systems. Specifically, replacing α by its optimal value (26) in expressions (11) and (14), the proportionality relation reads

$$\frac{\Gamma_u^*}{1-P} = \frac{\Gamma_D^*}{P} = \frac{C_D(C_u - a_u)}{PC_u + (1-P)C_D} - a_D. \tag{28}$$

This implies that when the eligibility probability P is either very low or very high, the Uplink and D2D respective performance will be highly unbalanced.

V. Performance of JBU scheme

In the JBU scheme, a single joint scheduler allocates successive Uplink RBs to each active flow at a given time, be it a flow of the direct D2D link or a flow of the cellular Uplink sub-system. Due to joint scheduling, the Uplink channel ensures fair sharing among all active flows on a time slot basis; we can thus model the Uplink performance by means of a single multi-class PS queue. This PS queue jointly handles $2I + I\times K$ classes of customers, namely:

(i) I classes of cellular Uplink flows generated in each area i, with local capacity $C_{u,i}$ and location probability p_i;

(ii) I classes of D2D flows transferred to the cellular network, and generated in each area i, with local capacity $C_{u,i}$ and location probability p_i;

(iii) $I\times K$ classes of D2D flows generated in each zone (i,k), with local capacity $C_{i,k}$ and location probability $p_i r_{i,k}$.

The per-class traffic intensities are the same as in Section III, while the local capacities are now not partitioned among cellular and D2D systems, so that the local loads now read

$$\varrho_{i,u} = \frac{p_i a_u}{C_{u,i}}, \quad \varrho_{i,D} = \frac{(1-P_i)p_i a_D}{C_{u,i}}, \quad \varrho_{i,k} = \frac{P_i p_i r_{i,k} a_D}{C_{i,k}}$$

for all $i \in \{1,...,I\}$ and $(i,k) \in \{1,...,I\}\times\{1,...,K\}$. The total load on this Uplink queue is then

$$\begin{aligned}\varrho &= \sum_{i=1}^{I}(\varrho_{i,u} + \varrho_{i,D}) + \sum_{1\leqslant i\leqslant I, 1\leqslant k\leqslant K}\varrho_{i,k} = \varrho_u + \varrho_D \\ &= \frac{a_u + (1-P_u)a_D}{C_u^0} + \frac{Pa_D}{C_D} = \frac{a_u + a_D}{C_{u,S}},\end{aligned}$$

where the equivalent capacity $C_{u,S}$ is given by

$$\frac{1}{C_{u,S}} = \frac{1}{a_u + a_D}\left(\frac{a_u + (1-P_u)a_D}{C_u^0} + \frac{Pa_D}{C_D}\right). \quad (29)$$

Under the stability condition $\varrho < 1$, the throughputs are obtained by applying formulas (3) and (4), which readily give $\Gamma_{i,u} = \Gamma_{i,D} = C_{u,i}(1-\varrho)$ and $\Gamma_{i,k} = C_{i,k}(1-\varrho)$ for each individual traffic class, together with

$$\Gamma_{u,S} = C_{u,S}(1-\varrho) = C_{u,S} - (a_u + a_D) \quad (30)$$

for the overall Uplink system. The "per-traffic type" throughputs can then be evaluated, as above, by

$$\Gamma_{u,u} = \frac{a_u}{\sum_i \mathbb{E}(N_{i,u})} = \frac{a_u}{\sum_i p_i a_u/\Gamma_{i,u}} = C_u^0(1-\varrho) \quad (31)$$

for the Uplink traffic flows,

$$\Gamma_{u,D} = \frac{(1-P)a_D}{\sum_i (1-P_i)p_i a_D/\Gamma_{i,D}} = C_u^0(1-\varrho)\frac{1-P}{1-P_u} \quad (32)$$

for the D2D flows transferred on the cellular Uplink, and

$$\Gamma_D = \frac{Pa_D}{\sum_{i,k} P_i p_i r_{i,k} a_D/\Gamma_{i,k}} = C_D(1-\varrho) \quad (33)$$

for the D2D flows on the D2D link.

