On the Compound Impact of Opportunistic Scheduling and D2D Communications in Cellular Networks

Arash Asadi
Institute IMDEA Networks
University Carlos III of Madrid
Leganes (Madrid), Spain
arash.asadi@imdea.org

Vincenzo Mancuso
Institute IMDEA Networks
University Carlos III of Madrid
Leganes (Madrid), Spain
vincenzo.mancuso@imdea.org

ABSTRACT

Opportunistic scheduling was initially proposed to exploit user channel diversity for network capacity enhancement. However, the achievable gain of opportunistic schedulers is generally restrained due to fairness considerations which impose a tradeoff between fairness and throughput. In this paper, we show via analysis and numerical simulations that opportunistic scheduling not only increases network throughput dramatically, but also increases energy efficiency and can be fair to the users when they cooperate, in particular by using D2D communications. We propose to leverage smartphone's dual-radio interface capabilities to form clusters among mobile users. We design simple, scalable and energy-efficient D2D-assisted opportunistic strategies, which would incentivize mobile users to form clusters. We use a coalitional game theory approach to analyze the cluster formation mechanism, and show that proportional fair-based intra-cluster payoff distribution brings significant incentive to all mobile users regardless of their channel quality.

Keywords

LTE; opportunistic scheduling; clustering; D2D.

Categories and Subject Descriptors

C.2.1 [**Computer-communication networks**]: Network Architecture and Design—*Wireless communication*

1. INTRODUCTION

Opportunistic schedulers have become a promising solution to cope with the mobile consumer traffic boom in cellular networks. This class of schedulers exploit multiuser diversity to reorder transmissions so that each user is served, with high probability, when it is in a relatively good channel state. However, the achievable opportunistic gain is restrained by user's fairness requirements, and by limited memory and computational resources at the base station. In fact, schedulers that achieve a good tradeoff between throughput and fairness, e.g., the renowned Proportional Fair scheduler (PF) [33], are too complex for the centralized architecture of cellular networks. In contrast, today's mobile devices are equipped with high processing power, large memory, and multiple radio interfaces. Therefore, mobile users may communicate with each other using 802.11 interfaces, e.g., using WiFi-Direct capabilities [31]. Particularly, multiple radios could be exploited for establishing cooperation via Device-to-Device (D2D) communications, e.g., to form clusters, as illustrated in Figure 1.

Figure 1: Cellular network with clusters of dual-radio mobiles.

In this paper, we propose to use the multi-radio capabilities of newly designed mobile devices to form cooperative clusters. The presence of clusters simplifies the scheduling operation at the base station and enables efficient, opportunistic radio resource utilization. Specifically, we design a multi-layer cluster-based scheduling mechanism in which the base station *schedules clusters* instead of users, while intra-cluster resource distribution is left to cluster members using D2D communications.

Note that scheduling clusters instead of users means that each cluster connects to the base station via a *cluster head*. However, unlike existing clustering approaches, we propose to select the cluster head *opportunistically* in each frame. The resulting scheduling mechanism consists of two elements: an algorithm to schedule clusters, and an algorithm to select cluster heads opportunistically within each cluster. We primarily focus on the cluster scheduling part and show that cluster heads can be chosen in a pure opportunistic way.

The main contributions of this paper are as follows: (i) we design a novel network architecture based on cooperative D2D communications, combining clustering techniques and opportunistic scheduling; (ii) we design two novel cluster scheduling algorithms, and present a mathematical analysis of these schemes; (iii) we provide an analytic model for user power consumption in the proposed architecture; (iv) we provide a game theory-based description of the clustering formation and revenue distribution; (v) via extensive numerical simulations, we evaluate the compound impact of opportunistic scheduling and D2D techniques within the proposed architecture, and show that we achieve dramatic throughput

MSWiM'13, November 3–8, 2013, Barcelona, Spain.

http://dx.doi.org/10.1145/2507924.2507929.

gain, extremely fair throughput distributions among mobile users, and high energy efficiency.

The remainder of this paper is organized as follows. In Section 2 we discuss the related work. The system model is presented in Section 3. In Section 4 we model the scheduling algorithms used in this work. We provide a novel model for power consumption of clustered dual-radio mobile users in Section 5. In Section 6 we evaluate the clustering gain with static topologies. In Section 7 we propose a coalitional game theory approach to form clusters and evaluate our clustering proposal in realistic dynamic topologies. Section 8 concludes the paper.

2. RELATED WORK

Since our proposal defines opportunistic *cluster* scheduling using cooperative communications, here we review the literature related to opportunistic scheduling and cooperative communications.

Opportunistic scheduling. Proposals on opportunistic scheduling are very mature. Indeed, a simplified version of PF is already implemented in 3G systems [33]. However, there are other schedulers proposed in the literature, which promise better throughput performance than PF, at the expenses of either fairness or increased complexity. Knopp and Humblet proposed the so called MaxRate opportunistic scheduler in [14], which always schedules the user in the best channel condition. Since opportunistic gain of MaxRate relies on multiuser channel diversity, opportunistic beamforming was proposed in order to increase the user diversity [27, 29]. MaxWeight [1] is another opportunistic scheduler that selects the user with the highest product of queue length and transmission rate. *Exp-rule* schedulers [26] are throughput-optimal schedulers that prioritize users based on an exponential formula using queue size and transmission rate of every user.

Our proposal leverages the MaxRate scheduler for communications between base station and cluster heads, thus achieving the maximum utilization of the airtime allotted to each cluster. Differently from existing (opportunistic) schedulers, we propose to schedule clusters as a whole, and show that *cluster* scheduling can be operated by means of simple and scalable schemes achieving high fairness, notwithstanding the use of MaxRate.

User cooperation. Research on cooperative communications in cellular networks has investigated different aspects of cooperation, such as base station-level cooperation [5] and user-level cooperation [13].

Relaying is an example of cooperation, in which mobile stations act as fixed base stations for long periods, hence increasing coverage without the need for setting up additional cells [15, 32]. Specifically, Wu *et al.* [32] propose to use dual-radio relay stations (cellular and ad-hoc) to perform load balancing among different cells. Their approach does not consume cellular radio resources for relaying but it is not channel-opportunistic and introduces extra delay.

Another example of user cooperation is represented by clustering techniques. Clustering has been well studied in wireless sensor networks for energy saving, routing and coverage improvement purposes [4, 10]. The work in [6, 12, 18, 23, 30] studied clustering in WLANs and cellular networks. In particular, Lin *et al.* [18] proposed clustering to form robust multi-hop networks. In [6, 23], mobile users are clustered to form a virtual antenna array which emulates a MIMO device via D2D communications. Furthermore, Dohler *et al.* [6] proposed to use a second wireless interface (e.g., bluetooth or WiFi) for intra-cluster communications. These proposals are either not suitable for cellular network protocols, or do not exploit opportunistic scheduling.

