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/TMC.2017.2689768, IEEE Transactions on Mobile Computing

Mode Selection and Resource Allocation in Device-to-Device Communications: A Matching Game Approach

S. M. Ahsan Kazmi, Nguyen H. Tran, *Member, IEEE,* Walid Saad, *Senior Member, IEEE,* Zhu Han, *Fellow, IEEE,* Tai Manh Ho, Thant Zin Oo, and Choong Seon Hong, *Senior Member, IEEE*

Abstract—Device to device (D2D) communication is considered as an effective technology for enhancing the spectral efficiency and network throughput of existing cellular networks. However, enabling it in an underlay fashion poses a significant challenge pertaining to interference management. In this paper, mode selection and resource allocation for an underlay D2D network is studied while simultaneously providing interference management. The problem is formulated as a combinatorial optimization problem whose objective is to maximize the utility of all D2D pairs. To solve this problem, a learning framework is proposed based on a problem-specific Markov chain. From the local balance equation of the designed Markov chain, the transition probabilities are derived for distributed implementation. Then, a novel two phase algorithm is developed to perform mode selection and resource allocation in the respective phases. This algorithm is then shown to converge to a near optimal solution. Moreover, to reduce the computation in the learning framework, two resource allocation algorithms based on matching theory are proposed to output a specific and deterministic solution. The first algorithm employs the one-to-one matching game approach whereas in the second algorithm, the one-to many matching game with externalities and dynamic quota is employed. Simulation results show that the proposed framework converges to a near optimal solution under all scenarios with probability one. Moreover, our results show that the proposed matching game with externalities achieves a performance gain of up to 35% in terms of the average utility compared to a classical matching scheme with no externalities.

Index Terms—Resource allocation, D2D communication, Markov approximation, matching games with externalities, heterogeneous cellular networks.

1 Introduction

To efficiently cope with the rapid increase in wireless traffic, device-to-device (D2D) communications over wireless cellular networks has emerged as a promising technique to boost the capacity and coverage of tomorrow's 5G systems [1]–[3]. Using D2D communication, a D2D transmitter can directly transmit to the D2D receiver without routing its traffic through the cellular base station (BS). The use of D2D communications over cellular networks can significantly improve the network performance in terms of data offload [3], [4], content sharing/dissemination [5], [6], energy efficiency [7], [8], coverage extension [3], and improved spectrum efficiency [9]–[11]. However, reaping the benefits of D2D communications requires meeting significant challenges in terms of resource allocation and interference management [12]–[14].

One of the most critical challenges in D2D is to manage the interference stemming from the reuse of spectrum resources [1]. D2D links can use either the unlicensed spectrum (i.e, out-band) [10] or the licensed spectrum (i.e., in-band) [12] for transmission. In both cases due to spectrum reuse, the D2D transmission links can cause interference to other users in the network. We focus on the use of in-band spectrum (i.e., cellular resources) for D2D communication, as in-band D2D communication can provide better quality of service guarantees compared to the out-band spectrum [11]. Furthermore, in an in-band D2D communication, cellular resources can be allocated to D2D links in either an orthogonal manner, i.e., the D2D connections use reserved resources (the dedicated mode or overlay), or in a non-orthogonal manner, i.e., the D2D connections use same resources as the cellular connections (the shared mode or underlay). In this work, we adopt the *underlay (shared) mode* since it provides a much better spectral efficiency than the dedicated mode, particularly in dense networks.

Then, our challenge is to manage the interference stemming from the reuse of cellular resources between D2D links and regular cellular links. In such a D2D enabled network, both cross tier (i.e., between a D2D pair and cellular user) and co-tier (i.e., between two D2D pairs when in close proximity) interference can occur, which significantly degrades the network performance. Moreover, unlike classical approaches for resource allocation, in a D2D enabled system, the number of choices for allocating resources increases exponentially with the number of D2D pairs. Thus, centralized solutions [12], [13] can no longer cope with the massive overhead in terms of required computation and signaling. Therefore, an efficient resource allocation scheme is required that guarantees interference protection to cellular links and operates in a distributed fashion.

1.1 Related Works

Resource allocation in D2D networks has attracted significant recent attention and a comprehensive survey can be found in [11]. In particular, there has been a number of recent works [12]–[20], that focused on underlay D2D networks. For instance, in [12], the authors optimize the throughput over the shared D2D resources while meeting prioritized cellular service constraints. However, this work is based on a centralized approach that requires significant overhead and is not tailored to the dense nature of D2D networks. In [13], a practical and efficient interference-aware resource allocation scheme is presented for D2D enabled networks. In [12] and [13] resource allocation in D2D communication is completely base-station (BS) controlled. This centralized control can lead to significant overhead for a dense D2D network [3]. Indeed, device-centric architectures are more suitable for dense D2D networks in which a user device is at least able to control his action based on its local information, thus distributing the control in the network [3].

A distributed scheme for resource allocation is studied in [14] to enable ad-hoc D2D networks during uplink transmission of the cellular system. Despite the resulting improvement in the system throughput, this approach requires significant message passing to operate in a distributed manner. In [15], joint power control and reuse partner selection is investigated and shown to have improved performance for D2D systems. Similarly, in [16], a tractable iterative solution is proposed for improving the energy and resource usage in a D2D network, using fractional programming. Moreover in [17], a comprehensive survey on the application of different game-theoretic models for D2D resource allocation problem is demonstrated. In [18],

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/TMC.2017.2689768, IEEE Transactions on Mobile Computing

a coalition game approach is proposed to solve the joint power and channel allocation problem in which D2D and cellular links act as the players. Similarly, in [19], a novel power and channel allocation scheme for a D2D enabled system is studied using matching theory to improve cellular network throughput. However, the works in [15], [16], [18], and [19] do not account for the presence of multiple D2D pairs on the same resource block, which can improve the overall system resource utilization, particularly in dense networks. Moreover, in existing works, such as in [14]–[16], [18], and [19], uplink resources for the D2D communication are considered due to ease of interference management. However, these existing works do not directly extend to the downlink due to the different system dynamics and interference characteristics. Furthermore, downlink is the dominating wireless traffic in 5G and beyond systems [21], thus, novel approaches are needed for the downlink resource reuse in an underlay D2D communication. Moreover, in most of the aforementioned works (except [12], [15], and [18]), a fixed resource sharing approach for D2D communication is considered, which cannot cope with the dynamic channel conditions and buffer status of D2D users.

The use of resource sharing can be an effective solution for interference mitigation in D2D communications. In D2D systems, resource sharing includes mode selection along with resource allocation. Using mode selection, the network can decide whether dedicated resources or shared resources are used for D2D communication. In existing works such as [12], [15] and [18] that consider joint mode selection and resource allocation, it has been observed that the *shared mode* can provide significant improvement in terms of network throughput compared to the dedicated mode, especially for dense networks. Moreover, a mixed mode approach in which D2D links can operate in multiple modes through resource multiplexing has also been studied in [20]. Typically, for mode selection, a binary mode selection variable can be used, where the decisions for the mode are taken at the BS subject to the D2D users' channel conditions and buffer status information. However, under dense deployment scenarios, this centralized control will incur excessive complexity and overhead on the BS. Moreover, a centralized solution for the joint mode selection and resource allocation in D2D enabled cellular systems is still an open issue. Therefore, distributed approaches for such joint problems will be needed. In order to address these shortcomings, one approach is to incorporate learning theory, which will be critical for future deployment of dense networks.

In general, the use of a Markov approximation framework is suitable for solving a number of combinatorial optimization problems with feasible learning features [22]. However, the solutions produced by this framework require complete network information, which may not be scalable with the network size [22], [23]. To address this limitation, the work in [24] and [25] presented a near optimal solution for a joint problem (i.e., user association and resource allocation) in heterogeneous cellular networks. Moreover in [26]–[28], other learning approaches are applied to address the resource allocation problem in D2D networks. These works achieved improved system performance by adding the learning aspect to D2D networks. However, these works have ignored the mode selection aspect for D2Ds, which can further improve network throughput performance.

1.2 Contributions And Organizations

The main contribution of this paper is to introduce a distributed scalable solution for a dense D2D network by jointly addressing the problems of mode selection, resource allocation, and interference management aspects. We propose a novel learning framework based on Markov approximation to address these issues. Unsupervised learning is used for mode selection and a two-sided matching game is incorporated to address the resource allocation aspects. The proposed matching game is shown to reduce the computation and configuration size in the framework while enabling a self-organizing and distributed control. Furthermore, we consider a practical scenario in which multiple D2D pairs are allowed to reuse the same resources simultaneously as long as the cellular transmission protection can be guaranteed. In summary, our key contributions include the following:

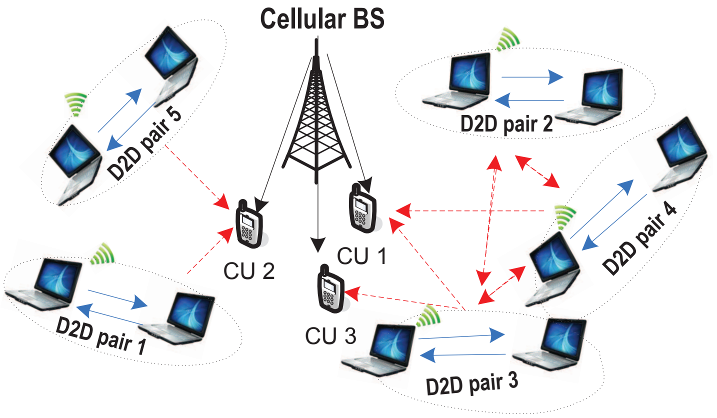


Figure 1: A downlink D2D communication system. The solid line shows the information links while the dashed line shows the interference links.

- First, we formulate the joint problem of mode selection, and resource allocation with an objective to maximize the utility of all D2D pairs subject to interference protection for cellular transmission. The formulated problem is a mixed-integer non-linear optimization problem that is NP-hard and requires exponential computation efforts to obtain the optimal solution.
- Second, to solve the joint problem, we propose a learning framework based on Markov approximation. Furthermore, we design an ergodic Markov chain and the transition probabilities, which makes the Markov chain converge to its stationary probabilities. Using these transition probabilities, we propose a novel two phase algorithm to perform mode selection and resource allocation in the respective phases. This distributed algorithm eventually converges to the near optimal solution in probability with a bounded performance gap between the optimal and converged solutions.
- Third, in order to reduce the computation and configuration size in Markov approximation, we propose two algorithms for resource allocation based on matching theory. Furthermore, we prove the stability, convergence, and optimality of the matching based resource allocation algorithms.
- Simulation results show the convergence, optimality gap, and utility gains achieved using the proposed framework. Results show that the framework converges to a near optimal solution. Moreover, our results show that the proposed matching game with externalities achieves a performance gain of up to 35% in terms of the average utility compared to a classical matching scheme with no externalities.

The rest of this paper is organized as follows. Section 2 presents the system model and problem formulation. Section 3 describes in detail how we map the proposed optimization problem into the learning framework and derive a distributed algorithm. Resource allocation via matching theory is discussed in Section 4. In Section 5, we present the simulation results analysis to validate the performance of our proposed solution. Finally, conclusions are drawn in Section 6.

2 System Model and Problem Definition

Consider the downlink of a cellular network consisting of a single BS and a set $\mathcal{K}$ of K D2D pairs located under its coverage, as shown in Fig. 1. The choice of downlink reflects the worst case interference scenario.[1] We use the index k_0 to indicate the BS. We let set $\mathcal{C}$ be the set of C cellular users. The BS and D2D pairs use the same set $\mathcal{R}$ of R orthogonal resource blocks (RBs).[2] For any given RB $r \in \mathcal{R}$, a predefined interference threshold I^r_{max} must be maintained for protecting the cellular users. Our system model is focused on a dense communication environment in which the density of the users is higher than the number of connections that a given BS can support (e.g., a football stadium). Typically, in such an environment, congestion occurs due to the high number of connections. Therefore, D2D communication can be used to improve the area spectral efficiency and increase the number of connected devices per shared RBs.

2.1 Resource Allocation and Link Model

In our model, the D2D transmissions are synchronized to the cellular transmissions. We assume that all transmitters (BS and D2D pairs) transmit using a fixed power [29] within the RB duration. However, each transmitter can have its individual value for the power budget. In addition, we assume that the transmit power of each transmitter is equally divided among its RBs and thus, the interference power is constant. The D2D pairs at each time slot need to determine which RB is feasible in order to maximize the utility of the system while protecting the cellular users. For RB allocation, we introduce the binary variables x^r_k:

$$x^r_k = \begin{cases} 1, & \text{if D2D pair } k \text{ is assigned RB } r, \\ 0, & \text{otherwise.} \end{cases}$$

The received signal to interference noise ratio (SINR) pertaining to the transmission of the D2D pair k over RB r with transmit power P^r_k is:

$$\gamma^r_k = \frac{x^r_k P^r_k g^r_k}{P^r_{k_0} g^r_{k_0,k} + \sum_{i \in \Omega_r, i \neq k} x^r_i P^r_i g^r_{i,k} + \sigma^2}, \tag{1}$$

where the RB gain over the link of D2D pair k is g^r_k, $g^r_{i,k}$ represents the RB gain between D2D pair i and D2D pair k, and $g^r_{k_0,k}$ is the RB gain from the BS to D2D pair k. $P^r_{k_0}$ and $P^r_i, \forall i \in \Omega_r$, represent the transmit powers of the BS and the other D2D pairs, respectively, and Ω_r is the set of D2D pairs which are using RB r. Note that, the set of D2D pairs Ω_r using RB r is updated dynamically. The noise power is assumed to be σ^2. Similarly, the SINR of cellular user c over RB r is given as:

$$\gamma^r_c = \frac{P^r_{k_0} g^r_{k_0,c}}{\sum_{i \in \Omega_r} x^r_i P^r_i g^r_{i,c} + \sigma^2}, \tag{2}$$

where $g^r_{k_0,c}$ and $g^r_{i,c}$ represent the RB power gains from the BS to cellular user c and D2D pair i to cellular user c, respectively. Note that $\sum_{i \in \Omega_r} x^r_i P^r_i g^r_{i,c}$ is the interference experienced by the cellular user c from a set of D2D pairs Ω_r that use RB r. Then, the data rate of any user $u \in \mathcal{K} \setminus \{k_0\} \cup \mathcal{C}$ on RB r is represented as follows:

$$R^r_u = W^r \log(1 + \gamma^r_u), \tag{3}$$

where W^r is the bandwidth of RB r.