We can appraise the performance gain provided by JBU by comparing (30), (31), (32) and (33) to expressions (20) corresponding to the maximized throughputs in the HBP scheme: it is easily observed that the former are always greater than the latter. Note finally that the stability condition $a_u + a_D < C_{u,S}$, translated by means of (29) in terms of the supremum of D2D traffic intensity that can be fed into the system, reads $a_D < C^*$ with C^* as given by (27), that is, the capacity ensuring maximum stability in the HBP scheme.

One can summarize the above properties by stating that JBU allows the system to benefit from the advantages of both schemes: the maximum capacity provided by optimized bandwidth partitioning, together with the throughput gain provided by joint scheduling.

VI. Numerical Results

In this section, we describe the cellular radio network framework and illustrate the performance of the different classes of flows obtained from the previously developed model.

A. Cellular network framework

The purpose here is to derive, from a realistic cellular network scenario, the necessary inputs to the performance model. We simulate a network with hexagonal cells, each cell with the BS at its center and a cell radius of 500 m. User devices, both D2D and cellular, are generated uniformly over each cell. We consider a urban outdoor scenario whose associated simulation parameters are specified in Table I.

The SINR (Signal to Interference plus Noise Ratio) is computed for either cellular or potential D2D communication

TABLE I
Cellular Network: Simulation Parameters

Environment	Urban
Context	Outdoor
Cell Radius	500 m
User Transmit Power	$P_{T,u} = 23$ dB
BS Transmit Power	$P_{T,BS} = 46$ dB
Number of Tx/Rx antennas	MIMO 2×2
Carrier Frequency	2 GHz
Bandwidth	10 MHz

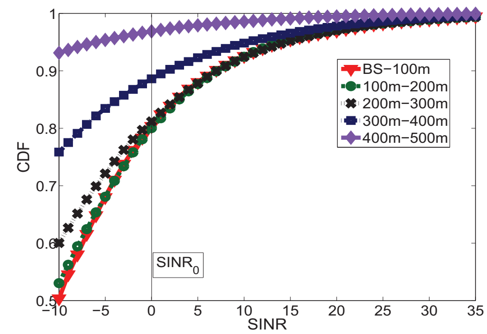

Fig. 2. D2D SINR Cumulative Distribution Function.

as $SINR = P_s/(P_I + P_n)$ where P_s represents the signal power, P_n the noise power and P_I the interference power. In particular, the signal power is $P_s = 10^{(P_T - PL(d))/10}$ where P_T denotes the transmit power and where the path loss $PL(d)$, function of the transmitter-receiver distance d, is given by the Winner model [1], [2] for D2D and by the ITU model [7] for cellular communications. We assume a scenario where the whole frequency band is allocated to the considered cell and reused in all adjacent cells, thus creating inter-cell interference. The interference power is then $P_I = \sum_{s'} P_{s'}$, where s' are the simultaneous signals in the network over the same bandwidth.

Given the latter cellular network framework, the parameters of the performance model can now be specified as follows:

(i) we presently consider five cellular concentric areas and five D2D concentric zones, that is, $I = J = K = 5$. Assuming a hexagonal area for cell and circular areas for D2D, simple geometrical considerations give us the values of probabilities $p_i (= q_j)$ and $r_{i,k}$, $1 \leqslant i, j, k \leqslant 5$;

(ii) concerning D2D communications, we compute the SINR for all UE pairs in the cell and plot in Fig. 2 the Cumulative Distribution Function (CDF) of the D2D SINR inside each area i. Each CDF then gives us the eligibility probability for a D2D user by setting a threshold $SINR_0$ below which the user is stated as non-eligible; we chose $SINR_0 = 0$ dB hence giving the corresponding $1 - P_i$, $1 \leqslant i \leqslant 5$, from the plots;

