Mobility Aware Relay Selection in 5G D2D Communication using Stochastic Model

Durgesh Singh and Sasthi C. Ghosh

***Abstract*—Challenges in relay selection for D2D communication arise due to the mobility of nodes which brings uncertainty in various network parameters. We first developed a network assisted stochastic integer programming (SIP) model to incorporate uncertainty which predicts the network parameters for upcoming time instance based on information available at current time instance. We converted SIP model to an equivalent deterministic mixed integer non-linear program (MINLP) model and proved its hardness result. By exploiting the constraints of MINLP, we developed a distributed greedy metric termed as connectivity factor (CF) which is calculated locally at each node on per-hop basis. It captures the nodes mobility and hence takes care of link reliability which in turn controls packet loss and delay. It can be computed in O(n) time, where n is the number of transmitters interfering with the given link. Our approach is applicable to any mobility model with relevant distributions of mobility parameters known. We constructed perceived graph based on CF values to devise network assisted and device controlled relay selection algorithms for given source-destination pairs. Extensive simulation results show significant improvements in packet loss and average end to end delay by our approach over a recent implementation of AODV based algorithm.**

***Index Terms*—Stochastic integer programming, Mobility of devices, Relay selection, 5G D2D communication, Packet loss and End to end delay.**

I. Introduction

DEVICE to device (D2D) communication in 5G is a paradigm shift in cellular networks which helps to meet challenges like seamless connection, high data rate and power savings arising due to exponential rise in mobile users in coming years [2]–[6]. In D2D communication, proximity users can directly communicate with each other with or without the help from the base station (BS). To implement D2D efficiently, one has to solve several issues like service and peer discovery, resource allocation, mode selection, channel quality estimation, relay selection and power usage [3]–[5], [7]–[10] among others. Parameters like connectivity, capacity of links, packet loss, delay and throughput are inherently dependent on temporal and spatial behavior of the mobile nodes. This shows that we must have to account for the dynamic nature of mobile nodes which brings stochastic elements in this type of networks.

We are considering the device-tier of 5G D2D cellular architecture as reported in [3] and shown in Figure 1. In this architecture, proximity devices or user equipments (UEs) can communicate directly among themselves *with* or *without* the help of a BS which are referred as *operator-controlled* (OC) and *device-controlled* (DC) respectively. In both OC and DC categories, UEs can either directly send data to the destination (DC−OC or DC−DC [3] shown in Figure 1(i) & 1(iii) respectively) or can send data through intermediate UEs acting as relays (DR−OC & DR−DC [3] shown in Figure 1(ii) & 1(iv) respectively) to further transfer the data to the destination.

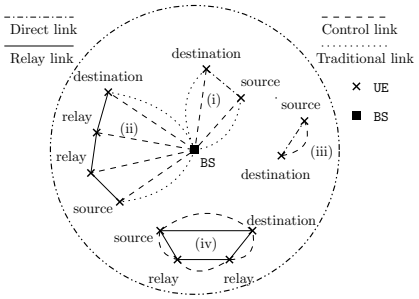

Fig. 1. Device-tier of the D2D architecture for 5G cellular networks [3]

In the relay selection problem appropriate relay nodes are chosen to transfer data packets from source to destination nodes when they are not in vicinity of each other. Here source node, destination node as well as the relay nodes all denote the UEs which may be mobile. One of the objective in the relay selection problem is to minimize the average delay while keeping the packet loss under control which is directly influenced by the selection of appropriate relay nodes. We assume that mobile nodes have the capability to establish links among themselves for D2D communication. They send each other some information in the form of packets over the communication channel. When the source and destination are not in vicinity of each other, the intermediate relay nodes receive the data packets from the sending node and forward it to an appropriate next hop node which can be another relay or the destination node. Time is discretized as t, $t+1$, $\cdots$, where Δt is the small time difference between t and $t+1$. The challenge arises due to the fact that nodes may change their position from a time instant t to the next time

Manuscript received xxxxxxxx xx, yyyy; revised xxxxxxxx, yyyy; accepted xxxxxxxx, yyyy. Date of publication xxxxxxxx xx, yyyy; date of current version xxxxxxxx xx, yyyy. The review of this paper was coordinated by Prof. xxxxxxxxxxxxxx. This paper was presented in part at the IEEE 16th International Symposium on Network Computing and Applications, Cambridge, MA, USA, October 2017 [1]. (Corresponding author: S. C. Ghosh.)

D. Singh and S. C. Ghosh are with the Advanced Computing & Microelectronics Unit, Indian Statistical Institute, Kolkata 700108, India (e-mails: durgesh.ccet@gmail.com, sasthi@isical.ac.in).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org

Digital Object Identifier xx.xxxx/TVT.xxxx.xxxxxxx

instant $t+1$. Thus data packets transferred at time t may get lost at time $t+1$ while they are still in transit, due to the mobility of nodes. We are considering that for a little time difference Δt, there can be a large impact on the distance and signal to noise plus interference ratio (SINR) after the link is established, depending on the velocity of the mobile users. A lot of research has been done to optimize one or more aforementioned objectives which considers only a snapshot of the network [11]–[16], which is inadequate when considering the velocity of nodes [17]. Since the mobility of nodes can lead to exponential performance degradation depending upon average users' speed, global optimum solution when available to the mobile nodes may become less useful [18]. Thus we need to develop a framework based on predicting the future information by capturing various mobility related parameters *locally*. Most of the works consider mobility *implicitly* [19], [20] and hence the performance is *tightly coupled* with the considered mobility model [17]. Hence there is a need to model the mobility of nodes *explicitly*. Note that different nodes may have different mobility patterns. This heterogeneity must also be captured while computing the SINR. Thus instantaneous SINR cannot be a good metric. Hence we need a metric which can predict channel parameters (e.g., SINR) for the upcoming time instance $t+1$ which should be computed at the current time t and should also consider the heterogeneous mobility of nodes. We have developed such a metric and shown its computations along with its computational efficiency in the analysis section. Specifically, our contributions can be summarized as follows:

- We have developed a stochastic integer programming (SIP) model to capture the dynamic network parameters by incorporating node's mobility explicitly (section III).
- We have converted the developed SIP to its equivalent deterministic mixed integer non-linear program (MINLP) and proved its NP-hardness (section IV).
- We have developed a distributed greedy metric termed as *connectivity factor* (CF) by exploiting the constraints of MINLP model, which captures the reliability and expected capacity of the links. This metric is computed locally at current time instance t which consists of the predicted information for the upcoming time instance $t+1$ with some probabilistic guarantee. We have shown that CF for a link can be computed in $O(n)$ time where n is the number of transmitters interfering to that link. CF is generic in nature which captures the heterogeneous mobility of devices and can be applied to both line of sight (LOS) and non-line of sight (NLOS) scenarios. To compute CF, either we need to know the distributions of mobility related parameters or knowing only the expectation and variance of these parameters is sufficient when the distributions are not known (section V).
- We have constructed a perceived graph using the CF metric. Perceived graph consists of the predicted information for the upcoming time instance $t+1$ which is computed at the current time instance t. Then we developed relay selection strategies for both the network assisted and the device controlled scenarios using perceived graph (section VI).
- We have validated the performance of our algorithm through extensive simulations and the results are compared with a recent algorithm based on AODV as described in [21] (section VII).

The related works are discussed in section II and conclusion is mentioned in section VIII.

II. Related Works

D2D communication in 5G is studied by several authors recently [2]–[5]. The authors in [3] gave a two-tier architecture of 5G D2D communication. In *device tier*, proximity devices communicate directly among themselves with or without the need of a BS. Whereas, in *macro tier*, devices take part in conventional cellular communication. In [4], [5], [7]–[10], [22], the authors discuss various issues and challenges in 5G D2D communication which are to be met for its successful implementation. Various optimization frameworks [11]–[14] have been proposed for solving the challenges in D2D communication. An optimization framework for resource allocation problem is given in [11], [12]. Performance of D2D communication using a stochastic geometry analysis is done in [13]. A simple 5G D2D network design has been developed in [14] as an MINLP to minimize the average network delay. A cross layer network assisted relay selection scheme for D2D communication in LTE-A is developed in [15]. Outage probability analysis of full duplex relay assisted D2D network is done in [16]. All these works consider a snapshot of the network and do not consider the mobility of the nodes adequately, thus they are more suitable for stationary D2D users as argued in [17]. Effects of user mobility in D2D network is considered in [23] by using stochastic geometry tools. Stochastic geometry has been used to analyze the transmission capacity for relay assisted D2D network [24]. The authors in [25] have used stochastic geometry for improving the coverage and spectral efficiency of millimeter wave (mmWave) D2D communication. The authors in [17] stated that optimization methods and stochastic geometry are not able to provide appropriate tools to capture the nodes' mobility and change in network topology adequately. The authors in [26] showed how chance constrained programming (CCP) is helpful in predicting uncertain scenarios with some guarantee of success for autonomous vehicles path planning and thus claiming the usefulness of CCP. The author in [27] considered various telecommunication problems where stochastic optimization is used as a method to tackle uncertainty due to mobility of nodes.

Mobility can play an important role to ensure the performance of several network related parameters [28], [29]. In [29], the authors surveyed the mobility models in vehicular adhoc network. Role of mobility on the behavior of energy and bandwidth efficiency has been discussed in [30] and node density has been considered to account for mobility. The authors in [31] concluded that mobility doesn't affect the performance of network much where mobile users are pedestrians. The authors in [32] discussed smart mobility solutions which reduces the signaling overhead for low speed mobile

users. However, mobility can lead to exponential performance degradation depending upon node's speed as mentioned in [18]. The authors in [18] gave a network assisted method where the global optimum solution may or may not be used by nodes based upon local information. This method is not applicable in the device-controlled mode, where nodes may move independently without any support from BS. The authors in [19], [20] considered the mobility of nodes in terms of the contact time where contact time denotes the duration for which two nodes are within the transmission range of each other. The authors in [20] developed a distributed transmission strategy where the nodes use the local information opportunistically to choose the next hop node based on the contact time distribution. They emphasized on transmission of single copy of a message to reduce the communication overhead. The authors in [17] argued that metric considering mobility implicitly (contact time) is tightly coupled with mobility models. Thus the performance of such metrics may be qualitatively different for various mobility models even when nodes have same average speed. They emphasized the importance of explicit mobility modeling which incorporates the user's movement directly into the optimization framework.