2.2 D2D Decision and Mode Selection model

Next, we present the models for D2D decision and mode selection used in our system. In the D2D decision model, each D2D pair acts based on its achieved utility. The action here represents the D2D decision to use a given mode or not. We assume that each D2D pair selfishly and rationally acts in a way that maximizes its utility. Moreover, each D2D pair has knowledge of its own utility functions. Therefore, each D2D pair only acts to maximize its own utility. A decision variable α_k is used to indicate if D2D pair k will follow a specific mode, as follows:

$$\alpha_k = \begin{cases} 1, & \text{if D2D pair } k \text{ uses the mode,} \\ 0, & \text{otherwise.} \end{cases}$$

This D2D decision model assists the BS in the mode selection process. For mode selection, we consider two modes that can be selected for RB allocation for the D2D pairs. Motivated by the resource utilization gain achieved by the reuse mode, we only employ the reuse mode in our model. However, we propose to classify the reuse mode for our network into two modes:

- *Partial reuse mode:* Only one D2D pair can be allocated to an RB currently in use by a cellular user, only if the interference is below a pre-defined threshold. By using this mode, there exists no co-tier interference (i.e., between D2D pairs). This mode is suitable for scenarios in which the number of D2D pairs is limited compared to the RBs or the D2D pairs are in close proximity with each other.
- *Full reuse mode:* A group of D2D pairs can share an RB only if the interference produced by this group is below the predefined threshold for protecting the cellular tier. However, by using this mode, co-tier interference will also occur. This mode is preferred in the scenario where there exist a large number of D2D pairs compared to RBs. Moreover, this mode can further enhance the RB efficiency, if co-tier interference is well handled.

However, in any given time slot only one mode will be activated for use in the network [20]. A binary variable y is defined to represent the two modes, controlled by the BS:

$$y = \begin{cases} 1, & \text{partial-reuse mode,} \\ 0, & \text{full-reuse mode.} \end{cases}$$

In contrast to previous works [12], [15], [18] and [20], in our model, the BS does not choose a mode for individual D2D pairs based on their channel conditions and buffer status. Here, the BS chooses a mode depending upon the utility achieved by the network. This significantly reduces the computational load since the BS will only need to calculate the utility of the network. However, to obtain the utility for the network, the D2D pairs and BS need to respectively learn which D2D users can be successfully admitted under which mode such that the global network utility is maximized.

2.3 Problem Formulation

Our goal is to maximize a utility function that captures the sum rate of the D2D pairs by selecting the optimal mode for communication, admitting the best D2D pairs, and properly reusing the RBs already occupied by the cellular tier. Therefore, we define the utility function of the D2D network as follows:

$$U(y, \boldsymbol{\alpha}, \boldsymbol{x}) = \sum\nolimits_{k \in \mathcal{K}} \sum\nolimits_{r \in \mathcal{R}} [y \alpha_k R^r_k + (1 - y) \alpha_k R^r_k]. \tag{4}$$

Here, we note that a D2D pair can only use a given RB if the interference level is less than the predefined interference threshold I^r_{max} set by the BS on each r. Moreover, the interference experienced by cellular user c over RB r from a D2D pair k is given by $I^r_k = \alpha_k x^r_k P^r_k g^r_{k,c}$. Note that the binary D2D decision α_k and RB allocation variables x^r_k ensure that we only account for the interference created by the D2D pair that use the given mode and is assigned the same RB. Then our considered joint mode selection and RB allocation (JMARA) problem can be stated as follows:

[1] The developed methodology can also be applied to the uplink case by simply considering the protection of cellular BS.

[2] One resource block can correspond to one sub-carrier of the OFDM-based LTE network.

$$\textbf{JMARA:} \underset{y,\boldsymbol{\alpha},\boldsymbol{x}}{\text{maximize}} \;\; U(y,\boldsymbol{\alpha},\boldsymbol{x}) \tag{5}$$

$$\text{s.t.} \;\; \sum\nolimits_{r\in R} x_k^r \leq 1, \;\; \forall k \in \mathcal{K}, \tag{6}$$

$$y\alpha_k I_k^r + \sum_{k=1}^{|\Omega_r|} (1-y)\alpha_k I_k^r \leq I_{\text{max}}^r, \;\; \forall r \in \mathcal{R}, \tag{7}$$

$$x_k^r \in \{0,1\}, \;\; \forall k \in \mathcal{K}, \forall r \in \mathcal{R}, \tag{8}$$

$$\alpha_k \in \{0,1\}, \;\; \forall k \in \mathcal{K}, \tag{9}$$

$$y \in \{0,1\}. \tag{10}$$

In **JMARA**, the first constraint (6) ensures that each D2D transmitter can be allocated to only one RB. The condition in (6) is used to better manage the interference stemming from D2D communications. The second constraint (7) ensures the protection of cellular user by keeping the interference produced by D2D transmitters below a predefined threshold under either partial-reuse mode ($y = 1$) or full-reuse mode ($y = 0$). Finally, the binary indicator variables for RB allocation x_k^r, D2D decision α_k, and mode selection y are represented by constraints (8), (9) and (10), respectively. The problem **JMARA** is a non-convex, integer problem, which is difficult to solve in practical settings with a large set of D2D pairs and RBs [30]. Thus, we adopt a Markov approximation [22], [23] framework to solve **JMARA** because of its ability to solve combinatorial problems, which will be presented in the next section.

3 JMARA via Markov Approximation

Our proposed solution framework is composed of two steps. The first step is to create a log-sum-exp approximation and the second step is to derive the Markov chain for our problem.

We let $f = \{y, \boldsymbol{\alpha}, \boldsymbol{x}\}$ be a network configuration and $\mathcal{F}$ be the set of all F feasible configurations defined by constraints (6) and (7). For ease of presentation, we let $U_f = U(y, \boldsymbol{\alpha}, \boldsymbol{x})$. Therefore, **JMARA** can be written as

$$\max_{f\in\mathcal{F}} \; U_f. \tag{11}$$

However, U_f in not differentiable. Thus, we transform (11) from a discrete function of f to an equivalent continuous function of p_f (i.e., an equivalent maximum weight independent set (MWIS) problem) as:

$$\begin{aligned} \max_{\boldsymbol{p}\geq 0} \quad & \sum_{f\in\mathcal{F}} p_f U_f \\ \text{s.t.} \quad & \sum_{f\in\mathcal{F}} p_f = 1, \end{aligned} \tag{12}$$

where p_f represents the probability of choosing configuration f, i.e., the weight of the configuration. p_f can be viewed as the fraction of the time a configuration f is activated. Note that, both problems given in (11) and (12) have the same optimal value [22]. However, (12) is still challenging to solve due to the combinatorial nature of the variables. Next, to solve this combinatorial problem, we use the Log-sum-exp Approximation.

3.1 Step 1: Log-sum-exp Approximation

The Log-sum-exp function is a convex and closed function [22] mainly used by machine learning algorithms as a smooth approximation of the max function. Therefore, we interpreted it as a differentiable approximation of the max function given in (11) [30, pp. 72]. Hence, we have:

$$\max_{f\in\mathcal{F}} U_f \approx g_\beta(U_f) = \frac{1}{\beta} \log \left[\sum_{f\in\mathcal{F}} \exp(\beta U_f) \right], \tag{13}$$

where β is a positive constant. Furthermore, the approximation gap is upper-bounded by F, where F is the size of the set $\mathcal{F}$, and $U_{\text{max}} = \max\limits_{f\in\mathcal{F}} U_f$, and then the approximation accuracy will be [30]:

$$0 \leq |U_{\text{max}} - g_\beta(U_f)| \leq \frac{1}{\beta} \log F. \tag{14}$$

Clearly, as $\beta \to \infty$, $\frac{1}{\beta}\log F \to 0$, which renders the approximation exact. The following problem is equivalent to solving the log-sum-approximation in (13) [22], [30]:

$$\begin{aligned} \max_{\boldsymbol{p}\geq 0} \quad & \sum_{f\in\mathcal{F}} p_f U_f - \frac{1}{\beta} \sum_{f\in\mathcal{F}} p_f \log p_f \\ \text{s.t.} \quad & \sum\nolimits_{f\in\mathcal{F}} p_f = 1, \end{aligned} \tag{15}$$

where the first term in (15) represents the MWIS objective and the second term represents the entropy term. We can obtain the optimal probability distribution p^* by solving the Karush-Khun-Tucker (KKT) condition for the above problem [30], given as follows $\forall f \in \mathcal{F}$:

$$p_f^*(U_f) = \frac{\exp(\beta U_f)}{\sum\limits_{f'\in\mathcal{F}} \exp(\beta U_{f'})} = \frac{1}{\sum\limits_{f'\in\mathcal{F}} \exp(\beta(U_{f'} - U_f))}, \tag{16}$$

where $(U_{f'} - U_f)$ is the difference in utilities. The optimal solution in (16) presents an implicit solution for (15) that differs from (12) by an entropy term $-\frac{1}{\beta}\sum_{f\in\mathcal{F}} p_f \log p_f$. Furthermore, the solution to (15) requires complete information of $\mathcal{F}$, which is typically unknown due to a large computational space. Thus, to find $\mathcal{F}$, a computationally exhaustive approach is needed, which is not practical.

3.2 Step 2: Markov Chain (MC)

The solution given in (16) is not practical since complete information on all feasible configurations $\mathcal{F}$ is required, which is not possible as discussed in Section 3.1. Hence, we view (16) as a Markov chain. To this end, each configuration f corresponds to a state with (16) being its stationary distribution. Then, the goal is to derive the Markov chain for the problem given in (15) and reach to the optimal stationary distribution given in (16) that represents its solution. From [22], it is shown that there exists at least one continuous-time time-reversible ergodic Markov chain with stationary distribution $p_f^*(U_f)$ for any probability distribution of the product form $p_f^*(U_f)$ presented in (16).

In order to construct a time-reversible Markov chain with stationary distribution $p_f^*(U_f)$, we let configuration f, $f' \in \mathcal{F}$ be the states of a time-reversible ergodic Markov chain and let $q_{(f\to f')}$ and $q_{(f'\to f)}$ denote the nonnegative transition rates from states $f \to f'$ and $f' \to f$, respectively. Then, the following two conditions are sufficient for the Markov chain design [22]:

- any two states are accessible from each other.
- the local balanced equation satisfies (17), $\forall f, f' \in \mathcal{F}$,

$$\begin{aligned} p_f^*(U_f)\; q_{(f\to f')} &= p_{f'}^*(U_{f'}) q_{(f'\to f)}, \\ \exp(\beta U_f) q_{(f\to f')} &= \exp(\beta U_{f'}) q_{(f'\to f)}. \end{aligned} \tag{17}$$

This balance equation is useful because it eliminates the need for complete information of all possible configurations $\mathcal{F}$. Any $q_{(f\to f')}$ and $q_{(f'\to f)}$ values can be used for the design of the algorithm as long as (17) is satisfied. Therefore, we limit the number of configurations to f and f', i.e., $\mathcal{F} = \{f, f'\}$. We set the conditional probabilities as the transition rates, i.e., $q_{(f'\to f)} = p_{f|\{f,f'\}}^*(U_f)$ and $q_{(f\to f')} = p_{f'|\{f,f'\}}^*(U_f')$. Hence, we obtain

$$\begin{aligned} p_{f|\{f,f'\}}^*(U_f) + p_{f'|\{f,f'\}}^*(U_f') &= 1, \\ q_{(f'\to f)} + q_{(f\to f')} &= 1. \end{aligned} \tag{18}$$

Thus, by solving (17) and (18) we obtain the transition probabilities as a logistic function of utility difference as

$$q_{(f\to f')} = (1 + \exp[\beta(U_f - U_{f'})])^{-1}, \tag{19}$$

$$q_{(f'\to f)} = (1 + \exp[\beta(U_{f'} - U_f)])^{-1}. \tag{20}$$

These transition probabilities are used to derive the Markov chain towards the optimal solution in (16). However, we cannot design a

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/TMC.2017.2689768, IEEE Transactions on Mobile Computing

