D2D underlaid cellular networks with user clusters: Load balancing and downlink rate analysis

Chiranjib Saha and Harpreet S. Dhillon

Abstract—**In this paper we develop a comprehensive analytical framework for a cellular network enhanced with in-band device-to-device (D2D) communication capability where the D2D links reuse the downlink resources of the cellular links. The locations of the cellular base stations (BSs) are modeled as a Poisson point process (PPP). The user positions are modeled as a Thomas cluster process (TCP) to capture the inherent tendency of users to be located at close proximity to each other. The bandwidth allocated to D2D transmission is fixed and to cellular transmission is dynamic dependent on the load (the number of users) served by the macro BS. The users inside a cluster can either establish a D2D connection with a potential D2D transmitter (Tx) containing the file of interest within the same cluster if it is located closer than certain distance threshold or otherwise connects to the cellular network. By increasing this distance threshold, cellular traffic can be offloaded to D2D connections which in turn increases the interference from simultaneously active D2D Txs. We characterize the downlink rate coverage probability for this setup. Our analysis shows that there exists an optimum distance threshold for which rate coverage is maximized.**

Index Terms—**Stochastic geometry, Thomas cluster process, clustered D2D network, rate coverage, load balancing.**

I. Introduction

Increasing proliferation of internet-enabled mobile devices has resulted in exponential increase in mobile data traffic over the last decade. Owing to its capability of allowing direct communication between two proximate devices, D2D network has emerged as an attractive option for handling this data deluge. In particular, these direct links can be used by the proximate users to exchange data (e.g., popular media files, local information for ambient-aware proximity-based applications like social networks, public safety) that would have otherwise been downloaded from the cellular network [1]. Such devices therefore get offloaded from the cellular network to D2D connections. Although this offloading may increase interference from simultaneously active D2D Txs, the downlink data rate experienced by the users may improve since (i) D2D connections have shorter link distances, (ii) macro BSs with reduced load can allocate more resources to the remaining users. Thus D2D communication not only opens up a host of new proximity based service opportunities but may also enhance the network performance. These benefits have lead to its standardization in LTE Release-12 [2]. While D2D communication can in principle be allocated bands orthogonal to the cellular bands (out-of-band), due to significant spectrum shortage current research thrust is on in-band D2D communication *underlaid* on existing cellular network, where the D2D and cellular user equipments (DUEs and CUEs) reuse the same spectral resources [3]. This will also be our focus.

Due to its popularity, D2D has received a lot of attention in the recent past. While it is not possible to discuss all the prior art here, interested readers are advised to refer to [3] for a comprehensive survey. The more relevant works for this paper are the ones that focus on the *macroscopic* performance analysis of D2D networks using tools from stochastic geometry. The early works in this direction were mainly based on a simple spatial model in which D2D Txs were assumed to be located according to a Poisson point process (PPP) and the intended D2D receivers (Rxs) were located at a fixed or random distance from D2D Txs [4]–[6]. More recently, however, focus has shifted on developing a more sophisticated model in which *serving* D2D Tx for a given D2D Rx is not fixed *a priori*. First proposed in [7], this model captures the clustering behavior of D2D devices using Poisson cluster process. This general methodology has since then enabled the analysis of several aspects of D2D networks, e.g., see [7], [8], and even the analysis of user-centric deployments of heterogeneous cellular networks [9]. In this paper, we will use this approach to study downlink rate distribution and load balancing in D2D enhanced cellular networks with clustered user distributions.

Contributions and outcomes. We propose a new tractable model for cellular network underlaid with D2D connections. We model the locations of the users as TCP and the cellular macro BSs as an independent PPP. Each user either operates in D2D mode if the file of interest is available in one of the potential D2D Txs within a circular proximity region around itself or connects to the nearest macro BS. Note that increasing the radius of the proximity region–termed as *distance threshold*–is equivalent to biasing the fraction of cellular traffic to D2D connections. After the exact characterization of the distributions of interferences from the macro BSs and D2D Txs, we derive approximate overall downlink rate coverage probability. As a key intermediate step, we derive the average number of users served by the macro BS for which we need to obtain the average number of points of a TCP falling in a typical cell of the Poisson-Voronoi tessellation. Our analysis reveals several key system design insights. For instance, it demonstrates the existence of an optimum distance threshold for which the rate coverage is maximized.

II. System Model

A. Network Model

We consider a cellular network consisting of macro BSs underlaid with D2D Txs. The spatial distribution of the BSs

The authors are with Wireless@VT, Department of ECE, Virginia Tech, Blacksburg, VA, USA. Emails: {csaha, hdhillon}@vt.edu. The support of the US National Science Foundation (Grants CCF-1464293, CNS-1617896, IIS-1633363) is gratefully acknowledged.

is modeled by a homogeneous PPP $\Phi_{\rm m}$ of density $\lambda_{\rm m}$. All the BSs have equal downlink transmit power $P_{\rm m}$. We assume that the D2D Txs and UEs are distributed according to TCPs $\Phi_{\rm d}$ and $\Phi_{\rm u}$ respectively with same parent PPP $\Phi_{\rm p}$ with density $\lambda_{\rm p}$, where $\Phi_{\rm p}$ is independent of $\Phi_{\rm m}$. Thus if $\mathbf{X} \in \Phi_{\rm p} \subset \mathbb{R}^2$ denotes the location of a cluster center and $\mathbf{Y}_{\rm u}$ ($\mathbf{Y}_{\rm d}$) denotes the location of a UE (D2D Tx) with respect to $\mathbf{X}$, then $\mathbf{Y}_{\rm u}$ and $\mathbf{Y}_{\rm d}$ are Gaussian random vectors with $\mathbf{Y}_{\rm u}, \mathbf{Y}_{\rm d} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_2)$ ($\mathbf{I}_2$ denotes 2×2 identity matrix). Also we denote the number of elements per cluster of $\Phi_{\rm u}$ by $N_{\rm u}$, where $N_{\rm u}$ is a Poisson random variable with mean N. All D2D Txs in $\Phi_{\rm d}$ are assumed to transmit at constant power $P_{\rm d}$. In the sequel, we perform downlink analysis for a *typical UE* which is selected uniformly at random from an arbitrary cluster–also termed as a *representative cluster*–from $\Phi_{\rm u}$.

