Towards D2D-Enhanced Heterogeneous Networks

Francesco Malandrino, Claudio Casetti, Carla-Fabiana Chiasserini
Politecnico di Torino, Italy

Abstract
In this paper, we examine upcoming 5G networks where the support of device-to-device (D2D) communication is expected to be a key asset for operators and users alike. Firstly, we argue the need to functionally integrate D2D and infrastructure-to-device (I2D) modes. Next, we address practical issues such as integrated resource scheduling of D2D communication within heterogeneous networks, proposing an extension of the proportional fairness algorithm, which we call multi-modal proportional fairness (MMPF). We evaluate the impact of D2D in a two-tier scenario combining macro- and micro-coverage, finding that, although I2D retains a clear edge for general-purpose downloading, D2D is an appealing solution for localized transfers as well as for viral content.

I. Introduction
These are exciting times for cellular networks. 4G networks are now being deployed and offered to customers in several countries all over the world, and 5G ones are brewing. 5G is not only an evolution of current network generations, but, more significantly, a revolution in the ICT field: it will enable efficiently new ultra reliable, dependable, secure, privacy preserving and delay critical services to everyone and everything, such as cognitive objects and cyber physical systems.

Among the new features foreseen by 5G networks, device-to-device (D2D) transfers may have a prominent role, and our purpose here is to assess whether, and to which extent, they can be successfully integrated in 5G networks.

The adoption of D2D transfers is driven by four main use cases [1]:

- safety applications and disaster scenarios;
- novel, commercial proximity services (ProSe) scenarios;
- network traffic offloading;
- industrial automation and machine-to-machine communication [2].

The focus of our work is on the use of D2D for commercial applications and, in particular, for data traffic offloading. The typical example is “flash crowds” [3]: thousands of users in a small area, perhaps attending a football game and suddenly becoming interested in the same content, e.g., a replay clip. Some users may download the content through D2D, thus partly relieving the infrastructure of its load. Furthermore, through reduced interference and extended coverage, D2D may lead to increased network capacity.
Cellular networks are not the first technology to allow D2D transfers: similar techniques, aiming at roughly the same goals, such as ad-hoc mode in IEEE 802.11, have existed for decades – but have never become mainstream. Technical and non-technical issues, from driver support to security concerns, have always hindered their widespread adoption.

The first, important question we have to ask ourselves is then: why should we expect D2D to be successfully ushered into cellular networks? For decades, they have been working – quite successfully, in fact – in an infrastructure-centric fashion: users send data to base stations, base stations send data to users. Should we dare depart from such a reliable, tested working scheme in exchange for some performance improvement and dire technical challenges?

The second, related question: do not small-cell techniques serve the same goals as D2D, i.e., reduced interference, extended coverage, and network offloading? They certainly allow operators to retain the familiar infrastructure-centric operation mode. Does this mean that we can just use small-cells, and disregard D2D? Are there use-cases and scenarios where D2D is more appropriate? More interestingly, do we need to choose between small-cells and D2D, or can the two paradigms coexist? And if they do, at which cost in terms of complexity and overhead?

It is important to stress that we are not investigating here the usefulness of D2D itself, but, rather, the opportunity for its integration in cellular networks. Although D2D is now part of 3GPP standards, this does not necessarily imply that it will be implemented by vendors and operators, nor that it will be widely used in services and applications.

We explore these fundamental questions in Sec. II, where we focus on the integration of D2D in cellular networks. Then, in Sec. III, we introduce a system architecture for D2D support. Sec. IV is devoted to the problem of radio resource sharing in systems where both the D2D and infrastructure-to-device (I2D) paradigms (the latter including small cells) are implemented. Sec. V presents our reference scenario, and Sec. VI our numeric results. Finally, in Sec. VII we draw our conclusions.

II. D2D in heterogeneous cellular networks

In this section, we discuss two important, preliminary issues. First, we analyze how D2D communication should be integrated in cellular networks. Then, we discuss whether D2D and small-cell technologies are alternative or complementary to each other.

