Resource Allocation for D2D Communications with Partial Channel State Information

Anushree Neogi, Prasanna Chaporkar and Abhay Karandikar

***Abstract*—Enhancement of system capacity is one of the objectives of the fifth generation (5G) networks in which device-to-device (D2D) communications is anticipated to play a crucial role. This can be achieved by devising efficient resource allocation strategies for the D2D users. While most of the works in resource allocation assume full knowledge of the channel states, transmitting it in every time slot reduces the system throughput due to extra control overhead and leads to wastage of power. In this paper, we address the problem of D2D resource allocation with partial channel state information (CSI) at the base station (BS) and ensure that the interference from the D2D users do not jeopardize the communications of cellular users (CUs). With partial CSI, existing algorithms determine the Nash equilibrium in a distributed manner, whose inefficiency in maximizing the social utility is well known as the players try to maximize their own utilities. This is the first work in the D2D resource allocation field in which within a game theoretic framework, an optimal D2D resource allocation algorithm is proposed which maximizes the social utility of the D2D players such that a social optimum is attained. Each D2D player with the help of the BS learns to select the optimal action. We consider the channel to exhibit path loss. Next, we consider both a slow and fast fading with CU mobility and propose two heuristic algorithms. We validate the performance of our proposed algorithms through simulation results.**

I. Introduction

Device-to-device (D2D) communication is envisioned to be a promising component of the fifth generation (5G) wireless networks. It explores the possibility of short range communication between two D2D users in an underlay Long Term Evolution (LTE) network so that they can reuse the radio resources already allocated to cellular users (CUs) through efficient resource allocation algorithms. The CUs are treated as primary users whose quality of service should not get affected due to the transmissions of D2D users. Since the network is able to accommodate more users, the network capacity gets enhanced.

A. Motivation

Most of the previous works on D2D resource allocation assume perfect channel state information (CSI) [1] - [3]. This implies that the base station (BS) as a central resource allocation unit has knowledge of all the channel gains among the different communicating entities. From this knowledge, the BS can determine the achievable rates of all the communication links in order to determine an optimal resource allocation algorithm that is centralized. The channel gains from a CU and a D2D transmitter to the BS are known because every user transmits its channel quality indicator (CQI) to the BS either periodically or aperiodically [4]. However, the BS cannot determine the CQI between a D2D transmitter-receiver pair and also that between a CU and a D2D receiver. For a fast fading channel, even if these channel gains are known through some technique, conveying them to the BS in every time slot requires significant control overhead and power. Thus, partial CSI is a limiting factor to the BS's resource allocation capability which makes this a challenging problem.

With partial CSI, resource allocation problems are generally modeled game theoretically and distributed resource allocation algorithms are designed. In such systems the decisions that are taken by the BS when perfect CSI is available are now taken by the user equipments (UEs) through a learning algorithm. This reduces the computational load at the BS considerably. Such systems are generally modeled as non-cooperative, in which players take actions to enhance their own utilities. The resulting Nash equilibrium solution to the resource allocation problem fails to achieve a socially optimum solution since the best response of every player may not maximize the overall social utility. The inefficiency of the Nash equilibrium is highlighted by the well acknowledged econometric principle, "*Tragedy of the Commons*", named after the celebrated work of Hardin [5] and can be quantified through the price of anarchy.

B. Related Work

We give a brief overview of some of the research works pertaining to D2D resource allocation, assuming perfect CSI, in a game-theoretic framework. In [6], the authors employ a Stackelberg game theoretic framework for joint power control and channel allocation. The existence and uniqueness of the Stackelberg equilibrium (subgame perfect Nash equilibrium) is analyzed. The Stackelberg model is also used in [7] in which the authors state that striving to reach a global maximum is a difficult mathematical feat and is also computationally complex. Thus, they propose a low complexity two-stage algorithm that addresses the power and resource allocation problems separately. However, the proposed algorithm is sub-optimal. A reverse iterative combinatorial auction framework is proposed in [8] but the iterative nature of the algorithm increases the communication overhead.

With partial CSI, the application of game theory to D2D resource allocation problems is limited. In [9], the authors have proposed a distributed power allocation scheme which is

The authors are with Indian Institute of Technology Bombay, Mumbai, Maharashtra, 400076, India. Email: (anushreen,chaporkar,karandi)@ee.iitb.ac.in. This work is funded by the Department of Electronics and Information Technology (DeitY), Government of India through a grant on "5G Research and Building Next Gen Solutions for Indian Market".

modeled as a non-cooperative game. They have shown that a unique Nash equilibrium exists. Though their proposed method does not require global CSI, it requires the knowledge of local CSI. With local exchange of information, the signaling overhead increases considerably. Moreover, for multiple D2D pairs, they have not addressed the existence of the Nash equilibrium. In [10], based on the Stackelberg game model, a distributed iterative scheme of resource allocation is proposed. However, due to the iterative nature of their proposed algorithm, signaling overhead is substantial. The authors of [11] propose a centralized graph based resource allocation scheme and model the power control problem as an exact potential game. They show that although a pure strategy Nash equilibrium maximizes the potential function but it does not necessarily imply that the social utility of the D2D players is also maximized.

C. Contributions

In this paper, we take a departure from the usual notion of selfish players and employ game theoretic principles such that every player adjusts its action so as to obtain an allocation strategy that maximizes the social utility or the sum throughput of the D2D players. Moreover, it is also ensured that the signal-to-interference-plus-noise-ratio (SINR) requirements of the CUs are satisfied. We assume that the CUs have already been allocated resources by the BS. The D2D players take their actions in a distributed manner and transmit it to the BS, based on which the BS decides on the final allocation that maximizes the social utility of the D2D players. This allocation is nothing but an assignment or mapping of CUs' resources to the D2D players. A feasible allocation is one which does not hamper CU communications. Our work is motivated by [12], in which the authors consider the optimal association problem of users to base stations. They consider a game theoretic model and propose a distributed algorithm to maximize the social utility of all the users.

The main contributions of our work are as follows:

- Our work differs from [12] in that our algorithm is not a completely distributed algorithm. In [12], due to a lack of supervision, multiple players choose the same BS. This results in a decrease in the network throughput. However, this problem does not arise in our case because of the intervention of the BS which allocates orthogonal resources to the D2D players. Thus, multiple players cannot share the same CU's resources, otherwise it results in interference among the D2D players which would decrease the D2D network's throughput. Moreover, since the BS assigns orthogonal resources, the search space for the optimal strategy gets restricted. Hence, the rate of convergence of the algorithm to the optimal strategy decreases in comparison to [12].
- In [12], the utility of a player is the rate that it observes when it transmits using the resources of a CU. However, in our paper, we propose a different way of designing the utility which is based on a frame structure.
- Further, the authors of [12] have not considered any kind of randomness in the system model, which is a challenging problem. We have considered both slow fading as well as fast fading with mobility of CUs. With this randomness into consideration, we have proposed two heuristic utility calculation methods which achieve good performance. In the field of D2D communications, ours is the first resource allocation algorithm that aims at maximizing the social utility of all the D2D players with partial CSI.

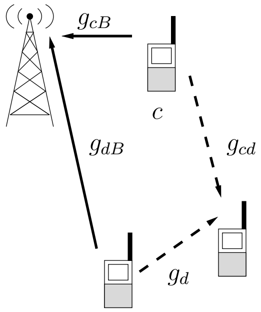

Fig. 1: Resource reuse between a CU and a D2D pair.

