




Analysis of Opportunistic Relaying and Load Balancing Gains Through V2V Clustering

Saadallah Kassir , Gustavo de Veciana , Nannan Wang , Xi Wang, *Member, IEEE*, and Paparao Palacharla

Abstract—Recent advances in digitization, artificial intelligence, and wireless networking are revolutionizing the traditional transportation networks and the automotive industry, leading for instance to the widespread adoption of ride-sharing services and the emergence of self-driving vehicles. As a result, a considerable rise in the volume of infotainment data consumed by the vehicles' passengers is expected. To support such traffic loads, we explore leveraging vehicles as part of the cellular infrastructure. This paper investigates and quantifies the performance gains that using Vehicle-to-Vehicle (V2V) communication to form relay clusters would enable, including (1) improving opportunistic channel access to the cellular infrastructure by relaying the cluster's traffic through the vehicle seeing the best channel quality; and (2) balancing traffic loads across cells through cluster multihoming. We introduce a stochastic geometric model allowing us to provide a deeper understanding of the possible gains associated with cluster-based opportunistic relaying and its sensitivity to the system parameters, e.g., base station density, vehicle density on the roads, etc. We develop a natural network utility maximization problem to serve as a baseline to evaluate the performance of a simpler distributed cluster management algorithm which we show to be near-optimal. Overall the results suggest that 10 – 20× gains in the vehicles' shared rate can be achieved along with significant fairness improvements, while other mobile devices also benefit from the proposed network architecture.

Index Terms—Load balancing, opportunism, vehicular network, vehicle-to-vehicle communication.

I. INTRODUCTION

TRANSFORMATIVE shifts are underway in the automotive industry that are expected to have a substantial impact on future wireless networks, including: (1) the emergence of the *Transportation-as-a-Service* business model, and increased use of ride sharing fleets; (2) the development and expected adoption of driverless car technologies. These two major disruptions in transportation paradigms are expected to result in considerable shifts in the wireless traffic seen in cellular networks. Indeed, they induce the emergence of a new class of users, namely

commuters who are no longer required to actively drive and are free to work/play while on the road. These passengers are expected to consume massive amounts of *infotainment* data that include accessing online services such as video streaming, video conferencing, VoIP, and online gaming, all of which are associated with large data-rates and high throughput connectivity requirements. It is therefore critical to rethink current wireless networks operations, and more particularly the way vehicles can access and make the most of the available wireless resources.

Among the proposed solutions to support the connectivity requirements associated with this class of highly mobile users, leveraging Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) links (commonly called V2X) possibly operating in millimeter-wave (mmWave) frequency bands, and deploying Road Side Units (RSUs) close to the vehicles are among the most promising ones, see, e.g., [1]. On the one hand, the “ubiquitous” availability of V2X connectivity offers the prospect of enabling new vehicle-based services, e.g., data relaying and caching, that could also reduce the traffic loads on traditional cellular networks. On the other hand, RSU deployment comes at substantial costs for network operators, and solutions leveraging the existing cellular infrastructure may therefore be preferable.

In this paper, we embrace these changes by focusing on leveraging vehicle clustering using the V2X technology to provide improved cellular connectivity for infotainment content delivery to vehicle passengers (as opposed to delay-sensitive safety data). Vehicle clusters have indeed been shown to be particularly efficient at sharing information across vehicular networks, see, e.g., [2]–[5]. The central challenge of this work is to develop an understanding of the performance and tradeoffs of vehicular-based wireless architectures, when taking into account the roles of the vehicle clusters on the roads and the cellular network geometry. We consider a setting wherein clusters of well connected vehicles share possibly multihomed connectivity to the cellular infrastructure, i.e., one or more Base Stations (BSs) can transmit data to a cluster of vehicles which can in turn relay data to the appropriate vehicle. This leads to two types of benefits which we discuss next.

Opportunistic throughput gains: The first benefit stems from the significant throughput gains achievable through opportunistic relaying to vehicle clusters. For example, as shown in Fig. 1, rather than sending directly to a vehicle v_4 at the cell edge, a BS b_1 can send data to a nearby vehicle v_1 and the cluster can then use high capacity V2V connectivity to relay data to v_4 . Given the order of magnitude differences in the peak capacity of nearby users compared to edge users in a typical cell, as long

Manuscript received 28 April 2022; revised 20 May 2022; accepted 23 May 2022. Date of publication 26 May 2022; date of current version 19 September 2022. This work was supported by the affiliates of the 6G@UT center within the Wireless Networking and Communications Group at The University of Texas at Austin and the NSF under Grant ECC-1809327. The review of this article was coordinated by Prof. Pascal Lorenz. (Corresponding author: Saadallah Kassir.)

Saadallah Kassir and Gustavo de Veciana are with the The University of Texas at Austin, Electrical and Computer Engineering Department, Austin, TX 78712 USA (e-mail: skassir@utexas.edu; deveciana@utexas.edu).

Nannan Wang, Xi Wang, and Paparao Palacharla are with Fujitsu Network Communications, Richardson, TX 75082 USA (e-mail: nannan.wang@fujitsu.com; xi.wang@fujitsu.com; paparao.palacharla@fujitsu.com).

Digital Object Identifier 10.1109/TVT.2022.3178129

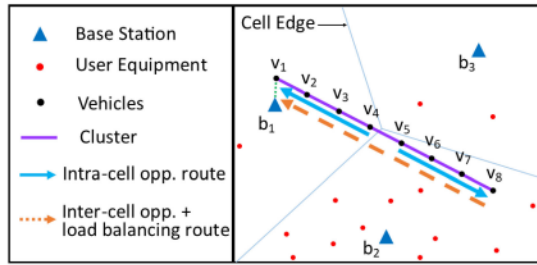


Fig. 1. Example of an eight vehicle cluster traversing two cells shared by other User Equipments.

as V2V capacity is plentiful the potential of such cluster-based cooperative relaying is extremely high. When v_4 and v_1 lie in the same cell we refer to this as *intra-cell opportunism*, and if b_1 uses relay v_1 to forward data to a vehicle in another cell (say v_5) we call this *inter-cell opportunism*. This approach might be particularly relevant in mmWave based infrastructures, whose short range and susceptibility to obstructions make efficient deployment challenging. By leveraging cluster-based relaying one can exploit spatial diversity to find Line of Sight (LoS) channels to BSs, providing improved coverage, throughput and reliability, see [1].

Load balancing gains: The second benefit comes from enabling load balancing across neighboring cells. For example, as shown in Fig. 1, the traffic destined to a cluster of vehicles spanning multiple BS cells, e.g., b_1 and b_2 , can be delivered through either one or both of the BSs, depending on their current loads. For instance, b_1 and b_2 currently have 5 and 15 users/vehicles in their respective cells but by using inter-cell cooperative relaying the loads could be shifted so that they each serve 9 and 11 respectively. Such an approach can help reduce the variability across cell loads, hence diminishing the spatial and temporal variability in users' perceived shared rate. This is especially important in the context of 5G where small-cells cover limited regions and thus might see higher relative load variability. Moreover, although the link between the cluster and the lightly-loaded BS b_1 may be weaker than the link to b_2 , the former will be able to allocate more wireless resources to the cluster, benefiting not only the vehicles, but also other devices associated to b_2 as fewer users are contending for channel access.

Paper's contributions and organization: The main objective of this paper lies in modeling and analyzing the potential benefits of vehicle cluster-based opportunistic relaying in terms of shared rate gains and improved fairness. More specifically, this paper makes five major contributions.

First, we present a model to study the performance of cellular networks leveraging V2V-clustering. The model captures the essential features and tradeoffs associated with this technique, while leveraging tools and results already established in the field of stochastic geometry.

Second, we study analytically the sources of intra-cell opportunistic gains, providing additional insight on the benefit of V2V-clustering, and providing tools to assess the performance gains.

Third, we formulate a network-level (centralized) fairness oriented resource allocation and load balancing optimization problem, allowing us to capture the full gains associated with intra-cell and inter-cell opportunism, as well as load balancing through BS multihoming.

Fourth, we propose a cluster-level (distributed) and computationally efficient load balancing algorithm that greedily and locally re-associates vehicles to BSs. We assess its performance by comparing it to the network-level fairness optimal algorithm, and policy which only leverages intra-cell opportunism. We then argue that our cluster-level resource management algorithm is suitable for real-time dynamic allocation compared to a centralized network-level solution, and may be preferable despite its sub-optimality. Finally, we discuss technical challenges associated with V2V cluster-based relaying, such as incentive mechanisms, the impact on packet delays, as well as real-time cluster management challenges.

This paper extends prior work [6], which characterized the gains associated with V2V-clustering under stronger network geometry assumptions and a simpler wireless link model. The improved model in this paper allows us to provide a deeper understanding of the character and extent of the performance gains via a complementary mathematical analysis of the origins of opportunistic gains, and to give further insight on optimal network operations in more realistic network environments.

The paper is organized as follows. In Section II we present an overview of related work. In Section III we propose our stochastic geometric network model and we present in Section IV the associated analysis geared at understanding the roots of intra-cell opportunistic gains. In Section V, we introduce a centralized network-level and a distributed cluster-level resource allocation and user association algorithms leveraging intra-cell opportunism, inter-cell opportunism, and load balancing. Section VI presents simulation results for a variety of scenarios suggesting $10 \times -20 \times$ shared rate gains along with significant improvements in shared rate fairness. In Section VII, we present a critical outlook on V2V-clustering by highlighting major technical challenges associated with the proposed network architecture. Finally Section VIII concludes the paper.

II. RELATED WORK

Extensive research efforts have recently been devoted to investigating the benefits of opportunistic relaying in cellular networks and cell association load-balancing. We present next an overview of relevant related work in both directions.

