This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2691470, IEEE Transactions on Vehicular Technology

Optimal Resource Allocation in Multicast Device-to-Device Communications Underlaying LTE Networks

Hadi Meshgi[1], Dongmei Zhao[1] and Rong Zheng[2]

[1]Department of Electrical and Computer Engineering, McMaster University
[2]Department of Computing and Software, McMaster University

***Abstract*—In this paper, we present a framework for resource allocations for multicast device-to-device (D2D) communications underlaying the uplink of an LTE network. The objective is to maximize the sum throughput of active cellular users (CUs) and feasible D2D multicast groups in a cell, while meeting a certain signal-to-interference-plus-noise ratio (SINR) constraint for both the CUs and the D2D groups. We formulate the general problem of power and channel allocation as a mixed integer nonlinear programming (MINLP) problem where one D2D group can reuse the channels of multiple CUs and the channel of each CU can be reused by multiple D2D groups. Distinct from existing approaches in the literature, our formulation and solution methods provide an effective and flexible means to utilize radio resources in cellular networks and share them with multicast groups without causing harmful interference to each other. The MINLP problem is transformed so that it can be solved optimally by a variant of the generalized Bender decomposition (GBD) method with provable convergence. A greedy algorithm and a low-complexity heuristic solution are then devised. The performance of all schemes is evaluated through extensive simulations. Numerical results demonstrate that the proposed greedy algorithm can achieve close-to-optimal performance, and the heuristic algorithm provides good performance, though inferior than that of the greedy, with much lower complexity.**

I. Introduction

Device-to-Device (D2D) communication is a technology component for Long Term Evolution-Advanced (LTE-A) of the Third Generation Partnership Project (3GPP) [1]. In D2D communication, cellular users (CUs) in close proximity can exchange information over a direct link rather than transmitting and receiving signals through a cellular base station (BS). D2D users communicate directly while remaining controlled under the BS. Compared to routing through a BS, CUs at close proximity can save energy and resources when communicating directly with each other. Moreover, D2D users may experience high data rate and low transmission delay due to the short-range direct communication [2]. Reducing the network load by offloading cellular traffic from a BS and other network components to the direct path between users is another benefit of D2D communication. In addition, D2D communications can enhance user experience at cell edges [3] or relay traffic for users experiencing poor channel conditions [4], [5] [6]. Other benefits and usage cases of D2D communication are discussed in [7].

The majority of the literature in D2D communications uses the cellular spectrum for both D2D and cellular communications, also known as in-band D2D [8]. Generally, in-band D2D falls in two categories, underlay and overlay [9]. Underlay in-band D2D can improve the spectrum efficiency of cellular networks by reusing cellular resources. Its main drawback lies in the mutual interference between D2D and cellular transmissions. Thus, efficient interference management and resource allocation are necessary [10], [11]. The overlay in-band D2D avoids the interference issue by dedicating part of the cellular resources to D2D communications. In this case, designing a resource allocation scheme is crucial to maximize the utilization of dedicated cellular resources [12]. Other works consider out-of-band D2D communications so that the cellular network performance is not affected by D2D communications [13]. Out-of-band D2D communication faces challenges in coordinating the communication over two different bands because usually D2D communication happens on a second radio interface (e.g., WiFi Direct and Bluetooth) [14].

Most existing work in D2D resource allocation targets the unicast scenario where a single or multiple D2D pairs reuse the resources of CUs. In [8], the authors consider throughput maximization where by allowing D2D communication to underlay the cellular network, the overall throughput in the network can be increased compared to a case where all D2D traffic is relayed by the cellular network. Some other works such as [14], [15] consider D2D communication reliability while guaranteeing a certain level of SINR or

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2691470, IEEE Transactions on Vehicular Technology

outage probability. Improving the sum rate is also the objective for the work in [16] and [17], where game theoretical methods are used for the D2D users to compete for cellular network resources. The works in [18], [19], [20], [21] consider both throughput and reliability. The work in [18] considers one CU and one D2D pair, and throughput is maximized subject to spectral efficiency and energy constraints. Multiple D2D users and CUs are considered in [19] and [20] for maximizing the total throughput. This is done by solving a mixed integer and nonlinear programming (MINLP) resource allocation problem in [19] and designing a maximum weight bipartite matching scheme in [20]. In [21], a simple pairing algorithm is proposed for the problem of sharing CU resource with D2D links.

Multicast D2D transmissions, where the same packets for a user equipment (UE) are sent to multiple receivers, are important for scenarios such as multimedia streaming and device discovery. Specially, D2D multicast communications are required features in public safety services like police and ambulance [1]. Compared to communicating with each receiver separately in unicast D2D, multicast D2D transmission reduces overhead and saves resources. However, unlike the more commonly studied unicast D2D (see e.g. [18], [20]), multicast D2D has its own challenges. Within a multicast group, the data rates attainable at different receivers are different because of the diverse link conditions between each receiver and the transmitter. A common approach is to transmit at the lowest rate determined by the user with the worst channel condition in the group to ensure that the multicast services can be provided to all users. As a result, the transmission rate tends to decrease with number of receivers in the multicast group.

As discussed in [22] there are lots of works in multicast scheduling and resource allocation for OFDMA-based systems. They can be broadly classified into two types: single-rate and multi-rate transmissions. In single-rate broadcast, the BS transmits to all users in each multicast group at the same rate irrespective of their non-uniform achievable capacities, whereas in multirate broadcast, the BS transmits to each user in each multicast group at different rates based on what each user can handle. All of the works mentioned in [22] targeted cellular networks where the multicast transmitter is the BS. However, in multicast D2D, UEs are multicast transmitters and the quality of service (QoS) requirements for both the D2D links and the cellular links should be satisfied.

The problem of resource management for D2D multicast communication was addressed in our previous work [23], where the power and channel allocation problem for D2D multicast communication is formulated for a special case where each D2D group can reuse the channel of one CU and the channel of each CU can be reused by at most one D2D group. In [23], since each cellular channel can be allocated to at most one D2D group, the radio resources in the cellular network may not be efficiently utilized, and the D2D groups cannot fully exploit the available channel resources to achieve higher transmission rates. A baseline multicast model is proposed in [24] for overlaying in-band D2D communications, and important multicast metrics like coverage probability, mean number of covered receivers and throughput are analyzed. In [25], a single D2D multicast group that can reuse at most one cellular channel in underlay mode is studied, and a resource allocation scheme based on cognitive radio is proposed to reduce interference and improve system performance.

In this paper, we consider a general scenario of multicast D2D communications underlaying a cellular network, where each D2D group can reuse the uplink channels of multiple CUs and the channel of each CU can be reused by multiple D2D groups. The main contributions of this work are summarized as follows:

- The problem of joint power control and channel allocations is formulated as an MINLP that maximizes the aggregated rate of all CUs and D2D groups. Meanwhile, a minimum SINR constraint is imposed to guarantee the QoS requirements for both CUs and D2D groups. In the MINLP, the transmission powers are continuous variables, and the integer variables are binary for channel allocations.
- The MINLP is decomposed it into a primal problem and a master problem, where the former corresponds to the original problem with fixed binary variables, and the latter is derived through nonlinear duality theory using the Lagrange multipliers obtained from the former.
- A variant of the generalized Bender decomposition (GBD) is applied to optimally solve the MINLP iteratively.
- Inspired by the decomposed problems of the MINLP, a greedy algorithm is proposed, which has much lower complexity than the GBD-based method but achieves very close-to-optimal performance.
- A low-complexity heuristic solution is then devised which trades off computational complexity with performance. This heuristic algorithm extends the heuristic algorithm presented in [23] to the general scenario.
- An exact solution to the MINLP is proposed for a special case where each D2D group can reuse the channel of at most one CU and each CU can share its channel with at most one D2D group.

The remainder of the paper is organized as follows. In Section II, the system model is described and the problem of power and channel allocation for underlay multicast D2D communication is formulated. Section III describes

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2691470, IEEE Transactions on Vehicular Technology

the generalized Bender decomposition method to solve the general problem. The matching-based optimal resource allocation for one special case is presented in Section IV, and the greedy and the heuristic algorithms are presented in Section V. Numerical results are demonstrated in Section VI, and Section VII concludes the paper.

II. System Model and Problem Formulation

We study resource allocation for D2D group communications underlaying uplink (UL) transmissions in LTE networks. UL resource sharing is considered since reusing downlink resources is more difficult and less effective than reusing uplink resources in the worst case of a cellular network where all channels are occupied by the cellular users, as demonstrated in [26]. Consider K D2D multicast groups coexisting with M CUs as shown in Fig. 1. Consider that there are M channels, each occupied by one CU. Our work is to emphasize the benefit of managing mutual interference between the D2D and CU transmissions by coordinating the transmission power and channel reuse allocations when the D2D multicast groups underlay the cellular network. Therefore, we only consider the cellular channels that have already been used by the CUs. Allocating channels that are not used by the CUs to the D2D groups does not need to consider the mutual interference with the CUs and are not included in the formulation below. We use $m \in \mathcal{M} = \{1, 2, \ldots, M\}$ to index both the mth CU and the channel it occupies, and $k \in \mathcal{K} = \{1, 2, \ldots, K\}$ to index the kth D2D group. We consider a single cell scenario and assume that advanced intercell interference mitigation is applied on top of our scheme [18], [27], and we consider to reuse the channel of CUs that experience sufficiently good SINRs. For the CUs that experience strong intercell interference, their channels will either not be reused for D2D transmissions or be reused for D2D transmissions but contribute little to the sum rate. Therefore, these channels are not considered for the objective of maximizing the sum rate. Within a D2D group, there is only one user that multicasts messages to the remaining users. Each D2D user only belongs to one D2D group. Multicast groups can be formed during the device discovery process. As low mobility is considered for the D2D users, overhead required for maintaining the groups is low. We use $\mathcal{D}_k$ to represent the set of D2D receivers in the kth multicast group, and $|\mathcal{D}_k|$ is the total number of receivers in the group. As a special case, when $|\mathcal{D}_k| = 1$, the scenario becomes unicast. We consider applications that require best effort rates and delay tolerable services. However, a minimum rate may be required in order to make the data useful at the receiver side. When the interference condition is poor and such a minimum rate cannot be achieved for a given D2D group, the D2D multicast temporarily stops,

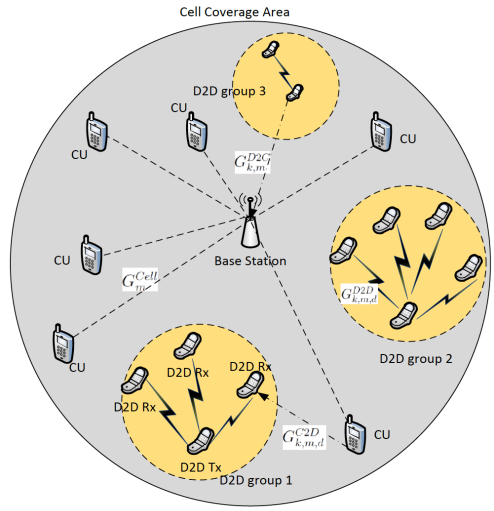


Fig. 1: System Model

i.e., no channel is allocated to the group. Maximizing the aggregated rate of all the CUs and D2D groups at all time opportunistically makes the best use of the current channel conditions. Similar objectives of maximizing the sum rate of D2D and CU transmissions have been considered in [9], [19], [20], [17] for different system models as summarized before. For applications that require high reliability at all time, the main objective is to guarantee a minimum rate at all channel conditions while the channel resource may not be fully utilized.