There is a need to develop metrics that should address the dynamic parameters due to mobility of nodes, not only for the network assisted case but also for the purely distributed case. There are several metrics described in literature and few to mention are in [33]–[35] to select the next hop, but they do not consider the mobility of nodes adequately. The authors in [36] argued that most of the studies give emphasis on network assisted approach and develop system centric metrics. Hence they developed a user centric metric but it also does not capture the mobility factor into account. SINR is a user centric metric which can be used to decide if a link should be active or not. But the authors in [37] discussed that SINR calculated in snapshot of network cannot be a good metric to predict link reliability due to the reason of mobility of nodes. Thus, it is required that mobility should be taken into account while computing the SINR values for the upcoming time instances. Distance between nodes is one of the parameters involved in computations of SINR. Distribution of distance between nodes is considered and analyzed in [23], [25] using stochastic geometry tools. The analysis based on distance distribution might dwell to an impractical solution when devices follow heterogeneous mobility settings with different speed ranges [38]. Supporting this argument, the authors in [39] studied and validated the impact of non-uniformly distributed users on the performance of 5G ultra dense networks. In our work, we have developed a distributed greedy metric which incorporates heterogeneous mobility of nodes into account explicitly. Our proposed metric not only takes care of the reliability of the link for the upcoming time instances but also predicts the variation in SINR and hence capacity of the link under consideration.

III. Stochastic Integer Programming Formulation of the Problem

We consider a service region served by a single BS where mobile devices or UEs are sending information to each other in the form of packets using wireless communication channels. We assume that mobile devices are moving independently and have the capability to establish links among themselves for D2D communications. Time is discretized and a mobile device may change its position from a time instant t to the next time instant $t+1$, where Δt is the small time difference between t and $t+1$. A graph $G^t = (N^t, E^t)$ is used to represent the mobile network at a particular time instant t, where N^t is the set of all nodes representing mobile devices and E^t is the set of all edges which represents links between the nodes which are in transmission range of each other at time t. There is a single static BS which is connected to all user devices in the given service region. Each link (i,j) at time t has a capacity C_{ij}^t for sending the data packets. Links may not be symmetrical due to presence of interference and uncertain wireless channels due to node's mobility. We are considering the general case, where Γ is set of source, destination and demand triplets $< s_\zeta, r_\zeta, b_\zeta >$. Here b_ζ is the demand supplied by source node s_ζ to destination r_ζ. Let $F_{ij}^t(\zeta)$ be the flow from node i to j for the source-destination pair $\zeta \in \Gamma$ at time t, an integer quantity which represents the number of packets transferred. Let δ_{ij}^{t+1} be the conditional probability that link (i,j) is active at time $t+1$ given that the link was active at time t. Let us define e_{ij}^t as a Boolean variable whose value is 1, if there is an edge between nodes i and j at time t, and 0, otherwise. Now δ_{ij}^{t+1} can be defined as follows:

$$\delta_{ij}^{t+1} = P\{e_{ij}^{t+1} = 1 | e_{ij}^t = 1\},\ \forall (i,j) \in E^t. \tag{1}$$

We assume that the network exhibits short range dependent characteristics. The average network delay t_s at each hop for the traffic exhibiting short range dependent characteristics can be computed as [14], [40]:

$$t_s \equiv \frac{1}{\lambda^t} \sum_{(i,j)\in E^t} \left[\frac{\lambda_{ij}^t}{\chi_{ij}^t - \lambda_{ij}^t} \right]$$

where $\lambda_{ij}^t = \sum_{\zeta\in\Gamma} F_{ij}^t(\zeta)$ is the average traffic on link (i,j) at time t, $\chi_{ij}^t = C_{ij}^t$ is the services capacity of link (i,j) at time t and $\lambda^t = \sum_{\zeta\in\Gamma} b_\zeta$ is the total demand in the network.

We will give a SIP model whose objective is to minimize average network delay per hop while keeping the link failure probability below a predefined threshold. Let α_{ij} be the threshold on allowable link failure probability for link (i,j). We define an optimization variable ω_{ij}^t, a Boolean variable which denotes those links (i,j) that are selected at time t for flow of traffic among possible set of available links in E^t. We can say that if at time t, $e_{ij}^t = 1$ and link (i,j) is selected for flow of packets, then $\omega_{ij}^t = 1$ else $\omega_{ij}^t = 0$. We also define a constant M which denotes the total number of available channels with the BS. The SIP formulation for a multiple source-destination pair is as follows:

$$Minimize\ t_s \tag{2}$$

$$\sum_{j:(i,j)\in E^t} F_{ij}^t(\zeta) - \sum_{i:(j,i)\in E^t} F_{ji}^t(\zeta) = \begin{cases} b_\zeta, & \text{if } i = s_\zeta \\ -b_\zeta, & \text{if } i = r_\zeta, \zeta \in \Gamma \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2893995, IEEE Transactions on Vehicular Technology

$$\sum_{\zeta\in\Gamma} F_{ij}^{t}(\zeta) \le E[C_{ij}^{t+1}]\cdot\omega_{ij}^{t}, \forall(i,j)\in E^{t} \quad (4)$$

$$F_{ij}^{t}(\zeta) \ge 0,\ \forall(i,j)\in E^{t},\ \zeta\in\Gamma \quad (5)$$

$$\sum_{(i,j)\in E^{t}} \omega_{ij}^{t} \le M \quad (6)$$

$$P\{e_{ij}^{t+1}=1|e_{ij}^{t}=1\} \ge (1-\alpha_{ij})\cdot\omega_{ij}^{t}, \forall(i,j)\in E^{t} \quad (7)$$

$$\omega_{ij}^{t}\in\{0,1\},\ \forall(i,j)\in E^{t} \quad (8)$$

The flow constraint (3) signifies the amount of flow coming in is same as the amount of flow going out at time t at a particular node except for source and destination nodes. The capacity constraint (4) signifies that total flow on link (i,j) at time t cannot exceed the expected capacity $E[C_{ij}^{t+1}]$ of that link at the next time instance $t+1$. Constraint (5) signifies that flow at a link cannot be negative. Constraint (6) signifies that there can be at most M links chosen for flow of traffic at time instance t as there are M available channels. Constraint (7) captures the conditional probability of link reliability for the next time instance $t+1$ given that it is active at the current time instance t. It checks if the link breakage probability satisfies the given threshold α_{ij}. Depending upon this constraint, we decide at time t to choose the appropriate link in the source-destination path for next time instance $t+1$. Hence this gives a guarantee over the link reliability probability.

In the next section, we will derive a way to convert the SIP to its equivalent MINLP. We will also prove that the converted MINLP is NP-hard.

IV. Converting SIP to its equivalent MINLP

A classical way of solving an optimization problem with probabilistic constraints is to first convert the probabilistic constraints into its equivalent deterministic constraints and then solve the optimization problem involving deterministic constraints only. Before converting constraint (7) to its equivalent deterministic form, we first need to take into the attenuation factors that affect the signal quality, which in turn affect the reliability of a link. The attenuation factors may be dependent on the device's mobility and other environmental factors which needs to be incorporated in our stochastic framework. Let us denote the overall attenuation by Ω_{ij}^{t}, which is a random variable capturing the randomness of the signal power and in turn the capacity of link (i,j) at time t, assuming noise is constant for the channel. We can say that if $\Omega_{ij}^{t}\le\gamma$, then then $e_{ij}^{t}=1$ else $e_{ij}^{t}=0$, where γ is a predefined threshold which determines the connectivity between two nodes. Now we will proceed to define the expectation $\mu_{ij}^{t}=E[\Omega_{ij}^{t}]$ and standard deviation $\sigma_{ij}^{t}=\sqrt{var(\Omega_{ij}^{t})}$, where $var(\Omega_{ij}^{t})$ is variance of Ω_{ij}^{t}. We assume these values are known for current time instance t. Assuming that link (i,j) is selected at time t ($\omega_{ij}^{t}=1$), constraint (7) can be re-written as:

$$P\{\Omega_{ij}^{t+1}\le\gamma|\Omega_{ij}^{t}\le\gamma\} \ge 1-\alpha_{ij}\ \ \forall(i,j)\ \in E^{t}. \quad (9)$$

Here Ω_{ij}^{t+1} represents the attenuation factor for link (i,j) in the next instance $t+1$. This is important because it gives a relationship of signal strength of mobile nodes for the upcoming time instance $t+1$ which is computed at the current time instance t.

Remark 1. *When two random variables X and Y are independent then we can say their conditional probability satisfies $P(X|Y)=P(X)$.*

Using remark 1 and the previous assumption that nodes are moving independently, we can say that Ω_{ij}^{t+1} and Ω_{ij}^{t} are independent and we can again rewrite constraint (9) as:

$$P\{\Omega_{ij}^{t+1}\le\gamma\} \ge 1-\alpha_{ij}\ \ \forall(i,j)\ \in E^{t}. \quad (10)$$

Using mean and variance of Ω_{ij}^{t+1}, we can convert probabilistic constraint (10) into its equivalent deterministic one using the approach stated in [41]. From constraint (10), we get,

$$P\left\{\frac{\Omega_{ij}^{t+1}-\mu_{ij}^{t+1}}{\sigma_{ij}^{t+1}} > \frac{\gamma-\mu_{ij}^{t+1}}{\sigma_{ij}^{t+1}}\right\} \le \alpha_{ij}. \quad (11)$$

Remark 2. *If X is a random variable with mean ν and variance s^2, then for any real number $c>0$, one-sided Chebyshev inequality (Cantelli's inequality) [42] can be stated as:*

$$P\left\{\frac{X-\nu}{s} > c\right\} \le \frac{1}{1+c^2}.$$

Using remark 2 with $c=\frac{\gamma-\mu_{ij}^{t+1}}{\sigma_{ij}^{t+1}}$, constraint (11) becomes

$$P\left\{\frac{\Omega_{ij}^{t+1}-\mu_{ij}^{t+1}}{\sigma_{ij}^{t+1}} > \frac{\gamma-\mu_{ij}^{t+1}}{\sigma_{ij}^{t+1}}\right\} \le \frac{1}{1+\left(\frac{\gamma-\mu_{ij}^{t+1}}{\sigma_{ij}^{t+1}}\right)^2}. \quad (12)$$

Now constraints (11) and (12) implies $\frac{1}{1+\left(\frac{\gamma-\mu_{ij}^{t+1}}{\sigma_{ij}^{t+1}}\right)^2} \le \alpha_{ij}$, which implies,

$$\mu_{ij}^{t+1}+\sigma_{ij}^{t+1}\sqrt{\frac{1-\alpha_{ij}}{\alpha_{ij}}} \le \gamma. \quad (13)$$

We can now replace probabilistic constraint (7) with its equivalent deterministic constraint (13) and get two new constraints:

$$\mu_{ij}^{t+1}+\sigma_{ij}^{t+1}\sqrt{\frac{1-\alpha_{ij}}{\alpha_{ij}}} \le \gamma+(1-\omega_{ij}^{t})\cdot L_{inf}, \forall(i,j)\in E^{t} \quad (14)$$

$$\mu_{ij}^{t+1},\sigma_{ij}^{t+1}\ge 0, \forall(i,j)\in E^{t}. \quad (15)$$

Here L_{inf} is a very large number representing positive infinity.