The organization of this paper is as follows. In Section II we present the system model which comprises of an underlay D2D network and the game theoretic modeling of the D2D resource allocation problem. In Section III, we propose an optimal resource allocation algorithm for the D2D players. We next discuss how the system state process is a finite state discrete time Markov chain (DTMC) in Section IV. In Section V, we prove the optimality of our proposed algorithm. We present an example to demonstrate its optimality in Section VI. Next, in Section VII we consider both slow fading and fast fading with CU mobility and propose two heuristic algorithms. We verify the performance of our proposed algorithms through simulations in Section VIII. In Section IX, we summarize our findings.

II. System Model

A. Network Model

We consider a single cell scenario with a BS and N_C CUs in uplink transmission mode. We assume that the BS has already allocated resources to these CUs. Consider N_D D2D pairs which have to be allocated the resources of these CUs. We denote the BS by B, a CU by c and a D2D pair by d. As shown in Fig. 1, the channel gain between a CU c to the BS B is denoted by g_{cB} and the channel gain between the D2D transmitter d to the BS B is given by g_{dB}. We assume that both these channel gains are known at the BS. Similarly, for a D2D pair d, the channel gain between its transmitter and receiver is given by g_d. The channel gain between a CU c and the D2D receiver is given by g_{cd}. We assume that a D2D receiver does not know the channel gains g_d and g_{cd}. We consider the channel to exhibit pathloss.

Let a CU's transmit power be P_C and that of a D2D user's be P_D. Given that a D2D user d is allocated the resources of a CU c, the SINR of the CU c at the BS is given by,

$$\gamma_c = \frac{P_C\, g_{cB}}{P_D\, g_{dB} + N_0}, \tag{1}$$

where N_0 is the average noise power. In case a D2D transmitter's interference power at the BS is very high when it is reusing the resources of the CU c and its SINR γ_c becomes less than a threshold γ_{tgt}, then it is not allocated this CU's resources. Then, this allocation is infeasible.

B. Game Theoretic Model

Let $\mathcal{G}$ be a strategic form game with N_D D2D players constituting a set $\mathcal{N}_D = \{d_1, d_2, \dots, d_{N_D}\}$. Each D2D player d can choose an action from a set of finite actions $\mathcal{A}_d$. Let the joint action set be $\mathcal{A} = \mathcal{A}_d^{N_D}$ and the utility function be $U_d : \mathcal{A} \rightarrow \mathbb{R}$.

As per the LTE standard, time is divided into subframes of 1 ms duration. We define a frame to be a time window consisting of N_D successive subframes. A subframe is indexed by $\tilde{n}$ while a frame by n. At the start of a frame n, the action taken by every D2D player d is to select a list $l_d(n)$ and transmit it to the BS. The list contains K possible CUs drawn from a set $\mathcal{N}_\mathcal{C} = \{c_1, c_2, \dots, c_{N_C}\}$ of available CUs to form a K-tuple. Hence, the total number of lists that a D2D player can generate is ${}^{N_C}P_K = L$. This set of L lists constitutes the action set $\mathcal{A}_d$ of a D2D player. Let the action profile be $\boldsymbol{l} = (l_1, l_2, \dots, l_{N_D}) \in \mathcal{A}$. We denote $\boldsymbol{l}_{-d}$ to be the action profile of the D2D players other than the D2D player d. Therefore, the action profile can also be written as $(l_d, \boldsymbol{l}_{-d})$. The utility of a D2D player can therefore be written as $U_d(\boldsymbol{l}) = U_d(l_d, \boldsymbol{l}_{-d})$.

In each subframe $\tilde{n}$ of a frame, as per the D2D players' lists, the BS decides on which CU's resources are to be allocated to the D2D players. Using these resources, each D2D player then starts transmitting. A D2D player d then observes a rate of $r_d(\tilde{n})$. We define the utility of a D2D player as follows.

Definition 1. (Utility) *The utility $r_d(n)$ of a D2D player d, calculated at the end of a frame n is the average of the rates obtained by it over all the sub-frames of a frame and is given by,*

$$r_d(n) = \frac{1}{N_D} \sum_{\tilde{n}=1}^{N_D} r_d(\tilde{n}). \tag{2}$$

We normalize the utility $r_d(n)$ such that it is strictly bounded between 0 and 1, [13].

Definition 2. (Social Utility) *The social utility is the sum of utilities of the D2D players obtained in every frame n and is given by $W_{\boldsymbol{l}}(n) = \sum_{d \in \mathcal{N}_\mathcal{D}} r_d(n)$.*

Our objective is to maximize the social utility (sum throughput) of the D2D players in every frame n, over all possible action profiles $\boldsymbol{l}$, to get the optimal action profile $\boldsymbol{l}^*(n)$, subject to a rate constraint of r_{tgt},

$$\begin{aligned} \boldsymbol{l}^*(n) &= \arg\max_{\boldsymbol{l}} \sum_{d \in \mathcal{N}_\mathcal{D}} r_d(n), \\ \text{s.t.} \quad r_d(n) &\geq r_{tgt}, \quad \forall d \in \mathcal{N}_\mathcal{D}. \end{aligned} \tag{3}$$

Thus, the optimal action profile $\boldsymbol{l}^*(n)$ is the set of socially optimal actions (Pareto efficient) taken by the D2D players such that the sum throughput of the D2D players is maximized.

III. Resource Allocation Algorithm

The resource allocation algorithm that we propose is a learning algorithm that guarantees convergence to an optimal action profile provided interdependence is ensured in the game. Interdependence implies that the utility of every player is affected by the choice of actions of other players. This in turn means that the change of utility of a player affects the utility of every other player.

Definition 3. (Interdependence) *A game $\mathcal{G}$ is interdependent if for every action profile $\boldsymbol{l} \in \mathcal{A}$ and for every proper subset of players $\mathcal{N} \subset \mathcal{N}_D$, there exists a player $d \notin \mathcal{N}$ and a choice of actions $\boldsymbol{l}'_\mathcal{N} \in \mathcal{A}_d^{|\mathcal{N}|}$ such that $r_d(\boldsymbol{l}'_\mathcal{N}, \boldsymbol{l}_{-\mathcal{N}}) \neq r_d(\boldsymbol{l}_\mathcal{N}, \boldsymbol{l}_{-\mathcal{N}})$.*

We consider the internal state variables of a D2D player to be its list $l_d(n)$, its utility $u_d(n)$ and another variable called its mood $m_d(n)$, based on which a D2D player decides on its action in every frame n. The mood can take two values: content (C) or discontent (D). The resource allocation algorithm which we will discuss next consists of the following stages: 1) list selection of a D2D player, 2) the BS's resource allocation rule 3) utility calculation for a D2D player and 4) its mood calculation method (refer to *Algorithm 1*).

A. List Selection

The list selection for every D2D player is done on a frame by frame basis. The mood of a D2D player in the previous frame helps it to select its list in the present frame. Let ϵ be the exploration/experimentation rate, such that $\epsilon > 0$ and k is a constant such that $k > N_D$ [13]. A D2D player whose mood is content can either decide to retain (exploit) its previous list with probability $1 - \epsilon^k$ or select (explore) a new list with probability $\epsilon^k/(L-1)$. It is more likely to retain its previous list rather than exploring other lists if the exploration rate ϵ^k is less than $1 - 1/L$. Thus, ϵ^k is an important design parameter that decides on the frequency of exploring versus exploiting lists. On the other hand, a discontent D2D player explores all the L possible lists with an equal probability of $1/L$. Steps 1-6 of *Algorithm 1* demonstrate the list selection method for a D2D player d. Once this is over for all the D2D players, every D2D player transmits its list to the BS. The BS then allocates the resources of the CUs to the D2D players as per the following allocation algorithm.