Define a set of binary variables $y_{k,m}$ with $y_{k,m} = 1$ if the kth D2D group reuses channel m, and $y_{k,m} = 0$ otherwise. In the general case, each D2D group splits its multicast traffic among maximally C_1 channels, and each channel can be reused by at most C_2 D2D groups, where $C_1 \leq M$ and $C_2 \leq K$. That is,

$$\sum_{m=1}^{M} y_{k,m} \leq C_1, \ \forall k \in \mathcal{K}, \tag{1}$$

$$\sum_{k=1}^{K} y_{k,m} \leq C_2, \ \forall m \in \mathcal{M}. \tag{2}$$

We further define $\beta_{k,m,d}^{D2D}$ as an indication of the channel quality for receiver d in the kth D2D group at channel m given by the ratio of the desired link gain to total power of the experienced interference as follows,

$$\beta_{k,m,d}^{D2D} = \frac{G_{k,m,d}^{D2D}}{P_{noise} + P_m^{Cell} G_{k,m,d}^{C2D} + \sum_{k' \neq k} P_{k',m}^{D2D} G_{k',m,d}^{D2D}}, \quad \forall k \in \mathcal{K}, \ m \in \mathcal{M}, d \in \mathcal{D}_k, \tag{3}$$

where P_{noise} is the aggregate power of background noise, $G_{k,m,d}^{D2D}$ is the link gain to D2D receiver d from the D2D transmitter in group k over channel m, $G_{k,m,d}^{C2D}$ is the link gain from CU m to D2D receiver d in group k, P_m^{Cell} is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2691470, IEEE Transactions on Vehicular Technology

the transmission power of CU m, $P_{k,m}^{D2D}$ is the transmission power of the kth D2D group transmitter at channel m, and $G_{k',m,d}^{D2D}$ is the link gain from the transmitter of D2D group k' to receiver d of D2D group k reusing channel m.

For the kth D2D group, its transmission condition in channel m is determined by the receiver with the worst condition. Define

$$\beta_{k,m}^{D2D} = \min_{d \in \mathcal{D}_k} \beta_{k,m,d}^{D2D}. \tag{4}$$

Then, the normalized transmission rate (bit/s/Hz) of the kth D2D group is given by

$$r_k^{D2D} = \sum_{m=1}^{M} y_{k,m} \log_2(1 + P_{k,m}^{D2D} \beta_{k,m}^{D2D}). \tag{5}$$

The aggregate transmission rate of the kth D2D group is given by

$$\begin{aligned} R_k^{D2D} &\leq |\mathcal{D}_k| r_k^{D2D} && (6)\\ &= \sum_{m=1}^{M} y_{k,m} |\mathcal{D}_k| \log_2(1 + P_{k,m}^{D2D} \beta_{k,m}^{D2D}). && (7) \end{aligned}$$

For CU m, its channel quality is given by

$$\beta_m^{Cell} = \frac{G_m^{Cell}}{P_{noise} + \sum_{k=1}^{K} y_{k,m} P_{k,m}^{D2D} G_{k,m}^{D2C}}, \tag{8}$$

where G_m^{Cell} is the link gain of CU m to the cellular BS, and $G_{k,m}^{D2C}$ is the link gain from the kth D2D transmitter to the cellular BS at channel m. Therefore, the normalized transmission rate for CU m is

$$R_m^{Cell} \leq \log_2(1 + P_m^{Cell} \beta_m^{Cell}). \tag{9}$$

There is a minimum SINR required for each D2D group and CU transmission that is set by the higher layer based on specific applications [28]. For the kth D2D group,

$$P_{k,m}^{D2D} \beta_{k,m}^{D2D} \geq y_{k,m} \gamma_{th}^{D2D}, \tag{10}$$

and for CU m,

$$P_m^{Cell} \beta_m^{Cell} \geq \gamma_{th}^{Cell}. \tag{11}$$

Note that we only consider the CUs whose SINRs are above the SINR threshold before adding D2D groups and (11) checks if the SINR threshold is satisfied after adding D2D groups.

Given these SINR threshold constraints, we can approximate the capacity in higher SINR regimes by removing the term "1" from the logarithm functions in both (7) and (9). The maximum power constraints for CUs and D2D groups, respectively, are given by

$$P_m^{Cell} \leq P_{\max}^{Cell}, \ \forall m \in \mathcal{M}, \tag{12}$$

and

$$\sum_{m=1}^{M} P_{k,m}^{D2D} \leq P_{\max}^{D2D}, \ \forall k \in \mathcal{K}. \tag{13}$$

The objective is to maximize the aggregate data transmission rate of all the D2D groups and CUs. Combining (1) – (13), we formulate the joint power control and channel allocation problem as follows,

$$\begin{aligned} \text{P1.} \quad & \max \left(\sum_{k=1}^{K} R_k^{D2D} + \sum_{m=1}^{M} R_m^{Cell} \right) && (14)\\ \text{s.t.} \quad & \beta_{k,m}^{D2D} \leq \beta_{k,m,d}^{D2D}, \forall k \in \mathcal{K}, m \in \mathcal{M}, d \in \mathcal{D}_k, && (15)\\ & y_{k,m} \in \{0,1\}, \ \forall k \in \mathcal{K}, m \in \mathcal{M}, && (16)\\ & \text{Constraints(1)–(3), (7)–(13).} \end{aligned}$$

Table I lists the parameters and variables used in the problem formulation. Clearly, P1 is a MINLP problem. In general, MINLP problems are NP-hard and thus no efficient polynomial-time solutions exist. In the general case, when C_1 and C_2 are arbitrary values, we will use GBD [29] to solve the problem optimally in the next section.

Based on the values of C_1 and C_2, several special cases exist. For example, when $C_1 = 1$ and $C_2 = 1$, each D2D group can reuse the channels of at most one CU and each CU can share their channels with at most one D2D group. Another special case of interest is when $C_2 = 1$. In this case, to increase the spectrum utilization, we allow each D2D group to reuse the resources of multiple CUs, but each CU cannot share its resource with more than one D2D group. Here, there is no interference among D2D groups and this setting is useful when the number of D2D groups is much less than the number of CUs. All the special cases can be resolved via GBD. However, it turns out that a polynomial algorithm can be devised when $C_1 = 1$ and $C_2 = 1$ as will be discussed in Section IV.

III. Generalized Bender Decomposition

The MINLP problem in P1 has the special property that when the binary variables ($y_{k,m}$'s) are fixed, the problem becomes a geometric programming problem with continuous variables ($P_{k,m}^{D2D}$'s and P_m^{Cell}'s), which can be transformed to a convex problem. A well-known solution to this type of problems is GBD [29]. However, non-trivial transformations are needed to ensure the separability of the problem with respect to the binary variables. This allows efficient solutions using GBD with guaranteed convergence. We next discuss the details of the proposed solution to P1.

A. Problem transformation

Let $\mathcal{X} = [P_{k,m}^{D2D}, P_m^{Cell}, R_k^{D2D}, R_m^{Cell}, \beta_{k,m}^{D2D}, \beta_m^{Cell}, k \in \mathcal{K}, m \in \mathcal{M}]$ represent the set of all continuous variables and $\mathcal{Y} = [y_{k,m}, k \in \mathcal{K}, m \in \mathcal{M}]$ represent the binary variables. We modify the constraints in problem P1 to separate binary variables $\forall y \in \mathcal{Y}$ from the continuous variables $\forall x \in \mathcal{X}$ and make the problem linear in terms

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2691470, IEEE Transactions on Vehicular Technology

TABLE I: Table of notations

Notation	Description
$\mathcal{M}$	Set of cellular users (CU)
$\mathcal{K}$	Set of D2D groups
$\mathcal{D}_k$	Set of receivers in kth D2D group
$\mathcal{A}$	Set of admissible or successful D2D groups
$y_{k,m}$	Binary variable, =1 if kth D2D group reuses CU m's channel, and =0 otherwise
C_1	Max. number of channels to be reused by a D2D group
C_2	Max. number of D2D groups sharing a CU channel
$G^{D2D}_{k,m,d}$	Link gain to D2D receiver d from D2D transmitter in group k at channel m
$G^{C2D}_{k,m,d}$	Link gain from CU m to D2D receiver d in group k
$G^{D2D}_{k',m,d}$	Link gain from the transmitter at D2D group k' to receiver d at D2D group k
G^{Cell}_{m}	Link gain of CU m to the cellular BS
$G^{D2C}_{k,m}$	Link gain from the kth D2D transmitter to the cellular BS at channel m
$P^{D2D}_{k,m}$	Transmission power of the kth D2D group transmitter at channel m
P^{Cell}_{m}	Transmission power of CU m
$\beta^{D2D}_{k,m,d}$	Channel quality of receiver d in the kth D2D group at channel m
β^{Cell}_{m}	Channel quality of CU m
R^{D2D}_{k}	Normalized transmission rate of the kth D2D group
R^{Cell}_{m}	Normalized transmission rate for CU m
R^{sum}	The summation of D2D and cellular throughput
γ^{D2D}_{th}	SINR threshold for all D2D groups
γ^{Cell}_{th}	SINR threshold for all CUs
$f_i(\lvert\mathcal{D}_{\mathcal{K}}\rvert)$	The complexity of solving problem Pi

of y's when the continuous variables are fixed. Problem P1 can be transformed to

$$\text{P2.}\quad \max_{x\in\mathcal{X},y\in\mathcal{Y}} f(x,y) = \max\left(\sum_{k=1}^{K} R^{D2D}_k + \sum_{m=1}^{M} R^{Cell}_m\right) \quad (17)$$

$$\text{s.t.}\quad \beta^{D2D}_{k,m} \le \frac{G^{D2D}_{k,m,d}}{P_{noise}+P^{Cell}_m G^{C2D}_{k,m,d}+\sum_{k'\neq k} P^{D2D}_{k',m} G^{D2D}_{k',m,d}},\quad \forall k\in\mathcal{K}, m\in\mathcal{M},\ d\in\mathcal{D}_k, \quad (18)$$

$$R^{D2D}_k \le \sum_{m=1}^{M}\left[|\mathcal{D}_k|\log_2(P^{D2D}_{k,m}\beta^{D2D}_{k,m}) + C(1-y_{k,m})\right],\quad \forall k\in\mathcal{K} \quad (19)$$

$$|\mathcal{D}_k|\log_2(P^{D2D}_{k,m}\beta^{D2D}_{k,m}) + C(1-y_{k,m}) \le C y_{k,m},\quad \forall k\in\mathcal{K}, m\in\mathcal{M}, \quad (20)$$

$$\beta^{Cell}_m \le \frac{G^{Cell}_m}{P_{noise}+\sum_{k=1}^{K} P^{D2D}_{k,m} G^{D2C}_{k,m}},\ \forall m\in\mathcal{M}, \quad (21)$$

$$\frac{P^{D2D}_{k,m}}{P^{D2D}_{\max}} \le y_{k,m} + \epsilon \le C P^{D2D}_{k,m},\ \forall k\in\mathcal{K}, m\in\mathcal{M}, \quad (22)$$

$$y_{k,m}\in\{0,1\},\ \forall k\in\mathcal{K}, m\in\mathcal{M}, \quad (23)$$

Constraints (1) − (2) and (9) − (13).