After converting the probabilistic constraint (7) in SIP formulation to its equivalent deterministic form, the final MINLP can be stated as: Minimize objective function (2) subject to the constraints (3)-(6), (14)-(15) and (8).

Proposition 1. *The MINLP formulated above is NP-Hard.*

Proof. Let us define a well known decision flow problem (ϱ) on graph $H(\hat{V},\hat{E})$ with source and sink vertices as $\hat{s}$ and $\hat{r}$, capacity $C(e)\in Z^{+}$ and price $p(e),\forall e\in\hat{E}$. Demand is $B\in Z^{+}$ and there is a bound $M'\in Z^{+}$ on the total price

of selected edges. Problem ϱ is defined as: is there a flow function $f : \hat{E} \rightarrow Z^+$ satisfying the following constraints:

$$\sum_{(i,j)\in\hat{E}} f(i,j) - \sum_{(j,i)\in\hat{E}} f(j,i) = 0,\ \forall\ i,j \in \hat{V} \setminus \{\hat{s},\hat{r}\} \quad (16)$$

$$\sum_{(i,\hat{r})\in\hat{E}} f(i,\hat{r}) - \sum_{(\hat{r},i)\in\hat{E}} f(\hat{r},i) = B \quad (17)$$

$$0 \leq f(e) \leq C(e) \quad (18)$$

$$\sum_{e:f(e)>0, e\in\hat{E}} p(e) \leq M'. \quad (19)$$

In [43], it is stated that problem ϱ is NP-complete even when $p(e) \in \{0,1\}\ \forall e \in \hat{E}$.

Let us consider a decision version of MINLP, where we want to find if such a flow exists or not which satisfies the respective constraints. Let us assume that constraints (14)-(15) always satisfied assuming γ is very large number. Flow constraints (3) in MINLP is equivalent to constraints (16) and (17) in problem ϱ. Also capacity constraint (4) and non-negativity constraint (5) in MINLP are equivalent to constraint (18) in problem ϱ. In problem ϱ, $p(e) \in \{0,1\}\ \forall e \in \hat{E}$ implies whether an edge is selected or not. Thus constraint (19) of problem ϱ implies number of selected edges is bounded by M' which is again equivalent to constraint (6) and (8) of our MINLP when $M' = M$. Hence we can say that problems ϱ can be reduced to the decision version of our MINLP and hence both are equivalent. This implies the developed optimization problem MINLP is NP-hard. □

Though our MINLP will give optimum solution but it will take considerable amount of time to produce solution because of the computational complexity of the problem. However, we can carefully exploit the constraints (4) (capacity constraint) & (14) (link breakage constraint) of MINLP to develop a greedy metric which can be used to produce a heuristic solution to transfer packets from source to destination. In the next section, we will derive the greedy metric from these two constraints and will show its efficient computation.

V. Greedy Metric & its Computation

A. Greedy Metric

Constraint (4) signifies the upper bound on the expected flow in the next time instance $t+1$, which is calculated at the current time instance t. We are considering this constraint because both communicating nodes are moving and thus interference dynamics from neighboring nodes are highly unpredictable which might reduce the data rate substantially. Hence higher capacity at time t may not imply that the capacity requirements will be satisfied at the next time instance $t+1$. Constraint (14) signifies that the probability of a link which was present at time t will also be present in time $t+1$. Lesser the value on the left hand side of constraint (14), higher is the probability of the link to be connected at time $t+1$. By combining these two constraints, for each link (i,j) at time $t+1$, we define the greedy metric termed as connectivity factor (CF) and denoted by Λ_{ij}^{t+1} as follows:

$$\Lambda_{ij}^{t+1} = \frac{E[C_{ij}^{t+1}]}{\left(\mu_{ij}^{t+1} + \sigma_{ij}^{t+1}\sqrt{\frac{1-\alpha_{ij}}{\alpha_{ij}}}\right)} \quad (20)$$

Since this metric is applied to each link (i,j), it is computed at each node locally using the link related information from their respective neighboring nodes. If Λ_{ij}^{t+1} has higher value, then link (i,j) has higher chance of being chosen. Let us denote the numerator and denominator of CF as $\Im_{ij}^{t+1}$ and $\wp_{ij}^{t+1}$ respectively for all link (i,j) at time $t+1$. The numerator $\Im_{ij}^{t+1}$ decides the time it will take to transfer the packets because it represents the expected capacity of the link. Denominator $\wp_{ij}^{t+1}$ signifies link breakage condition. For a link (i,j) to exist at time $t+1$ with probability $1-\alpha_{ij}$, $\wp_{ij}^{t+1} \leq \gamma$ must be satisfied. If $\wp_{ij}^{t+1}$ has lower value, then it implies lower chance of link breakage which in turn means higher reliability of the link's existence for the next time instance $t+1$. All of these values are computed at the current time instance t.

In the analysis of the following subsections, we will show how to compute efficiently the CF metric by incorporating the predicted information for the upcoming time instance $t+1$ on the link (i,j) under consideration. This analysis also shows how CF metric captures the local information when the mobile nodes can have heterogeneous mobility models.

B. Computations of μ_{ij}^{t+1}

To compute the components of CF value for a link, we will need to take into account the SINR values that each node receives from its neighbors because it is an important factor which gives information about the signal strength and link's capacity. Estimating SINR becomes critical due to mobility of nodes which effects the signal strength as well as interference from other nodes in proximity. We can say that two nodes i and j are connected at time t, if $S_{ij}^t \geq S_{th}$, where S_{ij}^t is the received SINR at time t and S_{th} is the threshold on minimum SINR required for a link to be considered as active. Since nodes i and j are moving, by the time they start communicating, the computed S_{ij}^t value may change by next time instance $t+1$. We need to model SINR and related parameters in terms of speed and direction of nodes which is developed in this and subsequent sections. We are giving a general network model which can take care of mobility of nodes irrespective of the mobility model chosen for the nodes.

We define attenuation factor as the inverse of SINR value between two links (i,j): $\Omega_{ij}^t = \frac{1}{S_{ij}^t}$ at time t, which is a random variable capturing the randomness of the signal power and in turn the capacity of link (i,j) at time t, assuming noise is constant for the channel. Now using this notation we can say that if $\Omega_{ij}^t \leq \gamma$, then there is an edge between nodes i and j, where $\gamma = \frac{1}{S_{th}}$. We will compute μ_{ij}^{t+1} of Ω_{ij}^t using the SINR expression at time t as stated below:

$$S_{ij}^t = \frac{Q_{ij}^t}{N_{th} + I_{ij}^t}. \quad (21)$$

Here Q_{ij}^t is the power received by node j from node i, I_{ij}^t is the interference received on link (i,j) at time t and N_{th} is the thermal noise which is assumed to be constant. Both Q_{ij}^t and I_{ij}^t depend on the path loss (PL) model. PL is a function of distance and shadowing factors which both can be random [44]. We use the PL model given by [44], [45] which can be stated as follows:

$$Q_{ij}^t = P_i^t \cdot K \cdot \left(\frac{d_0}{d_{ij}^t}\right)^{\rho} \cdot \psi \tag{22}$$

Here P_i^t is the power transmitted by node i at time t. Here, $K = G_t \cdot G_r \cdot \left(\frac{\lambda}{4\pi d_0}\right)^2$ is a unit-less parameter which depends on the antenna characteristics and the average channel attenuation, G_t and G_r are the antenna gains at the transmitter and receiver respectively which is assumed to be constant for the analysis, $\lambda = c/f$ is the wavelength, c is the speed of light and f is the frequency. K is assumed to be constant in the analysis. Here, d_0 is a reference distance for the antenna far-field, d_{ij}^t is the distance between nodes i and j at time t, ρ is the path loss exponent (PLE) and ψ is the shadowing random variable. Shadowing component is random and can have any well known distribution p_ψ (e.g., log normal distribution [44]). Note that the ρ and ψ can take values depending upon whether line of sight (LOS)or non-line of sight (NLOS) communication takes place. We define Q_{ij}^{t+1} as the power received by node j from node i at the next time instance $t+1$ as follows:

$$Q_{ij}^{t+1} = P_i^t \cdot K \cdot \left(\frac{d_0}{d_{ij}^{t+1}}\right)^{\rho} \cdot \psi \tag{23}$$

Here Q_{ij}^{t+1} is a time dependent random variable which depends upon d_{ij}^{t+1} as well as the shadowing effect.