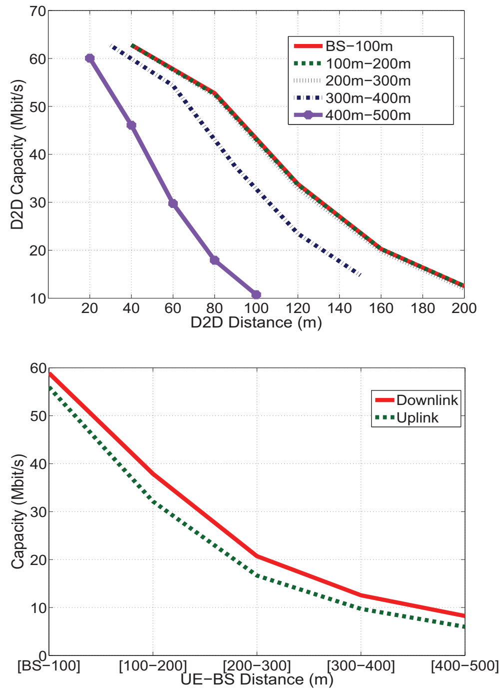

Fig. 3. D2D (a) and Downlink, Uplink (b) capacities.

(iii) the samples (Transmitter-Receiver distance, $SINR$) of the simulated D2D pairs enable us to derive a relation between mean SINR and mean D2D distance, dependent on the position of the receiver with respect to the BS. Such a deterministic relation then determines the D2D range as the maximum distance which corresponds to $SINR = SINR_0$. It turns out that the D2D range corresponding to $SINR_0$ decreases from about 200 m for the first three cellular areas to 150 m for the fourth one and 100 m for the fifth one;

(iv) the computed mean SINR values are then transformed into capacity values per RB using link level curves provided by 3GPP standards [2]. The obtained D2D ($C_{i,k}$), Uplink ($C_{u,i}$) and Downlink ($C_{d,j}$) capacities are plotted in Fig.3, as a function of the transmitter-receiver distance.

The nominal overall capacities are finally computed to give $C_d^0 = 14.1$ Mbit/s, $C_u^0 = 10.8$ Mbit/s and $C_D = 19.5$ Mbit/s; a higher capacity is thus potentially provided to D2D flows.

B. Impact of D2D eligibility

We first examine the impact of the eligibility probability P when considering this parameter as exogenous. Arbitrary values of $P \in [0, 1]$ are simulated as follows. Probability P actually depends on the user location in the cell; we thus have to determine I probabilities P_i. To replicate the behavior of Fig. 2, we set $I = 5$ and for each targeted value of P, we

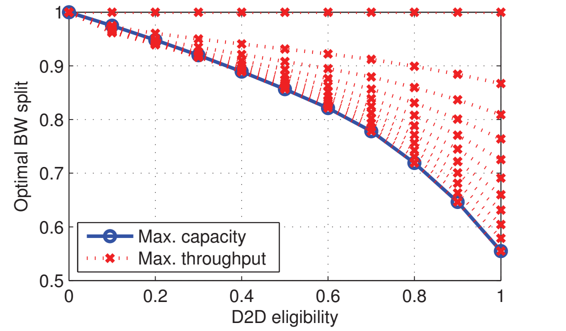

Fig. 4. Optimal BW partitioning parameters α^* and $\alpha^\#$ as functions of P (with top to bottom curves in increasing order of a_D from 0 to C^*).

approximately determine a set of 5 probabilities P_i, $1 \leqslant i \leqslant 5$, whose mean P is consistent with (6).

Fig. 4 shows the impact of P on the optimal bandwidth partitioning parameter α, either α^* for maximum stability region (or "D2D capacity") or $\alpha^\#$ for maximum Uplink throughput. In the latter case, since $\alpha^\#$ also depends on the offered D2D traffic, we consider a set of D2D traffic intensities regularly spaced between 0 and the maximal capacity C^*, see (27). When P grows from 0 to 1, α^* decreases from 1 to a_u/C_u^0, the load without D2D traffic. As the part of the Uplink spectrum allocated to the Uplink sub-system is close to the nominal Uplink load, a potential constraint is thus set on the Uplink performance. This explains why the performance of Uplink (sub-system and traffic) is much degraded when probability P is high, as will be observed in Subsection VI-C. Regarding the throughput-optimized scheme, $\alpha^\#$ lies between α^* and 1, thus allowing some margin in the bandwidth allocated to the Uplink and therefore leading to better performance.