where C is a very large number and $\epsilon > 0$ is a very small positive number. Constraint (18) combines constraints (3) and (15) in problem P1. Constraints (19) and (20) together are equivalent to constraint (7) in problem P1. In (19), when $y_{k,m} = 1$, the second term in the summand, namely, $1 - y_{k,m}$ is zero, and the sum of the two terms inside the summation is the same as the term inside the summation on the right-hand side in (7) for the same k and m. When $y_{k,m} = 0$, the second term in the summand (19) is a large number, and the constraint is automatically satisfied; while constraint (20) guarantees that the corresponding rate for the kth D2D group at channel m is zero when the channel is not allocated to the D2D group. The introduction of constraint (22) makes $P^{D2D}_{k,m}$ very small whenever $y_{k,m}$ is zero. This eliminates the binary variables $y_{k,m}$ in (8) and results in constraint (22). Meanwhile, when $y_{k,m}$ is zero, the middle part of (22) is a very small number, and having $P^{D2D}_{k,m}$ in both the left and right hand side of the inequality ensures that $P^{D2D}_{k,m}$ is a very small number but larger than zero. This condition is needed for the logarithm functions in (19) and (20) to be feasible. To this end, we have obtained in P2 a geometric MINLP problem with separable continuous and binary variables.

B. Solution using GBD

The basic idea of GBD is to decompose the original MINLP problem into a primal problem and a master problem, and solve them iteratively. The primal problem corresponds to the original problem with fixed binary variables. Solving this problem provides the information about the lower bound and the Lagrange multipliers corresponding to the constraints. The master problem is derived through nonlinear duality theory using the Lagrange multipliers obtained from the primal problem. The solution to the master problem gives the information about the upper bound as well as the binary variables that can be used in the primal problem in next iteration. When the upper bound meets the lower bound, the iterative process converges.

Primal problem: The primal problem results from fixing the $y_{k,m}$ variables to a particular 0-1 combination denoted by $y^{(i)}_{k,m}$, where i stands for the iteration counter. After replacing the variable $y_{k,m}$ with its current value in problem P2, the formulation for the primal problem at iteration i is given by

$$\text{P3.}\quad \max_{x\in\mathcal{X},y\in\mathcal{Y}} f(x,y^{(i)}) = \max\left(\sum_{k=1}^{K} R^{D2D}_k + \sum_{m=1}^{M} R^{Cell}_m\right) \quad (24)$$

$$\text{s.t.}\quad \beta^{D2D}_{k,m} \le \frac{G^{D2D}_{k,m,d}}{P_{noise}+P^{Cell}_m G^{C2D}_{k,m,d}+\sum_{k'\neq k} P^{D2D}_{k',m} G^{D2D}_{k',m,d}},\quad \forall k\in\mathcal{K}, m\in\mathcal{M},\ d\in\mathcal{D}_k, \quad (25)$$

$$R^{D2D}_k \le \sum_{m=1}^{M}\left[|\mathcal{D}_k|\log_2(P^{D2D}_{k,m}\beta^{D2D}_{k,m}) + C(1-y^{(i)}_{k,m})\right],\quad \forall k\in\mathcal{K}, \quad (26)$$

$$|\mathcal{D}_k|\log_2(P^{D2D}_{k,m}\beta^{D2D}_{k,m}) + C(1-y^{(i)}_{k,m}) \le C y^{(i)}_{k,m},\quad \forall k\in\mathcal{K}, m\in\mathcal{M}, \quad (27)$$

$$\beta^{Cell}_m \le \frac{G^{Cell}_m}{P_{noise}+\sum_{k=1}^{K} P^{D2D}_{k,m} G^{D2C}_{k,m}},\ \forall m\in\mathcal{M}, \quad (28)$$

$$\frac{P^{D2D}_{k,m}}{P^{D2D}_{\max}} \le y^{(i)}_{k,m} + \epsilon \le C P^{D2D}_{k,m},\ \forall k\in\mathcal{K}, m\in\mathcal{M}, \quad (29)$$

$$P^{D2D}_{k,m}\beta^{D2D}_{k,m} \ge y^{(i)}_{k,m}\gamma^{D2D}_{th},\ \forall k\in\mathcal{K}, m\in\mathcal{M}, \quad (30)$$

Constraints (9), (11), (12), (13).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2691470, IEEE Transactions on Vehicular Technology

Constraints (1)–(2) are no longer needed, constraints (25)–(29) are copied from (18)–(22), and (30) is the same as (10).

Since the optimal solution to this problem (if exists) is also a feasible solution to problem P1, the optimal value $f(x^*, y^{(i)})$ provides a lower bound to the original problem. In general, not all choices of binary variables lead to a feasible primal problem. Therefore, for a given choice of $y_{k,m}$'s, there are two cases for primal problem P3: feasible problem and infeasible problem. In the following, we consider each of these cases.

- Feasible Primal: If the primal problem at iteration i is feasible, its solution provides information on the transmission power of D2D and cellular transmitters, $f(x^*, y^{(i)})$, and the optimal multiplier vectors, $\lambda_q^{(i)}$, $q = 1, 2, \ldots, Q$ for the Q inequality constraints in Problem P3. Subsequently, using this information we can formulate the Lagrange function for all inequality constraints $G_q(x, y^{(i)}) \leq 0$ for $q = 1, 2, \ldots, Q$ as

$$L(x, y^{(i)}, \lambda^{(i)}) = f(x, y^{(i)}) + \textstyle\sum_{q=1}^{Q} \lambda_q^{(i)} G_q(x, y^{(i)}), \tag{31}$$

  where $\lambda^{(i)} = [\lambda_q^{(i)}, q = 1, 2, \ldots, Q]$.
- Infeasible Primal: If the primal problem is infeasible, to identify a feasible point we can formulate an l_1-minimization problem as

$$\text{P3.1.} \quad \min \textstyle\sum_{q=1}^{Q} \alpha_q \tag{32}$$
$$\text{s.t.} \quad G_q(x, y^{(i)}) \leq \alpha_q, q = 1, 2, ..., Q, \tag{33}$$
$$\alpha_q \geq 0, q = 1, 2, ..., Q. \tag{34}$$

  Note that if $\sum_{q=1}^{Q} \alpha_q = 0$, then P3 is feasible. Otherwise, the solution to this feasibility problem (FP) provides information on the Lagrange multipliers, denoted as $\bar{\lambda}_q^{(i)}$; the Lagrange function resulting from the FP at iteration i can be defined as

$$\bar{L}(x, y^{(i)}, \bar{\lambda}^{(i)}) = \textstyle\sum_{q=1}^{Q} \bar{\lambda}_q^{(i)} (G_q(x, y^{(i)}) - \alpha_q). \tag{35}$$

It is worth mentioning that two different types of Lagrange functions are calculated depending on whether the primal problem is feasible or infeasible. Also, the lower bound is obtained only from the feasible primal problem.

Master Problem: The master problem is derived from the non-linear duality theory [29]. The original problem P2 can be written as:

$$\max_{y \in \mathcal{Y}} \sup_{x \in \mathcal{X}} f(x, y) \tag{36}$$
$$\text{s.t.} \quad G_q(x, y) \leq 0, q = 1, 2, ..., Q.$$

Let also define set $\mathcal{V}$ as

$$\mathcal{V} = \{y : G_q(x, y)) \leq 0 \text{ for some } x \in \mathcal{X}\}. \tag{37}$$

Using the Lagrange function in (31) and duality theory, we obtain

$$\max f(x, y^{(i)}) = \max_{y^{(i)}} (\min_{\lambda^{(i)}} \sup_{x} L(x, y^{(i)}, \lambda^{(i)})) \tag{38}$$
$$= \max \eta \tag{39}$$
$$\text{s.t. } \eta \leq \sup_{x} L(x, y^{(i)}, \lambda^{(i)}), \ \forall \lambda \geq 0, \tag{40}$$
$$y^{(i)} \in \mathcal{Y} \cap \mathcal{V} \tag{41}$$

It is shown in [29] that a point $y \in \mathcal{Y}$ belongs also to the set $\mathcal{V}$ if and only if they satisfy the following system:

$$\inf_{x} \bar{L}(x, y^{(i)}, \bar{\lambda}^{(i)}) \leq 0, \ \forall \bar{\lambda}^{(i)} \in \Lambda, \tag{42}$$

where $\Lambda = \left\{\bar{\lambda}_q \geq 0, \sum_{q=1}^{Q} \bar{\lambda}_q = 1\right\}$. Substituting (42) for $y \in \mathcal{Y} \cap \mathcal{V}$ into (38) we can make the constraints over set $\mathcal{V}$ explicit and obtain the following master problem:

$$\text{P4.} \quad \max_{y^{(i)} \in \mathcal{Y}} \eta \tag{43}$$
$$\text{s.t.} \quad \eta \leq \sup_{x} L(x, y^{(i)}, \lambda^{(i)}), \ \forall \lambda^{(i)} \geq 0, \tag{44}$$
$$\inf_{x} \bar{L}(x, y^{(i)}, \bar{\lambda}^{(i)}) \leq 0, \ \forall \bar{\lambda}^{(i)} \in \Lambda, \tag{45}$$
$$\text{Constraints (1), (2).}$$