Now we will look into the denominator part of the SINR expression in equation (21) for the next time instance $t+1$: $N_{th}+I_{ij}^{t+1}$. Here I_{ij}^{t+1} is the interference seen at the link (i,j) at time $t+1$ which can be expressed as:

$$I_{ij}^{t+1} = \sum_{\forall k, k\neq i} Q_{kj}^{t+1}. \tag{24}$$

Now we can express Ω_{ij}^{t+1} as follows:

$$\Omega_{ij}^{t+1} = \frac{1}{S_{ij}^{t+1}} = \frac{N_{th} + \sum_{\forall k,k\neq i} Q_{kj}^{t+1}}{Q_{ij}^{t+1}}. \tag{25}$$

Assuming all devices transmit with the same power (i.e., $P_i^t = P,\ \forall i \in N^t$), the above expression can be simplified using equation (23) as:

$$\Omega_{ij}^{t+1} = \frac{\frac{N_{th}}{P\cdot K\cdot \psi} + \sum_{\forall k,k\neq i}\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}}{\left(\frac{d_0}{d_{ij}^{t+1}}\right)^{\rho}}. \tag{26}$$

Equation (26) is used to find out the expectation and standard deviation of Ω_{ij}^{t+1} for the next time instance $t+1$. This calculation is made at the current time instance t. Note that ψ depends upon material of medium and assumed to be independent of time and in turn independent of d_{ij}^{t+1}. Let us represent the numerator and denominator of equation (26) by Ψ_n and Ψ_d respectively. Ψ_n is made up of two components $\frac{N_{th}}{P*K*\psi}$ and $\sum_{\forall k,k\neq i}\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}$. Let us denote them by Ψ_{n1} and Ψ_{n2} respectively.

Remark 3. *If two random variables X and Y are independent then their measurable functions f(X) and g(Y) are also independent random variables.*

Proposition 2. *Ψ_n and Ψ_d are independent random variables.*

Proof. Ψ_n has two components Ψ_{n1} & Ψ_{n2}. Now both are independent to each other since random variables ψ and d_{kj}^{t+1} are independent. Since Ψ_{n1} is a function of ψ and Ψ_{n2} is a function of d_{kj}^{t+1}, we can say that Ψ_{n1} and Ψ_{n2} are also independent using remark 3.

Note that Ψ_d is the function of random variable d_{ij}^{t+1}. Also, Ψ_n consists of Ψ_{n1} and Ψ_{n2} which are functions of two independent random variables ψ and d_{kj}^{t+1} respectively. Now we can see that Ψ_{n2} is calculated for the values of $\frac{d_0}{d_{kj}^{t+1}}$ for all k where $k\neq i$. So we can say that d_{kj}^{t+1} is independent of d_{ij}^{t+1}. Hence Ψ_n and Ψ_d are independent random variables. □

Therefore using above proposition we can say that the random variables in the equation (26) are independent of each other.

Remark 4. *if X and Y are two independent random variables then $E[X/Y] = E[X]\cdot E[1/Y]$.*

Remark 5. *if $X_1, X_2 \cdots, X_n$ are random variables then $E[X_1+X_2+\cdots+X_n] = E[X_1]+E[X_2]+\cdots+E[X_n]$.*

By taking expectation on both sides of equation (26) and using remarks 4 and 5, we get:

$$\mu_{ij}^{t+1} = \left(\frac{N_{th}}{P\cdot K}\cdot E\left[\frac{1}{\psi}\right] + \sum_{\forall k,k\neq i} E\left[\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}\right]\right) \cdot E\left[\left(\frac{d_{ij}^{t+1}}{d_0}\right)^{\rho}\right]. \tag{27}$$

We will now find out $E\left[\left(\frac{d_{ij}^{t+1}}{d_0}\right)^{\rho}\right]$. Let us say nodes i and j have velocity V_i & V_j with known distributions f_{V_i} in $[V_i^{min}, V_i^{max}]$ and f_{V_j} in $[V_j^{min}, V_j^{max}]$ respectively. Their angle of movements θ_i & θ_j are assumed to have known distributions f_{θ_i} & f_{θ_j} in $[-\pi,\pi]$ respectively. The positions of nodes i and j at current time instance t are (x_i^t, y_i^t) & (x_j^t, y_j^t) respectively. Let us denote their positions at next time instance $t+1$ as (x_i^{t+1}, y_i^{t+1}) and (x_j^{t+1}, y_j^{t+1}) respectively. Then the distance d_{ij}^{t+1} between them at time $t+1$ can be expressed as:

$$d_{ij}^{t+1} = \sqrt{(y_j^{t+1}-y_i^{t+1})^2 + (x_j^{t+1}-x_i^{t+1})^2}$$

where, $x_j^{t+1} = x_j^t + V_j\Delta t\cos\theta_j$, $y_j^{t+1} = y_j^t + V_j\Delta t\sin\theta_j$, $x_i^{t+1} = x_i^t + V_i\Delta t\cos\theta_i$ and $y_i^{t+1} = y_i^t + V_i\Delta t\sin\theta_i$. Here Δt is the small time difference between t and $t+1$. Expression for d_{ij}^{t+1} is dependent on four independent random variables V_i, V_j, θ_i & θ_j. Since we know the probability distribution

functions (pdfs) of all of them, we can find $E\left[\left(\frac{d_{ij}^{t+1}}{d_0}\right)^{\rho}\right]$ as follows:

$$E\left[\left(\frac{d_{ij}^{t+1}}{d_0}\right)^{\rho}\right] = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{V_j^{min}}^{V_j^{max}}\int_{V_i^{min}}^{V_i^{max}} \left(\frac{d_{ij}^{t+1}}{d_0}\right)^{\rho} \cdot f_{V_i} \cdot f_{V_j} \cdot f_{\theta_i} \cdot f_{\theta_j} \; dV_i dV_j d\theta_i d\theta_j. \quad (28)$$

Using the similar approach, $E\left[\left(\frac{d_0}{d_{ij}^{t+1}}\right)^{\rho}\right]$ can also be calculated. Similarly, $E[\frac{1}{\psi}]$ can be calculated as follows:

$$E\left[\frac{1}{\psi}\right] = \int_{-\infty}^{\infty}\left(\frac{1}{\psi}\right)\cdot p_{\psi}. \quad (29)$$

Hence we can calculate the value of μ_{ij}^{t+1} as in equation (27).

C. Computations of σ_{ij}^{t+1}

We will use the similar approach as in previous subsection to calculate σ_{ij}^{t+1} of Ω_{ij}^{t+1} from its variance $var\{\Omega_{ij}^{t+1}\}$ as:

$$\sigma_{ij}^{t+1} = \sqrt{var\{\Omega_{ij}^{t+1}\}}. \quad (30)$$

By taking variance both sides in equation (26), we get:

$$var\{\Omega_{ij}^{t+1}\} = var\left\{\frac{\frac{N_{th}}{P\cdot K\cdot\psi} + \sum_{\forall k,k\neq i}\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}}{\left(\frac{d_0}{d_{ij}^{t+1}}\right)^{\rho}}\right\} \quad (31)$$

Remark 6. *if X and Y are two independent random variables then $var\{X/Y\} = E[X^2]\cdot E[1/Y^2] - E[X]^2\cdot E[1/Y]^2$.*

Remark 7. *if $X_1, X_2\cdots, X_n$ are independent random variables then $var\{X_1+X_2+\cdots+X_n\} = var\{X_1\}+var\{X_2\}+\cdots+var\{X_n\}$.*

As mentioned earlier, Ψ_n and Ψ_d are independent random variables. So using remark 6, equation (31) can be written as:

$$var\left\{\frac{\Psi_n}{\Psi_d}\right\} = \left\{E[\Psi_n^2]\cdot E\left[\frac{1}{\Psi_d^2}\right] - E[\Psi_n]^2\cdot E\left[\frac{1}{\Psi_d}\right]^2\right\}.$$

Since $\Psi_n = \Psi_{n1} + \Psi_{n2}$, using remark 4 and 5, we can write above expression as:

$$var\left\{\frac{\Psi_n}{\Psi_d}\right\} = \left\{\left(E[\Psi_{n1}^2] + E[\Psi_{n2}^2] + 2\cdot E[\Psi_{n1}]\cdot E[\Psi_{n2}]\right)\cdot E\left[\frac{1}{\Psi_d^2}\right] - \left(E[\Psi_{n1}]^2 + E[\Psi_{n2}]^2 + 2\cdot E[\Psi_{n1}]\cdot E[\Psi_{n2}]\right)\cdot E\left[\frac{1}{\Psi_d}\right]^2\right\}. \quad (32)$$

Substituting the values of Ψ_{n1}, Ψ_{n2}, Ψ_d and using remark 5, we can write above expression as:

$$var\left\{\frac{\Psi_n}{\Psi_d}\right\} = \left\{\left(\frac{N_{th}^2}{P^2\cdot K^2}\cdot E\left[\frac{1}{\psi^2}\right] + E\left[\left(\sum_{\forall k,k\neq i}\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}\right)^2\right] + 2\cdot\frac{N_{th}}{P\cdot K}\cdot E\left[\frac{1}{\psi}\right]\cdot\sum_{\forall k,k\neq i}E\left[\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}\right]\right)\cdot E\left[\left(\frac{d_{ij}^{t+1}}{d_0}\right)^{2\rho}\right] - \left(\frac{N_{th}^2}{P^2\cdot K^2}\cdot E[\frac{1}{\psi}]^2 + E\left[\sum_{\forall k,k\neq i}\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}\right]^2 + 2\cdot\frac{N_{th}}{P\cdot K}\cdot E\left[\frac{1}{\psi}\right]\cdot\sum_{\forall k,k\neq i}E\left[\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}\right]\right)\cdot E\left[\left(\frac{d_{ij}^{t+1}}{d_0}\right)^{\rho}\right]^2\right\}. \quad (33)$$

The term $E\left[\left(\sum_{\forall k,k\neq i}\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}\right)^2\right]$ in right hand side of equation (33) is quadratic and hence requires $O(n^2)$ time for computation. All other terms require $O(n)$ computation time. Here n is the number of transmitters interfering with the link (i,j). We assume that simple mathematical computations like addition and multiplication require $O(1)$ time. We will now show that this quadratic term can also be computed in linear time as stated in proposition 3 below.