In general, D2D communications may be utilized for cooperative communications, e.g., packet forwarding, and relaying. In [7]and [17], the authors explore some applications of D2D communications in cellular networks such as P2P, multiplayer gaming, and multicast transmissions. Yu *et al.* [34] propose D2D communications in cellular networks for local traffic handling. In their view, D2D transmissions are meant to handle communications among two mobile devices, however users do not help each other to relay traffic to the base station. Also, all transmissions occur over the same interface as cellular communications, and D2D resources are allocated by the base station. Differently from our work, the existing proposals do not exploit opportunistic scheduling.

3. SYSTEM MODEL

We model downlink transmissions in a single LTE-like cell using a 20 *MHz* bandwidth in FDD mode [28]. There are N mobile users served either with legacy user-based schedulers or with our proposed cluster-based scheduling algorithms. Since we are interested in network capacity and fairness under heavy load conditions, we study the case of fully backlogged downlink flows. The total downlink capacity, in terms of transmission slots per frame, is indicated as S_{tot}, and the maximum achievable rate in the cell is 80.64 *Mbps*.[1]

Selection of Modulation and Coding Scheme (MCS). The downlink channel in between the base station and mobile node i is characterized by stationary Rayleigh fading. Therefore, the SNR can be described as a r.v. C_i with average SNR γ_i, so that pdf and CDF of the SNR have the following expressions, respectively:

$$f_i(z) = \frac{1}{\gamma_i} e^{-\frac{z}{\gamma_i}} u(z), \quad F_i(z) = \left[1 - e^{-\frac{z}{\gamma_i}}\right] u(z); \qquad (1)$$

where $u(z)$ is the unit step function.We assume that user channels are independently distributed but not identically, and the channel state information is available at the base station. Transmissions occur at different rates according to M available MCSs. We assume that the MCS for user i is selected as a function of the instantaneous SNR, i.e.:

$$MCS_i = k \iff C_i \in [th_k; th_{k+1}[\,, \; k = 1 \dots M; \qquad (2)$$

$$th_1 = 0; \quad th_p < th_q \iff p < q; \quad th_{M+1} = \infty.$$

Therefore, the probability that a scheduled user i receives a frame encoded according to the kth MCS is:

$$\pi_k^{(i)} = \int_{th_k}^{th_{k+1}} f_i(z)dz = e^{\frac{th_k}{\gamma_i}} - e^{\frac{th_{k+1}}{\gamma_i}}. \qquad (3)$$

The number of data bits transferred in one OFDMA symbol using the kth MCS is denoted by b_k, see Table 1. The table shows the list of possible MCS's with their corresponding SNR thresholds for LTE-like networks [25]. The implementation margin (IM) in Table 1 is a value that represents the effects of non-ideal receiver. For the sake of tractability, in this paper we assume that mobile users belong to

[1] We neglect LTE protocol overhead, and consider that all available downlink resources can be used for data traffic.

Table 1: Modulation and coding schemes

Modulation	Coding Rate	SNR (dB)	IM (dB)	SNR+IM (dB)	b_k(Bits per symbol)
No transmission	-	$-\infty$	-	$-\infty$	0
QPSK	1/8	-5.1	2.5	-2.6	0.25
	1/5	-2.9		- 0.4	0.4
	1/4	-1.7		0.8	0.5
	1/3	-1		1.5	0.67
	1/2	2		4.5	1
	2/3	4.3		6.8	1.3
	3/4	5.5		8.0	1.5
	4/5	6.2		8.7	1.6
16QAM	1/2	7.9	3	10.9	2
	2/3	11.3		14.3	2.66
	3/4	12.2		15.2	3
	4/5	12.8		15.8	3.2
64QAM	2/3	15.3	4	19.3	4
	3/4	17.5		21.5	4.5
	4/5	18.6		22.6	4.8

one of three predefined SNR *classes*, which correspond to *poor*, *average*, and *good* mean SNR. The designated SNR for different classes are chosen in a manner that the mean achievable rates for *poor*, *average*, and *good* users are 20%, 50%, and 80% of the maximum transmission rate achievable in the system, respectively. With the thresholds and MCS values reported in Table 1, the designated SNR values are 7 dB, 16 dB and 23 dB, respectively for *poor*, *average*, and *good* users. Note also that using non-homogeneous channel qualities allows us to evaluate the long-term system fairness under different (opportunistic) scheduling mechanisms.

Clustering. We assume that mobiles can form clusters and receive downlink traffic through the *cluster head*, i.e., the node enabled to exchange data with the base station (see Figure 1). Note that each node is a potential candidate to act as cluster head, and cluster heads are selected on a per-frame basis. A cluster consists of several mobiles that form an ad-hoc wireless network and that are in radio range with each other, e.g., a WiFi clique. Hence, all intra-cluster communications take place over the ad-hoc wireless network. After cluster formation, the base station is aware of clustering decisions. Thus, whenever a packet is destined to a cluster member, the base station simply sends it to the cluster head, which maximizes the throughput at that epoch. Thereby, the base station schedules clusters as if they were regular users. From a modeling perspective, a cluster can be considered as a user whose SNR is the highest of the SNR values of cluster members. As for intra-cluster resource sharing, unless otherwise specified, we assume that the extra throughput gained from clustering is equally distributed among users.

Example of cluster impact. The incentive behind clustering can be easily observed by comparing the user's channel state probabilities, as defined in Eq. (3). Indeed, the primary effect of clustering is to increase the probability of transmitting with higher MCS's, due to the opportunistic selection of the cluster head. Figure 2 illustrates the impact of clustering on channel state probabilities π_k. Here, a cluster is considered as a user whose channel state is the highest of the channel states among cluster members. As shown in the figure, cluster formation highly boosts the transmission rate of *poor* and *average* users, while the impact on *good* users is less significant. Nonetheless, we will show that with our proposal the throughput of *good* users can improve as well by a non-negligible quantity. Therefore, *good users* can

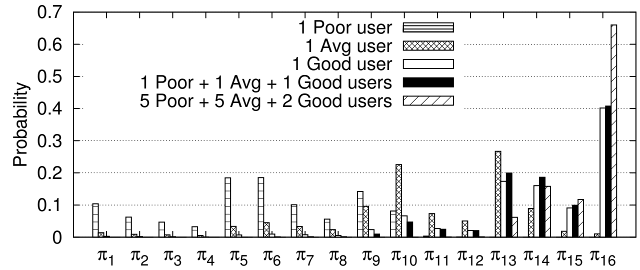


Figure 2: Impact of clustering on channel state probabilities.

be incentivized to help users with lower channel qualities. In practice, *good* users can be encouraged to participate in clustering by receiving an extra quota for the portion of traffic that they forward for others users. A more detailed discussion on clustering incentives and costs is provided later in Section 7.