The master problem P4 is similar to the original problem P2, but has two inner optimization problems that need to be considered for all λ and $\bar{\lambda}$ obtained from the primal problem in every iteration. Therefore, it has a very large number of constraints. Because of the separability of binary variables $\forall y \in \mathcal{Y}$ and continuous variables $\forall x \in \mathcal{X}$, and the linearity with regard to binary variables, we can adopt Variant 2 of GBD (V2-GBD) in [29]. It is proven in [29] that under the conditions for V2-GBD, the Lagrange function evaluated at the solution of the corresponding primal is a valid under-estimator of the inner optimization problem in P4. Therefore, the relaxed master problem can be formulated as,

$$\text{P5.} \quad \max_{y^{(i)} \in \mathcal{Y}} \eta \tag{46}$$
$$\text{s.t.} \quad \eta \leq L(x, y^{(i)}, \lambda^{(i)}), \ \forall \lambda^{(i)} \geq 0, \tag{47}$$
$$\bar{L}(x, y^{(i)}, \bar{\lambda}^{(i)}) \leq 0, \ \forall \bar{\lambda}^{(i)} \in \Lambda, \tag{48}$$
$$\text{Constraints (1), (2).}$$

The relaxed problem provides an upper bound to the master problem and can be used to generate the primal problem in the next iteration. The same procedure is then repeated until convergence. Over the iterations, the sequence of upper bounds are non-increasing and the set of lower bounds are non-decreasing. The two sequences are proven to converge, and the algorithm will stop at the optimal solution within a finite number of iterations [30]. Algorithm 1 summarizes the GBD procedure.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2691470, IEEE Transactions on Vehicular Technology

Algorithm 1 GBD Algorithm

First iteration, $i = 1$
Select an initial value for $y^{(i)}$, which makes the primal problem feasible.
Solve the primal problem in P3 and obtain the Lagrange function
$UBD^{(i)} = \infty$, $LBD^{(i)} = 0$
while $UBD^{(i)} - LBD^{(i)} > 0$ **do**
  $i = i + 1$
  Solve the relaxed master problem P5 to obtain η^* and y^*
  Set $UBD^{(i)} = \eta^*$
  Solve the primal problem P3 with fixed $y^{(i)} = y^*$
  if the primal problem is feasible **then**
    Obtain optimal solution x^* and the Lagrange function $L(x, y^{(i)}, \lambda^{(i)})$
    Set $LBD^{(i)} = \max(LBD^{(i-1)}, f^{(i)}(x^*, y^{(i)}))$
  else
    Solve the feasibility-check problem P3.1 to obtain the optimal solution x^* and the Lagrange function $\bar{L}(x, y^{(i)}, \bar{\lambda}^{(i)})$
  end if
end while

IV. Matching-based Optimal Resource Allocation for Single D2D Group per CU

In this section, we consider the MINLP problem in P1 for the special case $C_1 = 1$ and $C_2 = 1$. This case can be cast as a bipartite matching problem and thus can be solved polynomially. To formulate the bipartite problem, we divide P1 into two subproblems. In the first step, for each D2D group k and each CU m, we find their transmission power so that the sum throughput of the D2D group and the CU is maximized. If this problem is feasible, D2D group k is allowed to reuse the channel of CU m and is marked as a candidate partner in the second step; otherwise group k is excluded from the list of feasible partners. The second step is then to find the best CU partner for each D2D group among all feasible candidates so that the total throughput of all D2D groups and CUs is maximized.

1) Feasibility check and power allocation: In order to determine whether D2D group k can reuse channel m and to find the transmission power of the feasible D2D group and CU, we have problem P6 as follows:

$$\text{P6.} \quad \max \left(R_{k,m}^{D2D} + R_{k,m}^{Cell}\right) \tag{49}$$

$$\text{s.t.} \quad R_{k,m}^{D2D} = |\mathcal{D}_k| \log_2(P_{k,m}^{D2D} \beta_{k,m}^{D2D}), \tag{50}$$

$$R_{k,m}^{Cell} = \log_2 \left(P_m^{Cell} \beta_m^{Cell}\right), \tag{51}$$

$$P_{k,m}^{D2D} \beta_{k,m}^{D2D} \geq \gamma_{th}^{D2D}, \tag{52}$$

$$P_m^{Cell} \beta_m^{Cell} \geq \gamma_{th}^{Cell}, \tag{53}$$

$$\beta_m^{Cell} = \frac{G_m^{Cell}}{P_{noise} + P_{k,m}^{D2D} G_{k,m}^{D2C}}, \tag{54}$$

$$\beta_{k,m}^{D2D} \leq \frac{G_{k,m,d}^{D2D}}{P_{noise} + P_m^{Cell} G_{k,m,d}^{C2D}}, \ \forall d \in \mathcal{D}_k, \tag{55}$$

$$P_m^{Cell} \leq P_{\max}^{Cell}, \tag{56}$$

$$\textstyle\sum_{m=1}^{M} P_{k,m}^{D2D} \leq P_{\max}^{D2D}. \tag{57}$$

P6 is a reduced version of P1 by limiting it to only one D2D group and one CU with the objective of maximizing their sum throughput. Clearly, P6 is a geometric programming problem and can be transformed to a convex optimization problem using geometric programming techniques [31]. We solve problem P6 for all k and m pairs. Define a candidate channel set $\mathcal{C}_k$ for D2D group k. If the problem is feasible, D2D group k is admissible to channel m (i.e., eligible to use channel m), then m is added to $\mathcal{C}_k$. For $m \in \mathcal{C}_k$, denote the optimal throughput for the kth D2D transmitter and the mth CU as $R_{k,m}^{*D2D}$ and $R_{k,m}^{*Cell}$, respectively, and the optimal sum throughput as $R_{k,m}^{sum} = R_{k,m}^{*D2D} + R_{k,m}^{*Cell}$. For $m \notin \mathcal{C}_k$, we set $R_{k,m}^{*D2D} = 0$, $R_{k,m}^{*Cell} = \log_2 \left(\frac{P_{\max}^{Cell} G_m^{Cell}}{P_{noise}}\right)$, and thus $R_{k,m}^{sum} = R_{k,m}^{*Cell}$.

2) Maximizing total throughput: Given the maximum achievable throughput for each D2D group when reusing each cellular channel, to find the optimal channel allocation that maximizes the total throughput we have,

$$\text{P7.} \quad \max_{y_{k,m}} \textstyle\sum_{k=1}^{K} \sum_{m=1}^{M} y_{k,m} R_{k,m}^{sum} \tag{58}$$

$$\text{s.t.} \quad \textstyle\sum_{k=1}^{K} y_{k,m} \leq 1, \ \forall m \in \mathcal{M}, \tag{59}$$

$$\textstyle\sum_{m=1}^{M} y_{k,m} \leq 1, \ \forall k \in \mathcal{K}, \tag{60}$$

$$y_{k,m} \in \{0, 1\}, \ \forall k \in \mathcal{K}, \ m \in \mathcal{M}. \tag{61}$$

P7 is in effect the maximum weight bipartite matching problem, where the D2D groups and the cellular channels are two groups of vertices in the bipartite graph, and the edge connecting D2D group k and channel m has a weight $R_{k,m}^{sum}$. The Hungarian algorithm [32] can be used to solve the bipartite matching problem in polynomial time.

To determine the computational complexity, consider $M \geq K$ and the complexity of solving P6 is a function of the size of each D2D group, denoted as $f_6(|\mathcal{D}_\mathcal{K}|)$. Therefore, the time complexity of the matching-based optimal resource allocation is $\mathbf{O}(M \times K \times f_6(|\mathcal{D}_\mathcal{K}|)) + \mathbf{O}(M^3)$, where the first and second terms correspond to the computation time in the first and second steps, respectively.

V. Greedy and Heuristic Channel Allocation Algorithms

The MINLP problem in P1 is an NP-hard problem, and the computational complexity grows exponentially with the problem size in the worst case. In other words, GBD may

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2691470, IEEE Transactions on Vehicular Technology

converge in an exponential number of iterations. In this section, we first propose a greedy algorithm and then a heuristic solution to the general MINLP problem in P1.

Algorithm 2 Greedy algorithm

1: $\mathcal{M}$: Set of cellular users
2: $\mathcal{K}$: Set of all D2D groups
3: $e_{k,m} = 1,\ \forall k \in \mathcal{K}, m \in \mathcal{M}$
4: $Y = [y_{k,m} |\ y_{k,m} = 0,\ \forall k \in \mathcal{K}, m \in \mathcal{M}]$
5: $\mathcal{S} = \emptyset$
6: **while** $\sum_{k=1}^{K} \sum_{m=1}^{M} e_{k,m} \geq 1$ **do**
7: $\quad E = [e_{k,m} |\ e_{k,m} = 1,\ \forall k \in \mathcal{K}, m \in \mathcal{M}]$
8: $\quad T_{k,m}^{sum} = \sum_{m'=1}^{M} \log_2 \left(\frac{P_{\max}^{Cell} G_{m'}^{Cell}}{P_{noise}} \right),\ \forall k \in \mathcal{K}, m \in \mathcal{M}$
9: $\quad$ **for each** $e_{k,m} \in E$ **do**
10: $\quad\quad y_{k,m} = 1$
11: $\quad\quad$ **if** (k,m) is *Admissible* **then**
12: $\quad\quad\quad$ Solve P3 to find $P_{k',m'}^{D2D}$ and $P_{m'}^{Cell},\ \forall (k',m') \in [\mathcal{S} \cup (k,m)]$
13: $\quad\quad\quad$ **if** P3 is feasible **then**
14: $\quad\quad\quad\quad T_{k,m}^{sum} = \sum_{(k',m') \in [\mathcal{S} \cup (k,m)]} Z_{k',m'}$, where $Z_{k',m'} = y_{k',m'} |\mathcal{D}_{k'}| \log_2 (P_{k',m'}^{D2D} \beta_{k',m'}^{D2D}) + \sum_{m'=1}^{M} \log_2 \left(P_{m'}^{Cell} \beta_{m'}^{Cell} \right)$
15: $\quad\quad\quad$ **else**
16: $\quad\quad\quad\quad e_{k,m} = 0$
17: $\quad\quad\quad$ **end if**
18: $\quad\quad$ **else**
19: $\quad\quad\quad e_{k,m} = 0$
20: $\quad\quad$ **end if**
21: $\quad\quad y_{k,m} = 0$
22: $\quad$ **end for**
23: $\quad (k^*, m^*) = \arg\max_{\forall (k,m)} T_{k,m}^{sum}$
24: $\quad y_{k^*,m^*} = 1$
25: $\quad e_{k^*,m^*} = 0$
26: $\quad \mathcal{S} = \mathcal{S} \cup (k^*, m^*)$
27: **end while**

A. A greedy algorithm

Algorithm 2 shows the greedy resource allocation algorithm. The key idea of the greedy algorithm is that, in each iteration, it selects a CU and D2D group pair that maximizes the resulting sum throughput of all selected pairs. The algorithm terminates when no more pair can be included.