Remark 8. *If $var(A)$ and $E[A]$ are respectively the variance and expectation of a random variable A, then we can say that $E[A^2]$=$var(A)$+$E[A]^2$.*

Proposition 3. *$var\left\{\frac{\Psi_n}{\Psi_d}\right\}$ in equation* (33) *can be computed in linear time.*

Proof. We can see that the term $E\left[\left(\sum_{\forall k,k\neq i}\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}\right)^2\right]$ is making the computation of equation (33) quadratic by involving quadratic terms. Using remark 8 we can re-write this term as follows:

$$E\left[\left(\sum_{\forall k,k\neq i}\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}\right)^2\right] = var\left\{\sum_{\forall k,k\neq i}\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}\right\} + E\left[\sum_{\forall k,k\neq i}\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}\right]^2. \quad (34)$$

Using remarks 8, 7 and 5 we can re-write above as follows:

$$E\left[\left(\sum_{\forall k,k\neq i}\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}\right)^2\right] = \sum_{\forall k,k\neq i}\left\{E\left[\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{2\rho}\right] - E\left[\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}\right]^2\right\} + \left(\sum_{\forall k,k\neq i}E\left[\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}\right]\right)^2. \quad (35)$$

Now we can see that the right hand side of equation (35) can be solved in $O(n)$ time. This expression can be substituted in equation (33) to make its overall computation linear. □

Finally, σ_{ij}^{t+1} in equation (30) can be computed in linear time using equation (33).

D. Computations of $E[C_{ij}^{t+1}]$

The capacity of link (i,j) at the next time instance $t+1$ can be expressed using Shannon's theorem as:

$$C_{ij}^{t+1} = B\cdot\log_2(1+S_{ij}^{t+1}) \quad (36)$$

$$= B\cdot\log_2\left(\frac{N_{th}+I_{ij}^{t+1}+Q_{ij}^{t+1}}{N_{th}+I_{ij}^{t+1}}\right). \quad (37)$$

Here, B is the bandwidth of the signal, S_{ij}^{t+1} is the SINR on link (i,j) at time $t+1$. Taking expectation on both sides yield:

$$E[C_{ij}^{t+1}] = B \cdot E\left[\log_2\left(\frac{N_{th} + I_{ij}^{t+1} + Q_{ij}^{t+1}}{N_{th} + I_{ij}^{t+1}}\right)\right]. \quad (38)$$

Substituting equations (23) & (24) in equation (38), we get:

$$E[C_{ij}^{t+1}] = B \cdot E\left[\log_2\left(\frac{N_{th}}{P \cdot K \cdot \psi} + \sum_{\forall k}\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}\right)\right] - B \cdot E\left[\log_2\left(\frac{N_{th}}{P \cdot K \cdot \psi} + \sum_{\forall k, k\neq i}\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}\right)\right]. \quad (39)$$

Let us denote $X = \frac{N_{th}}{P\cdot K\cdot\psi} + \sum_{\forall k}\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}$ and $Y = \frac{N_{th}}{P\cdot K\cdot\psi} + \sum_{\forall k, k\neq i}\left(\frac{d_0}{d_{kj}^{t+1}}\right)^{\rho}$. Assume that there are n transmitters which are interfering to link (i,j). We can see that computing $E[\log_2(X)]$ requires computation of n different d_{kj}^{t+1} components corresponding to each of n interfering transmitters. Since computing each of d_{kj}^{t+1} requires evaluating a 4 dimensional integration, thus computation of $E[\log_2(X)]$ requires computing a $4n$ dimensional integration. Using the following remark, we can reduce the computation of $E[\log_2(X)]$ to n number of 4 dimensional integrations, which is computationally better than computing a $4n$ dimensional integration.

Remark 9. *If $Z \geq 1$ is a random variable with expectation $E[Z]$ and variance $var(Z)$ respectively, then $\ln(Z)$ around $E[Z]$ can be approximated using Taylor series as:*

$$\ln(Z) = \ln(E[Z]) + \frac{Z - E[Z]}{E[Z]} - \frac{(Z - E[Z])^2}{2 \cdot E[Z]^2} + \cdots$$

Neglecting higher order terms and taking expectation on both side in above yields,

$$E[\ln(Z)] \approx \ln(E[Z]) - \frac{var(Z)}{2 \cdot E[Z]^2}. \quad (40)$$

Now using remark (9), we can re-write $E[\log_2(X)]$ in terms of $\ln(E[X])$, $var(X)$ and $E[X]$, all of which can be computed in linear time. Similar argument holds for $E[\log_2(Y)]$ as well. Hence computation of $E[C_{ij}^{t+1}]$ takes linear time. Thus μ_{ij}^{t+1}, σ_{ij}^{t+1} and $E[C_{ij}^{t+1}]$ all can be computed in linear time.

Till now we have discussed how to compute the CF metric in a generic sense. For specific cases like mmWave communication, the G_t and G_r can be computed by incorporating the additional gain of the directional antenna arrays, as stated in the following remark. Accordingly the CF metric for such cases can be computed by using these values.

Remark 10. *Short range D2D communication involving mmWave channel usually employs directional antenna arrays with beam-forming capabilities which increases transmitter and receiver antenna gains (G_t & G_r) to compensate for high propagation loss [46]. For a uniform planar square antenna array composed of L_p elements, antenna gains can be written as [46]:*

$$G_x = \begin{cases} G_{ml} & \text{if } \theta \leq \phi/2 \\ G_{sl}, & \text{otherwise} \end{cases} \quad (41)$$

where $x = \{t, r\}$ is subscript for transmitter & receiver, $G_{ml} = L_p$ is the main-lobe gain, $G_{sl} = 1/\sin^2(3\pi/2\sqrt{L_p})$ is the side-lobe gain, $\phi = \sqrt{3}/\sqrt{L_p}$ is the beam-width, and $\theta \in [-\pi, \pi]$ is the angle off the boresight direction. Mobility of devices would make them go out of each other's range, hence at every such instance the alignment of transmitter-receiver pair needs to be done. However, this alignment overhead is in order of hundreds of micro seconds even for extremely narrow beams of width=1° as mentioned and validated in [47]. Since the data transmission time duration is in order of seconds, we can consider the alignment overhead to be negligible. So we are assuming that the transmitter-receiver pairs are perfectly aligned to obtain the maximum power gain as also done in [46]. In the simulation section, we have shown the effect of directionality on the end to end delay for different network loads in Figure 7.

VI. Relay Selection Strategy

In this section, we will first construct the perceived network graph based on Λ_{ij}^{t+1} values. Then we will discuss the associated relay node selection strategy to find the source-destination path for both network assisted as well as distributed cases. We will describe the relay selection for all four types of D2D communication as shown in Figure 1.

A. Perceived Graph

We will now calculate Λ_{ij}^{t+1} values in algorithm 1 using the values of μ_{ij}^{t+1}, σ_{ij}^{t+1} and $E[C_{ij}^{t+1}]$ computed in previous section. This algorithm uses constraint (13) to compute Λ_{ij}^{t+1} as shown in line 5. Note that the reliability with certain guarantee (α_{ij}) is ensured by the condition $\wp_{ij} \leq \gamma$ in this line. Each node runs this algorithm locally and shares Λ_{ij}^{t+1} values with the BS using channel state information (CSI) at current time instance t. These information are used at the BS to construct a *perceived graph* $G_{per}^{t+1}(N^{t+1}, \hat{E}^{t+1})$ as shown in algorithm 2. In G_{per}^{t+1}, $N^{t+1} = N^t$ and $\hat{E}^{t+1} \subseteq E^t$ such that $\Lambda_{ij}^{t+1} \geq 0$ & $(i,j) \in E^t$. Note that edges are not symmetric due to mobility and interference among nodes. It is also important to note that the perceived graph leverages the property of Λ_{ij}^{t+1} which keeps the delay and packet loss information for the upcoming time instance in a per hop basis. So G_{per}^{t+1} contains the information for the next time instance $t+1$ which is calculated at the current time instance t.

The perceived graph (G_{per}^{t+1}) can be used for both *network assisted* and *distributed* scenarios. In the network assisted (DR-OC & DC-OC) scenario, when the BS is acting as the controller for forming the links among the nodes, this graph can be used by BS to predict with some success guarantee (depending upon α_{ij}) to establish a source-destination path. Since Λ_{ij}^{t+1} values are calculated locally for each link (i,j) on a per hop basis, we can also use this metric in the distributed case in which nodes are moving in an adhoc manner without any BS assistance. Nodes in this setting will share the information among themselves as done in *route finding* phase of classical AODV protocol. In this scenario we assume that there are multiple route reply packets and the source node will wait for maximum of ϵ_t time after getting the first route

Algorithm 1: `CF_Computation` Algorithm

input : i, γ, α_{ij}, μ_{ij}^{t+1}, σ_{ij}^{t+1}, $E[C_{ij}^{t+1}]$

n_i; *//adjacent node set for current node i*
$\Lambda_{ij}^{t+1} \leftarrow (-1)$;
for $\forall j \in n_i$ **do**
  $\wp_{ij}^{t+1} := \left(\mu_{ij}^{t+1} + \sigma_{ij}^{t+1}\sqrt{\frac{1-\alpha_{ij}}{\alpha_{ij}}}\right)$;
  if $\wp_{ij}^{t+1} <= \gamma$ **then**
    $\Lambda_{ij}^{t+1} = \frac{E[C_{ij}^{t+1}]}{\wp_{ij}^{t+1}}$;
  end
end
return Λ_{ij}^{t+1};

Algorithm 2: Construct_Perceived_Graph Algorithm

input : $G^t(N^t, E^t)$, Λ_{ij}^{t+1}

$\hat{E}^{t+1} \leftarrow 0$;
for $\forall i \in \{1, 2, \cdots, N^t\}$ **do**
  for $\forall j \in \{1, 2, \cdots, N^t\}$ **do**
    if $(E^t[i][j] = 1)$ & $(\Lambda_{ij}^{t+1} \geq 0)$ **then**
      $\hat{E}^{t+1}[i][j] = 1$;
    end
  end
end
return $G_{per}^{t+1}(N^{t+1}, \hat{E}^{t+1})$;

reply packet to make decisions for the best path among all the available paths. This will essentially simulate a perceived graph which will give approximate solution restricted by ϵ_t time. This is elaborated in subsection VI-B2.

B. Route Selection Algorithm

1) Operator Controlled (`DR-OC & DC-OC`): In this mode, `BS` has complete topology information of the network and is responsible for finding a source-destination path. In `DC-OC`, data can be transferred directly to the destination when it is in vicinity of the source node. Otherwise, devices act as relays (`DR-OC`) and we need to choose the best relaying nodes along the path such that they minimizes overall packet loss and average delay.