B. Allocation Algorithm at the BS

At the beginning of every frame, the lists of all the D2D players are available at the BS. The BS follows two round robin (RR) sequences to prioritize the D2D players in every subframe and also across a frame. It first orders the D2D players in a RR sequence, $RR\_seq = d_1, d_2, \dots, d_{N_D}$. This is followed in the first subframe of a frame as shown in Fig. 2. In the next subframe, the RR sequence starts from 2 such that $RR\_seq = d_2, \dots, d_{N_D}, d_1$ and so on till the last subframe N_D when the RR sequence starts from d_{N_D}, as seen in Fig. 2. Thus, the selection of the first player changes as per this RR sequence in every subframe of a frame.

In every subframe, the BS ensures orthogonal allocation of

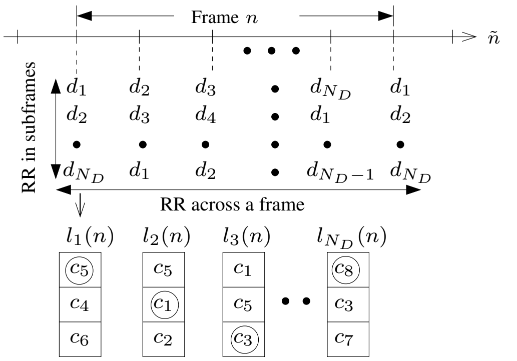
Frame n
$\tilde{n}$
RR in subframes
d_1 d_2 d_3 d_{N_D} d_1
d_2 d_3 d_4 d_1 d_2
d_{N_D} d_1 d_2 d_{N_D-1} d_{N_D}
RR across a frame
$l_1(n)$ $l_2(n)$ $l_3(n)$ $l_{N_D}(n)$
c_5 c_5 c_1 c_8
c_4 c_1 c_5 c_3
c_6 c_2 c_3 c_7
CU selection

Fig. 2: RR sequencing and the selection of CUs from the lists of all the D2D players in the first subframe of a frame.

CU resources as follows. It selects the first element (a CU) of a D2D player's list and allocates the CU's resources to it only if it is not allocated to the other players prior to it in the RR sequence followed in that subframe (see Fig. 2). If this element has already been assigned to another D2D player, the BS selects the next element of the D2D player's list till it reaches the end of the list. In case all the elements of its list are already assigned to the other players, it is not allocated any resource.

The RR across the subframes and way in which the utility is calculated enforces interdependence among the players as the priority of picking up a D2D player's list shifts from one D2D player to the next in each subframe. We explain this by considering the following example. As seen from Fig. 2, consider the first D2D player d_1 of the first subframe, which gets the first priority. Its selection of a CU, c_5 from its list $l_1(n)$ and in turn its observed rate is not affected by the choice of CUs (or lists) of other players because it gets the first precedence over others in the RR sequence followed in that subframe. Thus, its rate does not depend on the actions (lists) of others. As per *Definition 3*, interdependence should ensure that a player's utility depends on the actions of others. Now, with RR across the frame, in the second subframe, the priority shifts to the next D2D player d_2. Then the priority of d_1 becomes N_D. Its CU selection and hence its rate gets affected by the choice of CUs (or lists) of other players appearing before it in the RR sequence. Thus, when its utility is finally calculated as an average of the rates obtained by it in every subframe, the utility depends on the rate it achieves in every subframe and since the rate depends on the lists of other players, the utility depends on the actions of all the players. This is true for the other players also.

Algorithm 1 Resource Allocation Algorithm

```
List Selection $(m_d(n-1), l_d(n-1))$
if $m_d(n-1) == C$ then
    $l_d(n) \leftarrow i$,    w.p. $\frac{\epsilon^k}{L-1}$,    $\forall i \in \mathcal{A}_d \backslash l_d(n-1)$
    $l_d(n) \leftarrow l_d(n-1)$,    w.p. $1 - \epsilon^k$
else
    $l_d(n) \leftarrow i$,    w.p. $\frac{1}{L}$,    $\forall i \in \mathcal{A}_d$
end if

Allocation Algorithm at the BS $(\boldsymbol{l}(n))$
Initialize $RR\_seq \leftarrow [d_1\ d_2\ ...\ d_{N_D}]$
for all $\tilde{n} = 1\ to\ N_D$ do
    for all $j = 1\ to\ N_D$ do
        $index \leftarrow 1$
        $d \leftarrow RR\_seq[index]$
        Select D2D player $d$'s list
        $list\_index \leftarrow 1$
        while $list\_index \neq list\_length + 1$ do
            $c \leftarrow l_d(n, list\_index)$
            if $c$'s resources are not allocated to players before $d$ in $\tilde{n}$ then
                Assign $c$'s resources to $d$
                Break
            else
                $list\_index \leftarrow list\_index + 1$
            end if
        end while
        $index \leftarrow index + 1$
        if $\gamma_c < \gamma_{tgt}$ then
            $c$'s resources are not assigned to $d$
        end if
    end for
    Left shift $RR\_seq$ by 1
end for

Utility Calculation  $(c, r_d(\tilde{n}))$
$r_d(n) \leftarrow \frac{1}{N_D} \sum_{\tilde{n}=1}^{N_D} r_d(\tilde{n})$

Mood Calculation $(s_d(n-1), l_d(n), r_d(n))$
if $m_d(n-1) == C$ & $[l_d(n-1), r_d(n-1)] == [l_d(n), r_d(n)]$ then
    $m_d(n) \leftarrow C$
else if $r_d(n) \geq r_{tgt}$ then
    $m_d(n) \leftarrow C$, w.p. $\epsilon^{1-r_d(n)}$
    $m_d(n) \leftarrow D$, w.p. $1 - \epsilon^{1-r_d(n)}$
else
    $m_d(n) \leftarrow D$
end if
```

C. Allocation Feasibility Test

In order to ensure that CU communications are not hampered, in each subframe, the BS checks whether every CU's SINR decreases below γ_{tgt} or not, due to the interference from the D2D transmitters which have been allocated the resources of these CUs. For a CU whose SINR goes down below γ_{tgt}, the BS does not allocate its resources to the D2D player and the D2D player's rate $r_d(n)$ is zero in that subframe. After this feasibility test is over for all the CUs, the BS conveys to each D2D player which CU's resources are allocated to it. The mapping of CUs' resources to the D2D players is termed as the allocation profile. For example, from Fig. 2 if we assume that the elements c_5, c_1, c_3, ..., c_8 selected by the D2D players from its lists $l_1(n)$ to $l_{N_D}(n)$ pass the feasibility test, then the allocation profile is $(c_5, c_1, c_3, \dots, c_8)$.

D. Utility Calculation

In a subframe $\tilde{n}$, a D2D player transmits using the resources of the CU allocated to it by the BS and observes a certain rate $r_d(\tilde{n})$. At the end of a frame, it calculates its utility $r_d(n)$ as per Eqn. 2.

E. Mood Calculation

The mood of a D2D player is determined (refer to lines 31-38 of *Algorithm 1*) at the end of each frame from its previous state, its present list and utility. A content D2D player will remain content, if its present configuration does not not change with respect to its previous one, that is, $(l_d(n), r_d(n)) = (l_d(n-1), r_d(n-1))$. However this condition is violated if $l_d(n) \neq l_d(n-1)$ or $r_d(n) \neq r_d(n-1)$ or both. We next explain these cases. Suppose, in the present frame n, a content D2D player d decides to explore other lists. Then, its list $l_d(n)$ changes and $l_d(n) \neq l_d(n-1)$. Thus, its configuration changes. If this player decides to retain its previous list $l_d(n) = l_d(n-1)$, its present utility $r_d(n)$ can still change with respect to its previous utility $r_d(n-1)$. This is so because if some other player changes its list, then it would change the utility of this player because of interdependence. Therefore, as its configuration changes, the chances of becoming discontent or content depends on its present utility $r_d(n)$. If $r_d(n)$ is greater than the minimum required target rate r_{tgt}, it becomes content with probability $\epsilon^{1-r_d(n)}$. The chances of becoming content are more than that of becoming discontent, if $r_d(n)$ is sufficiently high. If $r_d(n)$ is less than r_{tgt}, then it becomes discontent with probability (w.p.) one.