C. Performance of the scheduling schemes

We here focus on the Uplink channel and, as concerns HBP scheme with two Uplink schedulers, address the cases where the bandwidth partitioning is optimized, either by setting $\alpha = \alpha^*$ or $\alpha = \alpha^\#$. In the considered scenario, we set the Uplink traffic intensity a_u to 6 Mbit/s. The throughputs for the different sub-systems and traffic types are shown as functions of the D2D traffic intensity in Figs. 5 and 6. Two D2D eligibility probabilities are considered: $P = 0.13$ and 0.8, where the first one is that deduced from the cellular framework of Subsection VI-A and the second one is simulated as described in Subsection VI-B. Note that the first value of P is rather low; this is due to the fact that the simulated devices are *a priori* uniformly distributed over the whole cell, thus making the eligibility for direct D2D transfer weak.

From these figures, we make the following observations:

• the throughputs perceived by the different traffic classes in the Uplink sub-system (i.e., Uplink traffic and D2D traffic)

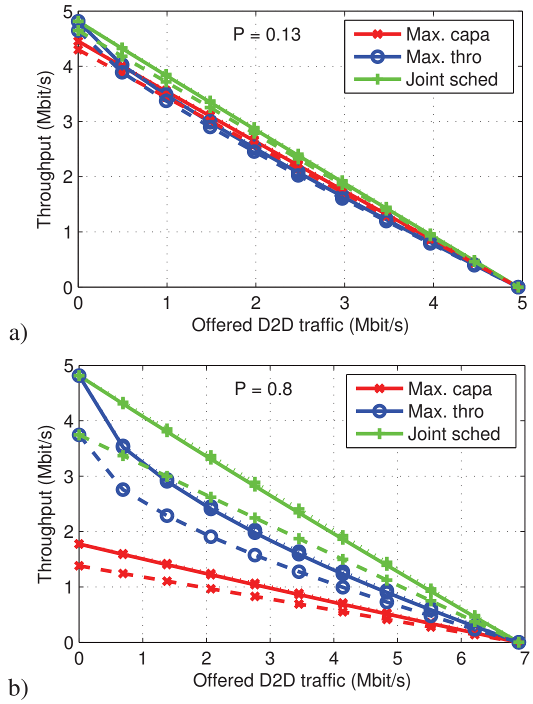

Fig. 5. Throughputs provided by the three scheduling schemes for the Uplink sub-system as functions of the D2D traffic intensity a_D for $P = 0.13$ (a) or $P = 0.80$ (b): Uplink traffic throughput $\Gamma_{u,u}$ (solid lines) and D2D traffic throughput $\Gamma_{u,D}$ (dashed lines).

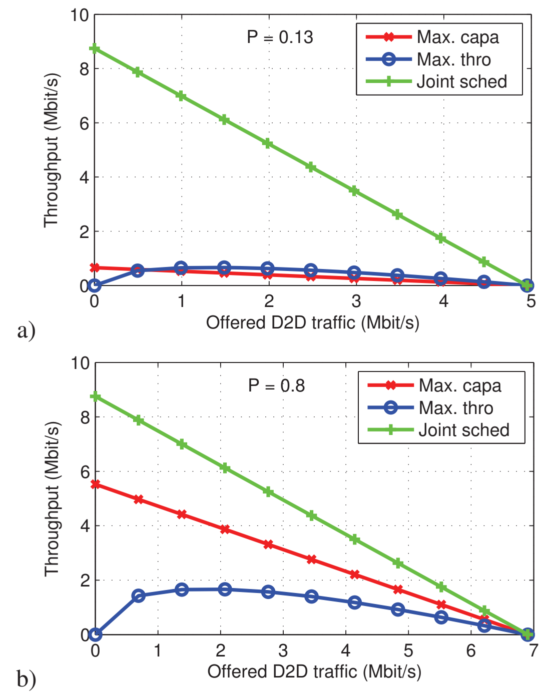

Fig. 6. Throughputs provided by the three scheduling schemes for the D2D sub-system, Γ_D, as functions of the D2D traffic intensity a_D for $P = 0.13$ (a) or $P = 0.80$ (b).

differ noticeably only when P is high, as shown by Fig. 5. The following discussions will thus not insist on these discrepancies and will speak of Uplink performance in general, for simplicity;

• when using HBP scheme, throughput performance is very contrasted between the two values of probability P. This is true, in particular, for the maximum capacity partitioning, which is consistent with our comments on Equ. (28). With this scheme, Uplink performance is much better than that of D2D link when P is low; the reverse holds when P is high (see the "Max capa" curves in plots a) and b), respectively);