If we choose a path with least average delay, it may not be the most reliable path, causing a lot of re-transmissions of data packets due to packet loss. If we choose a path with the highest reliability but with little consideration on the expected capacity, the selected path may cause higher delay. Hence we need to choose a path which considers both reliability and expected capacity which has precisely been considered by our `CF` metric (Λ_{ij}^{t+1}). We can see that choosing the *widest path* in the perceived graph takes care of both reliability and delay. The widest path P_W^ζ for a source-destination pair $\zeta \in \Gamma$ is the path having maximum *width*, where width of a path P^ζ is defined as $\min_{\forall(i,j)\in P^\zeta}\{\Lambda_{ij}^{t+1}\}$. The widest path P_W^ζ of the perceived graph $G_{per}^{t+1}(N^{t+1}, \hat{E}^{t+1})$ can be found using a variation of Dijkstra's shortest path algorithm which runs in $O(|\hat{E}^{t+1}| \cdot \log(|N^{t+1}|))$ time. We term the procedure to compute widest path as `WP()`. Procedure `WP()` returns P_W^ζ and the bottleneck capacity $min\_cap\_P_W^\zeta = \min_{\forall(i,j)\in P_W^\zeta}\{C_{ij}^t\}$.

So the overall relay selection algorithm can be described as follows. Suppose a sending node s_ζ wants to send b_ζ packets to r_ζ. If they are in vicinity of each other, then `DC-OC` is activated and the `BS` instructs s_ζ to directly transfer packets to r_ζ in the current time instance t. If the capacity of the link is not sufficient, then the communication takes place in the next time instance and so on until b_ζ packets are transferred. If s_ζ and r_ζ are not in vicinity of each other, then `BS` assists for `DR-OC` type of communication where intermediate nodes act as relays for source-destination path. `BS` constructs the perceived graph at every time instance t using algorithm 2 and uses `WP()` routine to find the corresponding widest path. Nodes in the chosen path are directed by `BS` to transfer data packets bounded by $min\_cap\_P_W^\zeta$ using `send_pkts_OC()` function. Acknowledgments of successful packet delivery are handled by this function. One or more links in this path may break while the packets are in transit leading to packet loss. Also, not all of the packets may be transferred to the destination due to capacity limitations of links in the path at the current time instance t. Hence an intermediate node i may posses the packets of the source node s_ζ after the time instance t ends. This procedure is shown in algorithm 3 which runs at `BS` for every time instance t until the destination receives all the packets.

Algorithm 3: Network assisted connectivity factor (`NCF`) Algorithm

input : $G^t(N^t, E^t)$, $\Gamma < s_\zeta, r_\zeta, b_\zeta >$

$G_{per}^{t+1}(N^{t+1}, \hat{E}^{t+1})$=$Construct\_Perceived\_Graph()$;
for $i = \{1, 2, \cdots, N^t\}$ **do**
  for $\forall \zeta \in \Gamma$ **do**
    if $(i \leftarrow\ sending\_node)$ **then**
      $(P_W^\zeta, min\_cap\_P_W^\zeta) \leftarrow WP(G_{per}^{t+1}, i, r_\zeta)$;
      $send\_pkts\_OC(P_W^\zeta, min\_cap\_P_W^\zeta)$;
      if $Packets\{r_\zeta\} == b_\zeta$ **then**
        remove ζ from Γ;
      end
    end
  end
end

2) Device Controlled (`DR-DC & DC-DC`): In this mode, nodes themselves control the link establishment as they are not assisted by `BS`. Since they don't have the knowledge of complete network topology, so finding widest path in distributed fashion is a challenge. We will leverage the path finding techniques in adhoc networks known as `AODV` [34] for this purpose. In `AODV` there is a route finding phase in which source node floods a route request (`RREQ`) packet in the network. When `RREQ` reaches destination, a route reply (`RREP`) packet is sent back to source node from destination forming a source-destination path. We are considering a vari-

ant of AODV described in [21] which assumes that multiple RREP packets arrive at the source in the increasing order of the delay. In our route selection scheme we are storing the width of a path in the corresponding RREP packet instead of the delay. The source node after receiving the first RREP packet, holds the transmission of data for maximum of ϵ_t time. This is to ensure that the source node gets the RREP packet for a path $P_W^{\zeta}(\epsilon_t)$ which has maximum width among all the RREPs arrived at the source node during ϵ_t time and it is selected for the communication. This essentially constructs a perceived graph in a distributed fashion. Note that $P_W^{\zeta}(\epsilon_t)$ might not be the best path globally which was selected in the operator controlled mode because here the path chosen will have the width restricted by ϵ_t time. The function call path_DC() is used to find $P_W^{\zeta}(\epsilon_t)$ which essentially calls the route finding phase routines RREQ and RREP. Λ_{ij}^{t+1} is input to this function which is calculated as described in algorithm 1.

So the overall relay selection algorithm can be described as follows. Suppose node s_ζ wants to send b_ζ packets to r_ζ. The data transmission can take place directly if they are in vicinity of each other (DC-DC). If s_ζ and r_ζ are not in vicinity of each other (DR-DC), node s_ζ will check if it has the corresponding $P_W^{\zeta}(\epsilon_t)$ path using path_exists_DC() function, otherwise it finds this path using path_DC() function. Packets are sent via $P_W^{\zeta}(\epsilon_t)$ in a distributed fashion using send_pkts_DC() function call which also handles the acknowledgments of the successful packet delivery. Here also packet loss may occur due to one or more links failure or the capacity limitations of the links in $P_W^{\zeta}(\epsilon_t)$. Hence an intermediate relay node i may posses some packets of node s_ζ which are left unsent to node r_ζ at the end of time t. These packets will be transferred to the destination at the next time instance $t+1$ and so on until destination receives all the packets. This procedure is shown in algorithm 4 which runs at every time instance t at every node which has data to be sent.

Algorithm 4: Device controlled distributed connectivity factor (DC-DCF) Algorithm

input : i, ϵ_t, Λ_{ij}^{t+1}, $\Gamma < s_\zeta, r_\zeta, b_\zeta >$

1 **if** $(i \leftarrow$ *sending_node*$)$ **then**
2 **for** $\forall \zeta \in \Gamma$ **do**
3 **if** $(i \neq r_\zeta)$ **then**
4 **if** $(path\_exists\_DC(i, r_\zeta))$ **then**
5 $send\_pkts\_DC(P_W^{\zeta}(\epsilon_t))$;
6 **end**
7 **else**
8 $P_W^{\zeta}(\epsilon_t) = path\_DC(i, r_\zeta, \Lambda_{ij}^{t+1}, \epsilon_t)$;
9 *goto Line* 4;
10 **end**
11 **end**
12 **end**
13 **end**

VII. Experiment and Results

A. Simulation Parameters

Simulation parameters are mostly taken from [48], [49]. Through simulations we are trying to analyze the strength of the proposed greedy metric and the associated algorithms. We are initially distributing 30 nodes uniformly in a 100×100 m^2 area. Nodes are moving independently following random walk mobility model with a speed uniformly distributed in $[0, V_{max}]$ m/s, where $V_{max} \in \{5, 8, 10, 12\}$ and their angle of motion is uniformly distributed in $[-\pi, \pi]$. Note that we have used random walk mobility model for computing μ_{ij}^{t+1}, σ_{ij}^{t+1} and $E[C_{ij}^{t+1}]$ in our simulation. However, similar approach can be extended for other mobility models as long as the pdfs of the corresponding mobility parameters are known. We are assuming that at any instance of time the number of node are always 30 in the simulation region and all the nodes are capable of participating in D2D communication. We have considered a single source-destination pair having a load of $L \in \{50, 100, 500, 1000\}$ packets, where a packet is of size 100 bytes. We have considered $\Delta t = \{0.5, 0.8, 1.0, 1.2\}$ s, admissible error rate $\alpha_{ij} = \{0.02, 0.10\}$ and $\epsilon_t = \{10, 30, 70, 100\}$ ms in our simulation. We have considered path loss model as given in [48] for non-line of sight (NLOS) scenario of an urban model for 61 GHz millimeter wave. The corresponding parameters are: path loss exponent ρ is 4.49, signal bandwidth is 20 MHz and standard deviation of log-normal shadowing component (σ_ψ) in dB is 4.0 dB. The maximum transmission power of a D2D user device is 24 dBm, the thermal noise density is -174 dBm/Hz [49] and hence $N_{th} = 2 \cdot 10^{-10.4}$ mW. We have used Shannon's capacity for each link (i, j) at time t to compute C_{ij}^t.

We have compared our algorithms NCF & DC-DCF with reliability based (RB-NA) & delay based (AODV) algorithms. The RB-NA approach is a network assisted algorithm where the path with the highest reliability is chosen. The reliability of a path P is defined as $\min_{(i,j)\in P} \frac{1}{\wp_{ij}^{t+1}}$ where $\wp_{ij}^{t+1}$ is the denominator of CF metric. The AODV in our experiment is based on the least delay path as discussed in [21].

B. Experimental Results & Analysis

We have run the experiment using the simulation environment mentioned in the previous subsection and taken the average of the results of around 30000 runs. We have written our own $C++$ custom code and run them on a GNU 4.8 compiler on Intel core $i7$ machine.

We have analyzed the effect of different parameters like Δt, V_{max}, L, ϵ_t and α_{ij} on packet loss and average end to end delay per packet. Figures 2, 3 and 4 depict results of varying Δt, V_{max} and L on the packet loss due to mobility keeping all other corresponding parameters fixed as shown in respective figures. Note that packet re-transmission is caused due to packet loss. Figure 5 depicts the result of varying L on the average end to end (E2E) delay per packet, keeping other parameters fixed as shown in this figure. Average E2E delay is defined as the average transmission time for a packet to reach the destination, ignoring the propagation and processing

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2893995, IEEE Transactions on Vehicular Technology

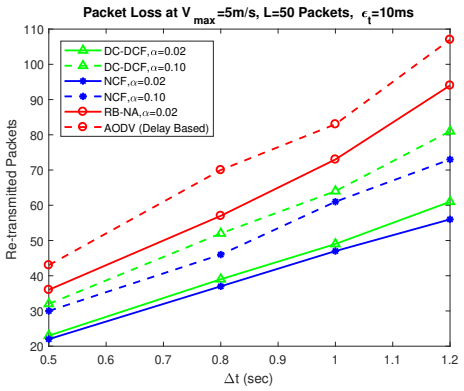

Fig. 2. Depiction of the effects of Δt on packet loss.