B. User Association

In this analysis, we assume that: (i) the UE will only connect to one of the D2D Txs within the same cluster, and (ii) the content of interest requested by the typical UE is available in a randomly selected D2D Tx of the representative cluster. Assumption (i) is justified since the clusters are typically separated in space which causes significant attenuation of the signals of D2D Txs from other clusters [7]. The low transmit power of D2D devices therefore makes it difficult for the devices across clusters to establish direct links with each other. Assumption (ii) makes sense because we do not know *a priori* where the content of interest for the typical UE is located. Note that even this simple setup significantly generalizes the Poisson dipole model-based setups typically used in the literature for D2D networks [4]–[6]. With some work, Assumption (ii) can actually be relaxed slightly by assigning to each D2D-Tx a probability of content availability which is proportional to file popularity distribution [1]. We delegate this discussion to the extended version of this paper due to space constraints.

We consider a *distance based association* policy for D2D transmission. In particular, each UE is allocated a potential serving D2D Tx containing the file of interest and belonging to the same cluster centered at $\mathbf{X} \in \Phi_{\rm p}$. A D2D link is established between the UE at $\mathbf{X}+\mathbf{Y}_{\rm u} \in \Phi_{\rm u}$ and the potential D2D Tx at $\mathbf{X}+\mathbf{Y}_{\rm d} \in \Phi_{\rm d}$ if $\|\mathbf{Y}_{\rm d} - \mathbf{Y}_{\rm u}\| < D$. If no D2D Tx is discovered in $b(\mathbf{X}, D)$, then the UE connects to the nearest macro BS in $\Phi_{\rm m}$, i.e., the BS that provides the maximum average received power.

C. Channel Model

We assume that all links between the transmitters (BS and D2D Tx) to the typical UE undergo distance dependent path-loss and Rayleigh fading. Thus, if $\mathbf{Z}_j$ is the location of a transmitter of type $j = \{\rm m, d\}$ where 'm' and 'd' stand for macro BS and D2D Tx, respectively, then the power received at the typical UE is $P_r = P_j h_j \|\mathbf{Z}_j\|^{-\alpha}$, where $\alpha > 2$ is the path-loss exponent and $\{h_j\}$ is an i.i.d. sequence of exponential random variables with unit mean modeling small scale fading. We also consider interference limited network and ignore the effect of thermal noise. We assume that the typical UE is at origin and its cluster center is at $\mathbf{X}_0$. The cluster of $\Phi_{\rm d}$ centered at $\mathbf{X} \in \Phi_{\rm p}$ is denoted as $\mathcal{B}^{\{\mathbf{X}\}}$. Conditioned on the fact that there exists an arbitrary D2D Tx at $\mathbf{Z}_{\rm d}^* \in b(\mathbf{0}, D) \cap \mathcal{B}^{\{\mathbf{X}_0\}}$ (we term this event as $S_{\rm d}$), the signal-to-interference ratio ($\mathtt{SIR}$) at the location of the typical UE can be expressed as:

$$\mathtt{SIR}_{\rm d} = \frac{P_{\rm d} h_{\rm d} \|\mathbf{Z}_{\rm d}^*\|^{-\alpha}}{\mathcal{I}_{\rm m,d} + \mathcal{I}_{\rm d,d}^{\tt intra} + \mathcal{I}_{\rm d}^{\tt inter}}, \tag{1}$$

where $\mathcal{I}_{\rm m,d} = \sum_{\mathbf{Z}_{\rm m}\in\Phi_{\rm m}} P_{\rm m} h_{\rm m} \|\mathbf{Z}_{\rm m}\|^{-\alpha}$, $\mathcal{I}_{\rm d,d}^{\tt intra} = \sum_{\mathbf{Z}_{\rm d}\in\mathcal{B}^{\{\mathbf{X}_0\}}\setminus\{\mathbf{Z}_{\rm d}^*\}} P_{\rm d} h_{\rm d} \|\mathbf{Z}_{\rm d}\|^{-\alpha}$ and $\mathcal{I}_{\rm d}^{\tt inter} = \sum_{\mathbf{Z}_{\rm d}\in\Phi_{\rm d}\setminus\mathcal{B}^{\{\mathbf{X}_0\}}} P_{\rm d} h_{\rm d} \|\mathbf{Z}_{\rm d}\|^{-\alpha}$ are respectively the interference from other BSs in $\Phi_{\rm m}$, D2D Txs of the representative cluster (termed as *intra-cluster D2D Txs*) and D2D Txs of other clusters (termed as *inter-cluster D2D Txs*).

When the typical UE connects to the BS (the event is denoted by $S_{\rm d}^{\rm c}$), the $\mathtt{SIR}$ can be expressed as

$$\mathtt{SIR}_{\rm m} = \frac{P_{\rm m} h_{\rm m} \|\mathbf{Z}_{\rm m}^*\|^{-\alpha}}{\mathcal{I}_{\rm m,m} + \mathcal{I}_{\rm d,m}^{\tt intra} + \mathcal{I}_{\rm d}^{\tt inter}}, \tag{2}$$

where, $\mathcal{I}_{\rm m,m} = \sum_{\mathbf{Z}_{\rm m}\in\Phi_{\rm m}\setminus\{\mathbf{Z}_{\rm m}^*\}} P_{\rm m} h_{\rm m} \|\mathbf{Z}_{\rm m}\|^{-\alpha}$ and $\mathcal{I}_{\rm d,m}^{\tt intra} = \sum_{\mathbf{Z}_{\rm d}\in\mathcal{B}^{\{\mathbf{X}\}}} P_{\rm d} h_{\rm d} \|\mathbf{Z}_{\rm d}\|^{-\alpha}$ and $\mathbf{Z}_{\rm m}^*$ denotes the location of the serving BS, i.e., $\mathbf{Z}_{\rm m}^* = \arg\max_{\mathbf{Z}_{\rm m}\in\Phi_{\rm m}} P_{\rm m} \|\mathbf{Z}_{\rm m}\|^{-\alpha}$.