• with maximum throughput partitioning, Uplink performance is always better than that of D2D link (this scheme has been designed to this end, indeed). The gap between them is particularly large when the probability P is low, and all the more when the D2D traffic intensity is very low, in which case the D2D throughput vanishes to 0 whatever the value of P;

• last, but not least, all plots show that the performance provided by JBU is always better than that provided by the two optimized schemes of HBP. Moreover, with JBU, there is no qualitative discrimination between Uplink and D2D link performance; the resource allocation to all concurrent flows is dynamic, so that the Uplink performance is not degraded and the different traffic flows perceive the maximum throughput enabled by the fair bandwidth sharing principle.

D. Throughput gain with JBU

We now provide a more precise evaluation of the gain provided by JBU scheme. Fig. 7 displays the ratio of throughput with JBU on throughput with HBP for the overall Uplink spectrum and D2D sub-system, as a function of α for $P = 0.8$ and $a_D = 4$ Mbit/s. The plot also shows the points where α is optimized, $\alpha = \alpha^{\#}$ or $\alpha = \alpha^*$.

The salient feature is that the gain provided by JBU is all the more important for D2D traffic (on the direct D2D link) when the BW partitioning for HBP is performed around the optimum maximizing the Uplink throughput. The curves also illustrate the danger to specify in HBP a value of α even slightly distinct from the optimal value $\alpha^{\#}$ (note the range for α is rather narrow due to stability constraints).

Within the numerical setting assumed here, the latter observations lead us to conclude that: (i) HBP may not be an adequate scheme even if the partitioning factor is optimized, either in a static way (maximum D2D capacity) or in a dynamic way (maximum Uplink throughput); (ii) with JBU, D2D users can make an efficient use of their potential capacity C_D, which is the very purpose of designing a direct D2D link.

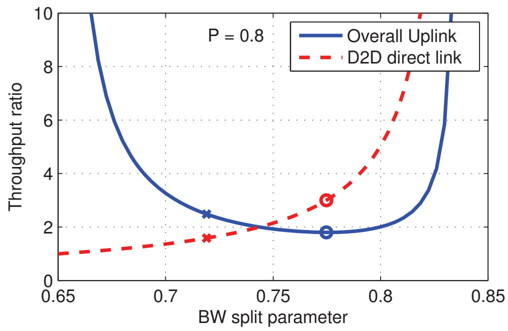

Fig. 7. Throughput gains provided by JBU vs. HBP, $\Gamma_{u,S}/\Gamma_{u,S}^{(\alpha)}$ for the overall Uplink and $\Gamma_D/\Gamma_D^{(\alpha)}$ for the D2D link, as functions of α for $P = 0.8$; optimal partitioning for HBP is marked by "o" and "x" for maximum Uplink throughput ($\alpha^{\#}$) and maximum D2D capacity (α^*), respectively.

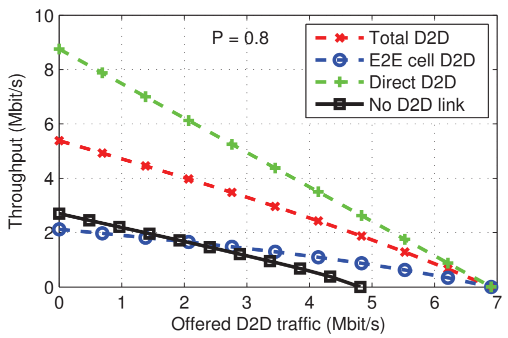

Fig. 8. Throughputs for the D2D traffic with JBU scheme and $P = 0.8$: Γ_D for the direct link, $\Gamma_{E,D}$ for the E2E cellular path, and $\Gamma_{T,D}$ for the total D2D traffic; comparison with D2D throughput $\Gamma_{E,D}$ when $P = 0$.