Many researchers have explored gains that can be achieved through opportunistic relaying in cellular networks [1], [7]–[14], commonly exhibiting gains in throughput, rate fairness, and/or outage probability. For instance, [10] proposes a promising software framework that leverages opportunism to improve by $2 \times$ the total throughput delivered by a WiFi-based WLAN network. Our work shows that much greater gains can possibly be achieved in large-scale cellular networks by leveraging V2V relaying to serve vehicle-bound users. Studies such as [11] analyze the opportunistic gain in the context of VANETs, and also show that opportunism improves the downlink throughput. The

focus is, however, on comparing the performance of different routing strategies, and proposing efficient relaying protocols rather than analyzing the gains associated with opportunism and load balancing. Although [11] studies RSU-based networks, it provides some valuable insight regarding exploiting opportunism, that can be applicable in cellular-network settings. The authors in [14] show that up to $5.7\times$ data rate gains can be achieved by the cell-edge users and $4.1\times$ gains for the median users via Device-to-Device (D2D) opportunistic relaying. In our work, we show that additional gains can be expected by considering the geometry of vehicle clusters on the roads in cellular networks as well as the role of load-balancing. In [1], the authors consider a simple wireless model (binary connectivity) and network model (infinite straight road with equispaced RSUs) to show evidence that V2V-clustering reduces the vehicles' rate variability. In this work, we show that the reduced variability (hence improved fairness) in the per-user shared rate is in reality coupled with shared rate gains when a more realistic system model consistent with 3GPP standards is adopted.

Another line of work explored the benefits of load balancing in wireless networks. Some proposed solutions include channel borrowing [15], cell breathing [16], [17], BS association biasing [18], [19], centralized dynamic inter-cell and intra-cell handovers [20], distributed user association policies under heterogeneous traffic [21], and combinations thereof. Our work proposes a novel load-balancing solution leveraging V2V connectivity among vehicles driving on a road network served by the cellular infrastructure.

Other works have also investigated the benefits of D2D-based load balancing. In [22], spectrum savings enabled by D2D-based load balancing across cells were investigated. In our work, we characterize instead the relaying benefits in terms of user shared-rate and fairness gains. In [23], the authors introduce an optimization framework to find the optimal load-balancing and routing strategies in D2D-relay-based networks, considering a sum-rate maximization objective, but do not study the potential fairness gains associated with D2D-relaying. In addition, in contrast to both of these works, we leverage tools from the stochastic geometry literature [24] to understand the role that the V2V-cluster relay geometry plays on their load-balancing ability, and hence, on the large-scale cellular network performance.

Other works have established the critical role that load balancing plays in improving mean user rate or improving a notion of fairness, see, e.g., [25]. This work has been extended to vehicular network settings where V2I and V2V links are used to offload traffic from one cell to another [26]–[28]. While these works exhibit the benefits of load balancing, they focus on defining routing strategies, rather than evaluating the resource allocation and the potential per-user rate gains that a load balancing scheme might generate.

The traditional approach to balance mobile users' loads across cells is via the formulation of an optimization problem, see e.g., [22], [23], [29] which in turn suggests appropriate scheduling algorithms, e.g., [23], [30], [31]. Other researchers propose learning-based solutions to determine effective association policies, see e.g., [32], but perhaps lack the development

of underlying insights useful towards the design of vehicular network-based relaying strategies.

Finally, some papers focused on studying multihomed load balancing schemes. For instance, [33] presents algorithms and experimental results showing how load balancing can improve the performance of multihop multihomed VANETs, but no network modeling and analysis was performed, and the study was mainly focused on uplink access, while we focus on multihomed downlink connectivity in this work.

This work is, to the best of our knowledge, the first one evaluating jointly opportunistic and load balancing gains by leveraging V2V cluster-based relaying to enhance the cellular infrastructure.

III. SYSTEM MODEL

In this section, we propose a system model enabling us to study gains associated with opportunism and load balancing in cellular networks enhanced by V2V cluster relaying.

A. Network Model

We consider a network where BSs are randomly placed on the plane according to a homogeneous Poisson Point Process (PPP) Φ_{BS} with intensity λ_{BS} , see e.g., [34]. Another independent homogeneous PPP Φ_M of intensity λ_M models the locations of the mobile User Equipment (UE). In this network, the road infrastructure is modeled as an arbitrary stationary line process Φ_R of line intensity λ_R meters of road per m^2 , and independent of Φ_{BS} . Conditioned on a realization ϕ_R of this road infrastructure, vehicles with a fixed Line-of-Sight V2V-communication range of d_R meters are dropped on the roads and form vehicle clusters.

Definition 1 (Vehicle Cluster): Given an arbitrary configuration of vehicles on a road, a vehicle cluster is a sequence of vehicles on the same road such that any two consecutive vehicles are within communication range d_R of each other.

It follows from this definition that a vehicle can only belong to a single cluster. Vehicles (hence clusters) are modeled as randomly distributed on the roads according to the following bursty-traffic model:

- Vehicle clusters consist of an independent random number of vehicles where the typical cluster size Z follows an arbitrary (but known) discrete finite-mean distribution.
- Vehicles are equispaced¹ within a cluster, and the inter-vehicular distance is fixed to be $d_V \leq d_R$ meters, consistent with Definition 1, i.e., clusters can be seen as chains of successive vehicles within communication range of each other. Thus, the random size Z of a typical cluster induces a random length L in meters from the first to the last vehicle, such that $L = (Z - 1)d_V$ meters. We use the convention that a cluster with one vehicle has length 0, a cluster with two vehicles has length d_V , etc.

¹While the equispaced model for vehicle placement within a cluster is idealized, our extensive experiments showed that for a given cluster length distribution the vehicle configuration within the cluster has an almost negligible impact on the network performance. Hence, we shall adopt the equispaced model to keep the subsequent analysis tractable.

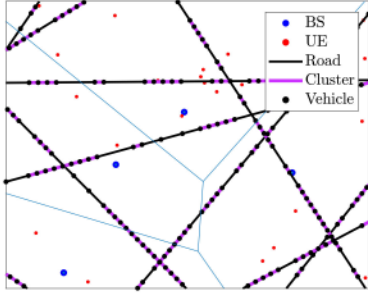


Fig. 2. Illustration of a random network realization, i.e., realizations of ϕ_{BS} and ϕ_M modeling the BS and UE locations, along with the induced $\mathcal{T}(\phi_{BS})$, and an arbitrary road infrastructure supporting randomly placed vehicle clusters.

- Clusters are dropped on roads in ϕ_R , such that the distance between two consecutive clusters on the same road is random. Specifically, the typical inter-cluster distance T between the last vehicle of a typical cluster and the first vehicle of the preceding one follows an arbitrary (known) distribution satisfying $T > d_R$ meters almost surely, consistent with Definition 1. This captures the requirement that two vehicles in different clusters cannot be within communication range of each other, otherwise they would be part of the same cluster.

For any given d_V and known $\mathbb{E}[Z] < \infty$ and $\mathbb{E}[T]$, one can characterize the vehicle density λ_V on a road in vehicles/meter as follows:

$$\lambda_V = \frac{\mathbb{E}[Z]}{(\mathbb{E}[Z] - 1) \cdot d_V + \mathbb{E}[T]} = \frac{\mathbb{E}[Z]}{\mathbb{E}[L] + \mathbb{E}[T]} \quad (1)$$

The above clustered vehicle model induces a spatial process Φ_V denoting the locations of the vehicles in the network, which is not a PPP. Each vehicle is assumed to also correspond to an active user, i.e., with full-buffer traffic. In the sequel we let $\phi_{BS} = \{b_i : i \in \mathbb{N}\}$ denote a realization of the PPP Φ_{BS} and refer to BSs directly through and their locations, e.g., b_i . This convention is adopted for all point processes. As shown in Fig. 2, the BSs in ϕ_{BS} induce a Voronoi tessellation $\mathcal{T}(\phi_{BS}) = \{\mathcal{T}^b(\phi_{BS}) \mid b \in \phi_{BS}\}$, where each BS b has an associated cell: $\mathcal{T}^b(\phi_{BS}) = \{x \in \mathbb{R}^2 \mid \|x - b\|_2 \leq \|x - b'\|_2, \forall b' \in \phi_{BS}\}$. Based on this tessellation of BSs' cells we define the following additional notation. The set of vehicles ϕ_V is partitioned such that $\phi_{V,c}$ denotes vehicles belonging to cluster c while ϕ_V^b denotes the set of vehicles in BS b 's cell. Similarly, the set of mobile UEs ϕ_M in the network is partitioned such that ϕ_M^b denotes the set of UEs in BS b cell. Finally, we let ϕ_C^b denote the set of clusters that include *at least one* vehicle in b 's cell, while $\phi_{BS,c}$ denotes the set of BSs containing at least one of cluster c 's vehicles in its cell.

Simulated Network Model: While our analytical results will be based on the above general network model, we shall adopt a more specific model for our simulation results, by relating the parameters d_V , λ_V , the distributions of the random variables Z and T , and the road process Φ_R in a consistent manner. Note that while Φ_R can be selected independently, the other parameters need to be carefully jointly chosen so as to satisfy all the model constraints, i.e., (1), and the communication range constraints.

To that end, we shall generate the road infrastructure as a Poisson Line Process (PLP) Φ_R of line intensity λ_R meters of road per m^2 , see [35], [36]. As for the cluster-generation parameters, we shall fix λ_V , and select the three other ones accordingly as follows:

- We model $Z \sim \text{Geometric}(e^{-\lambda_V d_R})$, which corresponds to the cluster size distribution as if the vehicles were distributed as a PPP on the roads and grouped if they were within communication range d_R [1].
- We let $d_V = \lambda_V^{-1} \left(1 - \frac{\lambda_V d_R e^{-\lambda_V d_R}}{1 - e^{-\lambda_V d_R}}\right)$ corresponding to the *mean* inter-vehicular distance within a cluster if the vehicles were distributed as a PPP on the roads, i.e., for $X \sim \text{Exp}(\lambda_V)$, $d_V = \mathbb{E}[X \mid X \leq d_R]$, see, e.g., [1].
- We model $T \sim \text{Exp}(\mu)$ such that $T > d_R$ almost surely, consistent with the corresponding distribution if the vehicles were distributed as a PPP on the roads, and corroborated by empirical observations in bursty-traffic settings [37]. The parameter μ is selected to be consistent with (1), i.e., $\mu = \lambda_V$.