D. Resource Allocation

For the downlink of all the macro BSs in $\Phi_{\rm m}$, we assume saturated dynamic resource allocation model, i.e., all the macro BSs always have some data to transmit to their associated users. Thus, assuming equal partition of the entire downlink resources among the CUEs (*roughly* the case in reasonable scheduling strategies over long term) associated with a macro BS, the downlink data rate achieved by a CUE is

$$\mathcal{R}_{\rm m} \triangleq \frac{W}{N_{\rm m}} \log(1 + \mathtt{SIR}_{\rm m}), \tag{3}$$

where W is the total available BW, $N_{\rm m}$ is the number of UEs *seen* by the tagged BS, also known as its load. For the in-band D2D transmission, we assume static partition of the bandwidth since it reduces additional overhead of the network to allocate resources to the D2D links. Thus the entire BW W is partitioned into $N_{\rm ch}$ orthogonal static channels and a D2D link is randomly assigned one of the channels. To improve area spectral efficiency, multiple D2D links within a cluster can reuse the same channel. Due to tractability reasons, we focus on the regime in which the load on a macro BS is much greater than $N_{\rm ch}$ (so that the channel allocated to the typical CUE overlaps in frequency with only one D2D channel). This is, however, not that unreasonable because the number of D2D connections in a given cluster will likely be much smaller than the load on a large macrocell. Hence, a CUE will observe interference from the D2D links of one channel reusing the same spectrum. For this setup, the rate achieved by a DUE is

$$\mathcal{R}_{\rm d} \triangleq \frac{W}{N_{\rm ch}} \log(1 + \mathtt{SIR}_{\rm d}). \tag{4}$$

Fig.1 provides a comprehensive illustration of the system model introduced in this section.

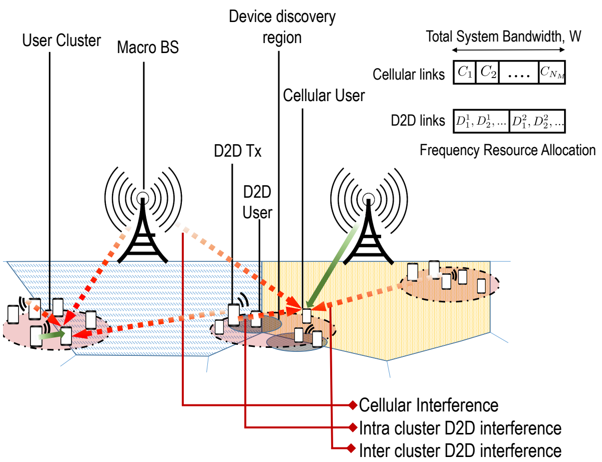

Fig. 1. Overview of the system model.

III. Rate Coverage

This is the main technical section of the paper where we derive the expression of rate coverage probability after characterizing the probabilities of association to D2D Tx and macro BS and the distribution of serving distances, interference and load distribution which are the key fragments of the analysis. The rate coverage probability, $\mathtt{P_r}$ is defined as the probability that the downlink data rate of a typical UE, denoted as $\mathcal{R}$ exceeds a certain threshold ρ. Hence $\mathtt{P_r} \triangleq \mathbb{P}(\mathcal{R} > \rho) =$

$$\mathbb{P}(\mathcal{R}_{\rm d} > \rho | S_{\rm d})\mathbb{P}(S_{\rm d}) + \mathbb{P}(\mathcal{R}_{\rm m} > \rho | S_{\rm d}^{\rm c})\mathbb{P}(S_{\rm d}^{\rm c}). \tag{5}$$

A. Association Probability and Serving Distance

Let $R_{\rm d}$ and $R_{\rm m}$ denote the distances from a randomly chosen D2D Tx within the representative cluster and the nearest macro BS from the typical UE. For this setup, the PDF of $R_{\rm m}$ is given by [10]:

$$f_{R_{\rm m}}(r_{\rm m}) = 2\pi\lambda_{\rm m}\exp(-\pi\lambda_{\rm m} r_{\rm m}^2), \quad r_{\rm m} > 0. \tag{6}$$

We denote the distance of the cluster center from the typical UE at origin as $V_0 = \|\mathbf{X}_0\|$. By definition of TCP, V_0 is a Rayleigh random variable with PDF and CCDF:

$$\text{PDF: } f_{V_0}(v_0) = \frac{v_0}{\sigma^2}\exp\left(-\frac{v_0^2}{2\sigma^2}\right), \quad v_0 > 0, \tag{7a}$$

$$\text{CCDF: } \bar{F}_{V_0}(v_0) = \exp\left(-\frac{v_0^2}{2\sigma^2}\right), \quad v_0 > 0. \tag{7b}$$

Let $\{u : u = \|\mathbf{X}_0 + \mathbf{Y}_{\rm d}\|\}$ be the sequence of distances from the typical UE to the intra-cluster D2D Txs. Then $\{u\}$ is a sequence of identically distributed but correlated random variables. However, conditioned on V_0, the sequence becomes i.i.d. The conditional distribution of $R_{\rm d}$ which is an element of $\{u\}$, given $V_0 = v_0$, is Rician [7, Lemma 1] with:

$$\text{PDF: } f_{R_{\rm d}}(r|v_0) = \tfrac{r}{\sigma^2}\exp\left(-\tfrac{r^2+v_0^2}{2\sigma^2}\right) I_0\left(\tfrac{rv_0}{\sigma^2}\right), \ r > 0, \tag{8a}$$

$$\text{CCDF: } \bar{F}_{R_{\rm d}}(r|v_0) = Q_1\left(\frac{v_0}{\sigma}, \frac{r}{\sigma}\right), \quad r > 0, \tag{8b}$$

where $I_0(\cdot)$ is the modified Bessel function of order zero and $Q_1(a,b) = \int_b^\infty \xi e^{-\frac{\xi^2+a^2}{2}} I_0(a\xi)\,{\rm d}\xi$ is the Marcum Q-function.