A. Integrating D2D in cellular networks

As discussed in [4] and [5], there are three main options concerning the integration of D2D in cellular networks: (i) whether D2D communication shall be network-controlled or not, (ii) whether it shall happen in-band or out-of-band, and (iii) whether it shall work in overlay or underlay fashion. For a conceptual framework concerning problems such as peer discovery, scheduling and resource allocation, see [6].

In the following, we discuss the most promising solutions, assuming that D2D-enhanced cellular networks will be network-controlled and operate in-band in an underlay fashion. As presented in [2], this is a fairly popular choice.
**Network-controlled** D2D essentially means that infrastructure nodes (base stations and control entities, as detailed in Sec. III) play a central role in establishing, arbitrating and managing D2D connections [7]. As explained in [1], they will provide the following, fundamental services: spectrum management, security, information brokering, and mobility management. The first two items in this list guarantee that D2D transfers do not translate into lower performance or poorer security. The last two imply that user equipment (UE) does not even have the burden of choosing between D2D and I2D. The opposite approach is represented by infrastructure-less networks, such as WiFi-direct or Bluetooth.

**In-band** D2D refers to the fact that device-to-device traffic uses the same, licensed frequencies as ordinary infrastructure-to-device traffic [7]. The main advantage of this approach is represented by the higher degree of control that operators retain on who transmits and how, which limits interference. Cooperation among users is also easier to enforce and check. Furthermore, terminals do not need to carry additional radio interfaces. The opposite approach, *out-of-band* D2D, is envisioned by those proposals that seek to offload cellular networks through other networks, e.g., 802.11 domestic access points with spare capacity.

**Underlay** refers to the fact that D2D communication has no part of the spectrum specifically reserved to it, as happens, e.g., in [5]. When networks operate in underlay fashion, D2D transfers share the same radio resources as those used by traditional cellular communications and are scheduled within the cellular bands in an opportunistic fashion. As a consequence of network-controlled, in-band operation, D2D transfers are *scheduled*, as mentioned earlier: this avoids the potential inefficiency of decentralized schemes based on the Carrier Sense Multiple Access (CSMA) paradigm. The *overlay* approach, instead, implies that separate radio resources are devoted to D2D and I2D, or to device-to-infrastructure (D2I) communication.

### B. D2D and small cells

**Small cell** is an umbrella term, covering several, quite different, technologies. In general, it refers to low-power, short-range, operator-owned nodes integrated in the cellular infrastructure to enhance its coverage and/or capacity. Such communication nodes go under names like pico- or micro-eNBs. The goals of small-cells are very similar to the ones of D2D, so one may think we just need to choose the most effective among the two and discard the other.

There is, however, a very important difference, and it concerns the *source* of the information being transmitted. With small cells, information is still downloaded from some remote server, i.e., on the Internet, and then transmitted to the user. With D2D, instead, the information moves directly from one user to another, either generated by the transmitting user itself, or previously received from a remote server or another user. With reference to the use cases described earlier, we can say that both D2D and small cells can be used for network offloading, but D2D is the most effective choice for ProSe scenarios. Later in this work, we study how the two can coexist in a real-world network.

### III. System overview and architecture

Owing to the new, complex tasks assigned to the cellular infrastructure, namely, arbitrating and scheduling D2D transfers, we envision that base stations (eNBs in LTE terminology) will be assisted by a
new kind of entity: area controllers. The main difference between these entities is the level at which they operate. eNBs are solely concerned with propagation and spectrum aspects. Area Controllers (ACs), instead, have a wider view of the network: they control a bigger area, and are in charge of content-aware decisions. More exactly:

- eNBs assign individual spectrum resources (PRBs in LTE) to pairs of communicating endpoints (i.e., themselves and a user, or a pair of users through D2D);
- ACs monitor the network state, including content demand and propagation conditions, and exploit such information to decide the paradigm (I2D or D2D) and the amount of resources a transfer should be assigned.

The system model, along with the way ACs, eNBs, and users interact with each other, is summarized in Figure 1.