B. Link Capacity Model

In our analysis, we will consider downlink transmissions. For the *traditional cellular* network, i.e., without V2V cluster relaying, we model the downlink capacity from BS $b \in \phi_{BS}$ to user $u \in \phi_V^b \cup \phi_M^b$ (i.e., to vehicle a mobile UE) for a given network realization as depending on the Signal-to-Interference-and-Noise-Ratio (SINR) given by:

$$\text{SINR}_u^b = \frac{p_{BS} \cdot H_u \cdot P_u^b}{I_u^b + \sigma^2}, \quad (2)$$

where p_{BS} is the BS transmission power, H_u models the independent Rayleigh fast-fading gain such that $H_u^b \sim \text{Exp}(1)$, P_u^b is a random variable modelling the path-gain of the link between b and u , I_u^b is the interference power seen by user u associated to BS b , σ^2 models the total thermal noise power over the allocated bandwidth. Letting d_u^b denote the distance between user u and BS b , we model LoS blocking for all the wireless links through the dual-slope distance-dependent binary random variable P_u^b purposed in the 3GPP standard [38]:

$$P_u^b = \begin{cases} k_{\text{LoS}} \cdot (d_u^b)^{-\alpha_{\text{LoS}}} & \text{w.p. } p_{\text{LoS}}(d_u^b), \\ \min \left[\begin{array}{l} k_{\text{LoS}} \cdot (d_u^b)^{-\alpha_{\text{LoS}}} \\ k_{\text{NLoS}} \cdot (d_u^b)^{-\alpha_{\text{NLoS}}} \end{array} \right] & \text{w.p. } 1 - p_{\text{LoS}}(d_u^b), \end{cases} \quad (3)$$

where k_{LoS} and k_{NLoS} capture the signal attenuation at a reference distance of 1 m for LoS and NLoS links respectively, while α_{LoS} and α_{NLoS} represent the respective path-loss exponents such that $\alpha_{\text{LoS}} \leq \alpha_{\text{NLoS}}$. Moreover, $p_{\text{LoS}}(\cdot)$ is a non-increasing non-negative function of the distance d_u^b , satisfying $p_{\text{LoS}}(0) \leq 1$. Finally I_u^b is the interference power seen by user u , originating from all the BSs except b , i.e.,:

$$I_u^b = \sum_{b' \in \phi_{BS} \setminus \{b\}} p_{BS} \cdot H_u^{b'} \cdot P_u^{b'}. \quad (4)$$

We assume for simplicity that a user u associated with BS b always observes interfering signals from other BSs through NLoS links, i.e., for all b' in $\phi_{BS} \setminus \{b\}$, we have $p_{\text{LoS}}(d_u^{b'}) = 0$.

Finally, the average transmission rate r_u^b from BS b to user u for a link of bandwidth w for a given network realization is modeled for simplicity by the Shannon ergodic rate:

$$r_u^b = w \cdot \mathbb{E}_{\{H_u^b\}_{b \in \phi_{BS}}} \left[\log_2 \left(1 + \frac{\text{SINR}_u^b}{\Gamma} \right) \right], \quad (5)$$

where Γ models the gap between the actual transmission rate and the Shannon capacity, modeling the joint effect of quantized modulation schemes, finite-length codes, channel estimation error due to vehicle mobility, etc., see [39], and the expectation is taken over all the fading terms. Note that we do not average the transmission rate over large-scale SINR variations such as blocking and link distance/vehicle mobility as we leverage opportunism with respect to these fluctuations.

In the *cluster-based opportunistic relaying* scenario, UEs see the same capacity as in the traditional cellular network setting. By contrast, the average downlink transmission rate from BS $b \in \phi_{BS}$ to any vehicle belonging to cluster $c \in \phi_C^b$ is modeled by

$$r_c^{b,*} = \max_{v \in \phi_{v,c} \cap \phi_v^b} r_v^b, \quad (6)$$

i.e., the transmission rate from b to a (relay) vehicle in cluster c and in BS b 's cell. In the sequel, we shall refer to this relay vehicle as the *cluster-head*, and a cluster may have multiple cluster-heads if it is multihomed. Note that we make cluster-head decisions based on the user average rates r_v^b as we envision the cluster-head selection decisions/handoffs to realistically occur on slower time-scales (on the order of hundreds of milliseconds to seconds) than the short channel coherence time associated with fading experienced by vehicles that may be moving at high velocity (on the order of milliseconds). We shall further make the following assumption which is in line with a setting where vehicles use high capacity V2V line of sight links, e.g., mmWave, to connect to the vehicles directly ahead and/or behind them in the same cluster.

Assumption 1: We assume intra-cluster V2V links have sufficiently high capacity (e.g., mmWave bands) so as to ensure they are not the bottleneck in relaying traffic to vehicles within clusters, and do not interfere with infrastructure transmissions (e.g., sub-6 GHz bands).

We envision the allocation of wireless resources for the V2V links to be in line with the 5G NR V2X Sidelink resource allocation schemes described in the 3GPP Release 16, i.e., either coordinated by the relevant BSs (mode 1) or uncoordinated, where vehicles access wireless resources from a resource pool preallocated by the BS (mode 2), see [40].

We summarize the notation introduced in this section and the rest of the paper in Table I.

IV. INTRA-CELL OPPORTUNISM PERFORMANCE ANALYSIS

As introduced in Section I, we are ultimately interested in analyzing both opportunism and load balancing gains. While the opportunism gain analysis can be tractable, the study of load-balancing gains is more complex as it is intrinsically related to the resource allocation policies used in the network. In this

TABLE I
TABLE OF NOTATION

Parameter	Definition	Units
Φ_{BS}	Base Station deployment point process	-
ϕ_{BS}	Set of Base Stations (realization of Φ_{BS})	-
λ_{BS}	Base Station density	BSs/km ²
Φ_M	Mobile UE placement point process	-
ϕ_M	Set of mobile UEs (realization of Φ_M)	-
λ_M	Mobile UE density	UEs/km ²
Φ_R	Road placement line process	-
λ_R	Road density	road km/km ²
Φ_V	Vehicle placement point process	-
ϕ_V	Set of vehicles (realization of Φ_V)	-
λ_V	Mean vehicle density on roads	vehicles/km
ϕ_C	Set of vehicle clusters	-
\mathcal{T}	Voronoi tessellation induced by Φ_{BS}	-
d_V	Inter-vehicle distance	m
d_R	V2V-communication range	m
Z	Typical cluster size	vehicles
L	Typical cluster length	m
T	Typical inter-cluster distance	m
p_{BS}	BS transmission power	dBm
H_u	Rayleigh fast-fading gain	dB
w	Channel bandwidth	MHz
k_{LOS}	LoS signal attenuation constant	dB
k_{NLOS}	NLoS signal attenuation constant	dB
α_{LOS}	LoS path-loss exponent	-
α_{NLOS}	NLoS path-loss exponent	-
σ^2	Thermal noise power	dBm
Γ	Shannon capacity gap	dB
D_u^b	Distance from BS b to user u	m
$D_c^{b,*}$	Effective distance from BS b to user c	m
p_{LOS}	LoS probability	-
d_{corr}	LoS correlation distance	m
P_u^b	Path-gain from BS b to user u	dB
I_u^b	Interference power seen by user u	dBm
SINR_u^b	Link SINR from BS b and user u	dB
r_u^b	Transmission rate from BS b to user u	bits/s
$r_c^{b,*}$	Transmission rate from BS b to cluster c	bits/s
s_u	Shared rate to user u	bits/s
$s_{u,*}^{intra}$	Shared rate to user u (intra-cell opp.)	bits/s
$s_{u,*}^{inter}$	Shared rate to user u (inter-cell opp.)	bits/s
π^b	Resource allocation vector of BS b	-

section, we focus on understanding the cause and effect of intra-cell opportunism solely, in a scenario where no load balancing is performed. In subsequent sections we will get a full picture of the intra-cell opportunism, inter-cell opportunism and load balancing gains. Recall that intra-cell opportunism considers data relaying only among vehicles that are in the same cluster and in the same BS cell. In other words, clusters are assumed to be artificially broken at the cell boundaries, creating logically independent sub-clusters. While this mechanism clearly reduces the benefits of V2V cooperation, a formal analysis of this setting allows for a better understanding of the origin of the gains associated with V2V cluster relaying.

The gains associated with intra-cell opportunism can be summarized in terms of two phenomena: 1) clustering allows vehicles to route traffic through the closest vehicle in their cluster to the BS, leveraging a higher SINR link; and 2), routing traffic through vehicles closer to the BS improves robustness to blocking, by increasing the probability of benefiting from a LoS wireless link. Although these two effects are closely related, we shall study them separately.