E. Throughput benefit of D2D

We finally discuss the compared performance of D2D traffic between the case where all traffic is transferred through the cellular network (thus corresponding to an eligibility probability $P = 0$) and the case where a fraction of this traffic is handled through the direct D2D link. We only consider JBU scheme, which provides the best performance. In Fig. 8, D2D throughputs are displayed for the direct link, for the E2E cellular path and for the whole D2D traffic, when $P = 0.8$ and $a_d = 8$ Mbit/s (with previous values for a_u and a_D). The E2E throughput is computed, similarly to $\Gamma_{E,D}^{(\alpha)}$ in (24), as $1/\Gamma_{E,D} = 1/\Gamma_{u,D} + 1/\Gamma_{d,D}$, and then the whole D2D traffic throughput as $1/\Gamma_{T,D} = P/\Gamma_D + (1-P)/\Gamma_{E,D}$.

We observe that the direct D2D link clearly offers a better throughput whatever the D2D traffic load; note also the significantly augmented value of the D2D capacity C^* (from 5 to 7 Mbit/s in the present case). Besides, the D2D traffic which follows the cellular End-to-End path perceives a very low throughput, close to that with no D2D link; it benefits, however, from the D2D capacity gain already mentioned.

VII. Conclusion

In this paper, we evaluated the flow-level performance of two scheduling schemes for D2D service on the Uplink, assuming that D2D traffic adds up to the cellular traffic. Our approach is original in that both D2D and cellular communications are scheduled, thus avoiding intra-cell interference issues, and in that we consider the alternative transfer of D2D flows on to the cellular links in case they are not eligible for direct D2D transfer. By means of multi-class Processor-Sharing queuing models discriminating traffic types and distinct area/zone capacity regions in the cell, we have shown that a joint scheduling on the Uplink dealing with both D2D and cellular traffic provides better performance than dedicated schedulers. While this result can be expected in a qualitative sense, as an instance of statistical multiplexing gain, we have here provided insightful models allowing to quantify this gain. Besides, joint Uplink scheduling has been shown to offer a significant throughput gain for the D2D communications, compared to the situation where no direct D2D link exists.

An evolved model accounting for the dependence between the Uplink and Downlink sub-systems could be the subject of future works. Further investigations on the performance of allocations schemes for D2D might also address the mobility of D2D pairs within the cell or the network.

References

[1] IST-4-027756 WINNER II D1.1.2 v1.2 WINNER II channel models. 11, 02 2008.

[2] 3GPP. Study on LTE Device-to-Device Proximity Services; Radio Aspects-TR 36.843.

[3] A. Asadi, Q. Wang, and V. Mancuso. A Survey on Device-to-Device Communication in Cellular Networks. *IEEE Communications Surveys and Tutorials*, 16:1801–1819, 2014.

[4] T. Bonald and A. Proutière. Wireless Downlink Data Channels: User Performance and Cell Dimensioning. In *Proc. of MobiCom'03*, San Diego, USA, Sept. 2003.

[5] S. E. Elayoubi, A. M. Masucci, J. Roberts, and B. Sayrac. Optimal D2D Content Delivery for Cellular Network Offloading. *MONET*, 22(6):1033–1044, 02 2017.

[6] P. Gandotra and R. K. Jha. Device-to-Device Communication in Cellular Networks. *J. Netw. Comput. Appl.*, 71(C):99–117, Aug. 2016.

[7] ITU-R. Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced - Report M.2135.

[8] M. Ji, G. Caire, and A. F. Molisch. Wireless Device-to-Device Caching Networks: Basic Principles and System Performance. *IEEE Journal on Selected Areas in Communications*, 34(1):176–189, 2016.

[9] S.-Y. Lien et al. Enhanced LTE Device-to-Device Proximity Services. *IEEE Communications Magazine*, pages 174–182, Dec 2016.

[10] A. Masucci, S. E. Elayoubi, and B. Sayrac. Flow level analysis of the offloading capacity of D2D communications. In *Proc. of 2016 IEEE WCNC*, pages 1–6, April 2016.

[11] P. Osti, P. Lassila, and S. Aalto. Performance of D2D Underlay and Overlay for Elastic Traffic. In *Proc. of MSWiM '16*, pages 270–277, New York, NY, USA, 2016. ACM.

[12] Y. Ozcan, C. Rosenberg, and F. Guillemin. A Benchmark for D2D in Cellular Networks: the Importance of Information. In *Proc. of IEEE PIMRC*, Montreal, Canada, Oct. 2017.