![Figure 1 System model, summarizing the entities involved in heterogeneous networks, and interaction diagram highlighting the role of each entity.](image)

eNBs update the AC on propagation conditions (1). The AC uses such information to update its policies (2). Upon a request from the users (3), the AC is also in charge of taking content-aware decisions (4), e.g., whether certain content should be downloaded through D2D or I2D. Although eNBs do have a better knowledge of propagation conditions than ACs, they are not aware of which content items are stored by each user, and thus which D2D transfers are possible. eNBs are, however, in charge of actual scheduling decisions (5).

Using its knowledge of the content being downloaded in the network, the AC can refine (6) its policies, e.g., by understanding that now content $c_1$ can be downloaded from user $u_1$ using D2D if needed. Such decisions are transmitted (8) to the eNB, which subsequently enacts them (9) using the most appropriate spectrum resources.

Operations performed by eNBs have, above all, to be fast: in LTE, PRBs are assigned as frequently as once per millisecond. Furthermore, being part of the access network, eNBs are oblivious of higher-level concerns: as an example, they cannot run content-aware scheduling algorithms. Such algorithms do, however, accept input parameters, e.g., which users should get their content through D2D. Conversely,
ACs are in an ideal position to take more complex decisions, exploiting more information and (possibly) taking longer to find the (ideally) optimal one. The interface between ACs and eNBs consists in the parameters of the fine-grained scheduling algorithms run by eNBs.

It is worth stressing that the time scales at which ACs and BSs operate are different. More exactly, while eNBs have to take decisions each millisecond, ACs can refine their policies over longer time intervals, accounting for more information and taking more accurate (and, if need be, computationally complex) decisions.

IV. Integrated I2D and D2D scheduling

This section is chiefly concerned with how D2D in LTE will happen. Specifically, we address spectrum and bandwidth aspects in Sec. IV.A, and scheduling in Sec. IV.B.

A. Spectrum, bands and resources

Unlike previous-generation networks such as 3G, LTE and its successors operate on several different carrier bands, anywhere between 700 MHz and 3.8 GHz in frequency and anywhere between 1.4 and 20 MHz in width. This allows flexible, scalable deployments, with wider, lower-frequency cells coexisting with smaller, higher-frequency ones. The availability of specific bands changes on a regional and country basis.

Details of the frequency and time structure of LTE are fairly complex, and are beyond the scope of this paper. Here we simply refer to physical resource blocks (PRBs). PRBs correspond to a set or resources in the time and frequency domains, namely, a bandwidth of 180 kHz for the duration of a time slot (0.5 ms), which corresponds to half a subframe. The subframe is the atomic scheduling unit for LTE networks. As an example, let us consider a bandwidth of 10 MHz for the downlink, which accommodates 50 PRBs per time slot, available for data traffic\(^1\). In downlink each PRB can carry up to 504 bits; then, the actual data rate depends on propagation conditions, interference and on the use of MIMO. From our viewpoint, scheduling simply means deciding how many PRBs shall be assigned to each pair of endpoints (an eNB and a UE, or two UEs) that need to communicate.

Given the uplink and downlink subcarriers, one may wonder where D2D would fit in such a picture. Indeed, the topic has been widely debated in standardization forums; the predominant orientation is to accommodate D2D traffic in uplink resources [4]. Reasons for this choice include lower interference and reduced impact on eNB-to-UE links. Additionally, the uplink spectrum is currently underutilized [4]. It is therefore more appropriate to use it for content downloading use cases (including flash crowds), where D2D links can effectively be used for network offloading.

B. Scheduling in heterogeneous networks

Scheduling in next-generation, heterogeneous networks is a complex task. Essentially, we have to assign a finite set of resources (PRBs) to a set of pairs of endpoints. These pairs of endpoints (i) may want to communicate in I2D or D2D fashion; (ii) differ in position, mobility and propagation conditions; (iii) aim

\(^1\) The remaining bandwidth is used as guarding bands.
at fetching different amounts of different content. The very metric to optimize is unclear: one may think that the total network throughput is a good candidate; however, this would mean disregarding fundamental fairness issues.