Algorithm 1 Learning Algorithm (LA)

```
1:  **initialize**:
    $\mathrm{i}^{[1]} \leftarrow 0,\ y^{[1]} \leftarrow \text{rand} \in \{0,1\},$
    $\Upsilon_k(1) \leftarrow 0,\ \omega_k \leftarrow 1,\ \beta_k(1) \leftarrow 1,\ \forall k.$
    $\alpha_k{}^{[1]} \triangleq [\emptyset]_{k\in\mathcal{K}},\ x_k{}^{[1]} \triangleq [\emptyset]_{k\in\mathcal{K}}.$
2:  **while** $t \neq T$ and $\mathcal{D} \neq \emptyset$ and $\omega_0 > 0$ **do**
    *Phase 1:* **Mode Selection**:
3:      **if** $\mathrm{i}^{[t]} \neq 1$ **then**
4:          $\{y,\alpha\}_k^{[t+1]} \leftarrow \begin{cases} \text{rand}\{y,\alpha\}_k & \text{with prob. } \omega_k\,, \\ \{y,\alpha\}_k^{[t]} & \text{with prob. } 1-\omega_k\,. \end{cases}$
5:      **else**
6:          Calculate $\nu \leftarrow q_{(f_k \rightarrow f_k{}')}$ using (21).
7:          $\{y,\alpha\}_k^{[t+1]} \leftarrow \begin{cases} \{y,\alpha\}_k^{[t-1]} & \text{with prob. } \nu, \\ \{y,\alpha\}_k^{[t]} & \text{with prob. } 1-\nu. \end{cases}$
8:          $\beta_k(n+1) \leftarrow \beta_k(n) * \beta_{step}.$
9:          $\Upsilon_k(n-1) \leftarrow \Upsilon_k(n).$
10:         $\Upsilon_k(n) \leftarrow \{y,\alpha\}_k^{[t+1]}.$
11:         **if** $\Upsilon_k(n-1) = \Upsilon_k(n)$ **then**
12:             $\omega_k \leftarrow \max\{0, \omega_k - \omega_{step}\}.$
13:     $\mathrm{i}^{[t+1]} \leftarrow 1 - \mathrm{i}^{[t]}.$
    *Phase 2:* **Resource Allocation**:
14:     **if** $y^{[t+1]} = 1$ **then**
15:         Run Alg. 2 for $\boldsymbol{\alpha}^{[t+1]}$ to obtain $\boldsymbol{x}^{[t+1]}$.
16:     **else**
17:         Run Alg. 3 for $\boldsymbol{\alpha}^{[t+1]}$ to obtain $\boldsymbol{x}^{[t+1]}$.
18:     Calculate utility $U_{k,f_k}^{[t+1]}$,$\forall\mathcal{D}$.
    *Phase 3:* **Update**:
19:     Update $y^{[t+1]}, \alpha_k^{[t+1]}, x_k^{[t+1]}$.
20:     **if** $\omega_k = 0$ **then**
21:         $\mathcal{D} \leftarrow \mathcal{D} \setminus \{k\}$.
```

distributed and scalable algorithm using (19) and (20), which use the global utility, i.e., U_f. Due to the distributed nature of the network, a player k is only aware of its own individual local utility U_{f_k} without additional signaling and overhead. Therefore, we define $U_{k,f_k} = U_k(m, \alpha_k, x_k)$ as the local utility for each player k. Then, we substitute the local utilities in (19) and (20) to obtain

$$q_{(f_k \rightarrow f_k')} = (1 + \exp[\beta_k(U_{k,f_k} - U_{k,f_k'})])^{-1}, \tag{21}$$

$$q_{(f_k' \rightarrow f_k)} = (1 + \exp[\beta_k(U_{k,f_k'} - U_{k,f_k})])^{-1}. \tag{22}$$

Hence, the Markov chain based on using these local utilities converges to a distribution $\widetilde{p}_f(U_f)$ instead of $p_f^*(U_f)$ given in (16). However, the gap between this distribution $\widetilde{p}_f(U_f)$ and the optimal $p_f^*(U_f)$ is also bounded [23], [24].

3.3 Learning Algorithm

Next, based on the analysis of the Markov chain in Section 3.2, we present the learning algorithm shown in Alg. 1 for solving the modeled Markov chain. The algorithm consists of three phases: (i) the *mode selection* phase (lines 3-13), (ii) *resource allocation* phase (lines 14-18), and (iii) *update* phase (line 19-21) as illustrated in Fig. 2. In *Phase 1*, we use unsupervised learning using the *logistic equations* given by (21) and (22). The learning approach uses properties from log-linear learning [31] and simulated annealing [32] for selection of the control variables action (i.e., y and $\boldsymbol{\alpha}$). For our scenario, the BS chooses the mode action and all D2D pairs choose their admission action. In *Phase 2*, resource allocation (i.e., $\boldsymbol{x}$) is performed (details of resource allocation are presented in Sec. 4), for the given mode and D2D pairs that decide to use this mode. Once the first two phases are executed, all control variables are updated in *Phase 3*.

In line 1 of Alg. 1, all the control variables and the auxiliary variables are initialized. We introduce the auxiliary variables as $\boldsymbol{\Upsilon_k} = [\Upsilon_1, ..., \Upsilon_{|n|}]$, $\mathrm{i}^{[t]}$, β_k and ω_k. Here, these auxiliary variables are used to control the mixing characteristics and stopping time for the underlying Markov chain. The vector Υ_k is used for convergence analysis, and $\mathrm{i}^{[t]}$ is an experimentation indicator that indicates whether or not experimentation takes place at time slot t. β_k controls the gap given in (14) and ω_k balances between exploration and exploitation rates. As explained earlier, as $\beta_k \rightarrow \infty$, the gap $\frac{1}{\beta_k} \log F_k \rightarrow 0$ and the β_k update control the mixing of the Markov chain [33], which can be either linear or geometric. We implement the geometric update (line 8), which gradually yields zero gap.

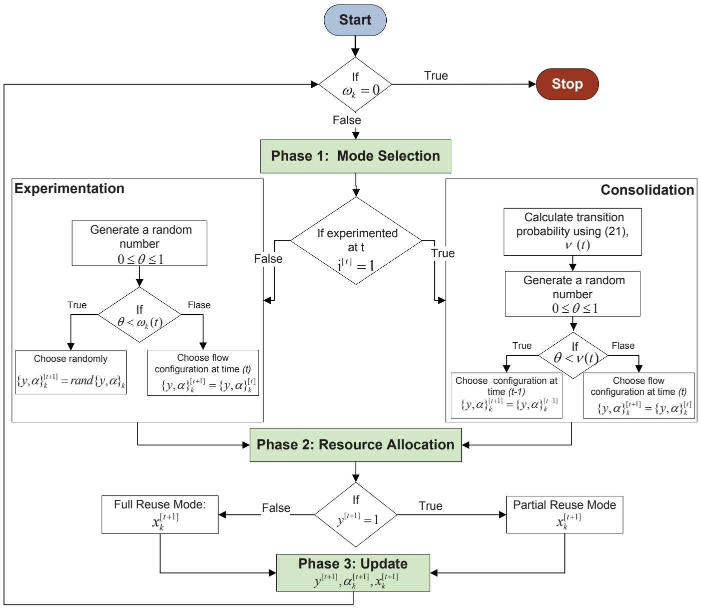

Start
If
$\omega_k = 0$
True
Stop
False
Phase 1: Mode Selection
Experimentation
Consolidation
If experimented at t
$\mathrm{i}^{[t]} = 1$
False
True
Generate a random number
$0 \le \theta \le 1$
If
$\theta < \omega_k(t)$
True
Flase
Choose randomly
$\{y,\alpha\}_k^{[t+1]} = rand\{y,\alpha\}_k$
Choose flow configuration at time (t)
$\{y,\alpha\}_k^{[t+1]} = \{y,\alpha\}_k^{[t]}$
Calculate transition probability using (21), $\nu(t)$
Generate a random number
$0 \le \theta \le 1$
If
$\theta < \nu(t)$
True
Flase
Choose configuration at time (t-1)
$\{y,\alpha\}_k^{[t+1]} = \{y,\alpha\}_k^{[t-1]}$
Choose flow configuration at time (t)
$\{y,\alpha\}_k^{[t+1]} = \{y,\alpha\}_k^{[t]}$
Phase 2: Resource Allocation
If
$y^{[t+1]} = 1$
False
True
Full Reuse Mode:
$x_k^{[t+1]}$
Partial Reuse Mode
$x_k^{[t+1]}$
Phase 3: Update
$y^{[t+1]}, \alpha_k^{[t+1]}, x_k^{[t+1]}$

Figure 2: Block diagram of learning algorithm (LA).

The learning algorithm starts by the BS (i.e., k_0) selecting a random mode $y^{[1]}$ (i.e., partial or full-reuse mode) when there exists no configuration (line 1). In *Phase 1*, the set of players $\mathcal{D} \subseteq \mathcal{K}$ either performs experimentation or consolidation. In experimentation, for time slot $t+1$, each player executes one of the two actions, i.e., a new random configuration is chosen (exploration) or it stays with the current configuration (exploitation) with probability ω_k or $1-\omega_k$, respectively (line 4). During consolidation, the current utility obtained at time slot t is compared with the previously achieved utility at time slot $t-1$ by all players. Then, each player probabilistically (i.e. with probability ν) chooses its action for time slot $t+1$. Furthermore, the actions that achieve the maximum utility have a higher probability to be chosen (lines 5-7). Furthermore, as the Markov chain moves towards convergence (line 11), we reduce the exploration rate by a constant step size (line 12). Note that all players are aware of the their own utility received, the configuration employed for the last two time slots, and whether or not they experimented in the last time slot.

After *Phase 1* is completed for time slot $t+1$, *Phase 2* starts. In this phase (details of this phase are presented in Section 4), based on the players actions, a resource allocation algorithm is executed to obtain the resource allocation vector $\boldsymbol{x}^{[t+1]}$ (lines 14-17). Then, the utility of configuration $U_{k,f_k}^{[t+1]}$ is evaluated for all players $\mathcal{D}$.

Finally, we update both the the control variables in *Phase 3* for the next time slot. Moreover, as the exploration rate ω_k approaches zero, we remove the player from learning, as it operates in the best configuration (lines 20-21). These three phases are repeated until an equilibrium is reached (line 2), i.e., the underlying Markov chain converges to the stationary distribution. Moreover, in our learning framework, the matching algorithm outputs a specific and deterministic solution for resource allocation. This matching outcome is then used in the learning framework as a joint configuration with D2D decision and mode selection. Since, the overall framework is based on an ergodic Markov chain, after a sufficiently large number of time slots T, it converges in probability to a near optimal solution [22], [24] and [25].

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/TMC.2017.2689768, IEEE Transactions on Mobile Computing

4 Resource Allocation via Matching

Once *Phase 1* of Alg. 1 is executed, we obtain the mode y as well as the D2D pairs $\boldsymbol{\alpha}$ that use the selected mode at time slot $t+1$. The next goal here is to perform RB allocation for the given mode and D2D pairs. For a given mode selection variable y, problem **JMARA** can be divided into two combinatorial problems, depending upon the value of y. In this section, we apply matching theory for solving these problems under two cases: the partial-reuse or the full-reuse modes. The motivation to apply matching theory for the RB allocation problem is its ability to tackle combinatorial problems and achieve a distributed solution [34], [35]. The benefits of matching theory come from the distributed nature of control in the system. Furthermore, matching theory allows each player (i.e., D2D pairs and RBs) to define its individual utilities depending upon its local information.

4.1 Case 1: Partial-Reuse Mode

In the partial reuse mode, i.e., $y = 1$, only one D2D pair can use an RB if the interference level is less than the predefined interference threshold $I^r_{\max}$ set by the BS. Then, we can state the following problem, as derived directly from **JMARA**:

$$\textbf{PR:} \underset{x^r_k \in \boldsymbol{x}}{\text{maximize}} \quad \sum_{r\in\mathcal{R}} \sum_{k\in\mathcal{K}} R^r_k \tag{23}$$

$$\text{subject to} \quad (6), (8),$$

$$I^r_k \leq I^r_{\max}, \ \forall r \in \mathcal{R}. \tag{24}$$

In **PR**, the objective is reduced to maximizing the sum-rate of all D2D pairs by assigning the RBs. The constraint given by (24) ensures the protection of cellular users by keeping the interference produced by the D2D transmitter below a predefined threshold. This allows the re-usability of an RB r to increase RB efficiency if the interference constraint can be maintained. Problem **PR** is still a combinatorial problem, and finding the solution becomes NP-hard, for a large set of D2D pairs and RBs in a practical amount of time [30]. Note that **PR** is desired to be solved in a distributed manner by each D2D pair such that it maximizes its own rate. Therefore, we use matching theory to map the problem **PR** into a matching game and then discuss the details of the solution in the following subsections.

4.1.1 Matching Game Formulation

We formulate the RB allocation as a two-sided matching game, then we define the utility and finally present a matching algorithm that can find a stable matching which is a key concept for a matching game.

We assume each D2D pair forms a set that can use a single RB. However, to use this RB, the interference produced by D2D pairs to RBs should be under the tolerable predefined interference level, i.e., constraint (24). Similarly, every RB also forms a set to accommodate a D2D pair among all the pairs. Therefore, our design corresponds to a *one-to-one matching* given by the tuple $(\mathcal{K}, \mathcal{R}, \succ_{\mathcal{K}}, \succ_{\mathcal{R}})$. Here, $\succ_{\mathcal{K}} \triangleq \{\succ_k\}_{k\in\mathcal{K}}$ and $\succ_{\mathcal{R}} \triangleq \{\succ_r\}_{r\in\mathcal{R}}$ represent the set of preference relations of D2D pairs and RBs, respectively. Formally, we define the matching as follows:

Definition 1. *A* matching μ *is defined by a function from the set* $\mathcal{K} \cup \mathcal{R}$ *into the set of elements of* $\mathcal{K} \cup \mathcal{R}$ *such that* $k = \mu(r)$ *if and only if* $r = \mu(k)$.

4.1.2 Preference Profiles of Players

Matching is performed by the two sets of players using preference profiles. For each player, the preference profile is used to rank the players of the opposite side. In the proposed game, the two sides, D2D pairs and RBs, will build their preference profiles by utilizing local information available at each side. The preference profile for the D2D pairs is based on the following preference function of the achievable data rate on RB r:

$$U_k(r) = W^r \log(1 + \gamma^r_k). \tag{25}$$

The intuition for such a preference function comes from the objective of problem **PR**, where each D2D pair wants to maximize its sum rate. Hence, each D2D pair ranks all the RBs r in a non-increasing order in its preference profile represented by $\mathcal{P}_k$. Note that an RB $r \in \mathcal{R}$ that produces a higher utility (consequently the data rate achieved by using the more preferred RB is higher) according to (25) will be preferred over an RB $r' \in \mathcal{R}$ by a D2D pair k, i.e., $r \succ_k r'$, for carrying out its transmission and will thus be placed higher in its preference profile.

Similarly, each RB r also needs to have a preference profile that ranks all the D2D pairs $k \in \mathcal{K}$ according to its preference function. By using a two-sided matching game for our problem so we can guarantee cellular tier protection by the RB defined preferences. This is important for the proposed game to guarantee (24). This is one of the main motivations for using a two-sided matching game for our problem. Moreover, the preference list for each RB is formed by the BS. The information required at the BS includes the power level of the D2D transmitters p_k, the predefined maximum interference threshold $I^r_{\max}$ for each RB, and the RB power gain between the D2D transmitter and cellular user $g^r_{k,c}$. The preference function is given by:

$$U_r(k) = \max\left(I^r_{\max} - I^r_k, \ 0\right). \tag{26}$$

According to this preference function, an RB gives less utility to a D2D pair k, which creates more interference. Additionally, all D2D pairs that violate (24) receive a zero utility and are ranked as the lowest in the preference profile of r. Furthermore, to calculate the ranking of each D2D pair, the BS for each r needs to calculate the interference I^r_k induced by the D2D pair k if an RB r is in use. As we assume the power levels of the D2D pair are fixed and known to the BS, the calculation of I^r_k only depends on the RB gain $g^r_{k,c}$.