Also, let $\{t : t = \|\mathbf{X} + \mathbf{Y}_{\rm d}\|,\ \mathbf{X} \in \Phi_{\rm p} \setminus \{\mathbf{X}_0\}, \mathbf{Y}_{\rm d} \in \mathcal{B}^{\{\mathbf{X}\}}\}$ be the sequence of distances from the typical UE at origin to the D2D Txs belonging to another cluster centered at $\mathbf{X}$. Conditioned on $\|\mathbf{X}\| \equiv V = v$, $\{t\}$ is again an i.i.d. sequence. If $R_{\rm t}$ denotes the distance from an arbitrary inter-cluster D2D Tx to the typical UE at origin, then $R_{\rm t}$ is an element of $\{t\}$ and its conditional distribution given $V = v$ is same as that of $R_{\rm d}$ given $V_0 = v$ [7, Lemma 2]. Hence,

$$f_{R_{\rm t}}(\cdot|V = v) \equiv f_{R_{\rm d}}(\cdot|V_0 = v). \tag{9}$$

Using these distributions, the probabilities of association to D2D Tx and macro BS are given in the following Lemma.

Lemma 1. *The association probabilities of a typical UE to a D2D Tx and the macro BS given that it is located at a distance $V_0 = v_0$ from the cluster center are given by:*

$$\mathbb{P}(S_{\rm d}|v_0) \triangleq \mathcal{A}_{\rm d}(v_0) = \int_0^D f_{R_{\rm d}}(r_{\rm d}|v_0)\,{\rm d}r_{\rm d}, \tag{10a}$$

$$\mathbb{P}(S_{\rm d}^{\rm c}|v_0) \triangleq \mathcal{A}_{\rm m}(v_0) = 1 - \mathcal{A}_{\rm d}(v_0), \tag{10b}$$

where $f_{R_{\rm d}}(\cdot|v_0)$ is given by (8a).

Since we have assumed that the typical UE connects to a random D2D Tx in $b(\mathbf{0}, D)$, $\mathcal{A}_{\rm d}(v_0) = \mathbb{P}(R_{\rm d} < D|v_0)$, from which (10) follows. The association probabilities to D2D Tx or macro BS are obtained by deconditioning over V_0, i.e., $\mathcal{A}_{\rm d} = \mathbb{E}_{V_0}[\mathcal{A}_{\rm d}(V_0)]$, $\mathcal{A}_{\rm m} = 1 - \mathcal{A}_{\rm d}$ where the distribution of V_0 is specified in (7).

In the next Lemma, we characterize the distance of the serving macro BS or the D2D Tx from the typical UE. Conditioned on $V_0 = v_0$ and $\mathbf{1}(S_{\rm d}(v_0)) = 1$, serving distance is denoted by $X_{\rm d} = R_{\rm d}$ given that $R_{\rm d} < D$. When $\mathbf{1}(S_{\rm d}(v_0)) = 0$, serving distance $X_{\rm m}$ is simply equal to $R_{\rm m}$.

Lemma 2 (Serving distance distribution)**.** *If a typical user is located at a distance $V_0 = v_0$ from the cluster center, the PDF of the serving distance conditioned on the event $S_{\rm d}(v_0)$ is*

$$f_{X_{\rm d}}(x_{\rm d}|v_0) = \frac{1}{\mathcal{A}_{\rm d}(v_0)} f_{R_{\rm d}}(x_{\rm d}|v_0), \qquad 0 \leq x_{\rm d} \leq D, \tag{11}$$

and the PDF of the serving distance conditioned on the event $S_{\rm d}^{\rm c}(v_0)$ is

$$f_{X_{\rm m}}(x_{\rm m}) = 2\pi\lambda_{\rm m} x_{\rm m}\exp\left(-\pi\lambda_{\rm m} x_{\rm m}^2\right), \quad x_m \geq 0, \tag{12}$$

where $f_{X_{\rm d}}(\cdot)$ and $\mathcal{A}_{\rm d}(v_0)$ are given by (8a) *and* (10a).

B. Interference distribution

We now derive the Laplace transforms of the distributions of of different interference components appearing in the expressions of $\mathtt{SIR}$ in (1) and (2).

Lemma 3. *The Laplace transform of interference from interfering of the macro BSs when the typical UE is served by its nearest macro BS at $X_{\rm m} = x_{\rm m}$ is:*

$$\mathcal{L}_{\mathcal{I}_{\rm m,m}}(s|x_{\rm m}) = \exp\left(-2\pi\lambda_{\rm m}\int_{x_{\rm m}}^\infty \frac{sP_{\rm m}u^{-\alpha}}{1 + sP_{\rm m}u^{-\alpha}}u\,{\rm d}u\right). \tag{13}$$

Proof. The proof follows on the same lines as [11, Theorem 1]. □

Lemma 4. *The Laplace transform of interference from all macro BSs when the typical UE is served by a D2D Tx is:*

$$\mathcal{L}_{\mathcal{I}_{\mathrm{m,d}}}(s) = \exp\left(-\pi\lambda_{\mathrm{m}}\frac{s^{2/\alpha}{P_{\mathrm{m}}}^{2/\alpha}}{\mathrm{sinc}(2/\alpha)}\right). \quad (14)$$