On the other hand, for a discontent D2D player d to become content, it has to explore all the lists uniformly randomly. If its rate $r_d(n)$ is greater than r_{tgt}, it becomes content with probability $\epsilon^{1-r_d(n)}$ or it will remain discontent.

IV. System State Process

We define the present state $\boldsymbol{s}_d(n)$ of a D2D player d as a 3-D tuple, $\boldsymbol{s}_d(n) = (l_d(n), r_d(n), m_d(n))$. The collection of states of all the D2D players constitute the system state $s(n) = (\boldsymbol{s}_1(n), \ldots, \boldsymbol{s}_{N_D}(n))$. Let the action profile in every frame n be $\boldsymbol{l}(n) = (l_1(n), \ldots, l_{N_D}(n))$, the utility profile be $\boldsymbol{r}(n) = (r_1(n), \ldots, r_{N_D}(n))$ and the mood profile be $\boldsymbol{m}(n) = (m_1(n), \ldots, m_{N_D}(n))$.

Lemma 1. *The system state process $s(n)$ is a finite state, aperiodic and irreducible DTMC.*

Proof. 1) A D2D player's utility $r_d(n)$ depends on the lists of the other players. The selection of the present list $l_d(n)$ of a D2D player in turn depends on its previous mood $m_d(n)$. Its present mood $m_d(n)$ is also determined from its previous state. Therefore, its present state $\boldsymbol{s}_d(n) = (l_d(n), r_d(n), m_d(n))$ depends on its previous state. Hence, the system state process $s(n)$ is a DTMC.
2) Let us rearrange $s(n)$ in terms of the action, utility and mood profiles as $s(n) = (\boldsymbol{l}(n), \boldsymbol{r}(n), \boldsymbol{m}(n))$. Consider the sets $\mathcal{A}$, $\mathcal{U}$ and $\mathcal{M}$ to contain all possible values of $\boldsymbol{l}(n)$, $\boldsymbol{r}(n)$ and $\boldsymbol{m}(n)$. Let the set $\mathcal{S}' = \mathcal{A} \times \mathcal{U} \times \mathcal{M}$ be the state space of all possible combinations of $\boldsymbol{l}(n)$, $\boldsymbol{r}(n)$ and $\boldsymbol{m}(n)$. We next prove that the system state space $\mathcal{S}$ is a subset of $\mathcal{S}'$.

Let the cardinality of the space of $\boldsymbol{l}(n)$, $\boldsymbol{r}(n)$ and $\boldsymbol{m}(n)$ be $|\mathcal{A}|$, $|\mathcal{U}|$ and $|\mathcal{M}|$. Since each of the N_D D2D players can choose any of the L possible lists, $|\mathcal{A}| = L^{N_D}$. We know from the notion of interdependence that the rate $r_d(n)$ of a D2D player d is a function of the action profile $\boldsymbol{l}(n)$, that is, $r_d(n) = U_d(\boldsymbol{l})$. Since multiple action profiles (lists) can generate the same allocation profile and thus the same utility profile $\boldsymbol{r}(n)$, $U_d(\cdot)$ is a many to one function. Thus, $|\mathcal{U}|$ is less than $|\mathcal{A}|$. Since the mood of every D2D player can take two possible values, $|\mathcal{M}| = 2^{N_D}$. Moreover, those system states in which at least one D2D player d's rate $r_d(n)$ is less than r_{tgt} and yet its mood $m_d(n)$ is content, cannot be part of the state space $\mathcal{S}$. Therefore, $\mathcal{S}$ is finite and is a subset of $\mathcal{S}'$.
3) Since self transitions are possible from every system state, the DTMC is aperiodic.
4) As the system states communicate with each other, they form a single recurrent communication class. Thus, the DTMC is irreducible. □

Since the DTMC is aperiodic and irreducible, it has a unique stationary distribution π^ϵ on the state space S. Let us denote the transition probability matrix by P^ϵ. We consider this DTMC P^ϵ, to be the perturbed version of the process P^0 with ϵ equal to zero and is a regular perturbation of P^0 [13], [14]. The transitions in P^0 occur in P^ϵ with high probabilities but with a very small probability some transitions in P^ϵ occur which would not have occurred in P^0. For all $\boldsymbol{s}_1$, $\boldsymbol{s}_2 \in \mathcal{S}$, the state transitions $P^\epsilon_{\boldsymbol{s}_1\boldsymbol{s}_2}$ approach $P^0_{\boldsymbol{s}_1\boldsymbol{s}_2}$ as ϵ tends to zero, i.e, $lim_{\epsilon \to 0} P^\epsilon_{\boldsymbol{s}_1\boldsymbol{s}_2} = P^0_{\boldsymbol{s}_1\boldsymbol{s}_2}$. We are interested in finding those states of P^ϵ called the stochastically stable states in which the system spends a large fraction of its time that maximizes the social utility, with the D2D players content and their rate constraints satisfied.

Theorem 1. *The stochastically stable states of a regular perturbed DTMC are the states $\boldsymbol{s}^* \in \mathcal{S}$, which satisfy the following conditions:*
1) The action profile $\boldsymbol{l}(n)$ should maximize the social utility $W_{\boldsymbol{l}} = \sum_{d \in \mathcal{N_D}} r_d(\boldsymbol{l})$ while satisfying the rate constraints of all the D2D players.
2) For each D2D player d, its rate $r_d(n)$ should be aligned to the action profile $\boldsymbol{l}(n)$ of all the D2D players in the system, that is, $r_d(n) = U_d(\boldsymbol{l})$.
3) In a stochastically stable state $\boldsymbol{s}^*$, all the D2D players must be content.

V. Proof of Theorem 1

In order to determine the stochastically stable states we define a few terms as follows.

Definition 4. (Resistance of Transition) *Let $P^\epsilon_{\boldsymbol{s}_1\boldsymbol{s}_2}$ be the transition probability from the system state $\boldsymbol{s}_1$ to $\boldsymbol{s}_2$. If $P^\epsilon_{\boldsymbol{s}_1\boldsymbol{s}_2} > 0$ for some ϵ, then there exists a unique real number $r_{\boldsymbol{s}_1\boldsymbol{s}_2} \geq 0$ called the resistance of transition from $\boldsymbol{s}_1$ to $\boldsymbol{s}_2$ such that $0 < \lim_{\epsilon \to 0} \frac{P^\epsilon_{\boldsymbol{s}_1\boldsymbol{s}_2}}{\epsilon^{r_{\boldsymbol{s}_1\boldsymbol{s}_2}}} < \infty$. The resistance $r_{\boldsymbol{s}_1\boldsymbol{s}_2}$ is zero if $P^\epsilon_{\boldsymbol{s}_1\boldsymbol{s}_2} > 0$.*

Definition 5. (i-Tree) *In a directed graph $\mathcal{G}$, an i-tree T_i is a spanning tree such that there is exactly one directed path from every vertex j to i, such that $j \neq i$.*

Definition 6. (Resistance of an i-Tree) *The resistance r^{T_i} of an i-tree T_i is the sum of resistances of its edges and is given by $r^{T_i} = \sum_j r_{ji}^{T_i}$.*

Definition 7. (Stochastic Potential) *In a directed graph $\mathcal{G}$, the stochastic potential γ_i of vertex i is the minimum i-tree resistance among the resistances of all possible i-trees.*

Definition 8. (Stochastically Stable States) *The stochastically stable states of a perturbed DTMC P^ϵ are the states which are contained in the recurrence classes of P^0 with the minimum stochastic potential [14].*

We now identify the different recurrence classes of P^0.