Here, we note that RB power gain $g^r_{k,c}$ can be estimated by cellular users and sent back to the BS by using the pilot signal or any standard RB estimation technique [29]. The total interference for each cellular user can be estimated as follows. All cellular users estimate the total received power and send this value to the BS. The BS can then calculate the interference induced by the D2D pair on RB r. Therefore, calculation of the interference only requires the standard RB estimation of $g^r_{k,c}$. In addition, signaling is only involved in sending these values from the cellular user to the BS, which only occurs once during the initialization phase. Once this information is acquired, I^r_k is calculated and the BS ranks each D2D pair k for each RB r in the preference profile of r represented by $\mathcal{P}_r$.

4.1.3 Resource Allocation Algorithm

We present the RB allocation algorithm based on the proposed matching game. The aim of this algorithm is to find a stable allocation that is a key solution concept in matching theory [36], [37] and can be defined as follows:

Definition 2. *A matching* μ *is* stable *if there exists no blocking pair* (k, r), *where* $k \in \mathcal{K}, r \in \mathcal{R}$, *such that* $r \succ_k \mu(k)$ *and* $k \succ_r \mu(r)$, *where* $\mu(k)$ *and* $\mu(r)$ *represent, the current matched partners of* k *and* r, *respectively.*

In our game, a stable solution ensures that no matched D2D pair would benefit from deviating from their assigned RB r with a new RB r'. The output of our algorithm is the RB allocation vector $\boldsymbol{x}$ of D2D pairs that maximizes the objective of the optimization problem **PR**, and the pseudo code is given in Alg. 2. The presented algorithm is guaranteed to converge to a stable allocation as it is a variant of the well-known deferred acceptance algorithm [36].

Alg. 2 has three phases namely, *the initialization phase, the matching phase* and the *RB allocation phase*. In the *initialization phase*, information on the active D2D pairs $\boldsymbol{\alpha}$ and local information required is attained to build the preference profiles (lines 1-3).

Algorithm 2 Partial Reuse-mode Resource Allocation Algorithm

```
1: ***Phase 1: Initialization***:
2: **input**: $\boldsymbol{\alpha}$, $\mathcal{P}_k$, $\mathcal{P}_r$, $\forall r, k$.
3: **initialize**: $t = 0$, $\mu^{(t)} \triangleq \{\mu(k)^{(t)}, \mu(r)^{(t)}\}_{k\in\mathcal{K}, r\in\mathcal{R}} = \emptyset$, ${\mathcal{L}_r}^{(t)} = \emptyset$ ${\mathcal{P}_k}^{(0)} = \mathcal{P}_k$, ${\mathcal{P}_r}^{(0)} = \mathcal{P}_r$, $I^r_{\text{max}}$, $\forall r, k$.
4: ***Phase 2: Matching***:
5: **repeat**
6:     $t \leftarrow t + 1$.
7:     **for** $k \in \mathcal{K}$, propose $r$ according to ${\mathcal{P}_k}^{(t)}$ **do**
8:         **while** $k \notin \mu(r)^{(t)}$ and $\mathcal{P}_k^{(t)} \neq \emptyset$ **do**
9:             **if** $I^r_{\text{max}} \geq I^r_k$ **then**
10:                **if** $k \succ_r \mu(r)^{(t)}$ **then**
11:                    $\mu(r)^{(t)} \leftarrow \mu(r)^{(t)} \setminus k'$.
12:                    $\mu(r)^{(t)} \leftarrow k$.
13:                    $\mathcal{P}'^{(t)}_r = \{k' \in \mu(r)^{(t)} | k \succ_r k'\}$.
14:                **else**
15:                    $\mathcal{P}''^{(t)}_r = \{k \in \mathcal{K} | \mu(r)^{(t)} \succ_r k\}$.
16:            **else**
17:                $\mathcal{P}'''^{(t)}_r = \{k \in \mathcal{K} | I^r_{\text{max}} \leq I^r_k\}$.
18:                ${\mathcal{L}_r}^{(t)} = \{\mathcal{P}'^{(t)}_r\} \cup \{\mathcal{P}''^{(t)}_r\} \cup \{\mathcal{I}'^{(t)}_r\}$.
19:                **for** $l \in {\mathcal{L}_r}^{(t)}$ **do**
20:                    ${\mathcal{P}_l}^{(t)} \leftarrow {\mathcal{P}_l}^{(t)} \setminus \{r\}$.
21:                    ${\mathcal{P}_r}^{(t)} \leftarrow {\mathcal{P}_r}^{(t)} \setminus \{l\}$.
22: **until** $\mu^{(t)} = \mu^{(t-1)}$.
23: ***Phase 3: Resource Allocation***:
24: **output**: $\mu^{(t)}$.
```

In the second phase *matching*, each unassigned D2D pair k proposes to its most preferred RB r according to $\mathcal{P}_k$ (lines 7-8). The BS determines the interference I^r_k produced and evaluates (24). If (24) is violated, the D2D pair k is rejected. Otherwise, the BS checks the preference ranking of the resource r. If ranked higher than the current match $(\mu(r)^t)$, the D2D pair k will be accepted. Otherwise, it will be rejected. Finally, all the rejected D2D pairs at iteration t, i.e., the set ${\mathcal{L}_r}^{(t)}$, are removed by both sides in order to update their preference profiles. The matching process is carried out iteratively until a stable match is found between both sides. The process will terminate when all the D2D pairs that can maintain the interference tolerance level are assigned to RBs or there are no more RBs to propose. The algorithm will converge when the matching of two consecutive iterations t remains unchanged (lines 4-22) [36]. The final stage is the *RB allocation* phase in which the matched D2D pairs are allowed to transmit on the matched RBs (lines 23-24).

Theorem 1. *The stable solution resulting from Alg. 2 is also a local maximum of the* ***PR*** *problem.*

Proof. Please see Appendix A. □

4.2 Case 2: Full-Reuse mode

In the full-reuse mode, i.e., $y = 0$, the BS allows a set of D2D pairs to reuse the RB with a cellular user in such a manner that this allocation does not violate the interference constraint, i.e., I^r_{max} set by the BS. Then, we can state the following problem:

$$\textbf{FR:} \underset{x^r_k \in \boldsymbol{x}}{\text{maximize}} \quad \sum_{r\in\mathcal{R}} \sum_{k\in\mathcal{K}} R^r_k \tag{27}$$

$$\text{subject to} \quad (6), (8),$$

$$\sum\nolimits_{k=1}^{|\Omega_r|} x^r_k P^r_k g^r_{k,c} \leq I^r_{\text{max}} \ . \tag{28}$$

Similar to problem **PR**, the objective in **FR** is to maximize the sum rate of all D2D pairs. However, in **FR**, the constraint given by (28) reflects reuse of the same RB by a set of D2D pairs Ω_r only if the interference is not violated (i.e., I^r_{max}) over RB r. The formulated problem **FR** is also a combinatorial problem and solving **FR** using classical optimization techniques is an NP-hard problem. Here, by relaxing some of the constraints, the complexity of **FR** will remain intractable for a sufficiently large set of RBs and D2D pairs. This motivates the use of matching theory.

4.2.1 Matching Game Formulation

Similar to the partial-reuse mode, in the full-reuse mode there are also two disjoint sets of agents, the set of RBs, $\mathcal{R}$, and the set of D2D pairs, $\mathcal{K}$. Each RB r has a strict, transitive, and complete preference profile $\mathcal{P}_r$ defined over D2D pairs, i.e., $2^{\mathcal{K}}$. Note that under the full-reuse mode, D2D pairs can operate on the same RB, which can cause severe interference to cellular users as well as other D2D pairs operating on the same RBs. This can be observed from (1), the SINR of a D2D pair k. From (6), it is given that each D2D pair can use a single RB. However, different D2D pairs can use the same resource to improve RB efficiency. Therefore in full-reuse mode, the preference profile $\mathcal{P}_k$ of D2D pairs is defined over the RBs, i.e., $\mathcal{R}$. Note that, other D2D pairs k' operating on that RB implicitly affect the preference ranking of the D2D pair k. Therefore, our design corresponds to the *one-to-many matching* given by the tuple $(\mathcal{K}, \mathcal{R}, \succ_{\mathcal{K}}, \succ_{\mathcal{R}})$. Here, $\succ_{\mathcal{K}} \triangleq \{\succ_k\}_{k\in\mathcal{K}}$ and $\succ_{\mathcal{R}} \triangleq \{\succ_r\}_{r\in\mathcal{R}}$ represent the set of preference relations of the D2D pairs and RBs, respectively. Formally, we define the matching as follows:

Definition 3. *A* matching μ *is defined on the set* $\mathcal{K}\cup\mathcal{R}$*, which satisfies for all* $r \in \mathcal{R}$ *and* $k \in \mathcal{K}$*:*

1) $|\mu(k)| \leq 1$ *and* $\mu(k) \in \mathcal{R} \cup \phi$,
2) $|\mu(r)| \leq q_r$ *and* $\mu(r) \in 2^{\mathcal{K}} \cup \phi$,
3) *If* $k \in \mu(r)$ *then* $\mu(k) = r$,
4) *If* $\mu(k) \in r$ *for RB* r *then* $\mu(r) = \mathcal{M}$,

where q_r denotes the quota of RB r, $\mathcal{M} \subset \mathcal{K}$ denotes the set of acceptable D2D pairs who prefer r, and $|\mu(.)|$ denotes the cardinality of the matching outcome $\mu(.)$. Then, the first two conditions here represent constraints given by (6) and (28), respectively, where q_r represents the total tolerable interference I^r_{max} of RB r. Note that, by using q_r, which represents the total tolerable interference, we can make a decision on the number of D2D pairs that can be allocated a given RB r without violating condition (28). Here, $\mu(k) = \phi$ means that k is not matched to any RB. Similarly, if $\mu(r) = \phi$, then there are no D2D pairs matched to RB r.

4.2.2 Preference Profiles of Players

Similar to the partial-reuse mode in the full-reuse mode, the agents on both sides need to rank each other using the preference profile. However, the preference profiles of D2D pairs here depend on the RBs as well as other D2D pairs assigned to that RB. Such interdependence relations are known in matching theory as *externalities* [34], and have important implications in the design of the proposed solution. Due to these externalities, an agent may continuously change its preference order in response to the formation of other agents and thus never reach a final RB allocation unless externalities are well-handled.

In order to build the preference profile of D2D pairs, each D2D pair calculates the achievable data rate for each RB and then ranks them in a descending order. The following preference function is used by each D2D pair:

$$U_k(r, \mu) = W^r \log(1 + \gamma^r_k). \tag{29}$$

Note that, channel gains in LTE-A system are acquired for sub-bands (i.e., group of RBs) rather than for each RB [38]. Then, each D2D pair k will have the same preference over that group of RBs, i.e., the RBs with same gains will result in the same achievable rate, thus, creating ties among these RBs in D2D's preference list. We can simply break all such ties in any arbitrary way and rank them in a strict order to achieve a stable allocation [39]. Thus, for any D2D pair k, a preference relation $\succ_k$ is defined over the set of RBs $\mathcal{R}$ such that, for any two RBs $i, j \in \mathcal{R}, i \neq j$, and two matchings μ and $\mu' \in \mathcal{K} \times \mathcal{R}, i = \mu(k), j = \mu'(k)$,

$$(i, \mu) \succ_k (j, \mu') \Leftrightarrow U_k(i, \mu) > U_k(j, \mu'). \tag{30}$$

Similarly, each RB r creates its preference profile by using the following preference function:

$$U_r(\mathcal{M}, \mu) = \max_i\{|\mathcal{M}_i| : I^r_{\mathcal{M}_i} \le I^r_{\text{max}}\}. \tag{31}$$

According to (31), each RB r chooses a subset of D2D pairs $\mathcal{M}$ such that the interference produced by $\mathcal{M}$ is less than the tolerable interference threshold I^r_{max}. This preference function maximizes the number of elements in $\mathcal{M}$, i.e., it maximizes the D2D pairs. Note that this allows the D2D pairs that produce the lowest interference to be preferred by RB r. The subset with the highest number of elements is the most preferred among all the feasible subsets and ranked accordingly. Moreover, for any RB r, a preference relation $\succ_r$ is defined such that for any two subsets of D2D pairs $\mathcal{M}, \mathcal{N} \in \mathcal{K}$, where $\mathcal{M} \neq \mathcal{N}$, and $\mathcal{M} = \mu(r), \mathcal{N} = \mu'(r)$:

$$(\mathcal{M}, \mu) \succ_r (\mathcal{N}, \mu') \Leftrightarrow U_r(\mathcal{M}, \mu) > U_r(\mathcal{N}, \mu'). \tag{32}$$

Once the matching game and preference profile of both agent sides have been defined, we now aim to find a stable RB allocation scheme for the proposed game. However, it is evident from (29) and (31) that our preferences are a function of the existing matching μ, and from (1), it is clear that the D2D pairs affect each other's performance through co-tier interference. Therefore, in the next subsection, we present a novel approach adopted to handle such externalities.

4.2.3 Preferences and Externalities

Next, we develop a novel approach to handle externalities in the proposed game and analyze its solution. In the proposed game, if D2D pair k is assigned to a RB r, it will produce interference with the cellular user as well as with the neighboring D2D pairs using the same RB r. Consequently, an agent (D2D pair) may change its preference order with regards to a given RB r in response to the action of other agents, i.e., D2D pairs k' that have been assigned to RB r. This may lead to a situation in which agents never reach a final allocation.