Proof. The proof follows on the same lines as [12, Theorem 1]. □

Before we derive the Laplace transforms of the distribution of interference caused by D2D Txs, it is important to characterize the number of interfering D2D Txs. Each UE in Φ_{u} is independently marked as CUE or DUE according to its connectivity with the probability $\mathcal{A}_{\mathrm{d}}$ and $(1-\mathcal{A}_{\mathrm{d}})$. Since the number of UEs per cluster is $N_{\mathrm{u}} \sim \mathrm{Poisson}(N_{\mathrm{m}})$, then the number of DUEs per cluster is $N_{\mathrm{d}} \sim \mathrm{Poisson}(\mathcal{A}_{\mathrm{d}}N_{\mathrm{m}})$ and the number of CUEs per cluster is $N_{\mathrm{c}} \sim \mathrm{Poisson}((1-\mathcal{A}_{\mathrm{d}})N_{\mathrm{m}})$ [13]. Note that each D2D link was randomly assigned a channel from a pool of N_{ch} mutually orthogonal channels. That being said, number of D2D links in each channel, call it N_{d} is i.i.d. with $N_{\mathrm{d}} \sim \mathrm{Poisson}\left(\mathcal{A}_{\mathrm{d}}N_{\mathrm{m}}/N_{\mathrm{ch}}\right)$. As stated in Section II-D, each CUE gets interference from D2D links of only one channel. Hence if N_{d}^{I} is the number of D2D interferers per cluster to a CUE, then $N_{\mathrm{d}}^{I} = N_{\mathrm{d}}$.

Lemma 5. *The Laplace transform of interference from all intra-cluster D2D Txs when the typical UE connects to the macro BS, conditioned on $V_0 = v_0$ is given by:*

$$\mathcal{L}_{\mathcal{I}_{\mathrm{d,m}}^{\mathtt{intra}}}(s|v_0) = \exp\left(\theta(\mathcal{M}(s,v_0)-1)\right), \quad (17)$$

where

$$\mathcal{M}(s,v_0) = \int_0^\infty \frac{P_{\mathrm{d}}sw^{-\alpha}}{1+P_{\mathrm{d}}sw^{-\alpha}} f_{R_{\mathrm{d}}}(w|v_0)\,\mathrm{d}w \quad \textit{and,} \quad (18)$$

$$\theta = \frac{\mathcal{A}_{\mathrm{d}}N_{\mathrm{d}}}{N_{\mathrm{ch}}}. \quad (19)$$

Proof. See Appendix A. □

Lemma 6. *The Laplace transform of interference from all intra-cluster D2D Txs when the typical UE connects to a D2D Tx, conditioned on $V_0 = v_0$ is given by:*

$$\mathcal{L}_{\mathcal{I}_{\mathrm{d,d}}^{\mathtt{intra}}}(s|v_0) = \frac{1}{\mathcal{M}(s,v_0)}\frac{\exp\left(\theta\mathcal{M}(s,v_0)-1\right)}{\exp(\theta)-1}, \quad (20)$$

where $\mathcal{M}(s,v_0)$ and θ are given by (18) *and* (19) *respectively.*

Proof. See Appendix B. □

Lemma 7. *The Laplace transform of the distribution of inter-cluster D2D interference i.e. the interference from all inter-cluster D2D Txs is given by: $\mathcal{L}_{\mathcal{I}_{\mathrm{d}}^{\mathtt{inter}}}(s) =$*

$$\exp\left(-2\pi\lambda_{\mathrm{p}}\int_0^\infty (1-\exp\left(\mathcal{M}'(s,v)\right)v\,\mathrm{d}v\right), \quad (21)$$

$$\textit{with, } \mathcal{M}'(s,v) = \int_0^\infty \frac{P_{\mathrm{d}}sw^{-\alpha}}{1+P_{\mathrm{d}}sw^{-\alpha}} f_{R_{\mathrm{t}}}(w|v)\,\mathrm{d}w, \quad (22)$$

where $f_{R_{\mathrm{t}}}(\cdot)$ is substituted from (9).

Proof. See Appendix C. □

C. Load Characterization

In this section, we will characterize the load (number of users) served by the macro BS nearest to the typical UE, also known as the *tagged BS*. When Φ_{u} is a PPP of UEs independently distributed with respect to Φ_{m}, the PMF of load on the tagged BS was derived in [14]. In this paper, we characterize the load on a tagged BS when Φ_{u} is a TCP, independent of Φ_{m}. The distribution of load in this case is difficult to obtain. We will derive an approximate expression of the mean load. The accuracy of this approximation will be verified in Section IV.

Lemma 8. *When the UEs are distributed as a TCP Φ_{u} independent of the macro BS distribution which is a PPP, Φ_{m}, the mean of the number of other UEs-except the typical user-associated with the tagged BS is*

$$\mathbb{E}[N_{o,\mathrm{macro}}] \approx 1.28\frac{\mathrm{N}\mathcal{A}_{\mathrm{m}}\lambda_{\mathrm{p}}}{\lambda_{\mathrm{m}}}. \quad (23)$$

Proof. See Appendix D. □

Conditioned on the fact that the typical user connects to the macro BS, the load on the tagged BS is obtained by adding one to the expression in (23). We present the result in the following Corollary.

Corollary 1. *The mean load on the tagged BS conditioned on the fact that the typical user operates in cellular mode can be written as:*

$$\bar{N}_{\mathrm{m}} \triangleq \mathbb{E}[N_{\mathrm{m}}] \approx 1 + 1.28\frac{\mathrm{N}\mathcal{A}_{\mathrm{m}}\lambda_{\mathrm{p}}}{\lambda_{\mathrm{m}}}. \quad (24)$$

The results following this approximation will be validated in Section IV.

D. Main Result

After expressions of the interference distribution and load are characterized, we present the rate distribution in the following Theorem.