4. CLUSTER-BASED SCHEDULING

In the following, we introduce a new class of schedulers, namely *cluster-based schedulers*. We consider the case in which the base station schedules N_c clusters instead of N normal users. This means that the base station decides which cluster has to be served, and then transmissions will be managed by the current cluster head.

Defining X_n as the SNR of cluster n, we have:

$$X_n \;=\; \max\{C_j, j : U_j \in CL_n\},\; n \in \{1 \dots N_c\}. \qquad (4)$$

The CDF of X_n can be readily computed considering that the random variables C_j are all independent:

$$F_{X_n}(z) = \prod_{j:U_j \in CL_n} F_j(z) = u(z) \prod_{j:U_j \in CL_n} \left(1 - e^{-\frac{z}{\gamma_j}}\right); \qquad (5)$$

the corresponding pdf, $f_{X_n}(z)$, can be obtained by derivation from $F_{X_n}(z)$. The adopted MCS, for each transmission, only depends on the instantaneous SNR of the best channel in the scheduled cluster, i.e., it only depends on X_n at the scheduling epoch:

$$\pi_k^{(CL_n)} = \int_{th_k}^{th_{k+1}} f_{X_n}(z)dz. \qquad (6)$$

CL(*WRR*)**: Cluster Weighted Round Robin.** Let us call N_n the number of members of cluster CL_n. With *CL*(*WRR*), clusters are scheduled according to a WRR mechanism, with each cluster CL_n having weight w_n. For instance, using $w_n = N_n/N$ would correspond to assign each cluster an amount of airtime proportional to the number of members in the cluster.

In this system, the per-cluster scheduling probability is exactly w_n, $n \in \{1 \dots N_c\}$, while the average symbol rate depends only on the selected MCS, which, in turn, only depends on the selected cluster. Since resources are allocated in WRR, the average throughput of each cluster only depends on the cluster weight and the probability distribution of the cluster's MCS:

$$E[T_{CL_n}] \;=\; w_n \sum_{k=1}^{M} \pi_k^{(CL_n)} b_k S_{tot}, \;\; n \in \{1 \dots N_c\}. \qquad (7)$$

CL(*MR*)**: MaxRate Between Clusters.** With this scheme, each frame is allotted to the cluster with the user experiencing the best SNR. The probability that cluster CL_n

is scheduled with the k-th MCS is:

$$P(CL_n|k) = Pr(X_n > Y_n|MCS_{CL_n} = k), \tag{8}$$

where $Y_n = \max\{Cj\}, j \notin CL_n$, and X_n represents the SNR of cluster n (see Eq. (4)). The CDF of X_n, subject to $MCS_{CL_n} = k$, is:

$$F_{X_n}(z|MCS_{CL_n} = k) = \\ = u(z - th_k)\frac{F_{X_n}(\min(z, th_{k+1})) - F_{X_n}(th_k)}{\pi_k^{(CL_n)}}. \tag{9}$$

Now we can rewrite (8) using the total probability formula, considering that X_n and Y_n are independent:

$$P(CL_n|k) = Pr(X_n > Y_n|MCS_{CL_n} = k) \\ = \int_0^\infty [1 - F_{X_n}(z|MCS_{CL_n} = k)]\, f_{Y_n}(z)dz. \tag{10}$$

Finally, the average per-cluster throughput can be computed using the following formula:

$$E[T_{CL_n}] = \sum_{k=1}^{M} P(CL_n|k)\pi_k^{(CL_n)} b_k S_{tot}. \tag{11}$$

5. POWER CONSUMPTION ANALYSIS

We derive the power consumption of mobiles from the empirical power models proposed for LTE and WiFi in [11] and [8]. The proposed models are desirable for two virtues: (i) they include the baseline power required to keep the interface up and running; (ii) they account for the variability of power consumption with transmission rate, and differentiate transmission from reception.

LTE consumption. Based on [11], the downlink power consumption of user i in the cellular network is:

$$W_{lte}^{(i)} = \beta_{lte} + \alpha_{rx} R_{rx}^{(i,lte)}, \tag{12}$$

where β_{lte} is the baseline power, α_{rx} is the downlink power consumption per $Mbps$, and $R_{rx}^{(i,lte)}$ is the average data rate received by user i over the LTE interface:

$$R_{rx}^{(i,\, lte)} = \sum_{k=1}^{M} \pi_k^{(i)} b_k t_k^{(i)}, \tag{13}$$

where $b_k t_k^{(i)}$ is the average data rate of user i when its MSC is k, b_k is the per-symbol throughput achieved with that MCS, and the value of $t_k^{(i)}$ represents resources allocated to user i, which depend on the scheduling policy adopted (see Section 4).

In particular, under cluster-based schedulers, only cluster heads receive packets from the base station, so that the power consumption of user i strongly depends on the probability of that user being the cluster head when it can use MCS k, namely $P_h^{(i|k)}$. Therefore, the fraction of resources allotted to user i when it is the cluster head and can use MCS k is:

$$t_k^{(i)} = S_{tot} P_h^{(i|k)} P(CL_n|k),\ i \in CL_n,\ n = 1 \ldots N_c, \tag{14}$$

where S_{tot} is the total number of symbols per frame transmitted by the base station, and $P(CL_n|k)$ is the probability that cluster CL_n is scheduled with the k-th MCS. Note that $P(CL_n|k)$ is equal to w_n for CL(WRR), while it is given by Eq. (8) for CL(MR).

Recalling that the mechanism for selecting *cluster heads* resembles a MaxRate scheduling policy, i.e., the cluster head is selected as the user i with the best channel within its cluster CL_n, we can compute $P_h^{(i|k)}$ with an expression similar to Eq. (10):

$$P_h^{(i|k)} = \int_0^\infty [1 - F_i(z|MCS_i = k)]\, f_{Y_i}(z)dz, \tag{15}$$
$$Y_i = \max_{j \in CL_n \setminus \{i\}} \{C_j\}, \quad \forall i \in CL_n.$$

As a remark, the total cluster throughput $E[T_{CL_n}]$, which is defined as the sum of cluster members' throughputs $E[T_i]$, is also equal to the sum of the data rates received by all cluster members over the LTE interface:

$$\sum_{i \in CL_n} E[T_i] = E[T_{CL_n}] = \sum_{i \in CL_n} R_{rx}^{(i,\, lte)}. \tag{16}$$

WiFi consumption. We use the accurate model proposed in [8], which accounts for the power required for packet processing as well as for transmission. The resulting power consumption of WiFi interfaces is:

$$W_{wifi}^{(i)} = \beta_{wifi} + \zeta_{tx}\tau_{tx} + \zeta_{rx}\tau_{rx} + \kappa_{tx}\lambda_{tx} + \kappa_{rx}\lambda_{rx}, \tag{17}$$

where β_{wifi} is the baseline power of WiFi, ζ_{tx} and ζ_{rx} represent the power consumptions due to transmission and reception, respectively; τ_{tx} and τ_{rx} are the fractions of time spent in transmission and reception, respectively; κ_{tx} and κ_{rx} are the power consumptions due to packet processing in transmission and reception, respectively; eventually, λ_{tx} and λ_{rx} are the packet rates, respectively in transmission and reception.