Lemma 2. *The recurrence classes of P^0 are: 1) the class D^0 consisting of all the system states in which every D2D player is discontent and 2) the set of all singleton classes $\mathcal{C}^0 = \{C^0\}$, where in each system state C^0 every D2D player is content.*

Proof. 1) In a system state of D^0, all the D2D players are discontent and they start exploring lists uniformly randomly. As their actions change, the action profile also changes. This results in a state transition. Since ϵ is equal to zero (refer *Algorithm 1*), irrespective of the fact whether its utility $r_d(n) \geq r_{tgt}$, a discontent D2D player remains discontent with probability one. Thus, when the process enters a system state in D^0, it can only transition to one of the states in D^0 as it cannot transition to any of the content states in $\mathcal{C}^0$. Hence, D^0 is a recurrence class of P^0.

2) Let us consider a state in C^0 in which every D2D player is content. Referring to *Algorithm 1*, if ϵ is zero, the D2D players retain their previous lists in the present frame with probability one. The action profile of the system therefore remains the same over frames. Due to RR allocation at the BS, the set of allocation profiles of a frame also remain the same over all the frames and thus the utility profile also remains the same. As there is no change in any of the D2D player's configuration, that is, $(l_d(n-1), r_d(n-1)) == (l_d(n), r_d(n))$, every content D2D player remains content. Thus, a system state in C^0 maintains its state because the action profile, the utility profile and the mood profile remains the same. These are therefore absorbing states and are therefore recurrent.

Next, let us consider a system state $\boldsymbol{s} \in \mathcal{S}$ in which some of the D2D players are content while others are discontent. Every discontent D2D player selects its list uniformly randomly and still remains discontent with probability one since ϵ is zero. Due to interdependence enforced by the BS, the utility of a content D2D player gets affected due to the change of actions of the discontent players. Thus, the content D2D player's utility changes, even though its list remains the same. Since its configuration changes, it becomes discontent with probability one. Gradually, all the content D2D players become discontent. Thus, the state $\boldsymbol{s'}$ transitions to one of the discontent states of D^0. Hence, all such system states $\boldsymbol{s'} \in \mathcal{S}$ are transient. □

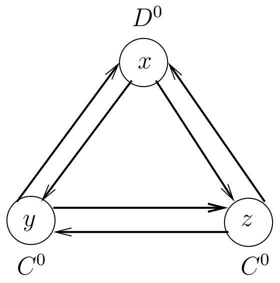

Fig. 3: Graph $\mathcal{G}$ to determine γ_{C^0} and γ_{D^0}.

With perturbation ϵ, we now consider the regular perturbed process P^ϵ and determine the stochastic potential of these recurrence classes. The transitions among the recurrence classes arise due to the perturbation ϵ. Let the stochastic potential of C^0 and D^0 be γ_{C^0} and γ_{D^0} respectively. The minimum resistance of transition from $C^0 \rightarrow D^0$, $D^0 \rightarrow C^0$ and $C^0 \rightarrow C^0$ can be determined as follows.

1) The transition from C^0 to a state in D^0 can occur when at least one content player decides to explore with probability $\epsilon^k/(L-1)$ and becomes discontent. Thus, content D2D players become discontent. Therefore, the minimum resistance of transition from C^0 to D^0 is k.

2) Similarly, the transition from a system state in D^0 to C^0 occurs when each D2D player explores lists uniformly randomly and becomes content with a probability of $\epsilon^{1-r_d(n)}$. Thus, the minimum resistance of transition from D^0 to C^0 is $\sum_{d \in \mathcal{N}_D}(1 - r_d(n))$.

3) The transition from a content state to another content state in $\mathcal{C}^0$, can occur when a content D2D player explores and its utility $r_d(n)$ is greater than r_{tgt}. Thus, the minimum resistance of transition in this case is $k + min_{d \in \mathcal{N}_D}(1 - r_d(n))$.

Let us consider the system state space $\mathcal{S}$ to consist of three recurrence classes: D^0 and two singleton classes in the set $\mathcal{C}^0$. We construct a directed graph $\mathcal{G}$ with these recurrence classes as its vertices and denote them by x, y and z. Fig. 3 shows the transitions or edges between the different recurrence classes.

Lemma 3. *The stochastic potential of a recurrence class C^0 is given by $\gamma_{C^0} = k(|\mathcal{C}^0| - 1) + \sum_{d \in \mathcal{N}_D}(1 - r_d(n))$.*

Proof. From graph $\mathcal{G}$ of Fig. 3, there are three possible *i-trees*, T_1, T_2 and T_3 that are rooted at vertex y. We represent them by a set of directed edges as follows: $T_1 = \{(x, y), (z, y)\}$, $T_2 = \{(x, z), (z, y)\}$ and $T_3 = \{(z, x), (x, y)\}$. The minimum resistances of these three *i-trees* are $r^{T_1} = r^{T_2} = k + min_{d \in \mathcal{N}_D}(1 - r_d(n)) + \sum_{d \in \mathcal{N}_D}(1 - r_d(n))$ and $r^{T_3} = k + \sum_{d \in \mathcal{N}_D}(1 - r_d(n))$. Thus, the minimum resistance *i-tree* is T_3 with resistance r^{T_3}, which is the stochastic potential of C^0. If the total number of states in $\mathcal{C}^0$ is $|\mathcal{C}^0|$, then the stochastic potential of C^0 is given by $\gamma_{C^0} = k(|\mathcal{C}^0| - 1) + \sum_{d \in \mathcal{N}_D}(1 - r_d(n))$. □

Lemma 4. *The stochastic potential of the recurrence class D^0 is given by $\gamma_{D^0} = k|\mathcal{C}^0|$.*

Proof. From graph $\mathcal{G}$, we note that there are three possible *i-trees* rooted at the vertex x, $T_1 = \{(y, x), (z, x)\}$, $T_2 = \{(z, y), (y, x)\}$ and $T_3 = \{(y, z), (z, x)\}$. The minimum resistance of these *i-trees* are, $r^{T_1} = 2k$ and $r^{T_2} = r^{T_3} =$

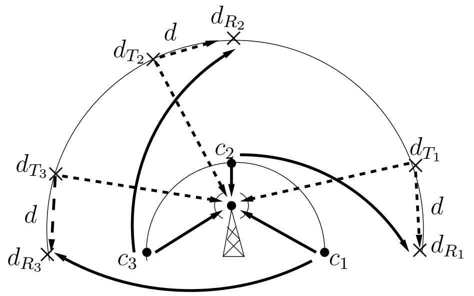

Fig. 4: Example demonstrating the optimal mapping of resources between the CUs and the D2D pairs.

$2k + min_{d\in\mathcal{N}_d}(1 - r_d(n))$. Thus, T_1 is the *i-tree* with the minimum resistance r^{T_1}, which is the stochastic potential of D^0. When the number of states in $\mathcal{C}^0$ is $|\mathcal{C}^0|$, the stochastic potential of D^0 is given by $\gamma_{D^0} = k|\mathcal{C}^0|$. □

The stochastic potential of C^0 is therefore less than that of D^0, i.e., $\gamma_{C^0} < \gamma_{D^0}$. Thus, the stochastically stable states are present in the recurrence classes of $\mathcal{C}^0$ which have the minimum stochastic potential.