Therefore, to build D2D pair preferences that can also handle the externalities, we propose the representation of the initial network as an interference graph. To deal with the externalities caused by neighboring D2D pairs, we use an approach similar to [40], [41]. In a graph, the nodes represent D2D pairs, and the edges indicate the interference between connected nodes. We assume that each D2D pair first evaluates its interfering neighboring D2D pairs. This can be done by assuming two D2D pairs i and k are connected by an edge that satisfies the following condition, i.e., the required signal ratio to the interference signal is below a threshold ζ_k:

$$\frac{P_k g^r_k}{P_i g^r_{i,k}} \le \zeta_k.$$

Here, ζ_k is the predefined thresholds of D2D pair k selected to determine the severity of the interference. This indicates that D2D pair k cannot share the same RB with D2D pair i if an edge exists. Once all the interfering D2D pairs are identified for each D2D pair, the D2D pairs send this set to the BS. We call this set as a conflict set for a D2D pair k and denote it as follows:

$$\mathcal{C}_k = \left\{ k' \in \mathcal{K} : \frac{P_k g^r_k}{P_i g^r_{i,k}} \le \zeta_k \right\}. \tag{33}$$

The main idea here is to restrict the reuse of RBs between D2D pairs who are very close to each other, as this will cause instability and will have an adverse effect on the network.

4.2.4 Resource Allocation Algorithm

In order to find a stable RB allocation scheme, first, we need to define the blocking pair. However, in our formulated game there is an additional challenge of dynamic quota, i.e., the BS allows a number of D2D pairs (with heterogeneous interference) to use each RB as long as the interference constraint on that RB is not violated. This heterogeneous interference of D2D pairs and dynamic quota of resources introduces new challenges in the game similar to [35] and [42]. Moreover, our formulated game has the additional challenge of externalities, which is not addressed in [35] or [42]. Therefore, the blocking pair for the formulated game with dynamic quota and externalities is defined as follows:

Algorithm 3 Full Reuse-mode Resource Allocation Algorithm

```
input: $\boldsymbol{\alpha}, \mathcal{P}^{(t)}_k, \mathcal{P}^{(t)}_r, \mathcal{C}_k, \forall r, k$.
initialize: $t = 0$, $\mu^{(1)} \triangleq \{\mu(k)^{(1)}, \mu(r)^{(1)}\}_{k\in\mathcal{K}, r\in\mathcal{R}} = \emptyset$, $I^r_{res}{}^{(1)} = I^r_{\text{max}}$,
  $\mathcal{J}_r{}^{(1)} = \emptyset$, $C^{(1)}_r = \emptyset, \forall r, k$.
$t \leftarrow t + 1$.
Update $\forall k$, $\mathcal{P}_k{}^{(t)}$ for given $\mu(r)^{(t-1)}$.
$\forall k \in \mathcal{K}$ with $r$ as its most preferred in $\mathcal{P}^{(t)}_k$.
while $k \notin \mu(r)^{(t)}$ and $\mathcal{P}^{(t)}_k \neq \emptyset$ do
    if $I^r_{res}{}^{(t)} < I^r_j$, then
        $\mathcal{P}'^{(t)}_r = \{k' \in \mu(r)^{(t)} | k \succ_r k'\}$.
        $j_{lp} \leftarrow$ the least preferred $k' \in \mathcal{P}'^{(t)}_r$.
        while $(\mathcal{P}'^{(t)}_r \neq \emptyset) \cup (I^r_{res}{}^{(t)} < I^r_j)$ do
            $\mu(r)^{(t)} \leftarrow \mu(r)^{(t)} \setminus j_{lp}$, $\mathcal{P}'^{(t)}_r \leftarrow \mathcal{P}'^{(t)}_r \setminus j_{lp}$.
            $I^r_{res}{}^{(t)} \leftarrow I^r_{res}{}^{(t)} + I^r_{j_{lp}}$ .
            if $I^r_{res}{}^{(t)} < I^r_j$ then
                $j_{lp} \leftarrow k$.
    else
        if $C^{(t)}_r = \{k' \in \mu(r)^{(t)} \cup C_k\} = \emptyset$ then
            $\mu(r)^{(t)} \leftarrow \mu(r)^{(t)} \cup k$, $I^r_{res}{}^{(t)} \leftarrow I^r_{res}{}^{(t)} - I^r_k$.
        else
            $D^{(t)}_r = \{k' \in C^{(t)}_r | k \succ_r k'\}$.
            for $j_{lp} \in D^{(t)}_r$ do
                $\mu(r)^{(t)} \leftarrow \mu(r)^{(t)} \setminus j_{lp}$.
                $I^r_{res}{}^{(t)} \leftarrow I^r_{res}{}^{(t)} + I^r_{j_{lp}}$.
            if $C^{(t)}_r = \{k' \in \mu(r)^{(t)} \cup C_k\} = \emptyset$ then
                $\mu(r)^{(t)} \leftarrow \mu(r)^{(t)} \cup k$, $I^r_{res}{}^{(t)} \leftarrow I^r_{res}{}^{(t)} - I^r_k$.
            else
                $j_{lp} \leftarrow k$.
    $\mathcal{J}_r{}^{(t)} = \{j \in \mathcal{P}_r{}^{(t)} | j_{lp} \succ_r j\} \cup \{j_{lp}\}$.
    for $j \in \mathcal{J}_r{}^{(t)}$ do
        $\mathcal{P}_j{}^{(t)} \leftarrow \mathcal{P}_j{}^{(t)} \setminus r$ $\mathcal{P}_r{}^{(t)} \leftarrow \mathcal{P}_r{}^{(t)} \setminus j$.
output: $\mu^{(t)}$.
```

Definition 4. *A matching μ is said to be* stable *if there exists no* blocking pair (k, r) *such that:*

a) $I^r_{res} \ge I^r_k$, $\quad k \succ_r \emptyset$, $r \succ_k \mu(k)$, and $\mu(r) \notin \mathcal{C}_k$,

b) $I^r_{res} < I^r_k$, $I^r_{res} + \sum_{k'\in\mu(r)} I_{k'}{}^r \ge I^r_k$, $\quad k \succ_r k'$, $r \succ_k \mu(k)$, and $\mu(r) \notin \mathcal{C}_k$,

where $I^r_{res} = I^r_{\text{max}} - \sum_{k\in\mu(r)} I^r_k$ represents the residual of the interference tolerance (remaining quota) on RB r. *The quota of an RB $r \in \mathcal{R}$ is filled when $I^r_{res} < I^r_k$ for a requesting $k \in \mathcal{K}$.* Definition 4 is based on the following intuition. Whenever a D2D pair k prefers an RB r over its assigned RB $\mu(k)$ that does not contain a conflicting D2D pair (i.e., $\mu(r) \notin \mathcal{C}_k$), if either: i) r has sufficient interference tolerance I^r_{res} and is willing to accept k (i.e., $k \succ_r \emptyset$), or ii) its quota is filled but it is able to accept k by rejecting some accepted D2D pairs which are ranked lower than k, then k and r can deviate from their assigned matching to form a blocking pair. A matching is stable only if there exist no blocking pairs.

In contrast to the partial reuse mode, here, the preference profile of the D2D pairs are interdependent with one another through the mutual interference terms, as seen in (1). Therefore, to achieve stability, a sufficient condition is that the formation of any new D2D-RB pair does not undermine the stability of existing matched D2D-RB pairs. By employing such a condition, the preference profile of currently matched D2Ds on an RB will remain unaltered even after this new pair formation. Stability in our solution ensures that after RB allocation, no matched pair (D2D-RB) in the network would benefit from replacing their assigned RB with a new better RB and vice versa.

Next, we present a novel and stable RB allocation algorithm. The algorithm starts by using the local information to build the preference

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/TMC.2017.2689768, IEEE Transactions on Mobile Computing

profiles (lines 1-3) similar to Alg. 2. At each iteration t, each D2D pair k, first calculates its utility and ranks all the RBs based on the previous matching $\mu(r)^{(t-1)}$ (line 4). Then, each D2D pair k proposes to the most preferred r, which can result in either of the two following cases.

The first case is when r does not have sufficient quota $I_{res}^{r}{}^{(t)}$ to accept k, and so r then finds the current matched D2D pairs that rank lower than D2D pair k according to $\mathcal{P}_r{}^{(t)}$ (lines 7-9). Each of the least preferred D2D pairs k' is sequentially rejected until either k can be accepted or there is no additional k' to reject (lines 10-12). If sufficient quota to accept k is not created, then k is also rejected and considered as the least preferred D2D pair represented by j_{lp} (lines 13-14).

The second case is when the quota of r is enough to accommodate k, in which it then checks the conflict set $\mathcal{C}_k$. If the conflict set is empty, the D2D pair k is accepted (lines 15-17). Otherwise, it removes all lower ranked conflicting D2D pairs compared to D2D pair k from its current matching (lines 18-22). If the conflict set is still non-empty, the D2D pair k is rejected and is considered as the least preferred j_{lp} (lines 23-26).

Finally, the least preferred D2D pair j_{lp} and all D2D pairs ranked lower than j_{lp} are removed from $\mathcal{P}_r{}^{(t)}$, and similarly these D2D pairs also remove r from their respective $\mathcal{P}_k{}^{(t)}$ (lines 27-29). *With this process, we guarantee that any less preferred D2D pair will not be accepted by that RB even if it has sufficient quota to do so, which is crucial for the matching stability of our design.* This process is repeated until the matching converges. The algorithm will converge when the matching of two consecutive iterations t remains unchanged.

Theorem 2. *Alg. 3 converges to a stable allocation.*

Proof. Please see Appendix B. □

The optimality property of the stable matching approach can be observed using the definition of weak Pareto optimality [43]. Let $U(\mu)$ denote the utility obtained by matching μ. A matching μ is weak Pareto optimal if there is no other matching μ' that can achieve a better utility, i.e., $U(\mu') \approx \sum_{r\in\mathcal{R}}\sum_{k\in\mathcal{K}} R_k^r(\mu') \geq U(\mu) \approx \sum_{r\in\mathcal{R}}\sum_{k\in\mathcal{K}} R_k^r(\mu')$. Formally, we state this as follows:

Definition 5. *A matching μ is weak Pareto optimal (PO) if there is no other matching μ' with $U(\mu') \geq U(\mu)$ [43].*

Theorem 3. *Alg. 3 produces a weak PO solution for the **FR** problem.*

Proof. Please see Appendix C. □

4.3 Computation Complexity and Implementation

In order to quantify the computational complexity of Alg. 2 and Alg. 3, first, we discuss the complexity of building the preference profile by both set of players (i.e., D2D pairs and RBs) that are the input to Alg. 2 and Alg. 3. Then, we discuss the running time of both algorithms. For each D2D pair k, the complexity of building the preference profile using any standard sorting algorithm is $O(R\log(R))$. Similarly the complexity of building the preference profile at the central BS for all RBs R is $O(KR\log(KR))$, where R and K represent the total number of RBs and D2D pairs, respectively. So, the input to Alg. 2 is $\eta = \sum_{k\in\mathcal{K}}|\mathcal{P}_k| + \sum_{r\in\mathcal{R}}|\mathcal{P}_r| = 2KR$, where $|\mathcal{P}|$ denote the size of preference profile $\mathcal{P}$. Moreover, Alg. 2 terminates after a finite number of iterations [36]. Under the worst case, when the preferences of all D2D pairs for all RBs are the same, it can be seen that the time complexity is *linear* in the size of input preference profiles (i.e., $O(\eta) = O(KR)$) [44].

In Alg. 3 to handle the externalities, at each iteration, all D2D pairs update their preference list (i.e., $O(R\log(R))$) based on the current matching. This is different from Alg. 2 whose preference list is updated only once during the initialization phase. Moreover, an additional input vector of the conflict set $\mathcal{C}_k$ will be added as an input with maximum size of $K-1$ (i.e., the worst case occurs when all D2D pairs are a member of the conflict set of all other D2D pairs). However, in general, the size of $\mathcal{C}_k$ will be far smaller than the total number (K) of D2D pairs in the network. Then, Alg. 3 input is equal to $\eta{=}\sum_{k\in\mathcal{K}}|\mathcal{P}_k|+\sum_{r\in\mathcal{R}}|\mathcal{P}_r|+\sum_{k\in\mathcal{K}}|\mathcal{C}_k| = 2KR+K(K-1)/2$. From Theorem. 1, we state that Alg. 3 terminates after a finite number of iterations. Then it can be stated that under worst case, the time complexity of Alg. 3 is also *linear* with respect to the size of input preference profiles (i.e., $O(\eta) = O(KR + \frac{K^2-K}{2})$). Thus, both algorithms show reasonable computational complexity for practical implementation.

4.4 Example Scenario

In this subsection, we provide a detailed discussion supported with examples for the RB allocation schemes. First, RB allocation using the partial reuse mode is discussed, i.e., Alg. 2. Then, we discuss the RB allocation process for the full-reuse mode i.e., Alg. 3. Moreover, we elaborate in detail the effect of externalities and their consequences if not well handled.