The WiFi power related parameters introduced in Eq. (17) are computed as follows: τ_{tx} is the ratio between the transmission rate over the WiFi interface and the achievable rate of the WiFi connection, i.e., for user i, we have $\tau_{tx}^{(i)} = R_{tx}^{(i,wifi)}/R_{wifi}$. Similarly, user i transmits $\lambda_{tx}^{(i,wifi)}$ packets per second over WiFi.

In order to compute $R_{tx}^{(i,wifi)}$ and $\lambda_{tx}^{(i,wifi)}$, we need to estimate the fraction of traffic which is received by user i over the LTE interface and is then forwarded to other users over the WiFi interface. To this aim, let us define δ_i as the ratio between the user's throughout $E[T_i]$ and the total cluster throughout $E[T_{CL_n}]$, and assume that the traffic distribution over a scheduling interval is the same as the long term distribution of throughputs within the cluster CL_n. Therefore, we have the following expressions for δ_i and $R_{tx}^{(i,wifi)}$:

$$\delta_i = \frac{E[T_i]}{E[T_{CL_n}]} = \frac{E[T_i]}{\sum_{j \in CL_n} R_{rx}^{(j,\, lte)}}, \tag{18}$$

$$R_{tx}^{(i,wifi)} = (1 - \delta_i) R_{rx}^{(i,\, lte)} = (1 - \delta_i) \cdot \sum_{k=0}^{M} \pi_k^{(i)} b_k t_k^{(i)}. \tag{19}$$

Note that $E[T_i]$ can be expressed in terms of throughput without clustering plus clustering gain, as follows:

$$E[T_i] = E[T_i^{\text{no cluster}}] + E[T_i^{\text{cluster gain}}]. \tag{20}$$

Therefore, both $E[T_i]$ and δ_i depend on the payoff allocation method adopted within the cluster, which will be further elaborated in Section 7.

As for $\lambda_{tx}^{(i,wifi)}$, we compute the average number of packets per second for the average packet size L_p:

$$\lambda_{tx}^{(i,wifi)} = \frac{R_{tx}^{(i,wifi)}}{L_p}. \tag{21}$$

Similarly, to derive $\tau_{rx}^{(i)} = R_{rx}^{(i,wifi)}/R_{wifi}$, we need an expression for the traffic received over the WiFi interface by user i in cluster CL_n when it is not the cluster head. To this aim, consider that $R_{rx}^{(i,wifi)}$ is a fraction δ_i of all the traffic delivered by LTE to the other cluster members:

$$R_{rx}^{(i,wifi)} = \delta_i \cdot \sum_{j \in CL_n \setminus \{i\}} R_{rx}^{(j,lte)}. \tag{22}$$

The corresponding value of $\lambda_{rx}^{(i)}$ is as follows:

$$\lambda_{rx}^{(i,wifi)} = \frac{R_{rx}^{(i,wifi)}}{L_p}. \tag{23}$$

Eventually, assuming that the achievable WiFi rate in the ad-hoc network is $R_{wifi} > \sum_{i \in CL_n} R_{tx}^{(i,wifi)}$, the power consumed by user i to transmit over the WiFi interface is computed by plugging $R_{tx}^{(i,wifi)}$, $R_{rx}^{(i,wifi)}$, R_{wifi}, $\lambda_{tx}^{(i,wifi)}$, and $\lambda_{rx}^{(i,wifi)}$ in (17):

$$W_{wifi}^{(i)} = \beta_{wifi} + \left(\zeta_{tx} + \frac{\kappa_{tx}}{L_p}\right) \frac{1-\delta_i}{R_{wifi}} \sum_{k=1}^{M} \pi_k^{(i)} b_k t_k^{(i)}$$
$$+ \left(\zeta_{rx} + \frac{\kappa_{rx}}{L_p}\right) \frac{\delta_i}{R_{wifi}} \sum_{j \in CL_n \setminus \{i\}} \sum_{k=1}^{M} \pi_k^{(j)} b_k t_k^{(j)}. \tag{24}$$

Total power consumption. Putting together the results for LTE and WiFi consumptions, the resulting total power consumption of a clustered user is as follows:

$$W_{tot}^{(i)} = \beta_{lte} + \beta_{wifi} + \left[\alpha_{rx} + \frac{1-\delta_i}{R_{wifi}} \left(\zeta_{tx} + \frac{\kappa_{tx}}{L_p}\right)\right]$$
$$\cdot \sum_{k=1}^{M} \pi_k^{(i)} b_k t_k^{(i)} + \left(\zeta_{rx} + \frac{\kappa_{rx}}{L_p}\right) \frac{\delta_i}{R_{wifi}} \sum_{j \in CL_n \setminus \{i\}} \sum_{k=1}^{M} \pi_k^{(j)} b_k t_k^{(j)}. \tag{25}$$

The first two terms in Eq. (25) represent the fixed power consumption due to activating the LTE and WiFi interfaces, respectively. The third term is the power spent to manage the overall cluster traffic when node i is the cluster head, and the last term is the power spent over the WiFi interface to receive packets from the cluster head when i is not the cluster head.

Energy efficiency. The overall power consumption with clusters increases due to the increased achievable throughput, which is fully utilized under the fully backlogged assumption we use. However, as we will show in Sections 6 and 7, clustering causes little increase in power consumption while enabling a much more efficient energy utilization. To evaluate the beneficial impact of clustering under fully backlogged traffic assumption, we use as metric the energy efficiency, i.e., the amount of data (bits) that can be transferred to the final user per energy unit (Joule), e.g., for user i, the energy efficiency is given by $E[T_i]/W_{tot}^{(i)}$.

6. PERFORMANCE OF STATIC CLUSTERS

The goal of this section is to provide a preliminary evaluation of throughput, fairness, and energy efficiency achievable with our clustering architecture and scheduling. For the sake of clarity, we consider a static scenario with a few clusters of different sizes, formed by users with heterogeneous average SNR.

We present results obtained via numerical simulation of the model presented in Sections 3 to 5. When we refer to the CL(WRR) scheduler, for simplicity we fix the WRR weights as proportional to cluster sizes, i.e., $w_n = N_n/N$. For power computation we use an average packet size $L_p = 1000\ B$ and average WiFi net rate $R_{wifi} = 28\ Mbps$. The evaluation scenario includes three clusters with fixed sizes, as shown in Figure 3. In this scenario, clusters C1, C2, and C3 have 5, 10 and 15 users, respectively. In each experiment, the SNR class of each user is chosen as *poor*, *average*, or *good* with the same probability. We repeat each experiment 2000 times, and we plot the average and standard deviation of the achieved results. The values which are used for power related parameters are reported in Table 2 and are derived from [8, 11].