The scheduling algorithm used in virtually all current networks is proportional fairness (PF). Users are given a priority that is directly proportional to the rate they can achieve, and inversely proportional to the amount of data they already transferred. The overall effect is to allow users with better propagation conditions to transmit more data (thus enhancing the global network capacity), without starving the others (thus guaranteeing a certain level of fairness).

Traditional PF is only concerned with which users shall be served, but not with how to do so. Indeed, in traditional cellular networks there is only one way to serve users, i.e., through the closest base station. In heterogeneous networks, however, we have two different problems:

- User service mode (i.e., through macro- or micro-eNBs, or via D2D);
- Resource allocation.

Not surprisingly, dealing with these two issues at the same time is substantially more complex than traditional scheduling. Many D2D-aware scheduling algorithms have been proposed for LTE networks [2][8][9], with different strategies and objectives. However, in this paper we are chiefly concerned with designing a practical scheduling scheme that can be implemented in cellular networks and with assessing the sheer impact of D2D on the performance of cellular networks. Thus, instead of presenting an optimized scheduling algorithm accounting for all the issues (and opportunities) of heterogeneous, D2D-enhanced networks, we propose an evolution of the PF algorithm and study its performance with and without D2D support. We name such algorithm Multi-Modal Proportional Fairness (MMPF).

C. The MMPF algorithm

As mentioned, scheduling in heterogeneous networks entails taking two decisions: which users we should serve, and how to do it. The first decision is the same as in traditional proportional fairness, and therefore it is taken in the same way. At each step, we select the user with the highest ratio between achievable rate and amount of already downloaded data. As for the second decision, we proceed as follows:

- if there is another user with the requested content within a distance $R_{\text{max}}$, use D2D;
- otherwise, use I2D.

The pseudocode of the MMPF algorithm is presented in Figure 2.
Note that, if I2D is selected, either macro- or micro-eNBs can be used, whichever has the highest received signal strength indicator (RSSI) (line 9). We may observe a lack of symmetry: we decide whether to use D2D based on distance, but then select macro- or micro-eNBs based on signal quality. The reason is eminently practical. While eNBs (and thus the AC) have many ways of estimating the location of users [10], there is no simple way they can obtain reliable information on the signal quality between two users. Estimation techniques do exist, but none is included in standards and all of them imply some overhead\(^2\). Similarly, using the threshold \(R_{\text{max}}\) is a somewhat coarse, hardly optimal, way of choosing between D2D and I2D. Other, more sophisticated approaches may yield a higher average throughput, a better fairness, or both. As an example, we may re-use the resources allocated for D2D, owing to the lower interference they suffer from. Recall, however, that our purpose is to assess the impact of D2D on the performance of cellular networks, and to which extent it can replace – or complement – small-cell solutions. In view of this, having an algorithm that closely resembles the de facto standard of today’s network is particularly appropriate and convenient.

It is important to point out that the first decision taken by MMPF, i.e., whether to use I2D or D2D, also impacts the part of spectrum – uplink or downlink – over which the transfer occurs. The original PF scheduling is applied independently for uplink and downlink; it follows that, in the uplink bandwidth, D2D downloads and D2I uploads will be scheduled together, i.e., will compete with each other. Therefore, the amount of upload traffic will have an impact on the performance of D2D downloads and vice versa.

Unlike the original PF algorithm, MMPF is run at the area controller (AC), as described in Sec. III. Indeed, the second decision, i.e., whether to use D2D or I2D, has to be content-aware, and must be taken by an entity with more complete information about the content available at each user. It is also important to stress that users are never requested to download a content item they are not interested in, for the sole purpose of acting as relays. Users are requested to share the content they have already downloaded but nothing more.