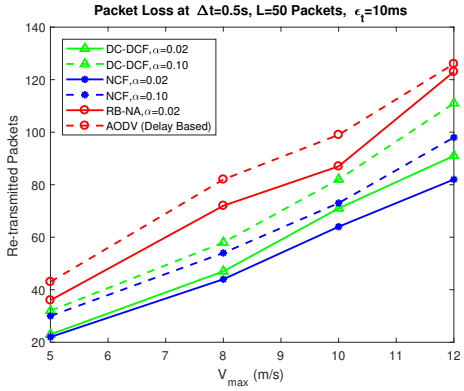

Fig. 3. Depiction of the effects of variation in maximum speed on packet loss.

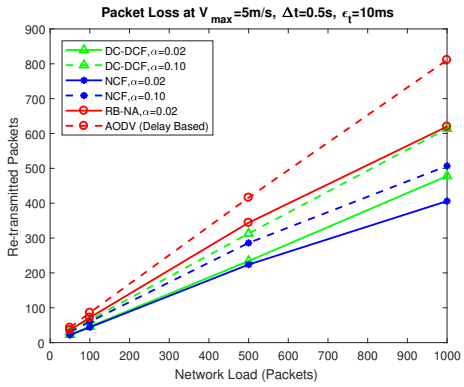

Fig. 4. Depiction of the effects of network load on packet loss.

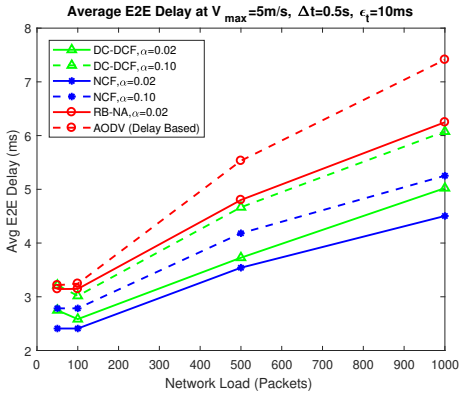

Fig. 5. Depiction of the effects of network load on average end to end delay per packet.

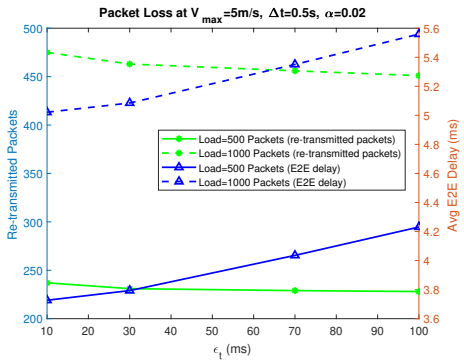

Fig. 6. Depiction of the effects of variation of ϵ_t on packet loss and average end to end delay per packet in DC−DCF for $\alpha = 0.02$.

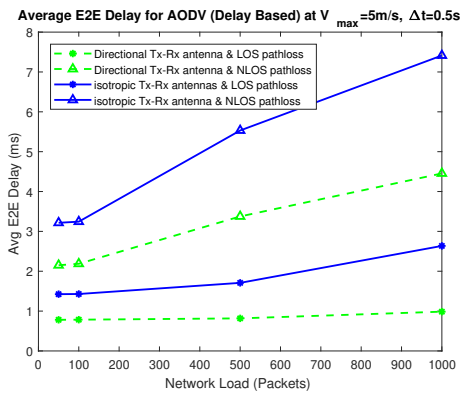

Fig. 7. Effects of isotropic & directional antennas for 61 GHz mmWave for both LOS & NLOS scenarios.

delays. Since we have considered only one source destination pair in our simulation study for simplicity, we have neglected the queuing delay.

The following observations can be made from Figures 2-4: 1) *effects of varying* α: NCF and DC−DCF with $\alpha = 0.02$ have lesser packet loss than NCF and DC−DCF with $\alpha = 0.10$ respectively. This is because the admissible error rate α plays an important role in determining the link reliability and which in turn affects the packet loss. Higher values of α implies lesser reliability on link existence for the next time instance as compared to lower value of α. 2) *Effects of assistance from the BS*: NCF with $\alpha = 0.02$ and $\alpha = 0.10$ have lesser packet loss than DC−DCF with $\alpha = 0.02$ and $\alpha = 0.10$ respectively. This is due to the fact that NCF is network assisted approach and BS have complete network topology information. Hence BS can select the best suitable source-destination path and thus gets a better performance as compared to DC−DCF which is a distributed approach, which lacks the complete network topology information. 3) Both NCF and DC−DCF outperform RB−NA and AODV due to the fact that later two algorithms respectively consider reliability and capacity factors individually, whereas both these factors are taken care by our algorithms.

It can be observed in Figure 5 that similar behavior as in case of packet loss happens in terms of average E2E delay for all the three points mentioned in previous paragraph. The reason is that additional time is required to compensate for the re-transmitted packets.

Additionally, we can observe the behavior specific to Figure 2: packet loss is increasing as Δt increases. This is due to the fact that the distance $(V\Delta t)$ between the nodes increases with Δt and hence the chance of packet loss also increases.

Also, Figure 3 depicts that increasing V_{max} causes more packet loss because the distance $V\Delta t$ between nodes increases with increasing speed. Higher speed causes the node to quickly go out of the range of the nearby nodes which are transmitting packets, hence causing higher packet loss.

Figure 4 depicts that the packet loss is increased as the number of packets transferred is increased. Higher the number of packets, more chances are to split the packets in multiple next hops, causing a higher chance of overall packet loss due to mobility of nodes.

It can be also observed in Figure 5 that the delay increases with the network load because the packet loss increases along with network load and thus the total time to compensate for the re-transmission of packets causes rise in average E2E delay.

Figure 6 shows the effect of increasing the wait time ϵ_t for DC−DCF on packet loss as well as average E2E delay for $L = \{500, 1000\}$ packets keeping other parameters fixed as shown. We can see the trade-off between delay and re-transmission of packets, as ϵ_t increases, packet re-transmission

is reduced but the average E2E delay increases and vice versa. This is because as ϵ_t increases, the devices get higher chance to select a better path which reduces average packet loss but it incorporates more delay because of higher ϵ_t.

From the results shown in Figures 2-6, we can see $\Delta t * V$ is an important factor in which both Δt and V have effects on packet loss which helps in choosing the best next hop node. High $\Delta t * V$ value causes higher packet loss. Higher speed V of nodes may end up in breakage of links causing more packet loss. If Δt is high, it causes more packets to be sent which in turn, may cause more packet loss. We need to look at an appropriate $\Delta t * V$ value such that the capacity of link is satisfied along with speed to ensure the packets loss is affordable. We cannot control the speed of nodes but we can surely control Δt value. In other words, we can determine Δt value at which our algorithms should be executed to ensure the packet loss is kept below a threshold. Also, in case of DC−DCF we have to choose a value of ϵ_t which have moderate packet loss probability as well as appropriate minimum delay as there is a trade-off between them.

Figure 7 shows the effect of using isotropic and directional antennas under both LOS and NLOS settings. For directional antennas, we have considered number of antenna elements $L_p = 4$ [46]. Thus we get main-lobe gain $G_{ml} = 4$, side-lobe gain $G_{sl} = 2$ and beam-width $\phi = 0.866$ using equation (41). Interference is caused due to the side-lobe power from the transmitting devices over the link under consideration. Using these values, we have computed the SINR. For NLOS, PL parameters are same as in previous results (ρ & σ_ψ are 4.49 & 4.0 dB respectively) and for LOS, ρ & σ_ψ are 1.88 & 1.2 dB respectively [48]. Figure 7 shows the end to end delay versus network load for both isotropic & directional antennas under both NLOS & LOS settings. It can be observe that the performance of directional antennas in LOS setting is overall best as expected. Similarly the performance of LOS setting is better than the NLOS setting.

VIII. Conclusion

We have developed an operator controlled SIP model to minimize the average network delay while keeping the packet loss under control by capturing the effect of mobility of nodes. After converting the SIP to its equivalent MINLP and proving its NP-hardness, we derived a distributed greedy metric CF by exploiting constraints of MINLP. CF computed at current time instance t captures the per-hop delay and reliability by incorporating channel uncertainty by modeling dynamic and unpredictable SINR for next time instance $t + 1$ using the nodes' mobility. We have shown the efficient computations of CF which is applicable for any mobility models as long as the pdfs of the corresponding mobility parameters are known. Using the locally computed CF values we constructed a operator controlled perceived graph at current time t which holds information for the upcoming time instance $t + 1$. This graph is used to find the appropriate source-destination path by using a widest path algorithm based on CF values. Similarly for device controlled case, this graph is constructed by waiting for ϵ_t time at the source node after it receives first RREP packet during route finding phase. Simulations have been performed to show that the proposed network assisted and distributed relay selection algorithms outperform the reliability and delay based approaches. Depending on the average velocity of nodes, we can tune the parameters Δt and ϵ_t such that both packet loss and delay are satisfied according to the need of an application.

References

[1] D. Singh and S. C. Ghosh, "A distributed algorithm for D2D communication in 5G using stochastic model," in *16th IEEE International Symposium on Network Computing and Applications*, pp. 459–466, 2017.

[2] M. Iwamura, "NGMN view on 5G architecture," in *81st IEEE Vehicular Technology Conference*, pp. 1–5, May 2015.

[3] M. N. Tehrani, M. Uysal, and H. Yanikomeroglu, "Device-to-device communication in 5G cellular networks: challenges, solutions, and future directions," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 86–92, 2014.

[4] C. I, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, "Toward green and soft: a 5G perspective," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 66–73, 2014.

[5] D. Feng, L. Lu, Y. Yuan-Wu, G. Y. Li, S. Li, and G. Feng, "Device-to-device communications in cellular networks," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 49–55, 2014.

[6] G. Fodor, S. Parkvall, S. Sorrentino, P. Wallentin, Q. Lu, and N. Brahmi, "Device-to-device communications for national security and public safety," *IEEE Access*, vol. 2, pp. 1510–1520, 2014.