Figure 3: Evaluation topology for static clusters.

We benchmark our proposal against Round Robin (RR) and Proportional Fair (PF) schedulers. With RR, each user is scheduled for an equal amount of resources (e.g., *equal time* or *equal rate*) [9]. Here we consider equal time rather than equal rate, since the throughput of equal rate decreases drastically in the presence of *poor* users. With RR, the throughput of each user is simply computed as $E[T_i] = \frac{1}{N}\sum_{k=1}^{M} \pi_k^{(i)} b_k S_{tot}, \forall i \in \{1 \dots N\}$. Since PF is an opportunistic scheduler that is already implemented in 3G systems, it is a good benchmark to evaluate our proposal. PF is a priority scheduler that uses the ratio of feasible data rate to average throughput at time t (i.e., $R_i(t)/\mu_i(t)$, $\forall i \in \{1 \dots N\}$) as priority function. In every frame the users with top n_s priorities are scheduled. Since no accurate model is available, in this paper the performance results of PF are obtained from a home grown C++ simulator, in which the averaging window is 100 frames, and only one user is scheduled in each frame..

Figure 4 illustrates the average user performance under different schedulers, for the three SNR classes we adopted. In Figure 4(a), we observe the throughput of a user when it belongs to the *poor*, *average* or *good* SNR class. Users receive the lowest throughput under RR because they are scheduled irrespective to their channel quality. Instead, PF has remarkably better performance in terms of throughput, due to its opportunistic nature. Nevertheless, PF is still considerably outperformed by cluster schedulers for all SNR classes. Between cluster schedulers, CL(MR) provides *good* users with higher throughput in comparison to CL(WRR), because the former is a pure opportunistic scheduler. Figure 4(b) shows that throughput gained with clustering produces increased power consumption of users. Specifically, while this power increment is negligible for *poor* and *average* users, *good* users incur slightly higher power increment ($\sim 180mW$, i.e., +14% with respect to RR) because they will play the role of cluster head more often. However, this power increase is limited and is not proportional to the

Table 2: Parameters used in the power model

LTE		WiFi				
β_{lte}	α_{rx}	β_{wifi}	ζ_{tx}	ζ_{rx}	κ_{tx}	κ_{rx}
1.29 $[W]$	51.97 $[nW/bps]$	0.28 $[W]$	0.46 $[W]$	0.44 $[W]$	0.11 $[mJ]$	0.09 $[mJ]$

throughput increase. Indeed, Figure 4(c) shows that the energy efficiency of all classes of users sensibly increases with clustering. CL(WRR) achieves the best energy efficiency for *poor* users, CL(MR) favors *good* users.

To show the impact of cluster sizes, we report in Figure 5(a) the aggregate throughput of C1, C2, and C3. For comparison, we also report in the figure the aggregate throughput achieved by cluster members if they where scheduled according to RR or PF. Therefore, results with RR, PF scale linearly with the cluster size. Similarly, CL(WRR) shows linearity, while the high variability of results for CL(MR) does not allow us to confirm or reject the hypothesis that CL(MR) scales linearly. This behavior is due to the fact that CL(MR), differently from CL(WRR), does not guarantee any minimum airtime to any cluster, so that clusters not including *good* user will receive little throughput.

Figure 5(b) sheds light on the aggregate throughout performance achieved by the considered scheduling schemes. The figure reports results for two cases: (i) all users have the same probability of being *poor*, *average* or *good*, and (ii) the probability of being *poor* is much higher than one of being *average*, which in turn is higher than the probability to be *good*. The figure also reports the upper bound for the downlink throughput. The aggregate throughput of RR and PF is outperformed by cluster-based schedulers. CL(MR) practically hits the upper bound, while the worst case for CL(WRR), i.e., when the number of *poor* users is predominant, outperforms RR and PF under their best performance.

So far, CL(MR) outperforms all other schedulers. However, considering fairness, CL(MR) is always the most unfair, especially when more *poor* users are present, while CL(WRR) performs like PF or better, see Figure 5(c).

To summarize the results reported in this section, we have observed that the clustering proposal not only increases the throughput and the energy efficiency, but it can also increase the fairness level. In particular, CL(WRR) achieves similar throughput and energy efficiency results as CL(MR), but it is much fairer. Therefore, the advantage of using CL(WRR) is fourfold: (i) it offers the possibility to gain a high throughput with respect to legacy RR and PF schedulers; (ii) it allows each cluster to exploit the clustering gain proportionally to its size; (iii) it provides nearly perfect fairness among users; (iv) energy efficiency is increased with respect to RR and PF.

7. CLUSTER FORMATION: A GAME THEORY APPROACH

The goal of this section is to provide a simple model for the cluster formation process, and shed light on the impact of clustering when users experience non-stationary channel qualities, e.g., due to mobility. Here, we consider the realistic case of users that can choose to join or leave a cluster, depending on their revenue expressed in terms of energy efficiency. We only consider the case of CL(WRR) cluster scheduling, which is the method that yields the best tradeoff between throughput, energy efficiency, and fairness (see Section 6). Since clustering yields increased throughput, we also discuss the impact of intra-cluster throughput distribution on fairness and energy efficiency.

The cluster formation in our proposed architecture can be modeled using *coalitional game theory* [21]. Coalitional games are a class of games studied in cooperative game theory and a coalition is simply a group of entities that agree on cooperating to increase their *social welfare*. In our case, a coalition is a group of users that agree on forming a cluster and acting as a single mobile user. Here, we use a simple dynamic coalition formation game that accounts for the basic cost to form a cluster—namely the power consumption—and for the basic clustering revenue—namely the increased throughput. Indeed, to determine which users should form a cluster and which users are better off alone, we use the energy efficiency as metric for the user's welfare.

7.1 Definition of the game

In the following, $U=\{u_1,\ldots,u_N\}$ denotes the set of users in the network and $S=\{S_1,\ldots,S_l\}$ is a partition of U, i.e., $\bigcup_{i=1}^{l} S_n=U$ and $S_n\cap S_j=\emptyset$ if $n\neq j$. The utility function $\nu(.)$ defines the value of a cluster S_n as:

$$\nu(S_n)=\begin{cases}\sum\limits_{k=1}^{M}\pi_k^{(S_n)}b_k t_j^{(S_n)} & \text{if } d_{S_n}\le d_{\mathrm{m}} \ \&\ \eta_i^{(S_n)}\ge\eta_i,\forall i\in S_n;\\ 0 & \text{otherwise;}\end{cases} \tag{26}$$

where $t_j^{(S_n)}$ is proportional to the resources allocated to the cluster S_n with the selected scheduler, see Eq. (14); d_{S_n} and d_{m} are the distances between the two farthest users in cluster S_n, and the maximum allowable distance among cluster members, respectively; $\eta_i^{(S_n)}$ and η_i are the energy efficiencies of user i when it joins cluster S_n and when it is not clustered, respectively. In particular, d_{m} accounts for the WiFi transmission range, and can be set to guarantee that any user inside a cluster can directly reach the rest of the cluster members. The constraint on the energy efficiency guarantees that users form a cluster only if energy efficiency increases.