Theorem 1. *The rate coverage probability $\mathtt{P_r}$ defined in* (5) *for a typical UE under the approximation of mean load is:*

$$\mathtt{P_r} = \mathbb{E}[\mathcal{A}_{\mathrm{d}}(v_0)\mathtt{P_{rd}}(v_0) + \mathcal{A}_{\mathrm{m}}(v_0)\mathtt{P_{rm}}(v_0)|V_0 = v_0], \quad (25)$$

where $\mathcal{A}_{\mathrm{d}}(v_0)$ and $\mathcal{A}_{\mathrm{m}}(v_0)$ are substituted from (10a) *and* (10b), *$\mathtt{P_{rd}}(v_0)$ and $\mathtt{P_{rm}}(v_0)$ from* (15) *and* (16) *at the top of the next page. Here, $\tau_{\mathrm{m}} = 2^{\rho\bar{N}_{\mathrm{m}}/W}-1$ and $\tau_{\mathrm{d}} = 2^{\rho N_{\mathrm{ch}}/W}-1$.*

IV. Results and Discussions

Validation of results: In this section, first we will validate $\bar{N}_{\mathrm{m}}$ i.e. the number of UEs associated with the tagged BS with simulation. For all simulations, we have taken $\lambda_{\mathrm{m}} = 10/(\pi 500^2)$, $\lambda_{\mathrm{p}} = \lambda_{\mathrm{m}}/5$, $\mathrm{N} = 10$, $\alpha = 4$, $W = 10$ MHz, $\rho = 256$ Kbps and $N_{\mathrm{ch}} = 10$. From Fig. 2, it is observed that the result presented in Corollary 1 remains very close to the actual simulation. Next, the accuracy of the approximate

$$\mathtt{P_{rd}}(v_0) = (\mathcal{R}_{\rm d}(v_0) > \rho | S_{\rm d}(v_0)) = \int_0^D \mathcal{L}_{\mathcal{I}_{\rm m,d}}\left(\frac{\tau_{\rm d} x_{\rm d}^{\alpha}}{P_{\rm d}}\right) \mathcal{L}_{\mathcal{I}_{\rm d,d}^{\tt intra}}\left(\frac{\tau_{\rm d} x_{\rm d}^{\alpha}}{P_{\rm d}}\middle| v_0\right) \mathcal{L}_{\mathcal{I}_{\rm d}^{\tt inter}}\left(\frac{\tau_{\rm d} x_{\rm d}^{\alpha}}{P_{\rm d}}\right) f_{X_{\rm d}}(x_{\rm d}|v_0)\,{\rm d}x_{\rm d}, \tag{15}$$

$$\mathtt{P_{rm}}(v_0) = (\mathcal{R}_{\rm m}(v_0) > \rho | S_{\rm d}^c(v_0)) = \int_0^\infty \mathcal{L}_{\mathcal{I}_{\rm m,m}}\left(\frac{\tau_{\rm m} x_{\rm m}^{\alpha}}{P_{\rm m}}\right) \mathcal{L}_{\mathcal{I}_{\rm d,m}^{\tt intra}}\left(\frac{\tau_{\rm m} x_{\rm m}^{\alpha}}{P_{\rm m}}\middle| v_0\right) \mathcal{L}_{\mathcal{I}_{\rm d}^{\tt inter}}\left(\frac{\tau_{\rm m} x_{\rm m}^{\alpha}}{P_{\rm m}}\right) f_{X_{\rm m}}(x_{\rm m})\,{\rm d}x_{\rm m}, \tag{16}$$

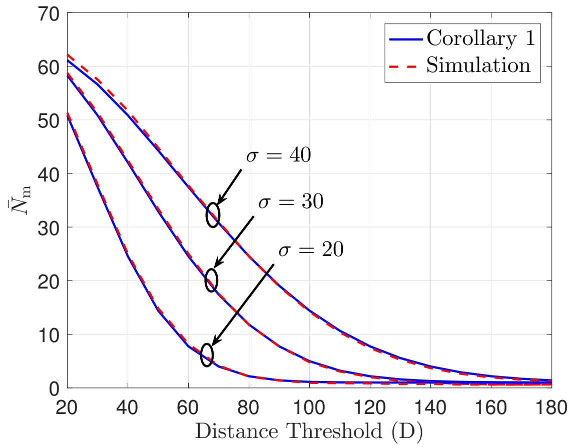

Fig. 2. Mean load on the tagged BS ($\bar{N}_{\rm m}$) as a function of distance threshold (D) for different cluster sizes (σ).

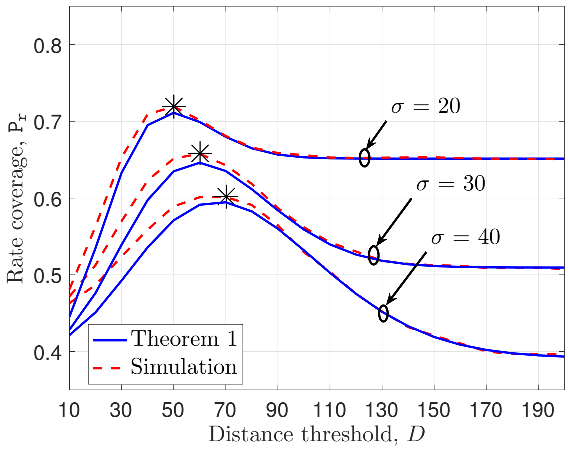

Fig. 3. Rate coverage probability ($\mathtt{P_r}$) as a function of distance threshold (D) for different cluster sizes (σ). The maxima is marked by asterisk ('$*$').

expression of $\mathtt{P_r}$ presented in Theorem 1 is validated with simulations. In Fig. 3, we have plotted $\mathtt{P_r}$ as a function of D for three different cluster sizes (σ). It can be observed that the simulation and analytical results follow the same trend and the gap between two curves is reduced as D increases. Note that $\mathtt{P_r}$ has two components $\mathcal{A}_{\rm m}\mathtt{P_{rm}}$ and $\mathcal{A}_{\rm d}\mathtt{P_{rd}}$ and we use mean load approximation for $\mathtt{P_{rm}}$ while the expression of $\mathtt{P_{rd}}$ is exact. As D increases the later dominates the former and the theoretical result approaches the results obtained by simulation. Nevertheless, the result is quite accurate for smaller values of D as well, especially given the fact that we used a simple *mean approximation* for the load.