[7] P. K. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, "Design considerations for a 5G network architecture," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 65–75, 2014.

[8] W. Lee, J. Kim, and S. W. Choi, "New D2D peer discovery scheme based on spatial correlation of wireless channel," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 10120–10125, Dec 2016.

[9] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Communications Magazine*, vol. 52, pp. 82–89, February 2014.

[10] F. Jameel, Z. Hamid, F. Jabeen, S. Zeadally, and M. A. Javed, "A survey of device-to-device communications: Research issues and challenges," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 2133–2168, 2018.

[11] G. Yu, L. Xu, D. Feng, R. Yin, G. Y. Li, and Y. Jiang, "Joint mode selection and resource allocation for device-to-device communications," *IEEE Transactions on Communications*, vol. 62, pp. 3814–3824, Nov 2014.

[12] M. Belleschi, G. Fodor, and A. Abrardo, "Performance analysis of a distributed resource allocation scheme for D2D communications," in *IEEE GLOBECOM Workshops*, pp. 358–362, Dec 2011.

[13] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, "Stochastic geometry study on device-to-device communication as a disaster relief solution," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 3005–3017, May 2016.

[14] S. Fowler, Y. Li, A. Pollastro, and S. Napoli, "Simple network design and power allocation for 5G device-to-device communication," in *19th IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks*, pp. 203–207, Dec 2014.

[15] R. Ma, Y. Chang, H. Chen, and C. Chiu, "On relay selection schemes for relay-assisted D2D communications in LTE-A systems," *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 8303–8314, Sept 2017.

[16] S. Dang, G. Chen, and J. P. Coon, "Outage performance analysis of full-duplex relay-assisted device-to-device systems in uplink cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 4506–4510, May 2017.

[17] A. Orsino, D. Moltchanov, M. Gapeyenko, A. Samuylov, S. Andreev, L. Militano, G. Araniti, and Y. Koucheryavy, "Direct connection on the move: Characterization of user mobility in cellular-assisted D2D systems," *IEEE Vehicular Technology Magazine*, vol. 11, pp. 38–48, Sept 2016.

[18] A. Orsino, A. Samuylov, D. Moltchanov, S. Andreev, L. Militano, G. Araniti, and Y. Koucheryavy, "Time-dependent energy and resource management in mobility-aware D2D-empowered 5G systems," *IEEE Wireless Communications*, vol. 24, pp. 14–22, Aug 2017.

[19] R. Wang, J. Zhang, S. H. Song, and K. B. Letaief, "Mobility-aware caching in D2D networks," *IEEE Transactions on Wireless Communications*, vol. 16, pp. 5001–5015, Aug 2017.
[20] Y. Han, H. Wu, Z. Yang, and D. Li, "A new data transmission strategy in mobile D2D networks-deterministic, greedy, or planned opportunistic routing?," *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 594–609, Jan 2017.
[21] M. Naseem and C. Kumar, "Congestion-aware fibonacci sequence based multipath load balancing routing protocol for MANETs," *Wireless Personal Communications*, vol. 84, pp. 2955–2974, Oct 2015.
[22] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Communications Surveys Tutorials*, vol. 16, no. 4, pp. 1801–1819, 2014.
[23] A. Omri and M. O. Hasna, "A distance-based mode selection scheme for D2D-enabled networks with mobility," *IEEE Transactions on Wireless Communications*, vol. 17, pp. 4326–4340, July 2018.
[24] Y. Yang, Y. Zhang, L. Dai, J. Li, S. Mumtaz, and J. Rodriguez, "Transmission capacity analysis of relay-assisted device-to-device overlay/underlay communication," *IEEE Transactions on Industrial Informatics*, vol. 13, pp. 380–389, Feb 2017.
[25] S. Wu, R. Atat, N. Mastronarde, and L. Liu, "Improving the coverage and spectral efficiency of millimeter-wave cellular networks using device-to-device relays," *IEEE Transactions on Communications*, vol. 66, pp. 2251–2265, May 2018.
[26] L. Blackmore, M. Ono, and B. C. Williams, "Chance-constrained optimal path planning with obstacles," *IEEE Transactions on Robotics*, vol. 27, pp. 1080–1094, Dec 2011.
[27] A. A. Gaivoronski, "Stochastic optimization problems in telecommunications," *Applications of stochastic programming*, vol. 5, pp. 669–704, 2005.
[28] R. Verdone and A. Zanella, "On the effect of user mobility in mobile radio systems with distributed DCA," *IEEE Transactions on Vehicular Technology*, vol. 56, pp. 874–887, March 2007.
[29] J. Harri, F. Filali, and C. Bonnet, "Mobility models for vehicular ad hoc networks: a survey and taxonomy," *IEEE Communications Surveys Tutorials*, vol. 11, no. 4, pp. 19–41, 2009.
[30] D. Wu, L. Zhou, Y. Cai, R. Q. Hu, and Y. Qian, "The role of mobility for D2D communications in LTE-advanced networks: energy vs. bandwidth efficiency," *IEEE Wireless Communications*, vol. 21, pp. 66–71, April 2014.
[31] Ó. Helgason, S. T. Kouyoumdjieva, and G. Karlsson, "Opportunistic communication and human mobility," *IEEE Transactions on Mobile Computing*, vol. 13, pp. 1597–1610, July 2014.
[32] O. N. C. Yilmaz, Z. Li, K. Valkealahti, M. A. Uusitalo, M. Moisio, P. Lundn, and C. Wijting, "Smart mobility management for D2D communications in 5G networks," in *IEEE Wireless Communications and Networking Conference Workshops*, pp. 219–223, April 2014.
[33] K. M. Sim and W. H. Sun, "Ant colony optimization for routing and load-balancing: survey and new directions," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 33, pp. 560–572, Sept 2003.
[34] C. E. Perkins and E. M. Royer, "Ad-hoc on-demand distance vector routing," in *Second IEEE Workshop on Mobile Computing Systems and Applications*, pp. 90–100, Feb 1999.
[35] N. Chakchouk, "A survey on opportunistic routing in wireless communication networks," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2214–2241, 2015.
[36] R. Tang, J. Zhao, H. Qu, and Z. Zhang, "User-centric joint admission control and resource allocation for 5G D2D extreme mobile broadband: A sequential convex programming approach," *IEEE Communications Letters*, vol. 21, pp. 1641–1644, July 2017.
[37] J. G. Andrews, S. Singh, Q. Ye, X. Lin, and H. S. Dhillon, "An overview of load balancing in hetnets: old myths and open problems," *IEEE Wireless Communications*, vol. 21, pp. 18–25, April 2014.
[38] A. Orsino, A. Ometov, G. Fodor, D. Moltchanov, L. Militano, S. Andreev, O. N. C. Yilmaz, T. Tirronen, J. Torsner, G. Araniti, A. Iera, M. Dohler, and Y. Koucheryavy, "Effects of heterogeneous mobility on D2D- and drone-assisted mission-critical MTC in 5G," *IEEE Communications Magazine*, vol. 55, pp. 79–87, February 2017.
[39] J. Ye, X. Ge, G. Mao, and Y. Zhong, "5G ultradense networks with nonuniform distributed users," *IEEE Transactions on Vehicular Technology*, vol. 67, pp. 2660–2670, March 2018.
[40] I. K. Son and S. Mao, "An reformulation-linearization technique-based approach to joint topology design and load balancing in FSO networks," in *Proceedings of the Global Communications Conference*, pp. 1–6, 2010.
[41] B. B. Nandi, A. Banerjee, S. C. Ghosh, and N. Banerjee, "Stochastic VM multiplexing for datacenter consolidation," in *Ninth IEEE International Conference on Services Computing*, pp. 114–121, June 2012.
[42] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*. Boston: McGraw Hill, fourth ed., 2002.
[43] M. R. Garey and D. S. Johnson, *Computers and Intractability; A Guide to the Theory of NP-Completeness*. New York, NY, USA: W. H. Freeman & Co., 1990.
[44] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
[45] S. Sun, T. S. Rappaport, M. Shafi, P. Tang, J. Zhang, and P. J. Smith, "Propagation models and performance evaluation for 5G millimeter-wave bands," *IEEE Transactions on Vehicular Technology*, vol. 67, pp. 8422–8439, Sept 2018.
[46] N. Deng and M. Haenggi, "A fine-grained analysis of millimeter-wave device-to-device networks," *IEEE Transactions on Communications*, vol. 65, pp. 4940–4954, Nov 2017.
[47] R. Congiu, H. Shokri-Ghadikolaei, C. Fischione, and F. Santucci, "On the relay-fallback tradeoff in millimeter wave wireless system," in *IEEE Conference on Computer Communications Workshops*, pp. 622–627, April 2016.
[48] A. Al-Hourani, S. Chandrasekharan, and S. Kandeepan, "Path loss study for millimeter wave device-to-device communications in urban environment," in *IEEE International Conference on Communications Workshops*, pp. 102–107, 2014.
[49] J. Lianghai, A. Klein, N. Kuruvatti, and H. D. Schotten, "System capacity optimization algorithm for d2d underlay operation," in *IEEE International Conference on Communications Workshops*, pp. 85–90, June 2014.

Durgesh Singh received the B.E. degree in Computer Science and Engineering from Chandigarh College of Engineering & Technology, Chandigarh, India, in 2009, the M.Tech. in Computer Science degree from the Indian Statistical Institute, Kolkata, India, in 2015. He is currently working towards the Ph.D. degree in Computer Science from the Indian Statistical Institute. His research interests include D2D communication, relay selection, wireless networks and mobile computing.

Sasthi C. Ghosh received the B.Sc. (Hons.) degree in Mathematics from the University of Calcutta, Kolkata, India, in 1996, the M.C.A. degree from the Jadavpur University, Kolkata, India, in 1999, and the Ph.D. degree in Computer Science from the Indian Statistical Institute, Kolkata, India, in 2004. He is presently an Associate Professor with the Advanced Computing and Microelectronics Unit of the Indian Statistical Institute. His current research interests include mobile computing and wireless networks, 5G device to device communications, handoff management in heterogeneous networks and wireless local area networks.