We consider Fig. 1 as our example for a D2D enabled system, where the dashed lines represent the interfering links. Note that the BS interferes with all D2D pairs, which is not shown in the figure. From Fig. 1, we consider that all D2D pairs choose to use the given mode (i.e., controlled by the vector $\boldsymbol{\alpha}$) so the two sides are $\mathcal{K} = \{k_1, k_2, k_3, k_4, k_5\}$, and $\mathcal{R} = \{r_1, r_2, r_3\}$. Let $P_{\mathcal{K}}$ and $P_{\mathcal{R}}$, represent the preference profile of all players as follows:

$$P_{k_1} = P_{k_5} = \{r_1, r_3, r_2\},\ P_{r_1} = \{k_1, k_2, k_5, k_4, k_3\},\ q_{r_1} = 1,$$
$$P_{k_2} = \{r_3, r_1, r_2\},\ P_{r_2} = \{k_5, k_4, k_2, k_1, k_3\},\ q_{r_2} = 3,$$
$$P_{k_3} = P_{k_4} = \{r_2, r_3, r_1\},\ P_{r_3} = \{k_4, k_2, k_5, k_1, k_3\},\ q_{r_3} = 1.$$

4.4.1 Partial-Reuse Mode

We first check the case when the partial-reuse mode (i.e., $y = 1$) is activated. Under this mode, there is no co-tier interference (no externalities among D2D pairs), thus we have a one-to-one matching scenario. From Alg. 2, all five D2D pairs propose to their respective preferred RBs simultaneously. Note that the BS manages the RBs preference profiles. From the preference profiles, we can see that k_1 and k_5 propose to r_1, k_2 proposes to r_3, and k_3 and k_4 propose to r_2 at time instant t. At t, we have:

$$\mu(r_1) = k_1, \quad \mu(r_2) = k_4, \quad \mu(r_3) = k_2.$$

Now at time instant $t+1$, the rejected D2D pairs k_3 and k_5 will update the preference by removing the RBs that have rejected them and then propose to the next best option, i.e., r_3 for both rejected D2D pairs. On receiving these proposals, r_3 compares its current match with the new proposals. It chooses the best among them (i.e., k_2) and rejects the rest (i.e., k_3, k_5). Now, the rejected pairs again update and propose until there are no more RBs to propose or all D2D pairs are matched. Finally, we have the following matching:

$$\mu(r_1) = k_1, \quad \mu(r_2) = k_5, \quad \mu(r_3) = k_2.$$

4.4.2 Full-Reuse Mode

Now consider the second case, i.e., the full-reuse mode ($y = 0$). As stated earlier, this is a one-to many matching. For ease of understanding, we assume each pair has a uniform interference (opposed to dynamic interference) on all RBs and a predefined quota for RBs (i.e., $q_{r_1} = 1, q_{r_2} = 3, q_{r_3} = 1$). Under this scenario, each D2D pair first identifies its conflict set and sends it to the BS. Note that this is done only once in the initialization phase. Additionally, this is important for handling the externalities as explained in Sec. 4.2.3. Considering Fig. 1, the conflict set using (33) is $C_{k_1} = \{\phi\}, C_{k_2} = \{k_3, k_4\}, C_{k_3} = \{k_2, k_4\}, C_{k_4} = \{k_2, k_3\}$, and $C_{k_5} = \{\phi\}$.

Similar to the first scenario, all D2D pairs propose to the most preferred RBs at time instant t and we obtain

$$\mu(r_1) = k_1, \quad \mu(r_2) = k_4, \quad \mu(r_3) = k_2.$$

This information is broadcast in the network by the BS. Note that, k_5 is rejected by r_1 due to the quota limitation $q_{r_1} = 1$, but k_3 is rejected by r_2 because $k_3 \in C_{k_4}$ (i.e., k_3 exists in the conflict set of a D2D pair k_4) and from P_{r_2}, we have $k_4 \succ_{r_2} k_3$. After receiving the current matching of all D2D pairs, we recalculate their respective utilities using (29) and re-rank all the RBs according to their utility. In this example, k_3 and k_4 change their preferences from $r_3 \succ_{k_i} r_1$ to $r_1 \succ_{k_i} r_3$ because $\mu(r_3) = k_2$ and $k_2 \in C_{k_i}$, where $i = 3, 4$. Hence, the new preference list, at time instant $t + 1$ is as follows:

$$\begin{array}{lll}
P_{k_1} = \{r_1, r_3, r_2\}, & P_{r_1} = \{k_1, k_2, k_4, k_3\}, & q_{r_1} = 0, \\
P_{k_2} = \{r_3, r_1, r_2\}, & P_{r_2} = \{k_4, k_2, k_1, k_5\}, & q_{r_2} = 2, \\
P_{k_3} = \{r_1, r_3\}, & P_{r_3} = \{k_4, k_2, k_5, k_1, k_3\}, & q_{r_3} = 0, \\
P_{k_4} = \{r_2, r_1, r_3\}, & & \\
P_{k_5} = \{r_3, r_2\}. & &
\end{array}$$

Now the rejected pairs, i.e., k_3 and k_5, propose to r_1 and r_3, respectively; k_3 and k_5 are rejected by r_1 and r_3 because $\mu(r_1) \succ_{r_1} k_3$ and $\mu(r_3) \succ_{r_3} k_5$ with $q_r = 0$. Again, all pairs update the preference profiles accordingly. k_3 and k_5 again propose at time instant $t + 2$ with the update preference list to r_3 and r_2, respectively. k_3 is again rejected because $\mu(r_3) \succ_{r_3} k_3$ and $q_{r_3} = 0$, but k_5 is accepted because $q_{r_2} = 2$ and $k_5 \notin C_{\mu(r_2)}$. Therefore, the final matching from Alg. 3 is

$$\mu(r_1) = k_1, \quad \mu(r_2) = k_4, k_5 \quad \mu(r_3) = k_2.$$

Note that k_3 has no more RBs to propose to and all the other D2D pairs are matched. Thus, the algorithm stops. Furthermore, we can observe that the spectral efficiency is improved by reusing the resources more in Alg. 3 (4 D2D pairs on 3 RBs) compared to Alg. 2 (3 D2D pairs on 3 RBs). However, Alg. 3 has an additional overhead due to coordination (i.e., conflict set information and matching update) compared to Alg. 2.

4.4.3 Full-Reuse Mode without Handling Externalities

Now consider the case where externalities are not handled. This means there is no conflict sets information available. Under this scenario, with the same initial quota information, k_1 and k_5 propose to r_1, k_2 proposes to r_3, and k_3 and k_4 propose to r_2 at time instant t. We obtain the following matching:

$$\mu(r_1) = k_1, \quad \mu(r_2) = k_3, k_4 \quad \mu(r_3) = k_2.$$

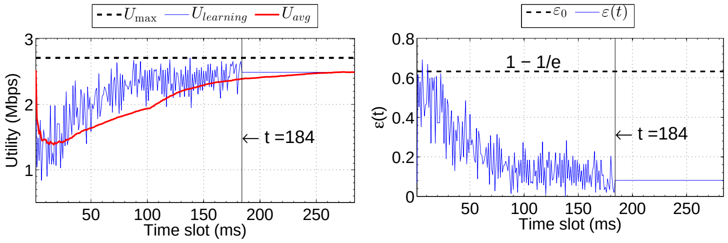
With this matching, the problem arises with $\mu(r_2)$, as both pairs when assigned to r_2 interfere with each other. This can reduce their actual utilities when compared to other RBs. Thus, they may be willing to switch to a new RB that provides them a higher utility. Assuming their second choice is better than their current match, then at time instant $t + 1$, the rejected pair k_5 and both unsatisfied pairs k_3 and k_4 propose once more to their best choices; they apply to r_3 in this example, and r_3 chooses k_4 due to the quota limitation. We then have

$$\mu(r_1) = k_1, \quad \mu(r_2) = \phi \quad \mu(r_3) = k_4.$$

With this assignment, we can see that both k_3 and k_4 prefer r_2 and that r_2 also prefers them to its current match. Both pairs will propose again in the next time instant and will be accepted. This brings us back to the initial case. Thus, under the case where externalities are not handled, these D2D pairs will always switch between their preferences and will never be able to converge to a stable allocation.

Table 1: Default Simulation Parameters [45]

Simulation Parameters	**Values**
Radius of MBS	500 m
Carrier frequency (f)	2 GHz
Frame Structure	Type 1 (FDD)
Transmission Time Interval (TTI)	1 ms
Total transmit power of BS	46 dBm
Total transmit power of D2Ds	23 dBm
System bandwidth	3 MHz
Bandwidth of each RB (W)	180 kHz
Number of subcarriers per RB	12
Neighboring subcarrier spacing	15 kHz
Modulation and coding scheme (MCS) [46]	QPSK: $1/12, 1/9, 1/6, 1/3, 1/2, 3/5$ 16QAM: $1/3, 1/2, 3/5$
Path loss (cellular link)	$128.1 + 37.6 \log(\mathrm{d})$, d[km]
Path loss (D2D links) [47]	$32.45 + 20 \log(\mathrm{f}) + 20 \log(\mathrm{d})$, f[MHz]
Shadow fading standard deviation [47]	3 dB
Proximity of D2Ds ($R2$)	random $\{20 \sim 30\}$ m
Thermal noise for 1 Hz at 20 ˚C	-174 dBm



(a) Real-time utility.

(b) Real-time performance gap.

Figure 3: Real-time performance of the learning scheme when $K = 20$ with system bandwidth 3 MHz.

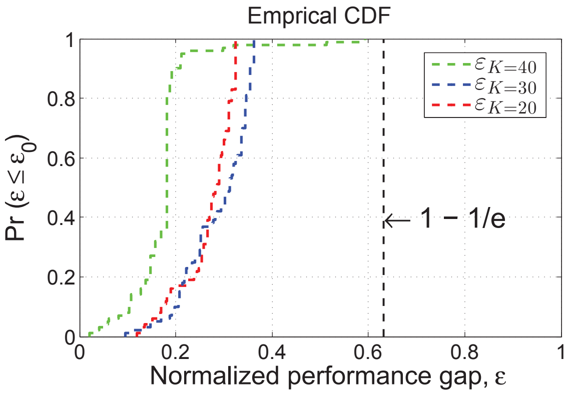


Figure 4: Normalized Gap (CDF)

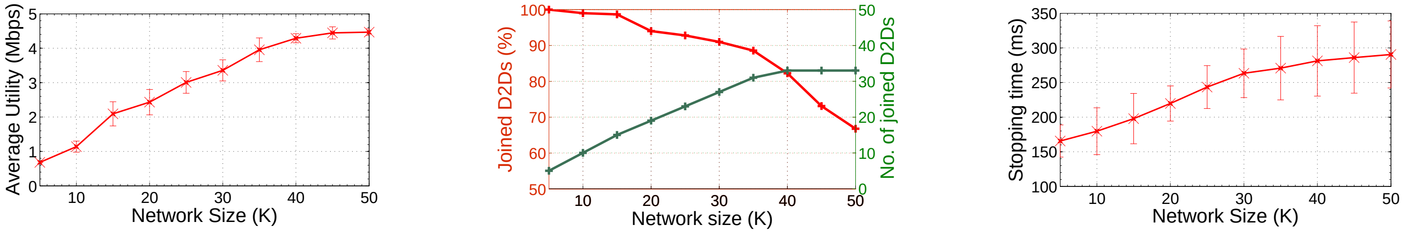
5 Simulation Results and Analysis

We consider a downlink system in which the BS is assumed to be deployed at a fixed location, and we randomly deploy C cellular users and K D2D pairs following a homogeneous Poisson point process (PPP). We assume the system bandwidth to be 3 MHz[3] which is occupied by the C cellular users. Moreover, we consider a full buffer model for all K D2D pairs. The main parameters used in our simulations are shown in Table 1 unless stated otherwise. These parameters are chosen according to the system model guidelines in [45]–[47]. Note that, all statistical results are averaged over 100 runs of random locations of D2D pairs, cellular users, and RB gains.

5.1 Simulation Results for Learning

In this subsection, we perform simulations to evaluate our proposed learning scheme. For this simulation, we first generate an instance of network with $K = 20$ D2D pairs. We then evaluate the following aspects of the learning scheme: the convergence of the learning scheme

[3] The methodologies developed in this paper can also be applied to any value of system bandwidth. The motivation for our choice (i.e., 3 MHz) is to analyze the performance under dense environment with peak network traffic and for the sake of simulation simplicity.

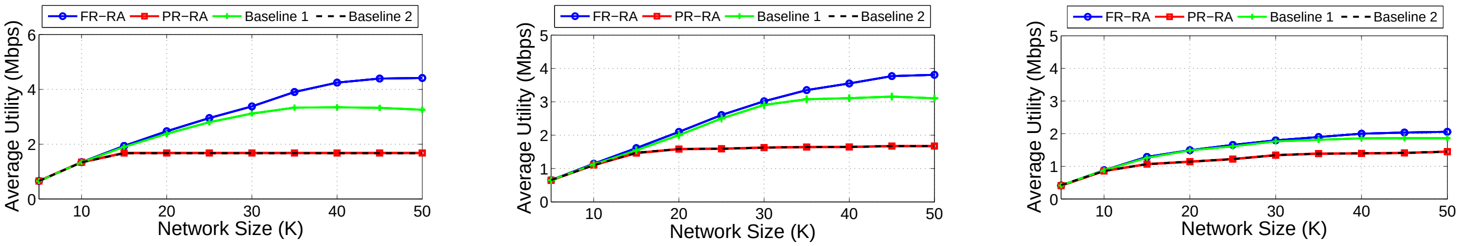


(a) Average Utility

(b) Average successfully joined D2Ds

(c) Average Stopping time

Figure 5: Performance of Learning scheme with varying network size. The error bars indicate 95% confidence intervals.



(a) $I^r_{\max} = -80$ dBm

(b) $I^r_{\max} = -100$ dBm

(c) $I^r_{\max} = -120$ dBm

Figure 6: Average utility under various tolerance levels.

and the normalized performance gap. Second, we generate instances of the network starting from $K = 5$ to $K = 50$. For this simulation, we run each instance 100 times to obtain the sample average of utility, the average number of successfully joined[4] D2D pairs in the system, and the average stopping time for convergence. Note that for these simulations, we assume the cellular-tier interference tolerance level to be fixed at $I^r_{\max} = -80$ dBm for all RBs. Finally, to evaluate our learning scheme, we define the normalized performance gap as follows:

$$\varepsilon(t) = 1 - \frac{U(t)}{U_{\max}}, \tag{34}$$

where $U(t)$ is the utility at time-slot t, and $U_{\max} = \max_{f \in \mathcal{F}} U_f$. We use the built-in simulated annealing functions in MATLAB to obtain optimal solution $U_{\max}$.

Fig. 3a shows the real-time utility values calculated using (4) along with its time average values, which are obtained by means of a sliding window. We observe that as the time slot increases, each D2D pair learns its possible configurations and chooses high utility configurations with high probabilities. Despite the fluctuations of the utility, the time average values show an increasing trend in Fig. 3a. This shows that the learning scheme converges in probability. However once the convergence is achieved, the configurations do not change, i.e., after time-slot 184. In Fig. 3b, we can see the corresponding performance gap calculated using (34), which has a descending trend with time. Furthermore, after a very short time-period (less than 20), we observe that the $\varepsilon(t)$ values becomes less than ε_0, where $\varepsilon_0 = 1 - 1/e$, which is the typical gap for randomized greedy algorithms [48].