It is also worth pointing out that, as the name suggests, MMPF is fair in the same sense the original proportional fairness algorithm is. As we can see in line 2, the next user to serve is selected taking into

\(^2\) Other simulations, when the quality of D2D links is perfectly known and used in MMPF, yielded essentially the same results as the ones we present here.
account the ratio between the rate of each user and the amount of data he/she already downloaded. As in the original algorithm, this does not imply that all users end up receiving the same amount of data – another, more intuitive, definition of fairness.

V. Reference scenario

We evaluate our solution in the two-tier scenario that is typically used within 3GPP for LTE network evaluation [11]. The scenario comprises a service network area of 12.34 km$^2$, covered by 57 macrocells and, unless otherwise specified, 228 microcells. Macrocells are controlled by 19 three-sector macro-eNBs; the macro-eNBs inter-site distance is set to 500 m. Micro-eNBs are deployed over the network area, so that there are 4 non-overlapping microcells per macrocell. A total of 3420 users are present in the area. In particular, in order to have a higher user density where microcells are deployed, 10 users are uniformly distributed within 50 m from each micro-eNB. The rest of the users are uniformly distributed over the remaining network area. Users move according to the cave man model [12], with average speed of 1 m/s.

According to current specifications [11], [13], we assume the following pairs of values for power and antenna height: (43 dBm, 25 m) for macro-eNBs, (30 dBm, 10 m) for micro-eNBs, and (23 dBm, 1.5 m) for UEs. All network nodes operate over a 10 MHz band at 2.6 GHz, thus there are a total of 50 PRBs to assign for each subframe. As already mentioned, the signal propagation is realistically modeled according to ITU specifications for urban environment [11], while the SINR is mapped onto peak throughput values using the experimental measurements in [14]. More precisely, the propagation loss for macro-cells, micro-cells and D2D is given by the following expressions [11]:

- $13.5 + 20 \log_{10} f_c + 39 \log_{10} d$
- $22.7 + 26 \log_{10} f_c + 36.7 \log_{10} d$
- $27.0 + 20 \log_{10} f_c + 22.7 \log_{10} d$

where $f_c$ is the carrier frequency and $d$ is the distance between transmitter and receiver. The experimental measurements of [14] are summarized in Figure 3, and refer to the case of 2x2 MIMO.

![Figure 3: Throughput versus SINR, based on the experimental results in [14].](image)
Users choose the content they request from a set of 21 different items, belonging to three categories: software updates, videos, or viral content [15]; their size, deadline and request rates are summarized in Table 1. Content items belonging to each category are chosen with uniform probability; similarly, the request rate is uniform throughout the simulation. Content deadlines imply that service requests are aborted after the deadline expires. Notice however that MMPF, just like original PF, does not account for such deadlines while scheduling traffic. We highlight that video and viral items have stricter constraints on delivery time. Additionally, the viral item is modeled as being in high demand to mimic content becoming suddenly popular through social networks because of “flash-crowds”. We assume that a scheme based on incentives is implemented and, thus, that users are willing to cooperate by providing content upon receiving a request.

Simulations are carried out through a custom simulator. The total simulation time is 30 s.

It is important to remark that the scenario we are addressing will necessarily only paint a part of the overall picture. Additional case studies involving out-of-coverage communication as well as interactive and conversational services should be addressed in future work for reasons of space.

### VI. Results

We vary the maximum distance allowed for D2D transfers, $R_{\text{max}}$, between 10 and 100 meters, and study the amount of data that is transferred through each of the possible paradigms – macrocells, microcells, and D2D. We are interested in investigating how D2D relates to the other paradigms, especially microcells. The results are depicted in Figure 4. First of all, we observe that most of the data flow through microcells; this is expected as this paradigm offers an excellent balance between short range, i.e., low interference, and high power. A smaller amount of data flows through macrocells, and a more limited one through D2D.