7.2 Cluster formation algorithm

The problem of finding optimal coalitions is NP-complete because it requires evaluating all possible partitions of the set of users U in the network [19]. Obviously, the existing base stations with limited computational resources are not able to handle an NP-complete problem involving a few tens of users.

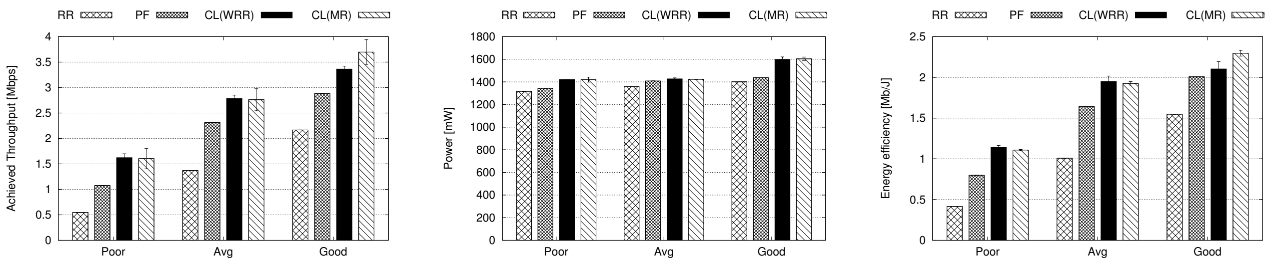
Hence, we adapt the simple *merge and split* algorithm to solve the coalition formation problem with low complexity [22]. Although merge and split is a trivial method for dynamic cluster formation, it was shown to be a good alternative when computational overhead is of concern [3,20]. The merge and split rules are defined as follows: merge any set $\{S_{a_1},..,S_{a_k}\}$ into a unique coalition (i.e., cluster), if the following inequality holds:

$$\sum_{i=1}^{k}\nu(S_{a_i})<\nu\left(\cup_{i=1}^{k}S_{a_i}\right). \tag{27}$$

Similarly, if the previous inequality does not hold for a coalition that can be described as $\cup_{i=1}^{k}S_{a_i}$, then split it into its components (i.e., split the big cluster into smaller clusters). The authors of [2] have proven that the merge and split algorithm terminates and the result is D_{hp}-stable, i.e., the system reaches a state in which there are no cluster members willing to perform a merge or a split operation.

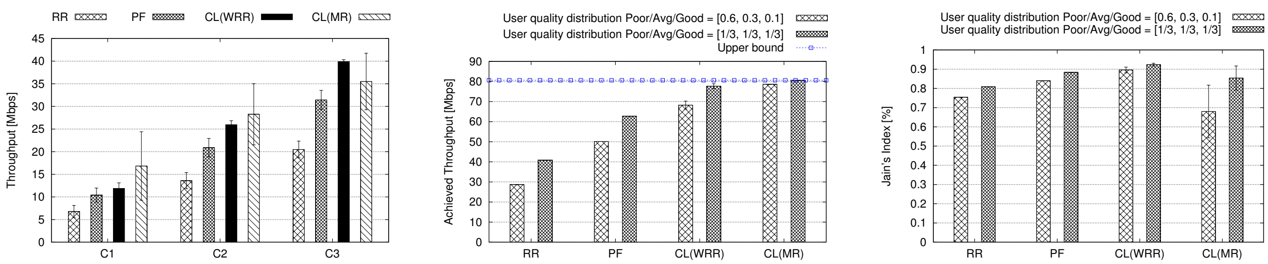
7.3 Payoff allocation

So far we have not investigated on how resources should be distributed among cluster members. We have generically



(a) Per-class per-user throughput (b) Per-class per-user power (c) Per-class per-user efficiency

Figure 4: Per-user per-class throughput and cost with equal user distribution (See Figure 3).



(a) Per-cluster throughput (b) Aggregate cell throughput (c) Jain's fairness indexes

Figure 5: Per-cluster and aggregate performance.

assumed that cluster revenue (i.e., the throughput gain) was equally shared by cluster members.

More in general, the *payoff* for a user is defined as the amount of throughput which is allotted to a member from the total cluster throughput. Formally, let $G \in S$ be a cluster of size $|G|$, and $\bar{x} = \{x_1, \ldots, x_{|G|}\}$ the payoff vector of members of G. A payoff vector is called *cost efficient* if $\sum_{i \in G} x_i = \nu(G)$ [20]. Of course, we are only interested in cost efficient payoff vectors.

As for the payoff distribution method, we chose to compare two mechanisms proposed in the literature, namely equal share and weighted share [20, 24]. These two mechanisms are simple and would allow us to easily illustrate how clustering can be made attractive for all classes of users.

Equal share. This is the simplest approach, in which the clustering gain is equally divided among members. The cost efficient payoff distribution used under equal share is formally expressed as follows:

$$x_i = \frac{\nu(G) - \sum_{j \in G} \nu(\{j\})}{|G|} + \nu(\{i\}), \quad i \in G. \tag{28}$$

Weighted share. Here, the cost efficient payoff distribution is computed based on the positive weights ω_i:

$$x_i = \frac{\omega_i}{\sum\limits_{j \in G} \omega_j} \cdot \left(\nu(G) - \sum_{j \in G} \nu(\{j\})\right) + \nu(\{i\}),\ i \in G. \tag{29}$$

As shown in Section 4, the clustering gain is mainly due to the presence of *good* users, whereas the channel state probability distribution of a cluster does not dramatically improve with the addition of a *poor* user (see Figure 2). Hence, equal share may not strongly motivate users with good channel quality to cluster with users with poor channel quality. In contrast, adjusting ω_i in Eq. (29), we can make sure that users with better channel quality receive enough incentive to cluster. Specifically, in our numerical simulation, we use values of ω_i which are proportional to the user non-cooperative throughput, so that the clustering gain share of each user is proportional to the user's throughput achieved without clustering.