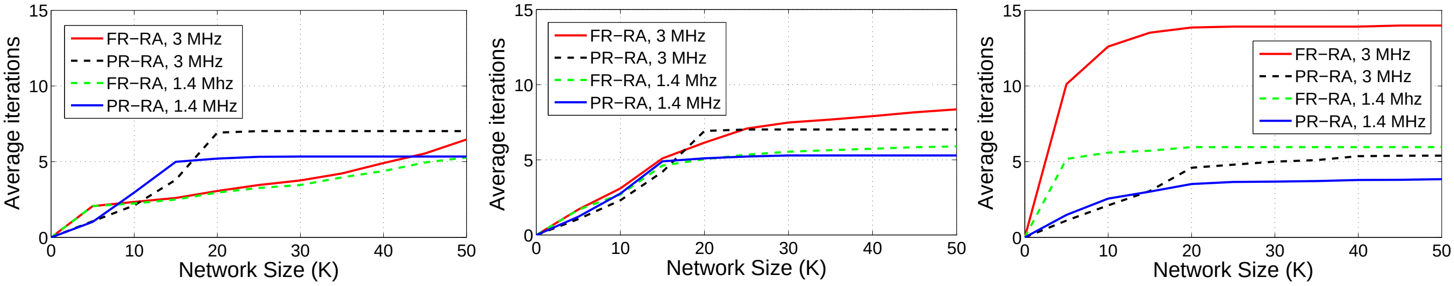
In Fig. 4, we test the normalized performance gap under three cases, $K = 20$, $K = 30$, and $K = 40$. It is observed that under all cases, the learning scheme converges to a near optimal solution. Additionally, when the ratio of the available RBs (i.e., 15 RBs with system bandwidth $3\,\text{MHz}$) to the number of D2D pairs satisfies ($\frac{R}{K} \geq 0.5$), the mode selection does not affect the gap and the normalized performance gap is below the randomized greedy algorithm gap (ε_0). However, if the ratio of available RBs to the number of D2D pairs is less than 0.5, (i.e., $\frac{R}{K} < 0.5$) (e.g., the $K = 40$ case), the impact of mode selection becomes apparent and increases the performance gap from the optimal. Still, as shown in Fig. 4, $\Pr\{\varepsilon \leq \varepsilon_0\} > 0.9$ for the majority of the time. This shows that the learning scheme selects the best mode of operation according to the network size the majority of the time, i.e., for a large network size ($K = 40$), the full-reuse mode is selected. Hence, we can infer that the network operates under the best configurations for most of the time.

Figs. 5a and 5b show the average utility achieved and fraction of successful joined D2D pairs for different network sizes, K. We observe that the utility increases with the network size despite a fixed number of RBs, i.e., $R = 15$. This is because according to the network size, the learning algorithm switches to the best suited mode, i.e., the partial-reuse mode for a small network size or the full-reuse mode for a larger network size. However, as the network size becomes larger ($K \geq 40$), the average utility approaches a saturation state due to limited RBs and the predefined $I^r_{\max}$ values. This trend is also evident in Fig. 5b, where the fraction of successfully joined D2D pairs decrease drastically after the saturation point (i.e., $K \geq 40$). In Fig. 5c, we evaluate the average stopping time for our learning scheme. It can be seen that for all network sizes, the learning scheme has a reasonable stopping time that increases sub-linearly with the network size. Moreover, it is observed that the stopping time has high confidence intervals which are a result of the mixing characteristic of the underlying Markov chain.

5.2 Simulation Results for Resource Allocation

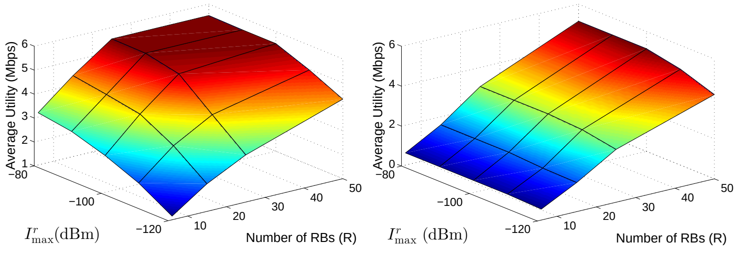
In order to evaluate the performance of the RB allocation schemes, first, we show the comparison in terms of average utility achieved by enabling the full-reuse and partial-reuse mode schemes under different network sizes (i.e., the number of Joined D2D users, K). Second, we evaluate the average utility for four different system bandwidth values, i.e., $1.4\,\text{MHz}$, $3\,\text{MHz}$, $5\,\text{MHz}$, and $10\,\text{MHz}$ for a fixed network size, i.e., $K = 50$. Finally, we show the average number of iterations

[4] Successfully joined D2D pairs represent the D2D pairs which choose to use the given mode and are also allocated RBs.



(a) $I^r_{\max} = -80$ dBm (b) $I^r_{\max} = -100$ dBm (c) $I^r_{\max} = -120$ dBm

Figure 8: Average number iterations vs. network size, for different tolerance levels.



(a) FR-RA Scheme (b) PR-RA Scheme

Figure 7: Average utility of the proposed FR-RA and PR-RA schemes under various tolerance levels with $K = 50$.

resulting for different network sizes. Note that, the performance of the RB allocation scheme depends upon the predefined max interference level $I^r_{\max}$ of the RB r. Therefore, we analyze the performance of RB allocation schemes with respect to three different maximum interference tolerance thresholds set by the cellular tier, $I_{\max} = -120, -100$, and -80 dBm [35], [49]. In our simulations for all D2D pairs K, we set the co-tier interference threshold to $\zeta_k = 10$ dB (i.e., between two D2D pairs).

We compare our proposed approaches with two other approaches: 1) The first approach (Baseline 1) is a distributed algorithm that is based on the one-to-many matching game, similar to our proposed algorithm for the full-reuse mode; however, no inter-tier interference among the D2D pairs is incorporated (i.e., without externalities). This approach aims to maximize the utility of all D2D pairs in the network while providing cellular tier interference protection. However, this approach is unstable due to the reasons discussed in Sec. 4.4.3. This benchmark algorithm is in line with some existing works used for RB allocation such as [35], [50], [51], 2) The second is a centralized approach (Baseline 2) that uses the Hungarian assignment method for RB allocation [52]. Results corresponding to the full-reuse mode and partial-reuse mode algorithms are denoted as "FR-RA", and "PR-RA", respectively.

In Fig. 6, the achievable utility by D2D pairs is shown with respect to three different $I^r_{\max}$ values for system bandwidth value of $3\,\text{MHz}$ (i.e., 15 RBs). In this simulation, we increase the network size (D2D pairs) and observed the average utility. First, we find that for the FR-RA and Baseline 1 schemes, the average utility increases as the network size grows. However, for Baseline 1, after the network size is sufficiently large (above 30 D2D pairs and higher), the utility starts to degrade. The reason for this performance degradation is as the network size increases, the inter-D2D interference also increases, which degrades the performance. A performance gain in terms of average utility up to 35%, 27%, and 13% under $I^r_{\max} = -80$, -100, and -120 dBm, respectively is observed by the FR-RA when compared to Baseline 1 for a network of 50 D2D pairs.

Second, the utility saturates as the network grows when $I^r_{\max} = -80$ and -100 dBm for the PR-RA and the Baseline 2 schemes. This is because of the limited amount of RBs (i.e., 15 in $3\,\text{MHz}$ of bandwidth) in the simulation, and both schemes allow a single D2D pair on an RB. Therefore, only the best one is allocated to the RB. Moreover, the performance of the PR-RA scheme and Baseline 2 is indistinguishable under all scenarios.

Third, it is observed from Figs. 6a, 6b, and 6c that the FR-RA scheme is highly affected by different $I^r_{\max}$ thresholds compared to the PR-RA scheme (i.e., at $I^r_{\max} = -120$ dBm, the utility drops to up to 52% of the utility obtained at $I^r_{\max} = -80$ dBm). This is mainly because the interference protection constraint becomes stricter and a smaller number of users can reuse the RBs in the FR-RA scheme, whereas in the PR-RA scheme, only one D2D pair is using the RB. Moreover, for a loose protection threshold (i.e., $I^r_{\max} = -80$ and -100 dBm), the FR-RA scheme yields a performance benefit of up to 158% and 123% compared to the PR-RA scheme, whereas for a tighter protection threshold, $I^r_{\max} = -120$ dBm, the performance gain is reduced to 36%. Finally, we can infer that for a network size of less than 15 D2D pairs, the performance of all the schemes are indistinguishable.

Fig. 7 compares the performance of the proposed FR-RA and PR-RA schemes. In this simulation, we fix the network size to 50 D2D pairs for four different system bandwidth values, i.e., $1.4\,\text{MHz}$ (6 RBs), $3\,\text{MHz}$ (15 RBs), $5\,\text{MHz}$ (25 RBs), and $10\,\text{MHz}$ (50 RBs) under different $I^r_{\max}$ values. It can be observed that under all $I^r_{\max}$ values, the average utility of the PR-RA scheme increases. This is because the unassigned D2D pairs are able to acquire RBs as the RBs in the system are increased. Moreover, we find that, the average utility for the FR-RA scheme almost saturates as the number of RBs increases in the system. The main reason for such an action is that under loose interference thresholds ranging from $I^r_{\max} = -80$ to $-$ 100 dBm, most of the D2D pairs get RBs assigned and under tight interference thresholds $I^r_{\max} = -120$ dBm, a few D2D pairs are allocated RBs while the rest are rejected.

Fig. 8 compares the average iterations versus the network size for two different system bandwidth values, i.e., $1.4\,\text{MHz}$, and $3\,\text{MHz}$. It can be observed that for a loose interference tolerance threshold level $I^r_{\max} = -80$ dBm (Fig. 8a), the proposed FR-RA scheme has a remarkable convergence time and does not exceed an average of 5 and 7 iterations for all network sizes for both $1.4\,\text{MHz}$, and $3\,\text{MHz}$ cases, respectively. This fast convergence time can be achieved due to the loose tolerance threshold level, as most of the D2D pairs are accepted at their initial proposals (line 15 of Alg. 3). Additionally, the average iterations increase with the network size because of the increase in inter D2D interference (i.e., less than 3 average iterations for a network size of 10 compared to 7 average iterations for a network size 50). However, the use of the PR-RA scheme under $I^r_{\max} = -80$ has a higher number of average iterations for both the $1.4\,\text{MHz}$ (less

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/TMC.2017.2689768, IEEE Transactions on Mobile Computing

than 6) and $3\,\mathrm{MHz}$ (less than 8) cases compared to the FR-RA scheme for all network sizes. In the PR-RA scheme, for a relatively loose $I^r_{\max}$ value, all the users meet the interference constraint (line 9 of Alg. 2). Then, to assign an RB to a D2D pair, all low ranked D2D pairs have to be analyzed and rejected (lines 10-15 of Alg. 2). This increases the average iterations even for a small network size (i.e., less than 15). However for a tighter $I^r_{\max}$ value (Fig. 8b and Fig. 8c), a number of D2D pairs will be initially rejected due to tighter interference constraint (line 9 of Alg. 2), which reduces the average iterations for a small network size. In the FR-RA scheme, at a tighter interference tolerance threshold level of $I^r_{\max} = -100$ dBm (Fig. 8b), the average number of iterations also increases as the network size increases, but does not exceed an average of 6 and 9 iterations for all network sizes with $1.4\,\mathrm{MHz}$ and $3\,\mathrm{MHz}$ bandwidth, respectively. Moreover, under $I^r_{\max} = -120$ dBm (Fig. 8c), the average iteration converges to 14 and 6 iterations for even a small network size (i.e., less than 5 D2D pairs) when bandwidth values of $3\,\mathrm{MHz}$ and $1.4\,\mathrm{MHz}$ are considered, respectively. This is because most of the D2D pairs are rejected by RBs due to the tight $I^r_{\max}$ (line 7 of Alg. 3). This then forces the pairs to propose to the next RBs, and consequently most of the D2D pairs re-propose until they are either accepted or rejected by all RBs in the system. Note that under all cases, the average number of iterations will always be less than the number of RBs. This can be achieved due to a completely distributed design of the FR-RA and PR-RA schemes.

6 Conclusion

In this paper, we designed a resource allocation framework for D2D communication over cellular networks by using Markov approximation and matching-game approaches. We considered two important aspects: mode-selection and resource block allocation for the performance of the D2D network. We used a learning framework based on Markov approximation in which we have designed a problem specific Markov chain that converges close to an optimal solution with probability one. Furthermore, we proposed novel resource allocation algorithms based on matching theory that can work within the proposed learning framework. These resource allocation algorithms help us obtain a stable resource allocation that is a locally optimal solution of an NP-hard resource allocation problem at each time slot of the Markov approximation process. Our framework has shown that it achieves a stable, distributed and scalable solution for the network. Simulation results have shown that the proposed framework convergence in probability, achieves interference protection and closely approaches the optimal solution. Furthermore, we have also validated the stability and convergence of the resource allocation algorithm.

References

[1] C. Xu, L. Song, and Z. Han, *Resource Management for Device-to Device Underlay Communication.* New York, NY, USA: Springer-Verlag, 2013.

[2] O. Semiari, W. Saad, S. Valentin, M.Bennis, and H. V. Poor, "Context-Aware Small Cell Networks: How Social Metrics Improve Wireless Resource Allocation," *IEEE Trans. on Wireless Comm.*, vol. 14, no. 11, pp. 5927–5940, Nov. 2015.

[3] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, pp. 74–80, Feb. 2014.

[4] S. Andreev, O. Galinina, A. Pyattaev, K. Johnsson, and Y. Koucheryavy, "Analyzing assisted offloading of cellular user sessions onto D2D links in unlicensed bands," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 1, pp. 67–80, Jan. 2015.