TABLE II
THEOREM 1 INTERMEDIARY VARIABLES; FOR $x \leq D \in \mathbb{R}$, $\theta \in [0, \frac{\pi}{2}]$

$\theta_0(d, x) = \sin^{-1}(x/d)$
$l_0(d, x, \theta) = d \cos(\theta) - \sqrt{x^2 - (d \sin(\theta))^2}$
$d_0(d, x, \theta) = d_V \cdot \lceil \frac{l_0(d, x, \theta)}{d_V} \rceil$
$r_0(d, x, \theta) = \sqrt{d^2 + d_0(d, x, \theta)^2} - 2d \cdot d_0(d, x, \theta) \cos(\theta)$
$a_0(d, x, \theta) = \pi r_0^2 - \left[r_0^2 \cos^{-1} \left(\frac{d_0^2 + r_0^2 - d^2}{2d_0 r_0} \right) + d^2 \cos^{-1} \left(\frac{d_0^2 + d^2 - r_0^2}{2d_0 d} \right) - \frac{\sqrt{(r_0 + d - d_0)(d_0 + r_0 - d)(d_0 + d - r_0)(d_0 + r_0 + d)}}{2} \right]$
$e_0(d, x, \theta) = e^{-\lambda_{BS} a_0(d, x, \theta)} \cdot \mathbb{1} \left\{ \left\lfloor \frac{l_0(d, x, \theta)}{d_V} \right\rfloor \neq \left\lfloor \frac{l_0(d, x, \theta) + 2\sqrt{x^2 - (d \sin(\theta))^2}}{d_V} \right\rfloor \right\}$

A. Clustering Reduces the Effective Distance

The most obvious benefit of intra-cell opportunism is the flexibility to route traffic through the vehicle that sees the best BS link in the cluster. A well known result in stochastic geometry characterizes the random distance between a typical user in a cellular network and its closest BS as following a Rayleigh distribution [24], under the assumption that the BSs are deployed according to a PPP. This result still applies to our framework despite the vehicles not following a PPP, as the roads and cluster generation processes are all independent of Φ_{BS} . Hence, we let the random variable D denote the distance between a typical vehicle in the network and its closest BS in the *traditional network* scenario, such that:

$$f_D(d) = 2\pi\lambda_{BS}de^{-\lambda_{BS}\pi d^2}, \forall d \in \mathbb{R}_+. \quad (7)$$

To quantify analytically the potential gains associated with intra-cell opportunism, we derive the distribution of the *effective distance* D^* between the typical vehicle and its closest BS in the *intra-cell cooperative network* scenario, i.e., the distance between the BS $b \in \phi_{BS}$ and the cluster-head associated with a typical vehicle's cluster $c \in \phi_C$. We have:

$$D^* = \max_{v' \in \phi_{v,c} \cap \phi_{b'}^c} D_{v'}^b. \quad (8)$$

We note that while (8) resembles (6), the closest vehicle to the BS in the cluster may not be the one relaying the data, as it may not provide the best link in the case where it is not in LoS link to the BS. Nevertheless, studying D^* allows us to quantify the potential of effective distance reduction gains through V2V relaying. Using the network geometry, one can develop an expression for the conditional c.d.f. of D^* given $D = d$.

Theorem 1 (Effective Distance Conditional c.d.f.): Given the distance $D = d$ between a typical vehicle and its closest BS, and the distribution for the typical cluster size Z , the c.d.f. of the distance between the closest vehicle in the typical vehicle's cluster to the serving BS is given by:

$$\mathbb{P}(D^* \leq x | D = d) = \begin{cases} 1, & \text{for } 0 \leq d \leq x, \\ \frac{2}{\pi \mathbb{E}[Z]} \int_0^{\theta_0(d, x)} e_0(d, x, \theta) \cdot \sum_{i=1+\lceil \frac{l_0(d, x, \theta)}{d_V} \rceil}^{\infty} \mathbb{P}(Z \geq i) d\theta, & \text{for } 0 \leq x \leq d, \end{cases} \quad (9)$$

and the associated variables and functions are given in Table II.

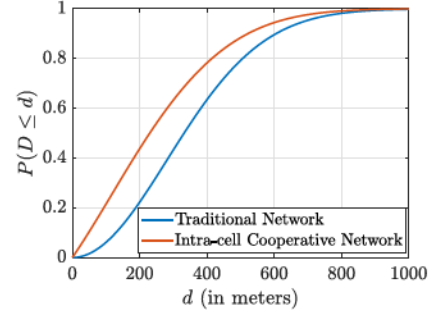


Fig. 3. Comparison of the c.d.f.'s of the effective distance between the typical vehicle and its attached BS, under the traditional and cooperative network scenarios, for $\lambda_{BS} = 2$ BSs/km², $\lambda_V = 30$ vehicles/km and $d_R = 100$ m.

A formal proof of this theorem is provided in Appendix A. One can then directly derive from Theorem 1 and (7) the (unconditional) distribution of D^* . Fig. 3 exhibits the reduction in the effective distance that the typical vehicle experiences in a traditional and cooperative relaying network by comparing the c.d.f.'s of D and D^* . As seen on the figure, a typical vehicle is expected to benefit from considerable gains associated with reduced effective distance to the tagged BS. For instance, 43% of the vehicles will be effectively within 200 m from their attached BS thanks to intra-cell cooperation, while only 23% would be within this range in a traditional network. Besides these considerable direct gains, this effective distance reduction between a typical vehicle and its associated BS induces another benefit that we study next.

B. Clustering Improves Robustness to Blocking

Another benefit of V2V cluster relaying is the flexibility to circumvent large blocking objects, such as buildings, that may interrupt LoS links and attenuate the received SINR. Indeed, vehicle clustering not only provides diversity through additional candidate links that have the potential to have a LoS to the BS, but closer vehicles are clearly more likely to benefit from such LoS links (see (3)). More precisely, if the probability that a typical vehicle a distance $D = d$ from its BS sees a LoS link is $p_{LoS}(d)$ in a traditional network setting, we denote the equivalent metric in the cooperative relaying setting by $p_{LoS}^*(d)$. In this formulation, a typical vehicle a distance $D = d$ from its BS is in a sub-cluster of size \tilde{Z} (containing vehicles in the same cluster and same cell as the typical vehicle) whose vehicles are at distances $\mathbf{D} = (D_1, D_2, \dots, D_{\tilde{Z}})$ from the BS, where there is an i such that $D_i = D$ almost surely. Now $p_{LoS}^*(d)$ has the following

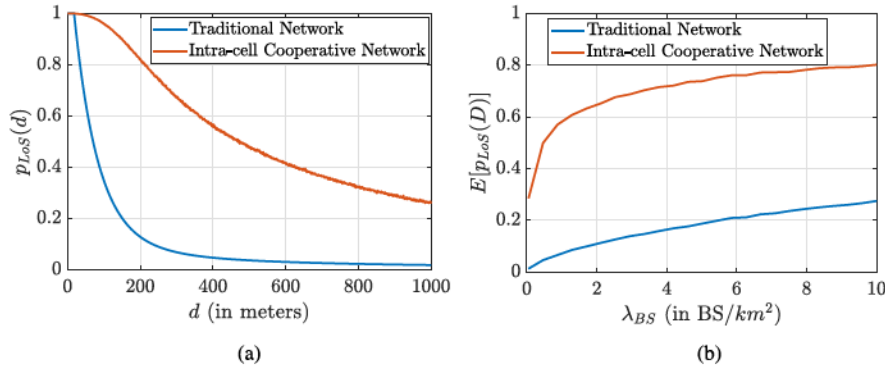


Fig. 4. Study of LoS probability for $\lambda_V = 30$ vehicles/km and $d_R = 100$ m. (a) LoS probability for a vehicle a distance d away from its closest BS, for $\lambda_{\text{BS}} = 2$ BSs/km². (b) LoS probability for a typical vehicle as a function of λ_{BS} .

form, assuming the LoS probabilities of vehicles within the same sub-cluster are conditionally independent given D :

$$p_{\text{LoS}}^*(d) = \mathbb{E}_{\Phi_{\text{BS}}, \Phi_V}^0 \left[1 - \prod_{i=1}^{\tilde{Z}} (1 - p_{\text{LoS}}(D_i)) \mid D = d \right] \quad (10)$$

where $\mathbb{E}^0[\cdot]$ denotes the Palm expectation with respect to the typical vehicle and its subcluster. Clearly, we have $p_{\text{LoS}}^*(d) \geq p_{\text{LoS}}(d)$, $\forall d$, i.e., the probability that at least one of the vehicles in the typical vehicle's sub-cluster sees a LoS link is always larger than the probability that the typical vehicle sees one without cooperation. We note that the conditional independence assumption is realistic when the inter-vehicle distance d_V is not too small. This ordering is exhibited in Fig. 4(a), which shows the LoS probability with and without V2V cluster relaying for a vehicle at a given distance from its closest BS, taking into account the random sub-cluster sizes. In the sequel, we adopt the following baseline $p_{\text{LoS}}(\cdot)$ function, as proposed in the 3GPP Release 14 standard in an Urban Macro-cell (UMA) environment [38]:

$$p_{\text{LoS}}(d) = \begin{cases} 1, & \text{if } d \leq 18 \text{ m,} \\ \frac{18}{d} + (1 - \frac{18}{d}) \cdot e^{-d/63}, & \text{if } d > 18 \text{ m.} \end{cases} \quad (11)$$

Fig. 4(a) shows a considerable improvement in the probability of benefiting from a LoS link. The gains are particularly significant for large values of d , i.e., for cell-edge vehicles. Vehicle clustering can then be seen as a mechanism that reduces the need to densify the network with BSs, hence reducing the infrastructure deployment costs as well as the mean interference power level in the network. This phenomenon is noticeable in Fig. 4(b) showing how $\mathbb{E}_D[p_{\text{LoS}}(D)]$ varies as a function of the BSs density λ_{BS} .

First, one observes that when the distribution of the distance D between a typical vehicle and its closest BS is taken into consideration, the probability of a LoS link increases 6-fold between the traditional and intra-cell cooperative network scenarios, for the network parameters selected in Fig. 4(a), i.e., $\lambda_{\text{BS}} = 2$ BSs/km². Second, vehicle clustering enables substantial savings in the density of BS needed to achieve a specific p_{LoS} level for a typical vehicle. For instance, to guarantee that a typical vehicle sees a LoS with probability 0.28, λ_{BS} needs to be equal to 10 BSs/km² in a traditional network, while the

same performance can be achieved with $\lambda_{\text{BS}} = 0.05$ BSs/km² in the cooperative setting for the selected network parameters. Hence, by using V2V cluster relaying, a network operator could achieve considerable savings in infrastructure deployment, when providing service only to vehicles.