In this algorithm, we first initialize all edges of a $K \times M$ bipartite graph, $e_{k,m}$, to one in line 3. The $K \times M$ assignment matrix $\mathbf{Y}$ is initialized to zero. $\mathcal{S}$ is the set of selected CU and D2D pairs that maximize the sum throughput and are initialized to zero at first. Matrix E includes all edges $(e_{k,m})$ with the value of one. The inner loop (lines 9-22) finds the sum throughput, $T_{k,m}^{sum}$, of all pairs in set $\mathcal{S}$ after an admissible pair (k,m) is added to $\mathcal{S}$. In line 11, to find if (k,m) is *admissible*, the algorithm checks constraints (1) and (2) for a given (k,m) pair. If either of these constraints is violated for the current (k,m), the procedure sets $e_{k,m}$ and $y_{k,m}$ to zero and moves to the next pair. Otherwise, the algorithm solves problem P3 and finds $T_{k,m}^{sum}$. In the outer loop, the pair (k^*, m^*) that maximizes $T_{k,m}^{sum},\ \forall (k,m) \in \mathcal{S}$ (line 23) is found and removed from E. The outer loop is iterated until $e_{k,m} = 0,\ \forall k \in \mathcal{K}$ and $m \in \mathcal{M}$.

Since a total of $\min\{M \times C_2, K \times C_1\}$ pairs can be found in the procedure, and in each iteration of the outer loop, only one such pair can be added, the computational complexity of the greedy algorithm is $\mathbf{O}(\min\{M \times C_2, K \times C_1\} \times K \times M \times f_3(|\mathcal{D}_{\mathcal{K}}|))$, where $f_3(|\mathcal{D}_{\mathcal{K}}|)$ is the complexity of solving P3 as a function of the size of each D2D group. The high complexity of the greedy algorithm mainly arises from the need to solve the optimization problem P3 up to $K \times M$ times to find the best pair in each iteration.

Algorithm 3 Heuristic algorithm

$\mathcal{M}$: List of cellular users in decreasing order of G_m^{Cell}
$\mathcal{K}$: List of all D2D groups
$G_{m,k}^{C2D} = \min_{d \in \mathcal{D}_k} G_{k,m,d}^{C2D},\ \forall k \in \mathcal{K}, m \in \mathcal{M}$
$G_{k,k'}^{D2D} = \min_{d \in \mathcal{D}'_k} G_{k,k',d}^{D2D},\ \forall k \in \mathcal{K}, m \in \mathcal{M}$
$y_{k,m} = 0,\ \forall k \in \mathcal{K}, m \in \mathcal{M}$
$P_m^{Cell} = P_{\max}^{Cell},\ \forall m \in \mathcal{M}$
$P_{k,m}^{D2D} = 0,\ \forall k \in \mathcal{K}, m \in \mathcal{M}$
$m = 1$
for each $m \in \mathcal{M}$ **do**
$\quad \mathcal{K}' = \{\forall k \in \mathcal{K} | \sum_{m=1}^{M} y_{k,m} < C_2\}$
$\quad$ **while** $\sum_{k=1}^{K} y_{k,m} < C_1$ or $\mathcal{K}' \neq \emptyset$ **do**
$\quad\quad k^* = \arg\min_{k \in \mathcal{K}'} \left(\sum_{k'=1}^{K} P_{k',m}^{D2D} G_{k,k'}^{D2D} + P_m^{Cell} G_{m,k}^{C2D} \right)$
$\quad\quad y_{k^*,m} = 1$
$\quad\quad$ Solve P3 to find $P_{k^*,m}^{D2D}$ and P_m^{Cell}
$\quad\quad$ **if** P3 is feasible **then**
$\quad\quad\quad$ D2D k^* transmits on channel m
$\quad\quad\quad y_{k^*,m} = 1$
$\quad\quad$ **else**
$\quad\quad\quad y_{k^*,m} = 0$
$\quad\quad$ **end if**
$\quad\quad \mathcal{K}' = \mathcal{K}' \setminus \{k^*\}$
$\quad$ **end while**
end for

B. A heuristic algorithm

Since the complexity of the greedy algorithm is high, we propose a heuristic algorithm with less complexity in Algorithm 3. In the following, we explain some intuition behind the algorithm.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2691470, IEEE Transactions on Vehicular Technology

To increase cellular and D2D throughputs, it is desirable to have higher SINR. From (3) and (7), it can be deduced that having smaller values of $G^{C2D}_{k,m,d}$ and $G^{D2D}_{k,k',d}$ reduces interference from CU m to D2D group k and from D2D group k to D2D group k', respectively, resulting in higher $\beta^{D2D}_{k,m}$ and D2D throughput. Furthermore, higher values of G^{Cell}_{m} lead to higher cellular throughput. Therefore, Algorithm 3 tries to pair up a CU that has a high link gain to the BS and a D2D group that has low interference to the CU.

Starting from $m = 1$, the outer loop in Algorithm 3 iterates through all CUs. For each m, the algorithm finds at most C_1 best D2D groups to share the channel m in the inner loop. Line 12 shows the criteria for choosing the D2D group that receives the minimum interferences from CU m and all other D2D groups using the same channel. In line 14, based on the current value of $y_{k,m}$, problem P3 is solved to find the optimal transmission power for each CU and D2D group. If P3 is feasible, D2D group k^* will reuse the channel m and we have $y_{k^*,m} = 1$, otherwise $y_{k^*,m} = 0$ in line 19. In both cases, k^* is removed from the D2D group list for the next iteration. The inner loop stops iterating after finding C_1 D2D groups for CU m or after at most K iterations. It is worth mentioning that each D2D group cannot reuse more than C_2 CUs. That is accomplished by introducing $\mathcal{K}'$ that keeps track of all D2D groups with less than C_2 assigned channels in line 10.

In this algorithm, problem P3 is solved $M \times C_1$ times in the worst case, and thus the complexity of the heuristic algorithm is $\mathbf{O}(M^2)+\mathbf{O}(M \times K \times f_3(|\mathcal{D}_{\mathcal{K}}|))$. This is much less than the complexity of the greedy algorithm. Note that this complexity analysis does not give the actual amount of time needed to run the algorithm, which is more important to determine whether or not the algorithm is acceptable in the real system. The exact amount of time for performing the algorithm also depends on the computing speed of the CPU. In addition, the delay requirements of the application and the mobility of the users (which determines the channel dynamics) can also affect the feasibility of the algorithm. We summarize the computational complexity of all the three solutions in Table II in the worst case.

TABLE II: Worst case complexity comparison

Algorithm	Worst Case Complexity
GBD	Exponential
Greedy	$\mathbf{O}(\min\{M \times C_2, K \times C_1\} \times K \times M \times f_3(\|\mathcal{D}_{\mathcal{K}}\|))$
Heursitic	$\mathbf{O}(M^2) + \mathbf{O}(M \times K \times f_3(\|\mathcal{D}_{\mathcal{K}}\|))$

C. Coordination and Overhead

Channel measurement is an indispensable component of resource allocations in D2D communications. Our proposed resource allocation algorithms are performed at the BS. The BS should first collect all required channel state information (CSI) in order to find the transmission power for all CU and D2D transmitters and allocate channels for each D2D group. It should then pass the values of the transmission power to individual transmitters, and the channel allocation information to the D2D groups. Collecting the CSI between a CU and the BS, between two D2D users, and between a CU and a D2D user can be performed during the device discovery process using the discovery signal. In order to reduce the overhead of reporting the involved CSI to the BS, the CSI feedback compression, signal flooding, and distance-based mechanisms can be utilized [17]. The overhead can be further reduced for short and low mobility user-to-user or user-to-BS links as the channel should have fewer taps and vary slowly.

TABLE III: Default Simulation Parameters

Parameter	Value
Cell radius (R)	1 km
Number of D2D receivers in each group	3
P_{noise}	-114 dBm
Pathloss exponent (α)	3
$P^{D2D}_{\max}$	20 dBm
$P^{Cell}_{\max}$	20 dBm
γ_{th} =γ^{Cell}_{th} =γ^{D2D}_{th}	10 dB
D2D cluster size(r)	50 m

VI. Performance Evaluation

We consider a single cell network as illustrated in Fig. 2, where cellular users are uniformly distributed in the cell. The distance-based path loss and slow Rayleigh fading are adopted as the channel model. The probability density function of the instantaneous link gain at any time is given by $f_G(x) = \frac{1}{\bar{G}}e^{-x/\bar{G}}$, for $x \geq 0$, where $\bar{G}$ is the average link gain between the transmitter and the receiver and can be calculated based on the distance-based path loss model. The proposed algorithms have been implemented in Matlab together with the CVX, a package for specifying and solving convex programs [33]. Default parameters used in the simulations are given in Table III. We run two sets of experiments to evaluate the performance of the proposed algorithms, namely, regularly placed D2D clusters and randomly placed D2D clusters. A larger M is used for the regular placement of users, since the results are collected based on one placement of the users. For the randomly placed users, each result is collected by averaging over a large number of different placements of users. Therefore, each result for the randomly placed users takes much longer time to compute than that for the regular placement. In order to keep the total simulation time to be reasonable, we have to use smaller M in the random placement.

a) Regularly placed D2D clusters: In Fig. 2, D2D groups are manually placed in six different locations and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2691470, IEEE Transactions on Vehicular Technology

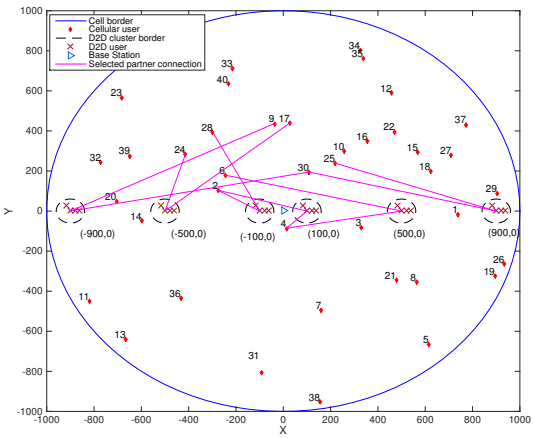


Fig. 2: Regularly placed D2D clusters in a cell, $C_1 = 2, C_2 = 2$, $M = 40$.