[5] A. Antonopoulos, E. Kartsakli, and C. Verikoukis, "Game theoretic D2D content dissemination in 4G cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 125–132, Jun. 2014.

[6] L. Militano, A. Orsino, G. Araniti, A. Molinaro, and A. Iera, "A Constrained Coalition Formation Game for Multihop D2D Content Uploading," *IEEE Tran. on Wireless Comm.*, vol. 15, no. 3, pp. 2012–2024, Mar. 2016.

[7] X. Lin, J. G. Andrews, and A. Ghosh, "Spectrum sharing for device-to-device communication in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 6727–6740, Dec. 2014.

[8] E. Datsika, A. Antonopoulos, N. Zorba, and C. Verikoukis, "Green cooperative device–to–device communication: A social–aware perspective," *IEEE Access*, vol. 4, pp. 3697–3707, Jun. 2016.

[9] P. Li, S. Guo and I. Stojmenovic, "A Truthful Double Auction for Device-to-device Communications in Cellular Networks," in *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 71-81, Jan. 2016.

[10] A. Asadi, V. Mancuso, "Network-assisted Outband D2D-clustering in 5G Cellular Networks: Theory and Practice," *IEEE Trans. on Mobile Comput.*, vol.PP, no.99, pp.1-1, 2016.

[11] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1801–1819, Fourth quarter 2014.

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

[13] 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 *Proc. IEEE Vehicular Technology Conference*, Barcelona, Spain, Apr. 2009.

[14] B. Kaufman, J. Lilleberg, and B. Aazhang, "Spectrum sharing scheme between cellular users and ad-hoc device-to-device users," *IEEE Trans. on Wireless Comm.*, vol. 12, no. 3, pp. 1038–1049, Mar. 2013.

[15] D. Feng, L. Lu, Y. Yuan-Wu, G. Y. Li, G. Feng, and S. Li, "Device-to-device communications underlaying cellular networks," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3541–3551, Aug. 2013.

[16] Y. Jiang, Q. Liu, F. Zheng, X. Gao, and X. You, "Energy efficient joint resource allocation and power control for D2D communications," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6119-6127, Aug. 2016.

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

[18] D. Wu, Y. Cai, R. Hu, and Y. Qian, "Dynamic distributed resource sharing for mobile D2D communications," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5417–5429, Oct. 2015.

[19] Y. Gu, Y. Zhang, M. Pan, and Z. Han, "Matching and cheating in device to device communications underlying cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2156-2166, Oct. 2015.

[20] H. Tang and Z. Ding, "Mixed mode transmission and resource allocation for d2d communication," *IEEE Trans. on Wireless Comm.*, vol. 15, no. 1, pp. 162–175, Jan. 2016.

[21] Ericsson, A. B. "Ericsson mobility report: On the pulse of the Networked Society", *Ericsson, Sweden, Tech. Rep. EAB-14 61078*, Jun. 2015.

[22] M. Chen, S. C. Liew, Z. Shao, and C. Kai, "Markov approximation for combinatorial network optimization," *IEEE Trans. on Information Theory*, vol. 59, no. 10, pp. 6301–6327, Oct. 2013.

[23] S. Zhang, Z. Shao, M. Chen, and L. Jiang, "Optimal distributed P2P streaming under node degree bounds," *IEEE/ACM Trans. on Networking*, vol. 22, no. 3, pp. 717–730, Jun. 2014.

[24] T. Z. Oo, N.H. Tran, W. Saad, J. Son, and C.S. Hong, "Traffic offloading via Markov approximation in heterogeneous cellular networks," in *IEEE/IFIP Network Operations and Management Symposium,* pp.52–60, Apr. 2016.

[25] T. Z. Oo, N. H. Tran, W. Saad, D. Niyato, Z. Han, C. S. Hong, "Offloading in HetNet: A Coordination of Interference Mitigation, User Association and Resource Allocation," in *IEEE Trans. on Mobile Comput.*, vol. PP, no. 99, pp. 1-1.

[26] S. Maghsudi, and S. Stanczak, "Channel selection for network-assisted D2D communication via no-regret bandit learning with calibrated forecasting," *IEEE Tran. on Wireless Comm.*, vol. 14, no. 3, pp. 1309–1322, Mar. 2015.

[27] S. Maghsudi and E. Hossain, "Multi-armed bandits with application to 5G small cells," in *IEEE Wireless Commun.*, vol. 23, no. 3, pp. 64-73, Jun. 2016.

[28] Z. Zhou, G. Ma, M. Dong, K. Ota; C. Xu, Y. Jia, "Iterative Energy-Efficient Stable Matching Approach for Context-Aware Resource Allocation in D2D Communications," in *IEEE Access*, vol.PP, no.99, pp.1-1, 2016.

[29] K. Son, S. Lee, Y. Yi, and S. Chong, "Refim: A practical interference management in heterogeneous wireless access networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1260–1272, Jun. 2011.

[30] S. Boyd, and L. Vandenberhe "Convex Optimization", *Cambridge University Press,* 2004.

[31] J. Marden and J. Shamma, "Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation," in *48th Annual Allerton Conference on Communication, Control, and Computing,* Monticello, Illinois, Sep. 2010.

[32] P. Laarhoven and E. Aarts, *Simulated Annealing: Theory and Applications.* New York, NY: Springer-Verlag, 1987.

[33] S. Kirkpatrick, "Optimization by simulated annealing: Quantitative studies," *Journal of statistical physics*, vol. 34, no. 5-6, pp. 975–986, Mar. 1984.

[34] Y. Gu, W. Saad, M. Bennis, M. Debbah, and Z. Han, "Matching theory for future wireless networks: fundamentals and applications," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 52–59, May 2015.

[35] S. M. Ahsan Kazmi, N. H. Tran, W. Saad, L. B. Le, T. M. Ho and C. S. Hong, "Optimized Resource Management in Heterogeneous Wireless Networks," in *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1397-1400, Jul. 2016.

[36] A. E. Roth, "Deferred acceptance algorithms: History, theory, practice, and open questions," *Int. J. Game Theory*, vol. 36, no. 3-4, pp. 537–569, Mar. 2008.

[37] D. Gale, and L. Shapley, "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, vol. 69, no. 1, pp. 9–15, Jan. 1962.

[38] E. Dahlman, S. Parkvall, and J. Sköld, *4G – LTE/LTEAdvanced for Mobile Broadband*, Academic Press, Apr., 2011.

[39] D. F. Manlove, *Algorithmics of Matching Under Preferences*. World Scientific, 2013.

[40] R. Zhang, X. Cheng, L. Yang, and B. Jiao, "Interference graph based resource allocation (InGRA) for D2D communications underlaying cellular networks," *IEEE Trans. Veh. Technol*, vol. 64, no. 8, pp. 3844–3850, Aug. 2015.

[41] R. Zhang, X. Cheng, Q. Yao, C.-X. Wang, Y. Yang, and B. Jiao, "Interference graph based resource sharing schemes for vehicular networks," *IEEE Trans. Veh. Technol*, vol. 62, no. 8, pp. 4028– 4039, Oct. 2013.

[42] H. Xu and B. Li, "Anchor: A versatile and efficient framework for resource management in the cloud," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 6, pp. 1066–1076, Jun. 2013.

[43] E. Jorswieck, "Stable matchings for resource allocation in wireless networks," in *Proc. of IEEE International Conference on Digital Signal Processing*, Greece, Jul. 2011.

[44] M. Hasan and E. Hossain, "Distributed resource allocation in 5G cellular networks," in *Towards 5G: Applications, Requirements and Candidate Technologies*, Hoboken, NJ, USA: Wiley, 2015.

[45] 3GPP, "Evolved universal terrestrial radio access (E-UTRA): Physical layer procedures, Release 11," *Tech. Rep. TS 36.213*, Dec. 2012.

[46] 3GPP TR 36.843, "Study on LTE Device to Device Proximity Services: Radio Aspects", Mar. 2014.

[47] Huawei, HiSilicon, "Channel model for D2D evaluations," *3GPP TSG RAN WG1 Meeting #73*, May. 2013.

[48] N. Buchbinder, and J. Naor, *The Design of Competitive Online Algorithms via a Primal-Dual Approach.*, Hanover, MA: NOW Publishers, 2009.

[49] A. Abdelnasser, E. Hossain, and D. I. Kim, "Tier-aware resource allocation in ofdma macrocell-small cell networks," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 695–710, Mar. 2015.

[50] A. Leshem, E. Zehavi, and Y. Yaffe, "Multichannel opportunistic carrier sensing for stable channel access control in cognitive radio systems," *IEEE J. Select. Areas Commun.*, vol. 30, no. 1, pp. 82–95, Jan. 2012.

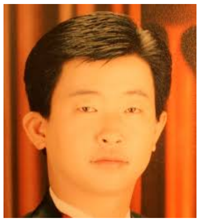
[51] Y. Wu, H. Viswanathan, T. Klein, M. Haner and R. Calderbank, "Capacity Optimization in Networks with Heterogeneous Radio Access Technologies," in *Proc. of IEEE Global Communication Conference*, Houston, TX, Dec. 2011.

[52] H. W. Khun, "The Hungarian method for the assignment problem," *Nav. Res. Logist. Quart.*, vol. 2, pp. 83–97, Mar. 1955.

S. M. Ahsan Kazmi received his Master's degree in Communication System Engineering from National University of Sciences and Technology (NUST), Pakistan, in 2012. Currently, he is pursuing his PhD degree from Kyung Hee University (KHU), South Korea, for which he was awarded a scholarship in 2014. His research interests includes includes radio resource management for HetNets, and software defined Networking for cellular networks.

Nguyen H. Tran (S'10-M'11) received the BS degree from Hochiminh City University of Technology and Ph.D. degree from Kyung Hee University, in electrical and computer engineering, in 2005 and 2011, respectively. Since 2012, he has been an Assistant Professor with Department of Computer Science and Engineering, Kyung Hee University. His research interest is to applying analytic techniques of optimization, game theory, and stochastic modeling to cutting-edge applications such as cloud and mobile-edge computing, data centers, heterogeneous wireless networks, and big data for networks. He received the best KHU thesis award in engineering in 2011 and best paper award at IEEE ICC 2016. He is the Editor of IEEE Transactions on Green Communications and Networking.

Walid Saad (S'07-M'10-SM'15) received his Ph.D degree from the University of Oslo in 2010. Currently, he is an Assistant Professor and the Steven O. Lane Junior Faculty Fellow at the Department of Electrical and Computer Engineering at Virginia Tech, where he leads the Network Science, Wireless, and Security (NetSciWiS) laboratory, within the Wireless@VT research group. His research interests include wireless networks, game theory, cybersecurity, and cyber-physical systems. Dr. Saad is the recipient of the NSF CAREER award in 2013, the AFOSR summer faculty fellowship in 2014, and the Young Investigator Award from the Office of Naval Research (ONR) in 2015. He was the author/co-author of five conference best paper awards at WiOpt in 2009, ICIMP in 2010, IEEE WCNC in 2012, IEEE PIMRC in 2015, and IEEE SmartGridComm in 2015. He is the recipient of the 2015 Fred W. Ellersick Prize from the IEEE Communications Society. Dr. Saad serves as an editor for the IEEE Transactions on Wireless Communications, IEEE Transactions on Communications, and IEEE Transactions on Information Forensics and Security.



Zhu Han (S'01-M'04-SM'09-F'14) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, in 1999 and 2003, respectively. From 2000 to 2002, he was an R&D Engineer of JDSU, Germantown, Maryland. From 2003 to 2006, he was a Research Associate at the University of Maryland. From 2006 to 2008, he was an assistant professor at Boise State University, Idaho. Currently, he is a Professor in the Electrical and Computer Engineering Department as well as in the Computer Science Department at the University of Houston, Texas. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid. Dr. Han received an NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the Journal on Advances in Signal Processing in 2015, IEEE Leonard G. Abraham Prize in the field of Communications Systems (best paper award in IEEE JSAC) in 2016, and several best paper awards in IEEE conferences. Currently, Dr. Han is an IEEE Communications Society Distinguished Lecturer.

Tai Manh Ho received the B.Eng. and M.S. degree in Computer Engineering from Hanoi University of Technology, Vietnam, in 2006 and 2008, respectively. He is currently a Ph.D. candidate at the Department of Computer Engineering, Kyung Hee University, Korea. His research interest includes radio resource management for wireless communication systems with special emphasis on heterogeneous networks.

Thant Zin Oo received the B.Eng. degree in electrical systems and electronics at Myanmar Maritime University, Thanlyin, Myanmar in 2008 and the B.S. degree in computing and information system from London Metropolitan University, U.K., in 2008, for which he received grant from the British Council. He is currently working towards Ph.D. degree in computer science and engineering from Kyung Hee University, Korea, for which he was awarded a scholarship in 2010. His research interests include wireless communications, and sustainable energy.

Choong Seon Hong (S'95-M'97-SM'11) received the B.S. and M.S. degrees in electronic engineering from Kyung Hee University, Seoul, South Korea, in 1983 and 1985, respectively, and the Ph.D. degree from Keio University, Minato, Japan, in 1997. In 1988, he joined Korea Telecom, where he worked on broadband networks as a Member of Technical Staff. In September 1993, he joined Keio University. He worked for the Telecommunications Network Laboratory, Korea Telecom, as a Senior Member of Technical Staff and the Director of the Networking Research Team until August 1999. Since September 1999, he has been a Professor with the Department of Computer Science and Engineering, Kyung Hee University. His research interests include future Internet, ad hoc networks, network management, and network security. He is a member of ACM, IEICE, IPSJ, KIISE, KICS, KIPS, and OSIA. He has served as the General Chair, a TPC Chair/Member, or an Organizing Committee Member for international conferences such as NOMS, IM, APNOMS, E2EMON, CCNC, ADSN, ICPP, DIM, WISA, BcN, TINA, SAINT, and ICOIN. In addition, he is currently an Associate Editor of the IEEE Transactions on Network and Service Management, International Journal of Network Management, and Journal of Communications and Networks and an Associate Technical Editor of the IEEE Communications Magazine.