C. Mean Shared Rate Gains Through Intra-Cell Opportunism

We now study the joint effect of the reduced effective distance among vehicles and their associated BSs, and the improved probability that they benefit from LoS links on the vehicles' mean shared rate. The vehicles' mean shared rate is defined to be the rate received by the vehicles after sharing the wireless resources amongst the vehicles and UEs. In the intra-cell cooperative network scenario, we consider a proportionally fair resource allocation scheme, i.e., all the users receive the same fraction of resources regardless of their link quality, and see a shared rate proportional to their average transmission rate. Hence, for a given network realization, vehicle v belonging to cluster c attached to BS b would receive in the traditional network a shared rate

$$s_v = \frac{r_v^b}{|\phi_V^b| + |\phi_M^b|}, \quad \forall v \in \phi_V^b, \quad (12)$$

and in the intra-cell cooperative network scenario a shared rate

$$s_v^{*,\text{intra}} = \frac{r_c^{b,*}}{|\phi_V^b| + |\phi_M^b|}, \quad \forall v \in \phi_V^b \cap \phi_{V,c}. \quad (13)$$

In addition, for a mobile UE m , s_m is defined similarly to s_v , and we have $s_m = s_m^{*,\text{intra}}$. Fig. 5 exhibits how both effects presented in this section can improve the mean shared rate by comparing s_v to $s_v^{*,\text{intra}}$, using the network parameters in Table III based on the 3GPP standard [38].

One observes that intra-cell opportunism leads to a considerable shared rate boost for the vehicles in the network, allowing them to experience higher shared rate especially when the roads are congested. At this stage, we emphasize that these considerable gains in mean shared rate result from the sole effect of intra-cell opportunism. In reality, clusters can cross cell edges (and are not artificially interrupted as assumed here for tractable analysis) leading to longer clusters providing even further opportunities for reduction in effective distance, and finding a LoS

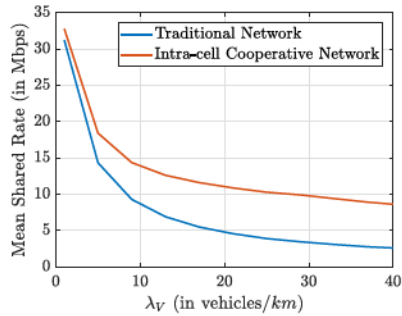


Fig. 5. Comparison of the vehicles' mean shared rate in the traditional and intra-cell cooperative scenarios as a function of λ_V , for the simulation parameters in Table III.

TABLE III
NETWORK SIMULATION PARAMETERS, BASED ON THE 3GPP RELEASE 14 MODELS [38]

Parameter	Value	Units
λ_{BS}	2	BSs/km ²
λ_M	10	UEs/km ²
λ_R	4.5	road km/km ²
d_R	100	m
p_{BS}	40	dBm
w	100	MHz
k_{LOS}	-34	dB
k_{NLOS}	-19.5	dB
α_{LoS}	2.2	-
α_{NLoS}	3.9	-
σ^2	$-199 + w _{dB}$	dBm
Γ	3	dB

link to the BS. As argued in the sequel, further gains are to be expected through effective network-level resource allocation strategies, allowing for instance the balancing of vehicular loads across cells, benefiting vehicles, but also mobile UEs that do not have any relaying abilities.

V. RESOURCE ALLOCATION ALGORITHMS FOR CLUSTER-BASED COOPERATIVE RELAYING NETWORKS

In order to study the full performance gains associated with V2V cluster relaying, i.e., leveraging both opportunism and load balancing, the wireless resource (e.g., time and/or bandwidth) allocation mechanisms need to be defined as the network performance depends considerably on the adopted policy. We no longer assume that clusters are interrupted at the cell edges, and we allow for cluster multihoming.

The resource allocation problem can be broken down into two sub-problems:

- 1) How should resources be allocated by BSs to serve the clusters and mobile UEs?
- 2) How should resources be shared amongst vehicles within each cluster?

A reasonable sharing strategy within a cluster is to divide resources equally amongst the associated vehicles. In our formulation, we will assume such a sharing policy, relegating the discussion of alternative strategies to Section VII. Below,

we introduce two different algorithms addressing the first sub-problem, while Section VI focuses on evaluating their respective performance.

A. Network-Level Fairness-Optimal Joint Rate Optimization

We first consider a centralized network-level joint optimization of opportunistic relaying and load balancing, aiming at fairly allocating the wireless resources among the mobile UEs and vehicle clusters. The optimization framework solves the first sub-problem for a given network configuration in a finite region. As described in Section III, we shall assume that a BS $b \in \phi_{BS}$ serves the set of mobile UEs ϕ_M^b in its cell, where for $m \in \phi_M^b$ the average transmission rate is r_m^b given in (5). Similarly, BS b may serve vehicles in any cluster $c \in \phi_C^b$ which would in turn perceive an average rate $r_c^{b,*}$, as defined in (6). However, as the BSs need to share their resources among the vehicle clusters and mobile UEs in their cells, each entity shall be served only for a fraction of the transmission time. Specifically, BS b decides on an allocation vector $\pi^b = (\pi_i^b : i \in \phi_M^b \cup \phi_C^b) \geq 0$, representing the fraction of time allocated to the mobile UEs and clusters it can serve, such that $\|\pi^b\|_1 = \sum_{i \in \phi_M^b \cup \phi_C^b} \pi_i^b = 1$, and we let $\pi = (\pi^b : b \in \phi_{BS})$. For a given allocation vector (potentially a function of the network realization), we define the shared rate $s_m^{*,inter}$ perceived by mobile UE m attached to BS b as

$$s_m^{*,inter} = \pi_m^b r_m^b, \quad \forall m \in \phi_M. \quad (14)$$

while vehicle v in cluster c multihomed through a set of BSs $\phi_{BS,c}$ perceives $s_v^{*,inter}$ such that

$$s_v^{*,inter} = \frac{1}{|\phi_{V,c}|} \sum_{b \in \phi_{BS,c}} \pi_c^{b,*} r_c^{b,*}, \quad \forall v \in \phi_{V,c}. \quad (15)$$

Recall that while mobile UEs are assumed to be served by only one BS, vehicle clusters can be *multihomed*, i.e., served by multiple BSs, explaining the summation in (15).

One way to improve the network users' QoS is to provide them with steady data rates. By ergodicity, this can be achieved by ensuring a fair distribution of resources among the mobile UEs, and the vehicles by selecting an appropriate resource allocation vector π . In general, there is a tradeoff between performance (measured in terms of mean shared rate per user) and fairness among the user allocations, see, e.g., [13]. One can define fairness in different ways, allowing one to control this tradeoff. For instance, a *max-min fair* resource allocation might be relevant for our scenario; but other fairness measures could also be used such as *proportional fairness*, that allocates resources proportionally to the link quality between the cluster and the BSs. In order to keep our framework as general as possible, we shall use α -fair utility functions, that model a range of fairness definitions via the parameter α , see [41]. For instance, proportional fair resource sharing corresponds to $\alpha = 1$, and max-min fair to a value of $\alpha \rightarrow \infty$. For each network user we posit an increasing concave utility function $\mathcal{U}_\alpha(\cdot)$ of its allocated shared rate s , where:

$$\mathcal{U}_\alpha(s) = \begin{cases} \frac{s^{1-\alpha}}{1-\alpha}, & \text{if } \alpha \geq 0, \alpha \neq 1, \\ \log(s), & \text{if } \alpha = 1. \end{cases} \quad (16)$$

With this notation in place, and given a network realization, the network utility maximization problem is given as follows:

$$\begin{aligned} & \max_{\pi} \sum_{v \in \phi_V} \mathcal{U}_\alpha(s_v) + \sum_{m \in \phi_M} \mathcal{U}_\alpha(s_m) \\ \text{s.t. } & \begin{cases} s_m = \pi_m^b r_m^b, & \forall m \in \phi_M, \forall b \in \phi_{BS}, \\ s_v = \frac{1}{|\phi_{V,c}|} \sum_{b \in \phi_{BS,c}} \pi_c^b r_c^{b,*}, & \forall v \in \phi_{V,c}, \forall c \in \phi_C, \\ \|\pi^b\|_1 = 1, \pi^b \geq 0, & \forall b \in \phi_{BS}. \end{cases} \end{aligned} \quad (17)$$

The above optimization problem is convex (as we maximize a concave objective function over a convex set) but may not have a unique optimizer [42]. Intuitively if there were a cycle of BSs linked by overlapping vehicle clusters it may be possible to shift resource allocations around the cycle while maintaining the same overall network utility.

While the proposed optimization framework allows the network to reach a proportional fairness-optimal resource allocation, one major issue associated with using such a network-level and centralized optimization algorithm is that it requires excessive computation, especially for large and congested networks. This is likely to hinder the ability to deploy and run such an algorithm in real-time, especially with highly mobile users such as vehicles that would impose a continual network re-optimization making such a solution impractical. In particular, most known algorithms to solve Problem 17 can find an ϵ -optimal solution within $\mathcal{O}(\delta/\epsilon^2)$ iterations, e.g., projected gradient ascent, or $\mathcal{O}(\delta/\epsilon)$, e.g., ADMM, where δ is the problem dimension, i.e., $\delta = |\phi_{BS}||\phi_C + \phi_M|$, see [43]. Note that the real algorithm complexity might be even worse as each iteration typically has a dimension-dependent per-iteration complexity. Clearly, as the network grows larger, the overall complexity of this problem also scales up, making it unusable in large-scale networks.

B. Cluster-Level Load-Balancing Algorithm

We propose in this section a decentralized algorithm solving the user-association and resource-allocation problems with a network-size independent complexity, allowing to spread the computations across multiple nodes in the network, while being less computationally intensive overall. As the vehicles are highly mobile, the optimal user association is likely to change quickly over time and it might be preferable to equip the network with an agile potentially sub-optimal algorithm, rather than a slow one that leads to an optimal yet obsolete solution. Unlike the centralized optimization framework we have proposed, this algorithm first solves the user-association problem and then allocates an equal amount of wireless resources to all the users associated to each BS. The idea is to have clusters asynchronously trigger REBALANCE, a re-association routine at random or periodic times.