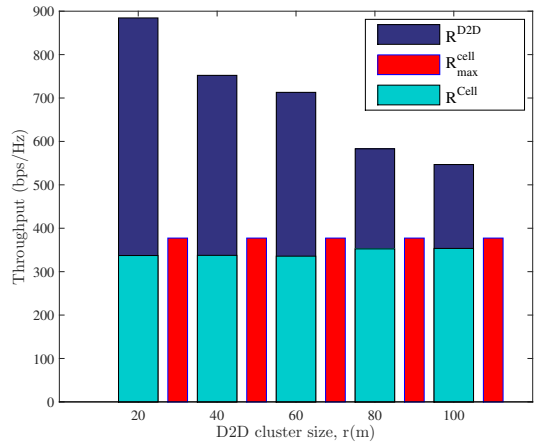


Fig. 3: Throughput comparison for different cluster sizes, $C_1 = 2, C_2 = 2$, $M = 40$, and $K = 6$.

D2D transmitters and receivers are placed in the fixed locations within each group with radius r. This scenario allows us to have a better understanding of the channel selection for D2D users and how it is impacted by geographical spacing. In the figure, D2D transmitters are labeled with their coordinates. The GBD algorithm finds the CU partner (or equivalent, the CU channel) for each D2D group among 40 CUs when $C_1 = 2$ and $C_2 = 2$. The straight lines in Fig. 2 connect D2D groups with their respective CU partners. As shown in the figure, the chosen CU partners, tend to be close to the base station to ensure the rate of the CUs. Meanwhile, the CU partners are away from the respective D2D users to reduce mutual interference between the CUs and the D2D users. As it can be seen in Fig. 2, all D2D groups found CU partners in this configuration. Note that even for CUs at the cell edges, their SINR constraints are satisfied as guaranteed by P1.

Fig. 3 compares the maximum cellular throughput (without D2D users), $R^{Cell}_{\max}$, the throughput of cellular users (with D2D users), R^{Cell}, and D2D throughput, R^{D2D}, defined as follows,

$$R^{Cell}_{\max} = \sum_{m=1}^{M} \log_2\left(\frac{P^{Cell}_{\max} G^{Cell}_m}{P_{noise}}\right), \tag{62}$$

$$R^{Cell} = \sum_{m=1}^{M} R^{Cell}_m, \tag{63}$$

$$R^{D2D} = \sum_{k\in\mathcal{A}} R^{D2D}_k, \tag{64}$$

where $\mathcal{A}$ is the set of D2D groups that are allowed to reuse at least one cellular channel. As can be observed in Fig. 3, the overall network throughput, $R^{sum} = R^{Cell} + R^{D2D}$, is greater than the maximum throughput before including D2D users, $R^{Cell}_{\max}$. With the introduction of D2D users, the overall throughput increases by 25% to 125%. This comes at the cost of reduced cellular throughput as $R^{Cell}_{\max} > R^{Cell}$ since adding D2D users causes interference to cellular users and decreases their throughput. However, the reduction is relatively small, compared to the D2D throughput. Moreover, although a larger D2D cluster size leads to lower D2D channel gain and lower D2D throughput, it does not affect the cellular throughput very much.

Fig. 4 shows D2D and sum rates versus C_1 for different values of C_2. Both rates increase with C_1 since the number of available channels for each D2D group increases and hence D2D rate increases. However, both the D2D and sum rates flatten out after a certain value of C_1. For instance, when $C_2 = 1$, each CU can serve at most one D2D group, and increasing C_1 does not increase the rate since there are not enough channels to allow all the D2D groups to reuse C_1 channels. Also, from this figure we see that cellular throughput, which is the difference between the sum rate and the D2D rate, decreases as C_1 increases. This is because of the fact that the interference from D2D groups on CUs increases with C_1. On the other hand, increasing C_2 increases the D2D and sum rate for higher values of C_1 since each CU can serve more D2D groups and hence there are more available channels for D2D groups. However, for lower values of C_1, since there are enough CUs in the cell to be reused by D2D groups, increasing C_2 does not change the D2D and sum rates significantly.

Fig. 5 shows the convergence of the GBD in Algorithm 1. As it is mentioned in this algorithm, in the first iteration $UBD^{(1)} = \infty$, $LBD^{(1)} = 0$. The second iteration starts with a initial value of $\mathbf{Y}$. It is shown in this figure the LBD results from solving primal problem and the UBD from solving master problem converge in iteration 5.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2691470, IEEE Transactions on Vehicular Technology

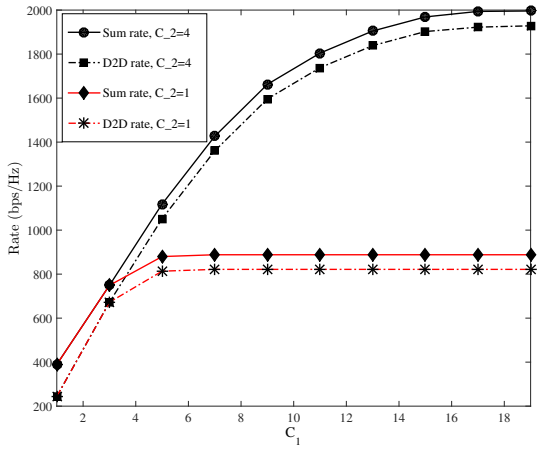


Fig. 4: Throughput comparison for different values of C_1 and C_2, $M = 20$, and $K = 6$.

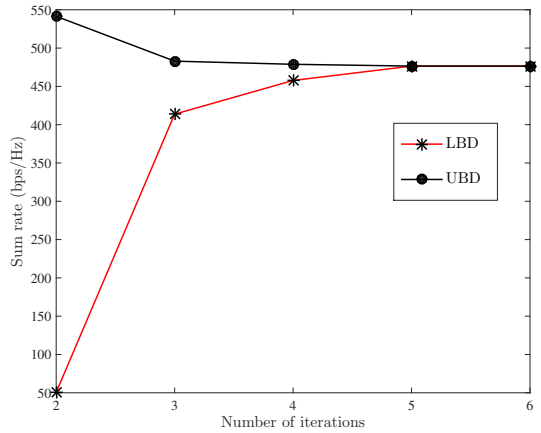
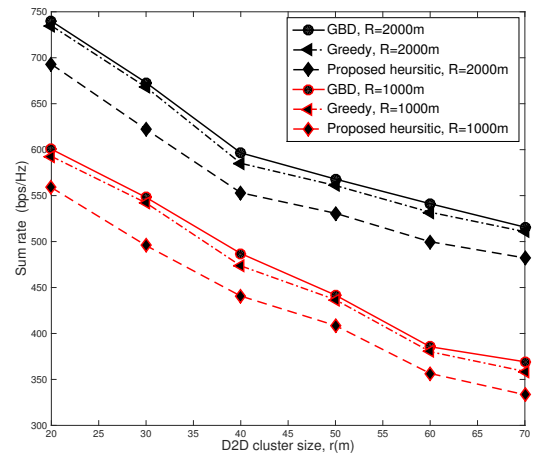
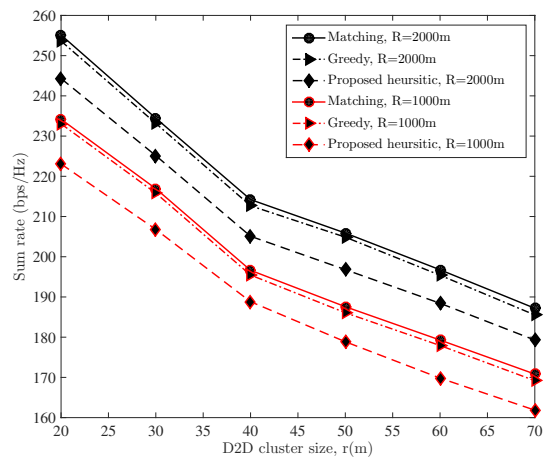


Fig. 5: Convergence of the GBD algorithm, $C_1 = 2$, $C_2 = 2$, $M = 20$, and $K = 6$.



(a) $C_1 = 4, C_2 = 3$



(b) $C_1 = 1, C_2 = 1$

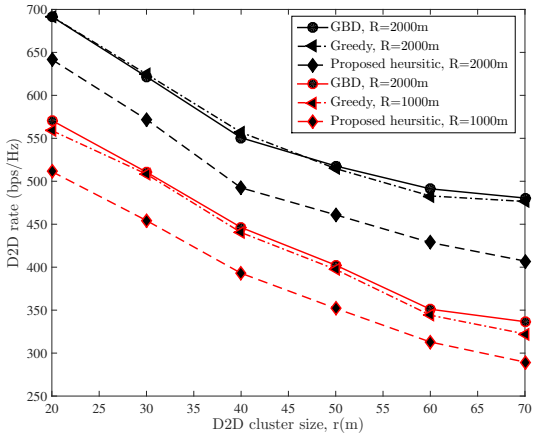
Fig. 6: Average sum throughput versus D2D cluster radius for different cell radii (R), $M = 10, K = 4$

b) Randomly placed D2D users: In the second set of experiments, we follow the clustered distribution model in [34], where clusters of radius r are randomly located in a cell and the D2D users in each group are randomly distributed in the corresponding cluster. Four metrics are used to evaluate the performance: the sum throughput, R^{sum}, the D2D throughput, R^{D2D}, and the success rate. The success rate is defined as the ratio of the number of D2D groups that found their CU partners ($|\mathcal{A}|$) and the total number of D2D groups.

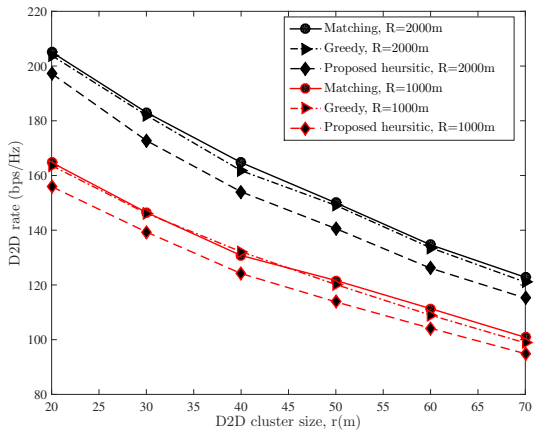
The results in this section have been generated for two sets of C_1 and C_2 values: in part (a) of all the figures, $C_1 = 4$ and $C_2 = 3$; and in part (b), $C_1 = 1$ and $C_2 = 1$. In the case of $C_1 = 1$ and $C_2 = 1$, both GBD and the matching-based algorithm return the same results since both are optimal. In [23], we have adapted the heuristic scheme in [19] for multicast D2D and compared it against proposed scheme when $C_1 = 1$ and $C_2 = 1$. Numerical results in [23] show that our proposed heuristic outperforms the resource allocation algorithm in [19], and thus evaluation of the heuristic in [19] is omitted here.