The optimal action profile $\boldsymbol{l}^*(n)$ is the one that minimizes γ_{C^0} and is given by,

$$\boldsymbol{l}^*(n) \in \arg\min_{\boldsymbol{l}\in\mathcal{A}} k(|\mathcal{C}^0| - 1) + \sum_{d\in\mathcal{N}_D} (1 - r_d(n)),$$

which is equivalent to maximizing the sum throughput,

$$\boldsymbol{l}^*(n) \in \arg\max_{\boldsymbol{l}\in\mathcal{A}} \sum_{d\in\mathcal{N}_D} r_d(n).$$

This completes the proof.

VI. Concept Illustration

We demonstrate through the following example how the social optimum maximizes the social utility of the D2D players where the social optimum is the optimum allocation profile and the social utility is the sum throughput of the D2D players. The social optimum maximizes the social utility such that at least one of the player's utility is the best with none of the other players' utilities any worse off either. We consider a simple topology to illustrate how the optimal allocation profile can be determined. In order to theoretically calculate it we assume that all the channel gains are known, from which we can determine the value of the maximum sum throughput of the D2D players. Next, we show through simulations that our proposed algorithm also gives the same value. We are then able to identify the actions/lists that the D2D players select which gives this maximum value of the sum throughput. From this information, we can then determine the stochastically stable states of the system.

1) Topology: Consider a topology with three CUs c_1, c_2 and c_3 positioned in a semi-circle of radius R_1 as shown in Fig. 4. Three D2D pairs d_1, d_2 and d_3 are positioned in an outer semi-circle of radius R_2.

2) Parameter Setting: We set $R_1 = 50$ m and $R_2 = 100$ m. Every D2D transmitter subtends an angle of 10 degree at the BS with respect to its receiver. The distance between a D2D transmitter and receiver can be calculated from Fig. 4 and is 17.43 m. Let $P_D = P_C$ be 10 mW, the path loss exponent be 2 and the average noise power N_0 be 0.1 mW. Let γ_{tgt} be equal to 0 dB.

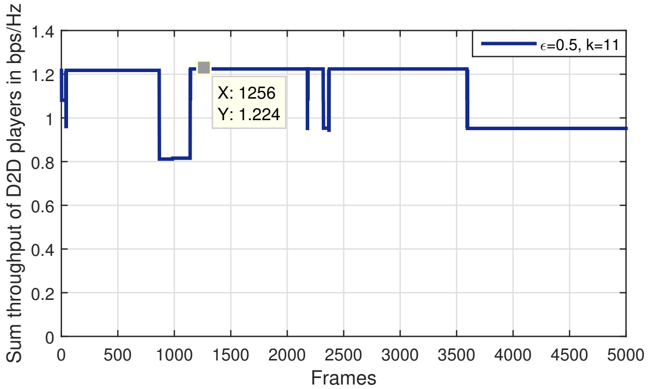

Fig. 5: Sum throughput of the D2D players over frames for the concept illustration example.

3) Analysis: Since the distances of the three CUs to the BS are R_1, the powers received from them at the BS are the same. The interference power received from the D2D transmitters at the BS is the same for all the D2D players because they are all equidistant from the BS. Hence, the SINRs of all the CUs are the same and we assume these to be higher than γ_{tgt}. Therefore, irrespective of the mapping of the resources of CUs to the D2D players, every allocation profile is feasible.

Now the best CU for a D2D player is one which is farthest from a D2D player's receiver. If CU c_1's resources are allocated to D2D pair d_3, then CU c_2's resources can be allocated to either d_1 or d_2. If c_2's resources are allocated to d_2, then d_2 faces the worst case interference from c_2. Thus, c_3's resources have to be allocated to d_1. Clearly, this allocation is not socially optimal because while d_1 and d_3 get their best choices of CUs, d_2 gets the worst CU. Thus, c_3's resources should be allocated to d_2 and c_2's resources should be allocated to d_1. Therefore, $a_1 = \{(d_1, c_2), (d_2, c_3), (d_3, c_1)\}$ is an optimal allocation profile. Note that the allocation profile $a_2 = \{(d_1, c_3), (d_2, c_1), (d_3, c_2)\}$ is also optimal. With allocation profile $a_1 = \{(d_1, c_2), (d_2, c_3), (d_3, c_1)\}$, the rates of the D2D players are calculated as $r_{d_1} = 0.4076$, $r_{d_2} = 0.4076$ and $r_{d_3} = 0.4089$ bps/Hz. Thus, the utility profile $\boldsymbol{r}(n)$ is (0.4076, 0.4076, 0.4089). Therefore, the maximum value of the sum throughput or the social utility is the sum of these rates which is 1.224 bps/Hz.

4) Verification: We simulate our proposed algorithm with the above topology and parameter setting. We set $\epsilon = 0.5$, $k = 11$ and obtain a plot of the sum throughput of the D2D players versus the number of frames as shown in Fig. 5. We observe that the maximum value of the sum throughput of the D2D players is 1.224 bps/Hz with the D2D players content. This matches our theoritical calculation. Two of the optimal action profiles are as follows: $\boldsymbol{l}_1^*(n) = ([2\ 3]^\mathsf{T},\ [3\ 1]^\mathsf{T},\ [1\ 2]^\mathsf{T})$ and $\boldsymbol{l}_2^*(n) = ([3\ 1]^\mathsf{T},\ [1\ 3]^\mathsf{T},\ [2\ 1]^\mathsf{T})$. When these lists are transmitted to the BS at the start of the frame, as per RR it determines

Algorithm 2 Threshold Based Utility Calculation

```
for all $\tilde{n} = 1 \; to \; N_D$ do
    $\boldsymbol{a}_d(\tilde{n}) \leftarrow c$
end for
$\boldsymbol{n}_d(c) \leftarrow \boldsymbol{n}_d(c) + 1$
$\boldsymbol{r}_a(c) \leftarrow \left(1 - \frac{1}{\boldsymbol{n}_d(c)}\right)\boldsymbol{r}_a(c) + \frac{1}{\boldsymbol{n}_d(c)} r_d(\tilde{n})$
if $\tilde{n} == mN_D^2$ then
    for all $c = 1 \; to \; N_C$ do
        if $\boldsymbol{u}_d(c) == 0$ or $abs(\boldsymbol{u}_d(c) - \boldsymbol{r}_a(c)) > \Delta$ then
            $\boldsymbol{u}_d(c) \leftarrow \boldsymbol{r}_a(c)$
        end if
    end for
    $r_d(\hat{n}) \leftarrow sum(\boldsymbol{u}_d(\boldsymbol{a}_d))/N_D$
end if
```

the optimal allocation profiles as $\{(d_1, c_2), (d_2, c_3), (d_3, c_1)\}$ and $\{(d_1, c_3), (d_2, c_1), (d_3, c_2)\}$. Note that these allocation profiles match our theoretically obtained ones. Thus, the optimal action profile, the utility profile and the mood profile constitute two stochastically stable states.

VII. Fading and CU Mobility

When the CUs are mobile and their location information is unavailable, devising efficient resource allocation algorithms for the D2D players is a challenging task. The mobility of a CU is characterized by its velocity and direction of movement. Therefore, the distance of a mobile CU to the D2D player's receiver changes with time. We consider slow fading due to shadowing as well as fast fading due to multi-path propagation. Thus, the D2D player's SINR and consequently its rate varies over subframes when it is allocated this CU's resources. Now with the same allocation profile or the same action profile, an infinite set of utility profiles are generated. The DTMC becomes infinite and all the states become transient. Therefore the steady state distribution π^ϵ does not exist. Thus, *Algorithm 1* fails to converge. We devise two novel utility calculation methods and modify *Algorithm 1* such that on an average the system performance does not degrade. The main concept behind these methods is to detect a significant change in the network topology or the channel variation that results in a change in the allocation profile of the D2D players. We next discuss these two methods.