This re-association routine consists in having each cluster c update how many of its set $\phi_{V,c}$ of vehicle based users should be served by each of the BSs in $\phi_{BS,c}$, i.e., the BSs serving cells crossed by c . In particular let $\mathbf{n}_c = (n_c^b \in \mathbb{Z}_+ : b \in \phi_{BS,c})$ where n_c^b denotes the number vehicles in cluster c served by BS b . Let $\mathbf{k}_c = (k_c^b \in \mathbb{Z}_+ : b \in \phi_{BS,c})$ where k_c^b denotes the number of *other* mobile UEs and vehicle based users BS b

is currently serving (i.e., excluding the ones in c). Finally we shall define $\mathbf{r}_c^* = (r_c^{b,*} \in \mathbb{R}_+ : c \in \phi_{BS,c})$ where $r_c^{b,*}$ denotes the highest transmission rate BS b can achieve amongst cluster c 's vehicles in its cell, as defined in (6).

When the cluster management update is engaged, it takes the current vectors \mathbf{k}_c and \mathbf{r}_c^* , and determines \mathbf{n}_c^* that maximizes the cluster-level utility, defined as

$$\mathcal{L}_{c,\alpha}(\mathbf{n}) = \sum_{b \in \phi_{BS,c}} n^b \cdot \mathcal{U}_\alpha \left(\frac{r_c^{b,*}}{n_c^b + k_c^b} \right) \quad (18)$$

such that

$$\mathbf{n}_c^* \in \arg \max_{\mathbf{n}} \left\{ \mathcal{L}_{c,\alpha}(\mathbf{n}) \mid \sum_{b \in \phi_{BS,c}} n^b = |\phi_{V,c}| \right\} \quad (19)$$

i.e., each cluster greedily maximizes the network utility function over its own the set of vehicles based on local information. Note the above assumes each BS allocates an equal fraction of time to each of its UEs and vehicles. As a final step, the vehicles in $\phi_{V,c}$ aggregate their resources in a common pool and redistribute them uniformly among themselves, in such a way that all the vehicles in a cluster perceive similar rate allocation. Finding the optimal cluster association vector \mathbf{n}_c^* (with respect to the cluster-level utility function) may require solving an NP-hard integer program, or attempting a brute force search over the set of weak integer compositions of $|\phi_{V,c}|$ into $|\phi_{BS,c}|$ parts, of cardinality $\binom{|\phi_{V,c}|+|\phi_{BS,c}|-1}{|\phi_{V,c}|}$, see [44]. While both options are computationally inefficient, we propose an alternative approach in Theorem 2. The proof of correctness of this algorithm is provided in Appendix B.

Theorem 2 (Sequential Association Solves the Cluster-level Maximization Problem): Consider a cluster c and let $\{\tilde{\mathbf{n}}_c^{(i)}\}_{i=0}^{|\phi_{V,c}|}$ be a sequence of vectors in $\mathbb{Z}_+^{|\phi_{BS,c}|}$, defined as:

$$\tilde{\mathbf{n}}_c^{(i)} \triangleq \arg \max_{\mathbf{n}} \left\{ \mathcal{L}_{c,\alpha}(\mathbf{n}) \mid \mathbf{n} = \tilde{\mathbf{n}}_c^{(i-1)} + \mathbf{e}_b, \forall b \in \phi_{BS,c} \right\}, \quad (20)$$

where \mathbf{e}_b is the b^{th} basis vector in $\mathbb{Z}_+^{|\phi_{BS,c}|}$ and $\tilde{\mathbf{n}}_c^{(0)} \triangleq \vec{\mathbf{0}} \in \mathbb{Z}_+^{|\phi_{BS,c}|}$, then $\mathcal{L}_{c,\alpha}(\mathbf{n}_c^*) = \mathcal{L}_{c,\alpha}(\tilde{\mathbf{n}}_c^{(|\phi_{V,c}|)})$.

Therefore, determining \mathbf{n}_c^* is a straightforward task of computational complexity $\mathcal{O}(|\phi_{BS,c}||\phi_{V,c}|)$, that is independent of the network size. Assuming BSs track k_c^b and r_c^b the data requirement to perform a cluster c update is $\mathcal{O}(|\phi_{BS,c}|)$. The REBALANCE routine is summarized in Algorithm 1.

For this algorithm, we envision vehicles follow a *dual-handoff* protocol wherein cluster-head handoffs, i.e., the decision to select a different cluster-head and BS handoffs, i.e., the decision of a cluster-head to associate to a different BS are taken separately, potentially on different time-scales. While one can rely on existing standardized protocols to manage the latter, the former needs further investigation. Clearly, in a dynamic network where vehicles and UEs are mobile, the cluster-head selection and cluster re-association routine REBALANCE need to be performed repeatedly, and the necessary rate of updates will depend on the vehicles' velocity. One mode of operation is to have clusters

Algorithm 1: REBALANCE(\mathbf{r}_c^* , \mathbf{k}_c).

Result: Solves for $\mathbf{n}_c^* \in \mathbb{Z}_+^{|\phi_{BS,c}|}$

- 1 Cluster c observes $\phi_{BS,c}$ and requests $k_c^b, \forall b \in \phi_{BS,c}$
- 2 Cluster c evaluates $\mathbf{r}_c^* \in \mathbb{R}_+^{|\phi_{BS,c}|}$
- 3 $\tilde{\mathbf{n}}_c^{(0)} = \mathbf{0} \in \mathbb{Z}_+^{|\phi_{BS,c}|}$
- 4 **for** $i=1$ to $|\phi_{BS,c}|$ **do**
- 5 $\tilde{\mathbf{n}}_c^{(i)} = \arg \max_{\mathbf{n}} \left\{ \mathcal{L}_{c,\alpha}(\mathbf{n}) \mid \mathbf{n} = \tilde{\mathbf{n}}_c^{(i-1)} + \mathbf{e}_b, \forall b \right\}$
- 6 **end**
- 7 $\mathbf{n}_c^* = \tilde{\mathbf{n}}_c^{(|\phi_{BS,c}|)}$
- 8 Cluster c informs all $b \in \phi_{BS,c}$ about updated \mathbf{n}_c^* .

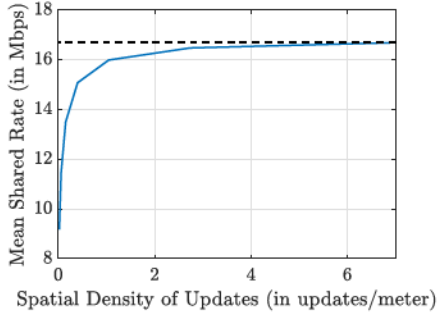


Fig. 6. Figure of a typical vehicle's mean shared rate under the distributed cluster-level algorithm as a function of the spatial density of updates, for the parameters in Table III, $\lambda_V = 30$ vehicles/km, and $d_{\text{corr}} = 30$ m. The dashed line shows the asymptote when the network users are static.

trigger updates after random exponential timeouts with given mean rate, proportional to the vehicles' velocity. In order to study the significance of the mean update rate, we performed time-domain simulations and we examined the *spatial density of updates*, i.e., the expected number of re-association updates performed by a cluster per meter traveled, and computed as the ratio of the cluster update rate and the cluster velocity. This metric is motivated by the fact that the network performance expressed, e.g., in terms of mean shared rate seen by a typical vehicle, is invariant to an increase in both the network users' velocities and the mean update rate by a common factor. To perform the time-domain simulations, we extend the network and wireless link models described in Section III by assuming a linear constant-velocity trajectory for all the clusters on their roads, while mobile UEs move in random linear directions in the network at a velocity set to be $10\times$ slower than the vehicles. In addition, to capture the spatial correlation in the wireless channels in a mobile setting, we define LoS_t to be the event that a user moving at velocity v sees a LoS link to its closest BS at a distance d_t at time t , we propose the following simple Markovian model capturing the temporal correlation in the users' LoS link probability, which is consistent with results presented in [45]:

$$p_{LoS}(d_t | LoS_{t-1}) = \beta \cdot p_{LoS}(d_t) + (1 - \beta) \cdot \mathbb{1}\{LoS_{t-1}\} \quad (21)$$

where $\beta = \min(\frac{v \cdot \Delta t}{d_{\text{corr}}}, 1)$, and d_{corr} models the LoS correlation distance. Fig. 6 exhibits the network performance in terms of the typical user's mean shared rate, as a function of the spatial density of updates. The key takeaway of Fig. 6 is that while the typical vehicle's mean shared rate improves as the re-association

rate increases and/or the cluster velocity decreases, reasonable mean cluster-head handoff rates (e.g., on the order of one update every 50 ms for clusters moving at medium to high velocity) are sufficient to ensure satisfactory performance.

VI. PERFORMANCE EVALUATION OF V2V CLUSTERING

In this section, we discuss Monte-Carlo simulations, aiming at providing additional insight regarding different resource allocation and network management strategies. We compare the performance of the two resource allocation algorithms presented in Section V, and quantify the full gains associated with V2V clustering, i.e., joint opportunism and load balancing gains.

Performance is evaluated via two metrics: (1) the mean gain in per-user shared rate, and (2), the Jain's index measure of fairness in the per-user shared rate. All the mean gains in shared rate are relative to the non-cooperative scenario, i.e., we evaluate the *average of the ratio* of the shared rate received by a network user in the cooperative setting over the shared rate that the same user would perceive if vehicles were not cooperating, or more formally the spatial averages of $\frac{s_v^*, \text{inter}}{s_v}$ and $\frac{s_m^*, \text{inter}}{s_m}$, i.e., averaged over all the vehicles and UEs in space, respectively. We present the results as a function of the cluster density λ_V representing how congested the roads are. We evaluated network performance in Fig. 7 for random network configurations based on the network parameters shown in Table III. Four settings were considered:

Traditional cellular: all users associate with their closest BS which shares its resources equally amongst them. This setting is used as a baseline for the mean shared-rate gain computations, and does not explicitly appear in Fig. 7(a) and 7(b).