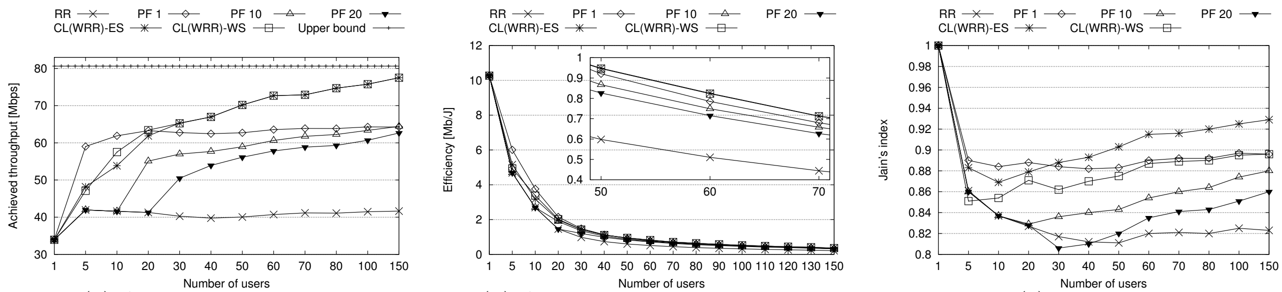
7.4 Numerical evaluation

So far, we have shown that clustering gives substantial throughput increment. We also defined the rules to form clusters and distribute the resources within the cluster. Our proposal might not be attractive for *good* users when the channel quality diversity of cluster members is high and most of the cellular transmissions are carried on by *good* users. Hence, power consumption of these users would increase. Now the question is: *Is there always enough incentive for the good users to participate in clustering?*

To answer this question, we simulate a network with variable number of users (from 1 to 150) with varying channel quality, randomly placed in a circular-shaped cell with 500 m diameter. The SNR class of a user is selected at random with a probability distribution that changes according to the distance from the base station. In order to have a more realistic scenario, we assume that mobile users move with an average pedestrian speed between 0 and 5 km/h [16]. A cluster can have a maximum diameter $d_{\mathrm{m}} = 150\ m$. Like in the static scenario, we use $L_p = 1000\ B$ and $R_{wifi} = 28\ Mbps$.

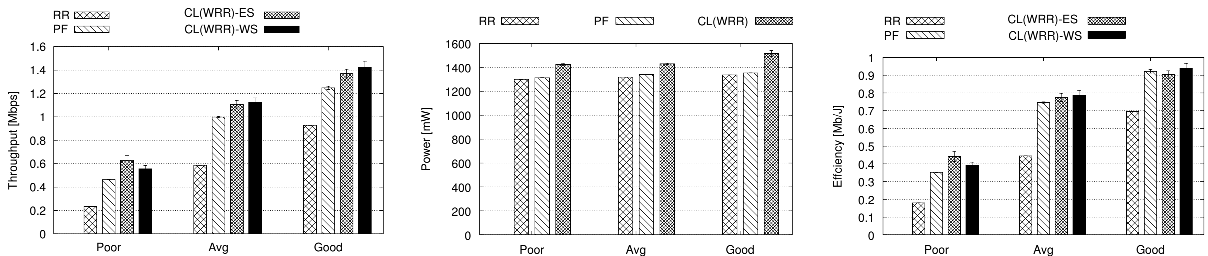
Figure 6 illustrates the performance metrics for different user population sizes. In the figure, we report results achieved with RR, PF, CL(WRR) with equal share, namely CL(WRR)-ES, and CL(WRR) with weighted share, namely CL(WRR)-WS. Additionally, we report results for PF when $n \geq 1$ users are scheduled per frame (PF n in the figure). We report this comparison since user-based schedulers allocate multiple users per frame, and it is indeed common to schedule tens of users per frame, even when opportunistic schedulers are adopted. However, RR and CL(WRR) are not affected by the number of user scheduled per frame, due to the assumption that user's channels are independent and stationary (and so are the channels of cluster heads).

In Figure 6(a), we can observe that the clustering gain raises with the number of users in the system, and, as soon as about 30 users are present, CL(WRR) achieves the high-



(a) Average throughput (b) Average per-user efficiency (c) Fairness

Figure 6: Throughput, efficiency and fairness under different scheduling mechanisms.



(a) Per-class per-user throughput (b) Per-class per-user power (c) Per-class per-user efficiency

Figure 7: Per-user per-class throughput, efficiency and power with 70 users.

est aggregate network throughput, which approaches the upper bound with a reasonable cell population size of 150 users. CL(WRR)-ES and CL(WRR)-WS do not sensibly differ over the aggregate network throughout. The throughput of PF reduces significantly as the number of scheduled users per frame increases. However, all PF curves converge, for high number of users, to a value well below the throughput of CL(WRR). In Figure 6(b), we can observe that the energy efficiency of CL(WRR)-WS is the best, and CL(WRR)-ES and PF 1 achieve similar efficiencies. Overall, the energy efficiency decreases with the number of users, due to the fact that each additional user incurs a minimum cost due to activating the network interfaces, while the cell capacity is upper bounded. However, e.g., with 70 users, the efficiency of CL(WRR) is higher than RR and PF 20 by $\sim 60\%$ and $\sim 15\%$, respectively. As regards fairness, Figure 6(c) shows that CL(WRR)-ES provides the highest fairness level, while CL(WRR)-WS achieves results comparable to the best results achieved by PF. The fairness improvement due to clustering with respect to RR and PF 10 or PF 20, which are realistic figures for PF performance, is remarkable.

We now zoom into the performance figures achieved by the three SNR classes. To this aim, in Figure 7 we fix the number of users to 70, which yields a reasonable cell size with reasonable opportunistic gain for PF and CL(WRR) schedulers. However, we verified that results for user populations larger than 30 are similar. In terms of throughput, Figure 7(a) shows that all users obtain the highest throughput under clustering schemes. In particular *good* users achieve more throughput with CL(WRR)-WS than CL(WRR)-ES, i.e., *good* users are more incentivized to form clusters with CL(WRR)-WS, at the expenses of *poor* users. In all cases, the incentive for clustering is high, especially if compared to RR throughputs. As for the power consumed by users, Figure 7(b) highlights that extra consumption due to clustering is quite limited, i.e., less than 15%, that is 200 mW in the worst case. However, Figure 7(c) confirms the superiority of cluster-based schemes also in terms of energy efficiency. In particular, *poor* users experience the best efficiency improvement, while *good* users experience similar performance as for the case of ideal PF scheduling with one user scheduled per frame.

8. CONCLUSIONS

In this paper, we proposed and modeled novel cluster-based scheduling schemes, which significantly improve the throughput of a cellular network (up to $\sim 50\%$), boost energy efficiency (+30% with realistic user-populations) and achieve fairness levels higher than PF. To show the impact of our proposal, we proposed a new model for the power consumption of dual-radio mobile devices using LTE and WiFi networking capabilities at the same time. Additionally, we used a game theory approach to model the cluster formation process in realistic network scenarios, in which throughput and fairness are boosted via user cooperation.

Our results show that CL(WRR)—which assigns resources to the clusters in weighted round robin and selects clusters heads opportunistically—achieves throughputs close to the maximum achievable. Furthermore, CL(WRR) outperform other opportunistic schedulers in terms of achieved fairness. Our numerical simulations confirmed that the compound impact of opportunistic scheduling and D2D techniques is beneficial for all users irrespective of their channel qualities, and dramatically improves energy-efficiency.