A. Threshold Based Utility Calculation

In this method, we modify *Algorithm 1* so that a frame is repeated N_D times to form a superframe which is indexed by $\hat{n}$. The list selection by every D2D player and their mood calculation occurs at epochs which are multiples of a superframe instead of a frame. As per *Algorithm 1* the list generation method for every D2D player and the allocation method at the BS remains the same. The utility calculation method will now differ, which we explain next (refer *Algorithm 2*).

In any subframe $\tilde{n}$, two pieces of information are available to the D2D player d: 1) the c^{th} CU whose resources are allocated to it and 2) its present rate $r_d(\tilde{n})$ using this CU's resources. In a continuum of subframes, a CU c's resource can be allocated several times to a D2D player d, though it may not be allocated exactly in two consecutive subframes. However, every time that it is allocated this CU, it observes a change in its rate as a result of the CU's mobility and the channel. Let us say that in the first frame of length N_D of a superframe, a D2D player is allocated the resources of different CUs. It stores this sequence of CUs in a vector $\boldsymbol{a}_d$ of length N_D. In a superframe, every time a D2D player d observes a certain rate with the CU c, it calculates a running average of all these rates using Monte Carlo (MC) averaging and stores it in a vector $\boldsymbol{r}_a$ of length N_C at its c^{th} index. It also stores the number of times that the CU c's resources are allocated to the D2D player d in a vector $\boldsymbol{n}_d$ of length N_C at its c^{th} index. This is required for the MC averaging of the rates that it observes with this CU over time. This averaging is also done for all the other CUs whose resources are allocated to it. It also maintains a vector $\boldsymbol{u}_d$ of length N_C which stores the average rates of $\boldsymbol{r}_a$ in it in the first superframe (initially). Then in the next superframe as $\boldsymbol{r}_a$ gets populated, an absolute difference of the elements of both the vectors $\boldsymbol{u}_d$ and $\boldsymbol{r}_a$ is calculated. If any of the resultants obtained exceed a threshold Δ, then its indices are mapped to the utility vector $\boldsymbol{u}_d$ which is now updated with the corresponding values of $\boldsymbol{r}_a$. Thus, the decision to update the entries of the utility vector $\boldsymbol{u}_d$ with that of $\boldsymbol{r}_a$ depends on Δ.

At the end of every superframe, a D2D player retrieves the rates from the vector $\boldsymbol{u}_d$ corresponding to the indices which are the contents of the vector $\boldsymbol{a}_d$. It then sums up these rates and divides it by N_D which is the utility $r_d(\hat{n})$ of the D2D player d at the end of every superframe $\hat{n}$. The mood of the D2D player is then determined as per the mood calculation method of *Algorithm 1* from its utility, the list selected by it in the present superframe and its previous state.

B. Intercept-Slope Based Utility Calculation

This method is non-parameterized unlike the threshold based algorithm which depends on Δ and is more robust (refer *Algorithm 3*). As in *Algorithm 1*, the list selection for every D2D player occurs at the beginning of each frame while its mood calculation occurs at the end of it. As discussed in the previous method, a D2D player can be allocated a CU's resources multiple times over subframes. A D2D player stores its previous rates observed with every CU and the corresponding time instants in two vectors $\boldsymbol{r}_p$ and $\boldsymbol{t}_p$ of length N_C each. In a subframe, if it is allocated a CU c, then it retrieves the value of the previous rate that it has observed with this CU and the previous time stamp from the vectors $\boldsymbol{r}_p$ and $\boldsymbol{t}_p$ respectively, from their c^{th} index. From these values and the rate $r_d(\tilde{n})$ that it observes with this CU and the time stamp $\tilde{n}$, it calculates the intercept-by-slope ratio corresponding to the c^{th} CU. This is stored in the c^{th} index of another vector $\boldsymbol{b}_a$ of length N_C. This computation is done in every subframe corresponding to the CU whose resources are allocated to it. It then sorts this vector $\boldsymbol{b}_a$ containing these intercept-by-slope ratios in descending order and stores it in $\boldsymbol{z}_2$. This is compared with the sorted order obtained in the previous subframe, stored

Algorithm 3 Intercept-Slope Based Utility Calculation

```
$(b, a) \leftarrow$ Linear regression $(\boldsymbol{r}_p(c), \boldsymbol{t}_p(c), \tilde{n}, r_d(\tilde{n}))$
$\boldsymbol{b}_a(c) \leftarrow b/a$
$\boldsymbol{z}_2(\tilde{n}) \leftarrow sort(\boldsymbol{b}_a)$
$\boldsymbol{r}_p(c) \leftarrow r_d(\tilde{n})$
$\boldsymbol{t}_p(c) \leftarrow \tilde{n}$
if $\boldsymbol{u}_d(c) == 0$ or $\boldsymbol{z}_2(\tilde{n}) \neq \boldsymbol{z}_1(\tilde{n}-1)$ then
    $\boldsymbol{u}_d(c) \leftarrow r_d(\tilde{n})$
end if
$w_d(\tilde{n}) \leftarrow \left(1 - \frac{1}{\tilde{n}}\right) w_d(\tilde{n}-1) + \frac{1}{\tilde{n}} \boldsymbol{u}_d(c)$
if $\tilde{n} == N_D$ then
    $r_d(n) \leftarrow w_d(\tilde{n})$
end if
```

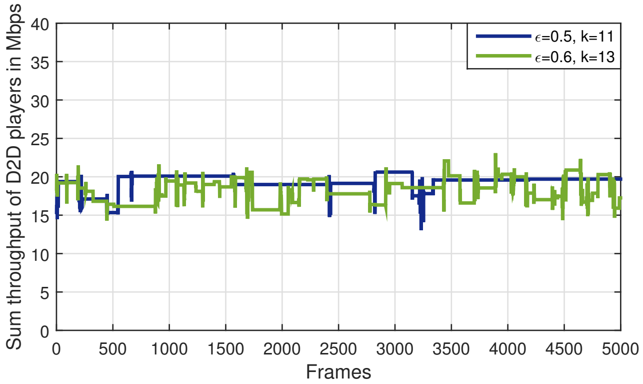

Fig. 6: Sum throughput of the D2D players over frames for *Algorithm 1*.

in another vector $\boldsymbol{z}_1$.

A change in the sorting orders indicate that the D2D player's observed rate with the CU c's resources in the present subframe has either increased or decreased with respect to the rates of the other CUs. This implies that the topology of the network or the channel could have changed significantly. Then the D2D player's present rate $r_d(\tilde{n})$ with the CU c is updated in the utility vector $\boldsymbol{u}_d$ of length N_C at its c^{th} index. If the sorting order remains the same, it discards the present rate. The vector $\boldsymbol{z}_2$ is then updated with the new sorting order. It then retrieves from the c^{th} index of the vector $\boldsymbol{u}_d$, the rate corresponding to the CU c. Over subsequent subframes of the frame, it continues to retrieve values from the indices of $\boldsymbol{u}_d$ corresponding to the CUs allocated to it in every subframe. It uses these values to calculate a running average (MC) $w_d(\tilde{n})$ till the end of the frame n, which is the utility $r_d(n)$.