Intra-cell opportunism: setting analyzed in Section III where only cluster-based intra-cell opportunism was exploited, and all users in a cell also receive an equal amount of resources.

Network Optimal (PF): corresponds to solving the centralized network utility maximization Problem 17 introduced in Section V where all users have log utilities, i.e., Proportionally Fair (PF) resource allocation.

Distributed algorithm: refers to the cluster-side distributed algorithm previously described where the Voronoi cells were used to initialize the association, before conducting multiple cluster re-balancing updates. The number of such updates was five times the number of clusters in the simulated area, the updated clusters were selected at random.

Note that only the last two exploit both opportunism and load balancing. Multiple takeaways can be extracted from Fig. 7. First, one can observe on Fig. 7(a) that a typical vehicle can see considerable gains in mean shared rate, regardless of the adopted algorithm, that may even reach $20\times$ gains for policies leveraging both opportunism and load balancing. Such gains arise from the substantial inequities in link capacities amongst vehicles at the cell-edge compared to the ones in LoS with the BS. V2V cluster-based relaying is therefore an effective mechanism to bridge this gap. One can clearly distinguish the gains associated with intra-cell opportunism reaching around $10.86\times$ with inter-cell and load balancing gains providing an additional $1.87\times$ gain factor when $\lambda_V = 40$ vehicles/km. We note that the reported

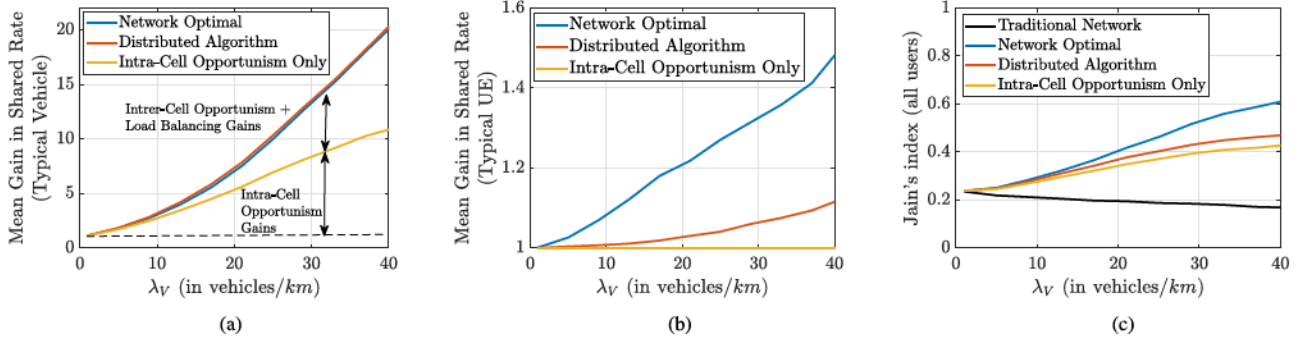


Fig. 7. Resource Allocation Algorithms Performance Comparison for the parameters in Table III. (a) Typical vehicle's gain in shared rate. (b) Typical UE's gain in shared rate. (c) Jain's index of per-user shared rate.

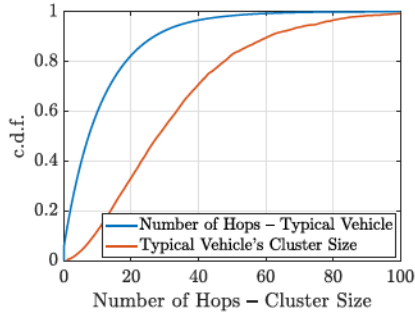


Fig. 8. C.d.f.s of the typical vehicle's cluster size and the number of hops experienced by the typical vehicle's packets, for the parameters in Table III.

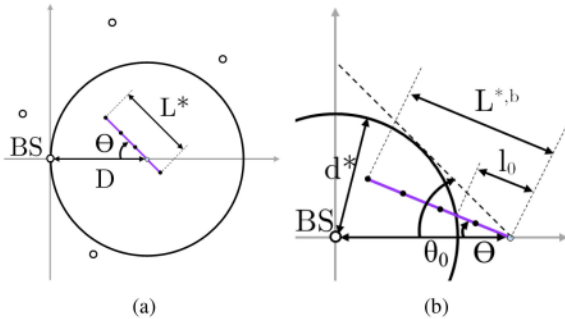


Fig. 9. Geometry of the typical vehicle's cluster and environment. (a) Geometry of vehicle cluster based opportunistic relaying. (b) Illustration of d^* , $\theta_0(d, d^*)$, $l^{*,b}$ and $l_0(d, d^*, \theta)$.

gains are under the PPP assumptions for the BS, UE, and clusters placement, and we expect the load balancing gains to be even more substantial under spatially bursty traffic (as may happen on roads), when load balancing across cells is the most needed. Similarly, all strategies lead to a considerable improvement in the users' shared rates Jain's index, as observable in Fig. 7(c), compared to the traditional cellular setting.

Second, we observe that our proposed cluster-level algorithm performs slightly better than the network fairness-optimal allocations in terms of mean gains in shared rate. In retrospect this is not surprising as the latter strategy optimizes for proportional fairness, rather than sum shared rate. This is reflected in Fig. 7(c), where the network optimization framework

logically outperforms the other resource allocation policies in terms on rate fairness.

Third, we observe on Fig. 7(b) that, on average, the mobile UEs also benefit from V2V relaying if the resource allocation policy leverages load balancing, although they are not actively relaying. While not all the UEs will benefit from the cooperative scheme, a typical UE will. Indeed, a typical UE is likely to belong to a large cell, that is expected to be highly loaded. This cell will benefit from load balancing by shifting the vehicles away, letting them to associate to smaller and less loaded neighboring cells. This mechanism frees up additional resources that can be shared with all the cell users (including the typical UE). In addition, UEs tend to benefit more from a network fairness-optimal resource allocation strategy, compared to the local cluster-based algorithm. This results from the fact that their perceived shared rate is explicitly taken into account in the problem formulation 17, whereas it only implicitly appears in 19, where clusters mainly consider the shared rate perceived by of their own vehicles while still attempting to balance the load across cells. The fact that the network fairness-optimal allocates more resources to the mobile UEs compared to vehicles also provides an additional explanation to the fact that this policy underperforms the cluster-level distributed strategy from the perspective of the typical vehicle's shared rate gain, but largely outperforms it in terms of per-user shared-rate fairness when both vehicles' and UEs' rates are taken into account, as seen in Fig. 7(c).

In summary, the adoption of the distributed cluster-level algorithm seems to be justified. Aside from its computational advantages, it has been shown to outperform the network fairness optimal algorithm in terms of shared rate gain seen by a typical vehicle in the network, although it does not provide as much gains to the mobile UEs. In addition, the distributed algorithm's fairness performance remains satisfactory as compared to the network fairness optimal strategy, making it a suitable solution for the resource allocation and user association problem.

Remark: In this work, we have only considered networks where the clusters are homogeneously distributed in space. In reality, these clusters may be spatially clustered, e.g., near highways and metropolitan areas. While the theoretical analysis of such networks is out of scope of this paper, we conjecture that while spatial correlation in the clusters' locations does not

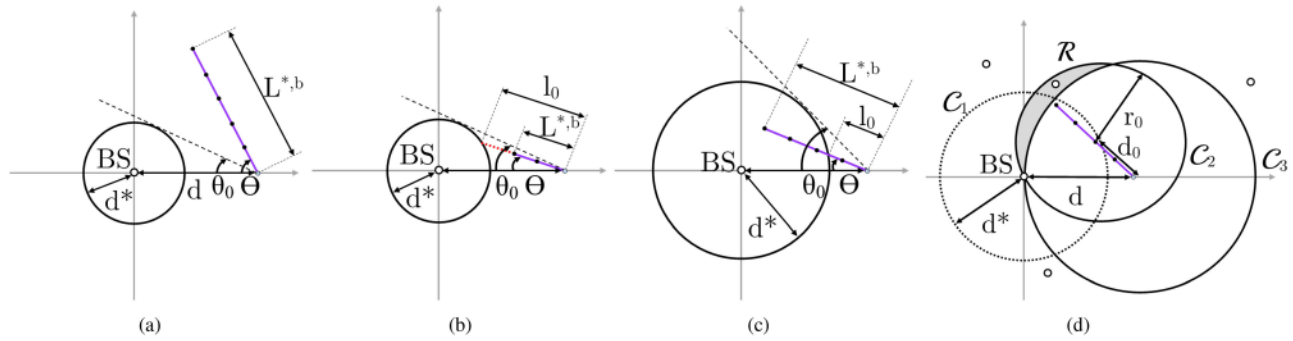


Fig. 10. Typical Vehicle's Cluster Configuration Analysis. (a) Case 1 Configuration. (b) Case 2 Configuration. (c) Case 3 Configuration. (d) Case 3 Region of Interest.

impact intra-cell opportunism gains, it may improve inter-cell opportunism gains as highly congested cells would lead to additional benefits from load-balancing.

VII. TECHNICAL CHALLENGES

We complement our theoretical and numerical analysis with additional technical challenges that would need to be examined if V2V cluster relaying was deployed.

A. Incentive Mechanism for Cluster Relaying

The first challenge that needs to be addressed is the design of an effective incentive mechanism for vehicles to be willing to route traffic for others in their cluster. Indeed, vehicles can be seen as selfish and greedy agents that need to be compensated for the additional transmission power and potentially additional overheads associated with routing and relaying.