Figs. 6 – 8 compare the performance of GBD, the greedy and the heuristic algorithms for different D2D cluster sizes (r) and different cell radii (R). From these figures, we observe that both the sum and the D2D throughput as well as the success rate decrease with the D2D cluster size. Since the channel gain of D2D link decreases when the cluster radius increases, more transmission power is required for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2691470, IEEE Transactions on Vehicular Technology

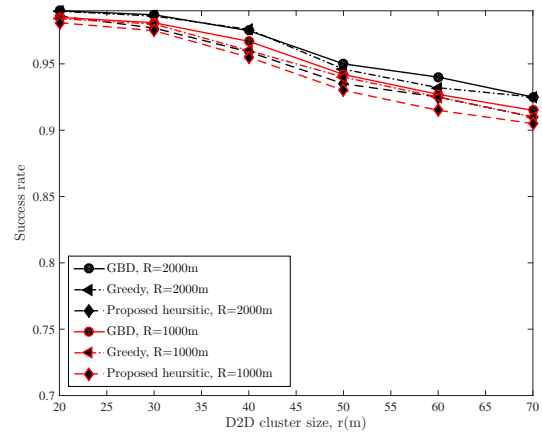


(a) $C_1 = 4, C_2 = 3$

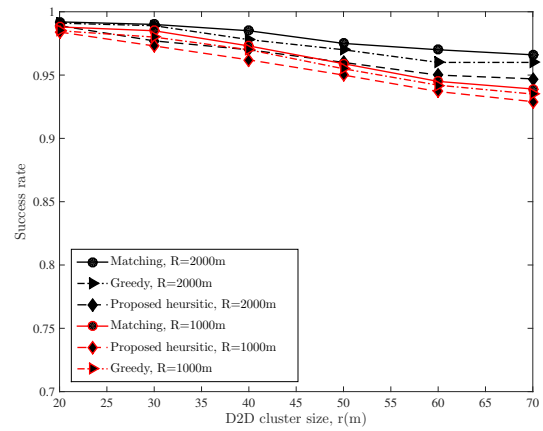


(b) $C_1 = 1, C_2 = 1$

Fig. 7: Average D2D throughput versus D2D cluster radius for different cell radii (R), $M = 10, K = 4$



(a) $C_1 = 4, C_2 = 3$



(b) $C_1 = 1, C_2 = 1$

Fig. 8: Average D2D success rate versus D2D cluster radius for different cell radii (R), $M = 10, K = 4$

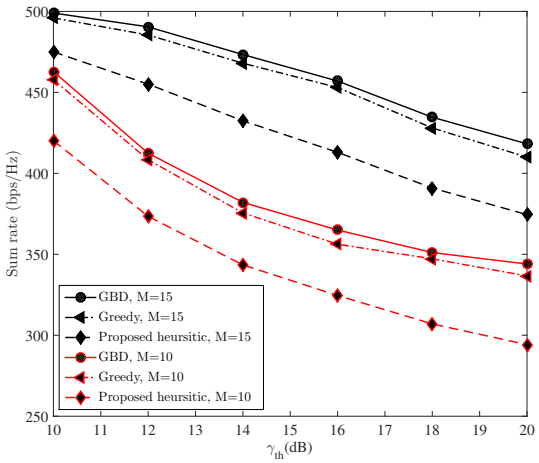
the D2D groups to satisfy the SINR threshold constraint. This in turn causes more interference to the reused CU partner. Furthermore, it is seen from these figures that the sum throughput, the D2D throughput and the success rate of all three algorithms increase with the cell radius. This is because increasing the cell radius increases the distance between the CUs and D2D receivers and also the average distance of individual nodes to the BS. Hence, the interference from CUs to D2D receivers and the interference from D2D transmitters at the BS is decreased. Recall that the D2D rate is the maximum throughput achieved by the admitted D2D groups. It is worth mentioning that increasing the cell size leads to reduction in the cellular throughput due to the decreased link gain between the CUs and the base station. However, with the current simulation parameters, R^{D2D} is the dominating part of the sum rate and, therefore, R^{sum} increases with the cell size in both parts (a) and (b).

It can be also seen from Fig. 6 that the optimal solutions, GBD algorithm for part (a) and matching-based algorithm for part (b), has the highest sum rates. In comparison, the greedy algorithm achieves close-to-optimal sum rate, while the heuristic algorithm has a lower sum rate compared to the other two algorithms, but it has the lowest complexity among them. Note that in Fig. 7, the D2D rate of the greedy algorithm exceeds that of the optimal solution for some D2D cluster sizes. This does not contradict the optimality of GBD since the objective of P1 is to maximize the sum rate not the D2D rate.

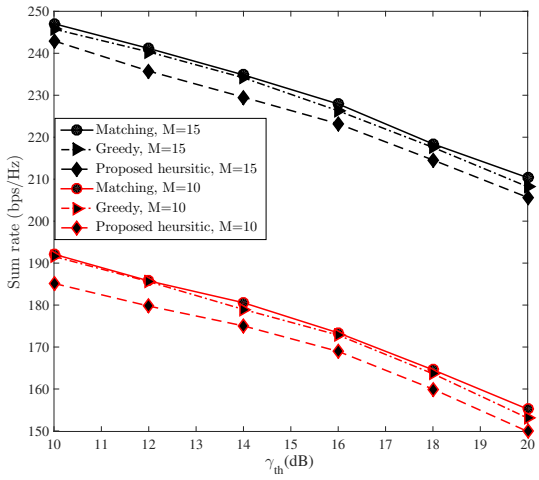
In Figs. 9 – 11 the performance of all proposed algorithms for different SINR thresholds ($\gamma_{th}^{D2D} = \gamma_{th}^{Cell} = \gamma_{th}$)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2691470, IEEE Transactions on Vehicular Technology

with different numbers of CUs (M) is shown. It is seen that increasing the SINR threshold leads to decreasing sum rates, D2D rates, and success rates since it limits the chances for D2D groups to find CU partners. It can be also observed that the total D2D throughput improves slightly with increasing number of CUs since there are more potential candidates for D2D groups to reuse.



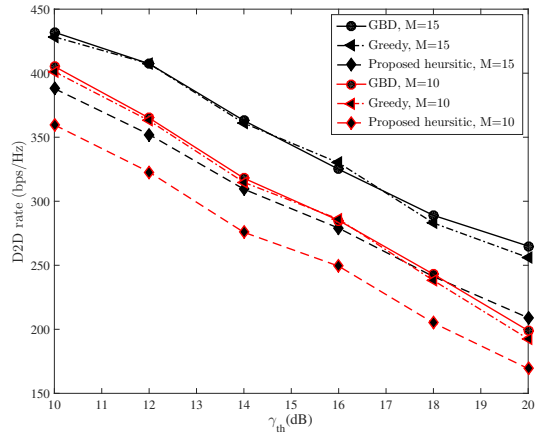
(a) $C_1 = 4, C_2 = 3$



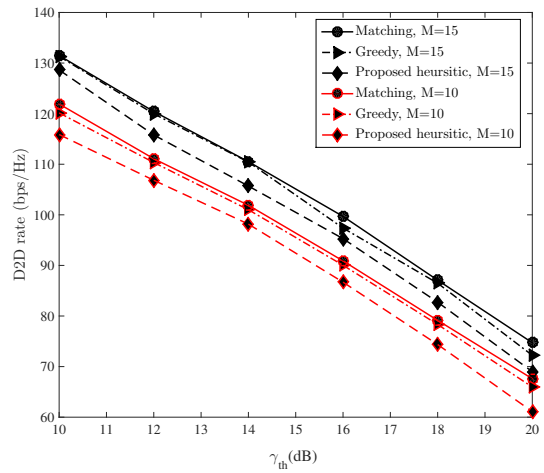
(b) $C_1 = 1, C_2 = 1$

Fig. 9: Average sum throughput versus γ_{th} for different number of cellular users (M), $R = 1000$ m, $K = 4$

The complexity of the GBD prevents us from obtaining the results for larger M, K, C_1 and C_2 values in a reasonable amount of time. In Figs. 12 and 13 we compare the greedy and the heuristic algorithms when the D2D cluster size varies, where $M = 25$, $K = 8$, $C_1 = 4$ and $C_2 = 3$. Fig. 12 shows the sum rate and the D2D rates using the two



(a) $C_1 = 4, C_2 = 3$

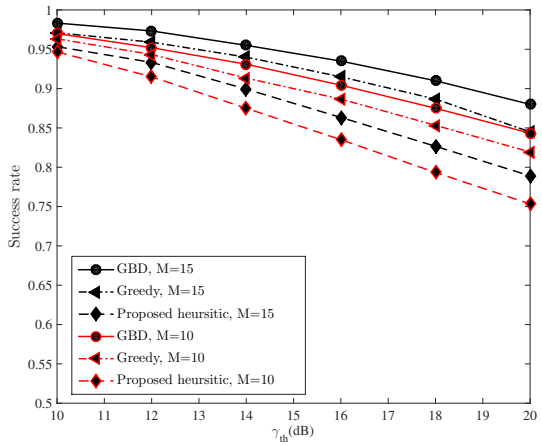


(b) $C_1 = 1, C_2 = 1$

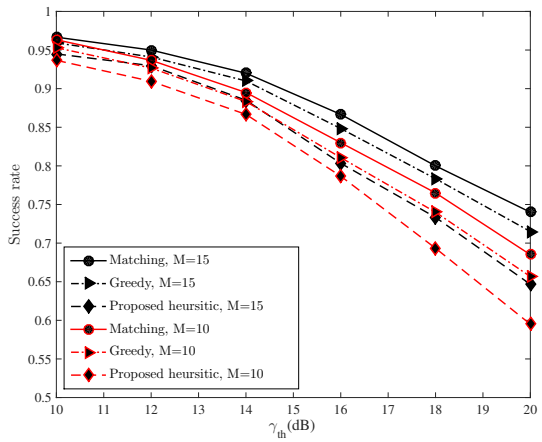
Fig. 10: Average D2D throughput versus γ_{th} for different number of cellular users (M), $R = 1000$ m, $K = 4$

algorithms, and Fig. 13 compares the success rates. It can be seen from Fig. 12 that although the heuristic underperforms the greedy algorithm in both sum rate and D2D rate, both algorithms achieve high D2D rate. Specifically, when the D2D cluster size is relatively small, say 10m, the D2D rate accounts for about 85% of the sum rate for both the greedy and the heuristic algorithms. This percentage becomes lower when the D2D cluster size becomes larger due to poorer D2D channel conditions. When the D2D cluster size is 70m, the D2D rate still accounts for about 80% of the sum rate for both algorithms. Note that such high D2D rates are obtained with the minimum SINRs guaranteed for the existing CUs. This demonstrates that our proposed algorithms can indeed support the D2D multicast with high rates without causing