9. ACKNOWLEDGEMENTS

The research leading to these results was supported by the CROWD project, under the European Union's Seventh Framework Programme (grant agreement n° 318115).

This publication was supported in part by the Comunidad de Madrid MEDIANET Project, Ref.: S2009/TIC-1468.

10. REFERENCES

[1] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting. Scheduling in a queuing system with asynchronously varying service rates. *Probability in the Engineering and Informational Sciences*, 18:191–217, Apr. 2004.

[2] K. Apt and T. Radzik. Stable partitions in coalitional games. *Arxiv preprint cs/0605132*, 2006.

[3] K. Apt and S. Witzel. A generic approach to coalition formation. 2006.

[4] S. Bandyopadhyay and E. Coyle. An energy efficient hierarchical clustering algorithm for wireless sensor networks. In *Proceedings of IEEE INFOCOM*, volume 3, pages 1713–1723, 2003.

[5] W. Choi and J. Andrews. The capacity gain from base station cooperative scheduling in a MIMO DPC cellular system. In *IEEE International Symposium on Information Theory*, pages 1224–1228, July 2006.

[6] M. Dohler et al. *Virtual antenna arrays*. PhD thesis, University of London, 2004.

[7] F. H. Fitzek, M. Katz, and Q. Zhang. Cellular controlled short-range communication for cooperative P2P networking. *Wireless Personal Communications*, 48(1):141–155, 2009.

[8] A. Garcia-Saavedra, P. Serrano, A. Banchs, and G. Bianchi. Energy consumption anatomy of 802.11 devices and its implication on modeling and design. In *Proceedings of ACM CoNEXT*, pages 169–180, New York, NY, USA, Dec. 2012.

[9] E. Hahne. Round-robin scheduling for max-min fairness in data networks. *IEEE Journal on Selected Areas in Communications*, 9(7):1024–1039, 1991.

[10] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In *Proceedings of System Sciences*, pages 10–pp, 2000.

[11] J. Huang, F. Qian, A. Gerber, Z. Mao, S. Sen, and O. Spatscheck. A close examination of performance and power characteristics of 4G LTE networks. In *Proceedings of ACM MobiSys*, pages 225–238, June 2012.

[12] E. Jon Vergara and N. Nadjm-Tehrani. Watts2share: Energy-aware traffic consolidation. In *Proceedings of IEEE GreenCom*, 2013.

[13] K. Khalil, M. Karaca, O. Ercetin, and E. Ekici. Optimal scheduling in cooperate-to-join cognitive radio networks. In *Proceedings of IEEE INFOCOM*, pages 3002–3010, Apr. 2011.

[14] R. Knopp and P. Humblet. Information capacity and power control in single-cell multiuser communications. In *Proceedings of IEEE ICC*, volume 1, pages 331–335 vol.1, June 1995.

[15] L. Le and E. Hossain. Multihop cellular networks: Potential gains, research challenges, and a resource allocation framework. *IEEE Communications Magazine*, 45(9):66–73, 2007.

[16] S. Lee, K. Kim, K. Hong, D. Griffith, Y. Kim, and N. Golmie. A probabilistic call admission control algorithm for WLAN in heterogeneous wireless environment. *IEEE Transactions on Wireless Communications*, 8(4):1672–1676, 2009.

[17] L. Lei, Z. Zhong, C. Lin, and X. Shen. Operator controlled device-to-device communications in LTE-advanced networks. *IEEE Wireless Communications*, 19(3):96–104, 2012.

[18] C. Lin and M. Gerla. Adaptive clustering for mobile wireless networks. *IEEE Journal on Selected Areas in Communications*, 15(7):1265–1275, 1997.

[19] D. Ray. *A game-theoretic perspective on coalition formation*. OUP Oxford, 2007.

[20] W. Saad, Z. Han, M. Debbah, and A. Hjorungnes. A distributed merge and split algorithm for fair cooperation in wireless networks. In *Proceedings of IEEE ICC Workshops*, pages 311–315, 2008.

[21] W. Saad, Z. Han, M. Debbah, A. Hjorungnes, and T. Basar. Coalitional game theory for communication networks. *IEEE Signal Processing Magazine*, 26(5):77–97, 2009.

[22] W. Saad, Z. Han, M. Debbah, A. Hjorungnes, and T. Basar. Coalitional games for distributed collaborative spectrum sensing in cognitive radio networks. In *Proceedings of IEEE INFOCOM*, pages 2114–2122, 2009.

[23] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity. Part I. System description. *IEEE Transactions on Communications*, 51(11):1927–1938, 2003.

[24] M. Sereno. Cooperative game theory framework for energy efficient policies in wireless networks. In *Proceedings of ACM e-Energy*, pages 1–9, May 2012.

[25] S. Sesia, I. Toufik, and M. Baker. *LTE–the UMTS long term evolution: from theory to practice*. Wiley, 2011.

[26] S. Shakkottai, R. Srikant, and A. L. Stolyar. Pathwise optimality of the exponential scheduling rule for wireless channels. *Advances in Applied Probability*, 36(4):1021–1045, 2004.

[27] N. Sharma and L. Ozarow. A study of opportunism for multiple-antenna systems. *IEEE Transactions on Information Theory*, 51(5):1804–1814, May 2005.

[28] Third Generation Partnership Project (3GPP). Physical layer procedures (Release 10) for Evolved Universal Terrestrial Radio Access (E-UTRA). 3GPP TR 36.213 v 10.5.0, Mar. 2012.

[29] P. Viswanath, D. Tse, and R. Laroia. Opportunistic beamforming using dumb antennas. In *IEEE Transactions on Information Theory*, page 449, June 2002.

[30] Q. Wang and B. Rengarajan. Recouping opportunistic gain in dense base station layouts through energy-aware user cooperation. In *Proceedings of IEEE WoWMoM*, pages 1–9, 2013.

[31] Wi-Fi Alliance Specification. Wi-Fi Peer-to-Peer (P2P) Specification v1.1, 2011.

[32] H. Wu, C. Qiao, S. De, and O. Tonguz. Integrated cellular and ad hoc relaying systems: iCAR. *IEEE Journal on Selected Areas in Communications*, 19(10):2105–2115, 2001.

[33] J. Yang, Z. Yifan, W. Ying, and Z. Ping. Average rate updating mechanism in proportional fair scheduler for hdr. In *Proceedings of IEEE GLOBECOM*, volume 6, pages 3464–3466, 2004.

[34] C.-H. Yu, K. Doppler, C. B. Ribeiro, and O. Tirkkonen. Resource sharing optimization for device-to-device communication underlaying cellular networks. *IEEE Transactions on Wireless Communications*, 10(8):2752–2763, 2011.