Optimum distance threshold: Note that association probabilities in (10) are a function of D, and increasing value of D offloads users from macro BS to D2D connections (see Fig. 2). While by aggressive offloading of cellular traffic to D2D, remaining CUEs get more share of the cellular BW, the interference from the simultaneously active D2D links also increases. It can be observed from Fig. 3 that there exists an optimum value of D for which $\mathtt{P_r}$ is maximized. It is also interesting to note that the optimum value of D increases with cluster size (σ). As cluster size increases, a typical user tends to associate the macro BS with higher probability as the potential Txs are located further apart. Therefore, the *optimum* fraction of traffic will be offloaded from the cellular network to D2D connections at a slightly higher value of D.

V. Conclusion

In this paper we developed a comprehensive framework for an in-band D2D underlaid cellular network where depending on the availability of a D2D Tx containing the file of interest within a distance threshold, a UE can either connect to D2D or cellular mode in its downlink. We capture the effect of user clustering by modeling the users as a TCP while the distribution of the macro BSs is assumed to be a PPP. For this setup we calculate the downlink rate coverage probability. Our analysis shows that there exists a value of the distance threshold for which the downlink rate coverage probability is maximized. This is achieved by offloading the optimum fraction of cellular traffic to D2D connections. To the best of our understanding, this is the first work on load balancing in clustered D2D networks. Immediate extensions include rate coverage analysis for user-centric small cell deployment in a cellular network underlaid with D2D connections [9], incorporating more realistic content availability models [1], and doing exact analysis for the load on macrocells when the user distribution is not uniform, as was the case in this work.

Appendix

A. Proof of Lemma 5

The conditional Laplace transform of the distribution of $\mathcal{I}_{\rm d,m}^{\tt intra}$ given $V_0 = v_0$ can be written as

$$\mathcal{L}_{\mathcal{I}_{\rm d,m}^{\tt intra}}(s|v_0) = \mathbb{E}\left[\exp\left(-s\sum_{k=0}^{N_{\rm d}^I-1} P_{\rm d} h_{\rm d} \|\mathbf{X}_0 + \mathbf{Y}_{{\rm d}(k)}\|^{-\alpha}\right)\right]$$

$$\stackrel{\text{(a)}}{=} \mathbb{E}\left[\prod_{k=0}^{N_{\rm d}^I-1}\frac{1}{1+P_{\rm d} s w_k^{-\alpha}}\right] \stackrel{\text{(b)}}{=} \mathbb{E}\left[\mathcal{M}(s,v_0)\right]^{N_{\rm d}^I} \tag{26}$$

$$= \sum_{n=0}^{\infty}[\mathcal{M}(s,v_0)]^n \frac{e^{-\theta}\theta^n}{n!},$$

where (a) follows from the expectation over $h_{\rm d} \sim \exp(1)$. Here $\{\mathbf{Y}_{{\rm d}(k)}\} \subset \mathcal{B}^{\{\mathbf{X}_0\}}$ is the sequence of interfering D2D Tx locations. Step (b) follows from the change of variable $\|\mathbf{X}_0 + \mathbf{Y}_{{\rm d}(k)}\| \to w_k$ followed by changing from Cartesian to polar co-ordinates and finally applying conditionally i.i.d. property of $\{w_k\}$ which has PDF given by Lemma 2. In the last step we have taken expectation over $N_{\rm d}^I \sim {\rm Poisson}(\mathcal{A}_{\rm d}N_{\rm d}/N_{\rm ch})$. The final result is obtained by simple algebraic manipulation.

B. *Proof of Lemma 6*

Let us assume that the serving D2D link is assigned a channel index $i = {\rm rand}[1, N_{\rm ch}]$. Then the number of D2D links in the i^{th} channel is atleast one. We can write the conditional Laplace transform of the distribution of the aggregate interference from all intra-cluster D2D Txs in the i^{th} channel given $V_0 = v_0$ as: $\mathcal{L}_{\mathcal{I}_{\rm d,d}^{\rm intra}}(s|v_0) =$

$$\mathbb{E}\left[\exp\left(-s\sum_{k=0}^{N_{\rm d}-1}P_{\rm d}h_{\rm d}\|\mathbf{X}_0+\mathbf{Y}_{{\rm d}(k)}\|^{-\alpha}\right)\Bigg|N_{\rm d}\geq 1\right]$$

$$\stackrel{\text{(a)}}{=} \mathbb{E}\left[[\mathcal{M}(s,v_0)]^{N_{\rm d}-1}\Big|N_{\rm d}\geq 1\right] = \frac{\mathbb{E}\left[[\mathcal{M}(s,v_0)]^{N_{\rm d}-1}, N_{\rm d}\geq 1\right]}{\mathbb{P}(N_{\rm d}\geq 1)}$$

$$= \frac{\sum_{n=1}^{\infty}[\mathcal{M}(s,v_0)]^{n-1}\frac{e^{-\theta}\theta^n}{n!}}{1-e^{-\theta}},$$

where (a) follows from the same sequence of steps performed in the proof of Lemma 4 upto (26). The final result is obtained by some algebraic manipulation.

C. *Proof of Lemma 7*

Again assuming the typical UE is connected to a D2D Tx in channel i, the Laplace transform of the distribution of the aggregate interference from all the co-channel inter-cluster D2D Txs is $\mathcal{L}_{\mathcal{I}^{\rm inter}}(s) =$

$$\mathbb{E}\left[\exp\left(-s\sum_{\substack{\mathbf{X}\in\Phi_{\rm p}\setminus\\ \{\mathbf{X}_0\}}}\sum_{\substack{\mathbf{Y}_{\rm d}\in\\ \mathcal{B}^{\{X\}}}}P_{\rm d}h_{\rm d}\|\mathbf{X}+\mathbf{Y}_{\rm d}\|^{-\alpha}\right)\right] \stackrel{\text{(a)}}{=} \mathbb{E}_{\Phi_{\rm p}}\left[\mathbb{E}[\mathcal{M}'(s,v)]^{N_{\rm d}}\right]$$

$$\stackrel{\text{(b)}}{=} \exp\left(-2\pi\lambda_{\rm p}\int_0^{\infty}\left(1-\left[\mathbb{E}[\mathcal{M}'(s,v)]^{N_{\rm d}}\right]\right)v\,{\rm d}v\right).$$

Up to step (a), the proof follows the same sequence of steps mentioned in the proof of Lemma 5 used to derive (26). Step (b) follows from the probability generating functional (PGFL) of PPP. The final result is obtained after some algebraic manipulation.