As the maximum range allowed for D2D increases, so does the amount of data transferred through it. This is essentially because a wider range directly translates into a higher number of potential sources, i.e., users carrying the needed content. More interestingly, increasing the range also increases the amount of data transferred through the other paradigms. Indeed, allowing more room for D2D has two positive effects:

- some data can be downloaded from closer sources, hence with higher quality and lower interference;

<table>
<thead>
<tr>
<th>Content class</th>
<th>Number of content items</th>
<th>Size [Mbit]</th>
<th>Deadline [s]</th>
<th>Request rate [copies/ms]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software updates</td>
<td>10</td>
<td>12</td>
<td>4</td>
<td>0.1</td>
</tr>
<tr>
<td>Video</td>
<td>10</td>
<td>3</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>Viral</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>50</td>
</tr>
</tbody>
</table>
• D2D transfers occur on uplink bands, thus downlink I2D traffic, hence interference, is reduced. Both imply that more downloads are completed within their deadline, which explains why the overall amount of transferred data increases.

![Figure 4](image)

**Figure 4** Baseline scenario: amount of transferred data through each paradigm.

In other words, by allowing D2D, not only we can move some traffic to the uplink spectrum, but we can also more effectively use the downlink one, increasing its capacity. Congestion is of course increased on the uplink bands. However, with most of the present (and future) cellular traffic being represented by downloads, this is an acceptable tradeoff.

**Table 2** Spectral efficiency for different transfer paradigms

<table>
<thead>
<tr>
<th></th>
<th>Macro</th>
<th>Micro</th>
<th>D2D</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF</td>
<td>0.96</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>MMPF</td>
<td>1.29</td>
<td>1.18</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Table 2 summarizes the spectral efficiency, i.e., the amount of data transferred through each PRB, for each traffic paradigm. We clearly see that by allowing D2D transfers we substantially increase the efficiency of macro- and micro-cell-based I2D as well.

Next, we try to assess whether D2D can complement, or replace, microcells. To this end, we reduce the number of microcells by 50%: this is not a disaster scenario, but it may account for, e.g., a less pervasive deployment due to economic concerns. The effect is summarized in Figure 5. The first, obvious observation is that there is much less traffic flowing through microcells. Such a loss is compensated partly by macro-cells, and partly by D2D, although the overall network capacity decreases. Thus, it cannot be said that D2D can fully replace micro-cells (and spare the costs of deploying such kind of infrastructure), nevertheless D2D can complement micro-cells, and offset a significant part of the effects of a reduced infrastructure deployment.
The above observation prompts another question: are there some content items that are more suitable than others for D2D delivery? The answer can be obtained by looking at what content gets transferred through each paradigm, for the baseline scenario with a 50-meter limit distance for D2D. From Figure 6, it can be observed that D2D is especially effective at transferring viral content. Indeed, viral content is highly popular and all requests for it happen in a fairly short interval of time, thus it is much more likely to find a neighbor device that can provide the requested content.

![Figure 5 Amount of transferred data for each paradigm, when the number of micro-cells is halved.](image)

![Figure 6 Data transferred through each paradigm for different content classes.](image)

In summary, we can say that D2D is an extremely appealing solution for proximity services and offloading of highly popular content. Although it cannot fully replace the ordinary I2D paradigm
(including microcells) for general-purpose downloading, it can be successfully integrated within heterogeneous, cellular-based networks. To answer the question we raised earlier about the coexistence between D2D and small-cell approaches, we can conclude that these two paradigms serve different, albeit partially overlapping, purposes, and can – should – profitably be integrated together in future cellular networks.

VII. Conclusions

The main contribution of our work was the analysis of integrated D2D and I2D scenarios. We discussed motivations for the coexistence and joint deployment of both transfer modes, as well as a practical solution for resource scheduling in a multimode context that extends the popular Proportional Fairness scheduling algorithm. Performance evaluation, in a typical scenario used for LTE network studies, provides new grounds for incorporating D2D in future releases of 3GPP standards, by showing its effective offloading potential in case of both localized and highly popular content.

With reference to the questions posed in the Introduction, it can be concluded that (i) D2D can indeed be profitably integrated within cellular networks, and (ii) it cannot fully replace small-cells but it is useful to complement them and to mitigate the effects of a reduced infrastructure deployment.

VIII. References