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2691470, IEEE Transactions on Vehicular Technology



(a) $C_1 = 4, C_2 = 3$



(b) $C_1 = 1, C_2 = 1$

Fig. 11: Average D2D success rate versus γ_{th} for different number of cellular users (M), $R = 1000$ m, $K = 4$

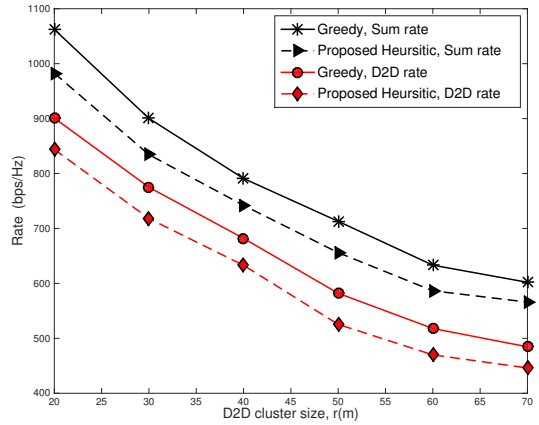


Fig. 12: Average D2D throughput versus D2D cluster radius, $C_1 = 4$, C_2=3, $M = 25$, $K = 8$, $R = 1000$ m

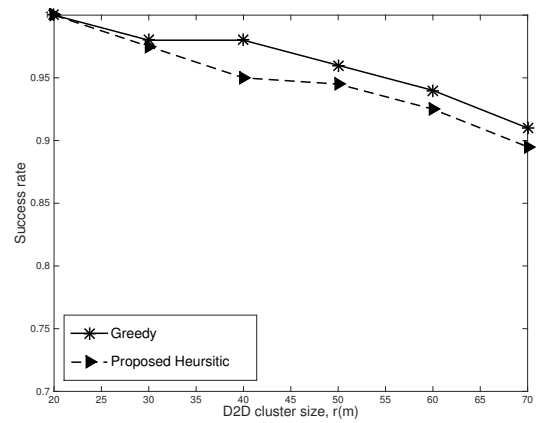


Fig. 13: Average D2D success rate versus D2D cluster radius, $C_1 = 4$, $C_2 = 3$, $M = 25$, $K = 8$, $R = 1000$ m

harmful interference to the CUs. Fig. 13 shows that both the greedy and the heuristic algorithms can achieve relatively high success rate in admitting the D2D multicast groups. This is consistent with the results with smaller M and K presented earlier.

VII. Conclusions

In this paper, we considered joint power and channel allocation for multicast D2D communications sharing uplink channels with the unicast CUs in a cellular network. To maximize the overall throughput while guaranteeing the QoS requirements of both CUs and D2D groups, we formulated an optimization problem and found the optimal solution using GBD. Then, we solved a special case when each D2D group can reuse the channels of at most one CU and each CU can share its channel with at most one D2D group, using maximum weight bipartite matching algorithm. Finally, a greedy algorithm and a low-complexity heuristic algorithm were also proposed. We performed extensive simulations with different parameters such as SINR threshold, cell size, D2D cluster size, and number of CUs. Results showed that the greedy algorithm has close-to-optimal performance in sum rate, D2D rate, and D2D success rate. In comparison, our proposed heuristic algorithm achieves lower sum rate and D2D rate than the greedy algorithm and similar success rate as the greedy algorithm with lower computational complexity. Meanwhile, we have observed that both the greedy and the heuristic algorithms achieve

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2691470, IEEE Transactions on Vehicular Technology

very high D2D rate, that accounts for 80% or higher of the sum rate. This indicates that our proposed algorithms can support very high throughput for D2D multicast that underlay the cellular channels. This conclusion is very important for supporting various services that may consume a significant amount of bandwidth if transmitted by the cellular BS directly. The optimum results obtained by using GDB to solve a transformed problem of the joint power and channel allocation problem provide a theoretical upper bound of the performance for this kind of system.

References

[1] "3rd generation partnership project; technical specification group SA; study on architecture enhancements to support proximity services (prose)," *TR23.703 V0.4.1, Release 12*, June 2013.

[2] X. Lin, J. Andrews, A. Ghosh, and R. Ratasuk, "An overview of 3GPP device-to-device proximity services," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 40–48, April 2014.

[3] A. T. Gamage, H. Liang, R. Zhang, and X. Shen, "Device-to-device communication underlaying converged heterogeneous networks," *IEEE Wireless Communications*, vol. 21, no. 6, pp. 98–107, 2014.

[4] Y. Cao, T. Jiang, X. Chen, and J. Zhang, "Social-aware video multicast based on device-to-device communications," *IEEE Transactions on Mobile Computing*, vol. 15, no. 6, pp. 1528–1539, 2016.

[5] B. Zhou, H. Hu, S.-Q. Huang, and H.-H. Chen, "Intracluster device-to-device relay algorithm with optimal resource utilization," *IEEE transactions on vehicular technology*, vol. 62, no. 5, pp. 2315–2326, 2013.

[6] J. Liu, S. Zhang, N. Kato, H. Ujikawa, and K. Suzuki, "Device-to-device communications for enhancing quality of experience in software defined multi-tier lte-a networks," *IEEE Network*, vol. 29, no. 4, pp. 46–52, 2015.

[7] L. Lei, Z. Zhong, C. Lin, and X. Shen, "Operator controlled device-to-device communications in LTE-advanced networks," *IEEE Wireless Communications*, vol. 19, no. 3, pp. 96–104, June 2012.

[8] K. Doppler, M. Rinne, C. Wijting, C. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to LTE-advanced networks," *IEEE Communications Magazine*, vol. 47, no. 12, pp. 42–49, Dec 2009.

[9] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Communications Surveys Tutorials*, vol. 16, no. 4, pp. 1801–1819, Fourthquarter 2014.

[10] T. Peng, Q. Lu, H. Wang, S. Xu, and W. Wang, "Interference avoidance mechanisms in the hybrid cellular and device-to-device systems," in *IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications, 2009*, Sept 2009, pp. 617–621.

[11] P. Janis, V. Koivunen, C. Ribeiro, J. Korhonen, K. Doppler, and K. Hugl, "Interference-aware resource allocation for device-to-device radio underlaying cellular networks," in *IEEE 69th Vehicular Technology Conference, (VTC) Spring*, April 2009, pp. 1–5.

[12] Y. Pei and Y.-C. Liang, "Resource allocation for device-to-device communications overlaying two-way cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3611–3621, July 2013.

[13] A. Asadi and V. Mancuso, "WiFi Direct and LTE D2D in action," in *IFIP Wireless Days (WD)*, Nov 2013, pp. 1–8.

[14] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklós, and Z. Turányi, "Design aspects of network assisted device-to-device communications," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 170–177, March 2012.

[15] H. Min, W. Seo, J. Lee, S. Park, and D. Hong, "Reliability improvement using receive mode selection in the device-to-device uplink period underlaying cellular networks," *IEEE Transactions on Wireless Communications,*, vol. 10, no. 2, pp. 413–418, February 2011.

[16] L. Song, D. Niyato, Z. Han, and E. Hossain, "Game-theoretic resource allocation methods for device-to-device communication," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 136–144, 2014.

[17] C. Xu, L. Song, Z. Han, Q. Zhao, X. Wang, X. Cheng, and B. Jiao, "Efficiency resource allocation for device-to-device underlay communication systems: a reverse iterative combinatorial auction based approach," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 348–358, 2013.

[18] C.-H. Yu, K. Doppler, C. Ribeiro, and O. Tirkkonen, "Resource sharing optimization for device-to-device communication underlaying cellular networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2752–2763, August 2011.

[19] M. Zulhasnine, C. Huang, and A. Srinivasan, "Efficient resource allocation for device-to-device communication underlaying LTE network," in *IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), 2010*, Oct 2010, pp. 368–375.

[20] D. Feng, L. Lu, Y. Yuan-Wu, G. Li, G. Feng, and S. Li, "Device-to-device communications underlaying cellular networks," *IEEE Transactions on Communications*, vol. 61, no. 8, pp. 3541–3551, August 2013.

[21] L. Wang and H. Wu, "Fast pairing of device-to-device link underlay for spectrum sharing with cellular users," *IEEE Communications Letters*, vol. 18, no. 10, pp. 1803–1806, 2014.

[22] R. Afolabi, A. Dadlani, and K. Kim, "Multicast scheduling and resource allocation algorithms for OFDMA-based systems: A survey," *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 240–254, Jan 2013.

[23] H. Meshgi, D. Zhao, and R. Zheng, "Joint channel and power allocation in underlay multicast device-to-device communications," accepted in *IEEE International Conference on Communications (ICC) 2015*.

[24] X. Lin, R. Ratasuk, A. Ghosh, and J. G. Andrews, "Modeling, analysis, and optimization of multicast device-to-device transmissions," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4346–4359, 2014.

[25] X. Wu, Y. Chen, X. Yuan, and M. E. Mkiramweni, "Joint resource allocation and power control for cellular and device-to-device multicast based on cognitive radio," *IET Communications*, vol. 8, no. 16, pp. 2805–2813, 2014.

[26] K. Doppler, M. P. Rinne, P. Janis, C. Ribeiro, and K. Hugl, "Device-to-device communications; functional prospects for LTE-advanced networks," in *IEEE International Conference on Communications Workshops*, 2009, pp. 1–6.

[27] S. Xu, H. Wang, T. Chen, Q. Huang, and T. Peng, "Effective interference cancellation scheme for device-to-device communication underlaying cellular networks," in *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*. IEEE, 2010, pp. 1–5.

[28] F. A. Onat, A. Adinoyi, Y. Fan, H. Yanikomeroglu, J. S. Thompson, and I. D. Marsland, "Threshold selection for snr-based selective digital relaying in cooperative wireless networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4226–4237, 2008.

[29] A. M. Geoffrion, "Generalized benders decomposition," *Journal of optimization theory and applications*, vol. 10, no. 4, pp. 237–260, 1972.

[30] C. Hua and R. Zheng, "Robust topology engineering in multiradio multichannel wireless networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 3, pp. 492–503, 2012.

[31] M. Chiang, "Geometric programming for communication systems," *Communications and Information Theory*, vol. 2, no. 1/2, pp. 1–154, 2005.

[32] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[33] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[34] B. Kaufman and B. Aazhang, "Cellular networks with an overlaid device to device network," in *42nd Asilomar Conference on Signals, Systems and Computers, 2008*. IEEE, 2008, pp. 1537–1541.