D. *Proof of Lemma 8*

The set of CUEs will form a TCP, call it $\Phi_{\rm u,c}$ with average number of points per cluster being ${\rm N}\mathcal{A}_{\rm m}$. If $\#N_0$ denotes the number of points of $\Phi_{\rm u,c}$ falling in the Voronoi cell of a typical BS in $\Phi_{\rm m}$ at origin, then, $\mathbb{E}[\#N_0] =$

$$\mathbb{E}\left[\sum_{\mathbf{u}\in\Phi_{\rm u,c}}\prod_{\mathbf{v}\in\Phi_{\rm m}}\mathbf{1}(\|\mathbf{u}\|\leq\|\mathbf{u}-\mathbf{v}\|)|o\in\Phi_{\rm m}\right]$$

$$\stackrel{\text{(a)}}{=} \mathbb{E}_{\Phi_{\rm u,c}}\left[\sum_{\mathbf{u}\in\Phi_{\rm u,c}}\mathbb{E}^!_{0,\Phi_{\rm m}}\prod_{\mathbf{v}\in\Phi_{\rm m}}\mathbf{1}(\|\mathbf{u}\|\leq\|\mathbf{u}-\mathbf{v}\|)\right]$$

$$\stackrel{\text{(b)}}{=} \mathbb{E}_{\Phi_{\rm u,c}}\left[\sum_{\mathbf{u}\in\Phi_{\rm u,c}}\exp\left(-\lambda_{\rm m}\int_{\mathbb{R}^2}(1-\mathbf{1}(\|\mathbf{u}\|\leq\|\mathbf{v}-\mathbf{u}\|)\,{\rm d}\mathbf{v}\right)\right]$$

$$= \mathbb{E}_{\Phi_{\rm u,c}}\left[\sum_{\mathbf{u}\in\Phi_{\rm u,c}}\exp(-\lambda_{\rm m}\pi\|\mathbf{u}\|^2)\right]$$

$$\stackrel{\text{(c)}}{=} \int_{\mathbb{R}^2}\exp(-\lambda_M\pi\|\mathbf{u}\|^2)\lambda(\mathbf{u})\,{\rm d}\mathbf{u} = \frac{\mathcal{A}_{\rm m}\lambda_{\rm p}{\rm N}}{\lambda_{\rm m}},$$

where (a) follows from the independence of $\Phi_{\rm m}, \Phi_{\rm u,c}$. Here '$o$' denotes the origin and $\mathbb{E}^!_{0,\Phi_{\rm m}}$ denotes the expectation over reduced Palm distribution of $\Phi_{\rm m}$. Step (b) is from Slivnyak's theorem and PGFL of PPP [10], (c) follows from Campbell's theorem. Density of $\Phi_{\rm u,c}$ is given by, $\lambda(\mathbf{u}) = \frac{\lambda_{\rm p}\mathcal{A}_{\rm m}{\rm N}}{2\pi\sigma^2}\int_{\mathbb{R}^2}\exp\left(-\frac{\|\mathbf{u}-\mathbf{q}\|^2}{2\sigma^2}\right){\rm d}\mathbf{q}$ [10, Section 6.4]. Now we know that in a Poisson Voronoi tescellation, the mean area of the Voronoi cell containing the origin is 1.28 times the mean area of a typical Voronoi cell [14]. We apply this same scaling factor to $\mathbb{E}[\#N_0]$ to obtain the desired result.

References

[1] M. Afshang, H. S. Dhillon, and P. H. J. Chong, "Fundamentals of cluster-centric content placement in cache-enabled device-to-device networks," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2511–2526, Jun. 2016.

[2] 4G Americas, "3GPP release 12 executive summary," Tech. Rep., Feb. 2015.

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

[4] H. ElSawy, E. Hossain, and M. S. Alouini, "Analytical modeling of mode selection and power control for underlay D2D communication in cellular networks," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 4147–4161, Nov. 2014.

[5] Q. Ye, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "Resource optimization in device-to-device cellular systems using time-frequency hopping," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5467–5480, Oct. 2014.

[6] H. Sun, M. Wildemeersch, M. Sheng, and T. Q. S. Quek, "D2D enhanced heterogeneous cellular networks with dynamic TDD," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4204–4218, Aug. 2015.

[7] M. Afshang, H. S. Dhillon, and P. H. J. Chong, "Modeling and performance analysis of clustered device-to-device networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4957–4972, Jul. 2016.

[8] M. Afshang and H. S. Dhillon, "Spatial modeling of device-to-device networks: Poisson cluster process meets poisson hole process," in *Proc., Asilomar*, Nov. 2015, pp. 317–321.

[9] C. Saha, M. Afshang, and H. S. Dhillon, "Enriched K-tier HetNet model to enable the analysis of user-centric small cell deployments," submitted to *IEEE Trans. on Wireless Commun.*, 2016. Available online: arxiv.org/abs/1606.06223.

[10] M. Haenggi, *Stochastic Geometry for Wireless Networks*. New York: Cambridge University Press, 2013.

[11] H.-S. Jo, Y. J. Sang, P. Xia, and J. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3484–3495, Oct. 2012.

[12] H. Dhillon, R. Ganti, F. Baccelli, and J. Andrews, "Modeling and analysis of K-tier downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 550–560, Apr. 2012.

[13] R. G. Gallager, *Stochastic processes: theory for applications*. Cambridge University Press, 2013.

[14] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in heterogeneous networks: Modeling, analysis, and design insights," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2484–2497, May 2013.