VIII. Results

In this section, we verify the performances of the our proposed algorithms. The simulation parameters are as follows. The macro cell radius is 250 m in which the CUs (N_C = 10) and the D2D transmitters (N_D = 10) are uniformly distributed. We assume the range of D2D communications to be 50 m. A D2D receiver is uniformly distributed around a D2D transmitter within this range. Note that the same topology is used for all the simulation results. The transmit power of a CU is $P_C = 250$ mW and that of a D2D player is $P_D = 1$ mW. We consider the LTE pathloss model, $PL = 128.1 + 37.6\,log\,(d)$. The UE and BS noise figures are 9 dB and 5 dB respectively. We assume that each CU is allocated one physical resource block (PRB) of 180 kHz bandwidth in each subframe. The SINR target for the CUs is $\gamma_{tgt} = 0$ dB. The minimum rate r_{tgt} for each D2D player is 50 % of the rates observed by them in their initial positions. The utilities of all the D2D players are normalized by a factor such that these lie in the range of 0 to 1. We set the length of the lists, K to 3.

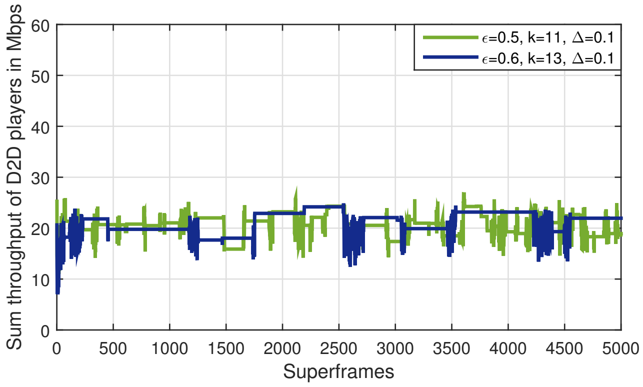

Fig. 7: Sum throughput of the D2D players over superframes with the threshold based utility calculation method.

A. *Without Fading and Mobility of CUs*

Fig. 6 demonstrates the performance of *Algorithm 1* with two different values of ϵ and k which are initially set to 0.5 and 11. As we increase them to 0.6 and 13, the exploration rate ϵ^k increases and the D2D players explore other lists more often. This results in their individual utilities to change more often due to a change in the action profile. This is reflected in the fluctuations observed in the sum throughput of the D2D players. The average sum throughput of the D2D players is approximately 18 Mbps.

B. *With Fading and Mobility of CUs*

We model slow fading with lognormal random variables of standard deviation 8 dB and for fast fading we model the channel gains as independent and exponentially distributed random variables with mean 1. The CU mobility is modeled as per the Random Way Point (RWP) model. The random direction in which a CU travels is uniformly distributed in the interval $[0\ 2\pi]$. We assume a constant pedestrian speed of 1 m/s for all the CUs.

When we incorporate the threshold based utility calculation method in *Algorithm 1*, we observe the following results with Δ equal to 0.1. As shown in Fig. 7, from a typical sample path realization of the sum throughput of the D2D players over superframes we observe that the average sum throughput of the D2D players is approximately 18 Mbps which is comparable to the optimal algorithm.

With the intercept-slope based method, which does not require any predefined parameter unlike the threshold based

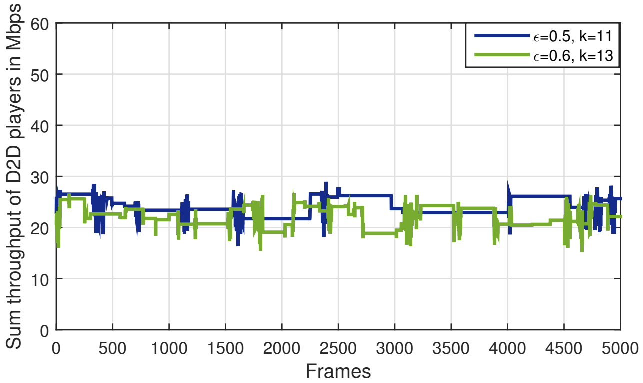

Fig. 8: Sum throughput of the D2D players over frames with the intercept-slope based utility calculation method.

method, we observe from Fig. 8 that the algorithm achieves a more robust performance as compared to the performance of the threshold based method. It shows less fluctuations before settling to a constant sum throughput. As we increase ϵ and k from 0.5 and 11 respectively to 0.6 and 13, we observe that the sum throughput of the D2D players fluctuates, as they explore other lists more frequently. Further, the average sum throughput of the D2D players is also around 18 Mbps.

IX. Conclusions

We have proposed an optimal resource allocation algorithm for D2D players in a game theoretic framework that ensures that the social utility is maximized, while ensuring that the communications of CUs are not hampered. We explain the notion of stochastically stable states and prove that it is in these system states that a D2D network's sum throughput is maximized. Our algorithm results in fast convergence to these states because the BS allocates resources orthogonally to the D2D players. We also propose two novel learning algorithms that show good performance in a fast fading channel with CU mobility.

References

[1] J. Wang, D. Zhu, C. Zhao, J. C. F. Li and M. Lei, "Resource Sharing of Underlaying Device-to-Device and Uplink Cellular Communications," *IEEE Commun. Lett.*, vol. 17, no. 6, pp. 1148-1151, June 2013.

[2] D. Feng, L. Lu, Y. Y. 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.

[3] W. Zhao and S. Wang, "Resource Allocation for Device-to-Device Communication Underlaying Cellular Networks: An Alternating Optimization Method," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1398-1401, Aug. 2015.

[4] S. Sesia, I. Toufik and M. Baker, "*LTE-The UMTS Long Term Evolution: From Theory to Practice*," 2nd ed., New York, NY, USA: Wiley, 2011.

[5] G. Hardin, "The Tragedy of the Commons", Science, vol. 162, no. 3859, pp. 1243-1248, Dec. 1968.

[6] F. Wang, L. Song, Z. Han, Q. Zhao and X. Wang, "Joint Scheduling and Resource Allocation for Device-to-Device Underlay Communication," in *Proc.* WCNC, pp. 134-139, Apr. 2013.

[7] X. Chen, R. Q. Hu and Y. Qian, "Distributed Resource and Power Allocation for Device-to-Device Communications Underlaying Cellular Network," in *Proc.* Globecom, pp. 4947-4952, Dec 2014.

[8] C. X. L. Song, Z. Han, Q. Zhao, X. Wang, X. Cheng and B. Jiao, "Efficiency Resource Allocation for Device-to-Device Underlay Communication Systems: A Reverse Iterative Combinatorial Auction Based Approach," *IEEE J. Sel. Areas. Commun.*, vol. 31, no. 9, pp. 348-358, Sept. 2013.

[9] H. H. Nguyen, M. Hasegawa and W. J. Hwang, "Distributed Resource Allocation for D2D Communications Underlay Cellular Networks," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 942-945, May 2016.

[10] R. Yin, C. Zhong, G. Yu, Z. Zhang, K. K. Wong and X. Chen, "Joint Spectrum and Power Allocation for D2D Communications Underlaying Cellular Networks," *IEEE Trans. Veh. Tech.*, vol. 65, no. 4, pp. 2182-2195, Apr. 2016.

[11] S. Maghsudi and S. Stanczak, "Hybrid Centralized-Distributed Resource Allocation for Device-to-Device Communication Underlaying Cellular Networks," *IEEE Trans. Veh. Tech.*, vol. 65, no. 4, pp. 2481-2495, Apr. 2016.

[12] M. Singh and P. Chaporkar, "An Efficient and Decentralised User Association Scheme for Multiple Technology Networks," in *Proc.* WiOpt, pp. 460-467, May 2013.

[13] J. R. Marden, H. P. Young and L. Y. Pao, "Achieving Pareto Optimality through Distributed Learning," in *Proc.* CDC, pp. 7419-7424, Dec. 2012.

[14] H. P. Young, "The Evolution of Conventions," *Econometrica*, vol. 61, no. 1, pp. 57-84, Jan. 1993.