To address this issue, two types of solutions can be proposed. First, the cluster-head vehicles are naturally incentivized to cooperate, when the network dynamics are considered. A cluster-head is indeed unlikely to keep this role continuously, as the vehicles are in motion and the channel qualities vary over time. A cluster-head is clearly better-off routing traffic for its cluster to benefit from future throughput improvement when its channel quality deteriorates. While this may not be necessary, a token-based mechanism can be implemented, as proposed in [46], where vehicles in each cluster would pay their cluster-head(s) through a virtual currency/tokens that can be redeemed at a later stage, when the latter can benefit from V2V-relaying.

Second, cooperation can be incentivised through more direct mechanisms. For instance, instead of equally dividing the wireless resources among all the vehicles in a cluster, they can be re-distributed in a way that rewards the cluster-heads, e.g., as a function of the rate each vehicle would have perceived without V2V cluster relaying. This solution has the advantage of providing instantaneous incentives to cluster-heads by providing them with additional throughput, but at the cost of negatively impacting the shared-rate fairness amongst the vehicles in the network. Additional benefits of instantaneous incentives over token-based have been discussed in [47].

Finally, we note that the incentive mechanism can also be designed to benefit vehicles that are not cluster-heads, but that are willing to forward other vehicles packets in the cluster, i.e., intermediate nodes. Suitable incentive mechanisms would include the ones providing rewards (wireless resources or tokens) proportionally to the volume of data forwarded.

B. Delay Management in Cluster Relays

The second challenge that would need to be accounted for would be the additional delays associated with packet routing, especially when the clusters are large. In this paper, we have assumed that the V2V links have very high capacity, hence leading to negligible transmission delays. However, this assumption may not always hold, and other types of delays may also be taken into account, such as packet processing and forwarding delays. Clearly, V2V-relaying is a promising technology that has been shown to provide considerable potential in terms of throughput, but it may be best fitted for applications that are not too delay sensitive.

One solution to partially address this issue would be to artificially break the clusters, e.g., by fixing a maximum cluster size to limit the overall packet delay to the furthest vehicle in the cluster. This would clearly impact the mean gain in shared-rate performance, but it would help providing better delay guarantees. Network operators can then tune the maximum cluster size parameter to control the tradeoff between mean throughput and packet delay.

Finally, we note that even though the clusters can have considerable sizes, they are also likely to be multihomed and traffic can be effectively routed from the closest serving BS, according to the association/resource allocation algorithm in effect. We show in Fig. 8 a comparison of the distributions of the typical vehicle's cluster size and the number of hops experienced by the typical vehicle's packets, for the cluster-level distributed algorithm presented in Section V. We observe that the typical vehicle experiences substantially fewer hops than its cluster size, and may not always leverage multihoming if the load across cells is very imbalanced. For instance, while the median typical vehicle's cluster size is 27 vehicles using the network parameters in Table III, the median number of packet hops for a typical vehicle is only 7 V2V hops.

C. Real-Time Cluster Management

The third practical challenge associated with V2V clustering is the ability to manage the cluster in real-time. As discussed in Section V, vehicles may be moving at high velocity, hence the cluster-head role and routing decisions will need to be frequently updated to ensure the best performance.

A perhaps more restrictive issue is that an efficient signaling protocol needs to be adopted to detect and adapt to changes in the vehicle-clustering/association. For instance, clusters may divide or merge over time, and BSs and cluster-heads need to be aware of such changes in real-time to adapt/re-optimize the resource allocation amongst users in the network. Two approaches could be used to resolve this issue. In the first solution, changes can be detected locally (e.g., by vehicles in the cluster) and the information can be propagated to the relevant cluster-head(s). For instance, cluster-heads of merging clusters can negotiate the best association strategy and propagate the information back to the relevant BSs. In the second solution, a centralized virtual controller (e.g., placed in an edge server) can aggregate information collected from vehicles to recompute the optimal re-association/vehicle clustering, and propagate the information back to the BSs and the vehicles. While the second solution may lead to better performance as more information is aggregated to take better decisions, the first solution may be preferable due to the simpler signaling required, and less computations, making it more reactive to sudden changes.

VIII. CONCLUSION

In this paper, we studied the gains associated with V2V cluster-based relaying by identifying and characterizing the sources of opportunistic and load balancing gains analytically and via simulations. We established a network-level optimization framework to associate users to BSs and ensure a fair wireless resource allocation scheme. The results show that shared-rate gains from V2V-relaying can exceed an order of magnitude, and are associated with shared-rate fairness improvements across network users, stemming from both opportunism and load balancing. We then used this framework as a benchmark to study the performance of a cluster-level, efficient, and distributed vehicle association and resource allocation algorithm. We show that excellent performance gains can be achieved via this policy, while being more convenient than the network-level algorithm when executed in real-time. Furthermore, V2V relaying has been shown to help providing improved service to all the network users, including mobile UEs that do not actively participate in the relaying scheme. While some technical challenges still need to be addressed, the considerable gains associated with V2V cluster relaying will motivate future development.

APPENDIX A PROOF OF THEOREM 1

Proof: Fig. 9(a) exhibits the geometry of vehicular cluster-based opportunistic relaying. The distance D between a typical vehicle on the x -axis and its closest BS assumed without loss of generality to be at the origin follows the distribution in (7). The typical vehicle belongs to a cluster of size Z^* vehicles whose

distribution corresponds to the size biased distribution of the typical cluster length Z such that $p_{Z^*}(z) = \frac{z \cdot p_Z(z)}{\mathbb{E}[Z]}$, for $z \in \mathbb{N}$, i.e., typical vehicles are more likely to belong to longer clusters. The typical vehicle's cluster size induces a typical vehicle's cluster length L^* (in meters), such that $L^* = (Z^* - 1) \cdot d_V$. The random orientation Θ of the typical vehicle's cluster (acute angle) is such that $\Theta \sim \text{Unif}[0, \frac{\pi}{2}]$ which is independent of Z^* , L^* and D .

To prove Theorem 1, we note that the location of a typical vehicle within its cluster is uniformly distributed, and "breaks" its cluster into two fragments. We denote as $Z^{*,b}$ and $L^{*,b}$ the size and length of the typical vehicle's cluster fragment "pointing in the direction of the BS," where there may be candidate opportunistic relays with better channels to the BS at the origin. The distribution of $Z^{*,b}$ is shown to be:

Lemma 1: Given a typical cluster size Z finite-mean distribution $p_Z(\cdot)$, the distribution of $Z^{*,b}$ is

$$p_{Z^{*,b}}(z) = \frac{\mathbb{P}(Z \geq z)}{\mathbb{E}[Z]}, \quad z \in \mathbb{N} \quad (22)$$

Proof: (Lemma 1) As the typical vehicle can be any one in its cluster with the same probability, i.e., $p_{Z^{*,b}|Z^*}(z|Z^* = i) = 1/i, z = 1, \dots, i$, the distribution of $Z^{*,b}$ is

$$p_{Z^{*,b}}(z) = \sum_{i=1}^{\infty} p_{Z^{*,b}|Z^*}(z|Z^* = i) \cdot p_{Z^*}(i), \quad \forall z \in \mathbb{N} \quad (23)$$

$$= \sum_{i=z}^{\infty} \frac{1}{i} \cdot \frac{i \cdot p_Z(i)}{\mathbb{E}[Z]}, \quad \forall z \in \mathbb{N} \quad (24)$$

$$= \frac{P(Z \geq z)}{\mathbb{E}[Z]}, \quad \forall z \in \mathbb{N} \quad (25)$$

□

The distribution of $L^{*,b}$ directly follows from the relation $L^{*,b} = (Z^{*,b} - 1) \cdot d_V$, giving

$$p_{L^{*,b}}(l) = \frac{P(Z \geq \frac{l}{d_V} + 1)}{\mathbb{E}[Z]}, \quad l = 0, d_V, 2 \cdot d_V, \dots \quad (26)$$

We seek to determine the distribution of the minimum distance D^* between a relay vehicle on the cluster of length $L^{*,b}$ and the BS at the origin with the additional requirement that the relay vehicle also belongs to the typical vehicle's cell, conditional on the distance $D = d$ between the typical vehicle and the BS. Note that $D^* \leq D$ almost surely, since a typical vehicle can of course receive data directly from its closest BS.

Fig. 9(b) exhibits the definition of two key functions of the geometry: (1) $\theta_0(d, d^*)$ the angle of the tangent to a disc of radius d^* , and (2) $l_0(d, d^*, \theta)$ which for $\Theta \leq \theta_0(d, d^*)$ is the length of the segment starting from $(d, 0)$ with angle θ to the disc of radius d^* .

With these definitions one can evaluate $\mathbb{P}(D^* \leq d^* | D = d)$ by identifying a partition $\mathcal{E}_1, \mathcal{E}_2$ and \mathcal{E}_3 corresponding to the three cases/events exhibited in Fig. A2 and given by :

- Case 1: $\mathcal{E}_1 = \{\Theta > \theta_0(d^*, d)\}$
- Case 2: $\mathcal{E}_2 = \{\Theta \leq \theta_0(d^*, d), L^{*,b} < l_0(d, d^*, \Theta)\}$
- Case 3: $\mathcal{E}_3 = \{\Theta \leq \theta_0(d^*, d), L^{*,b} \geq l_0(d, d^*, \Theta)\}$

In general we have from independence of Θ and $L^{*,b}$:

$$\begin{aligned} \mathbb{P}(D^* \leq d^* | D = d) &= \int_0^{\frac{\pi}{2}} \sum_{i=0}^{\infty} \mathbb{P}(D^* \leq d^* | D = d, \Theta = \theta, L^{*,b} = i \cdot d_V) \\ &\quad \times p_{L^{*,b}}(i \cdot d_V) f_{\Theta}(\theta) d\theta \end{aligned} \quad (27)$$

We consider the three cases individually.

Case 1: For a given d, d^* the critical angle, i.e., the angle of the tangent line to the circle of radius d^* , is given by

$$\theta_0(d, d^*) \triangleq \sin^{-1}(d^*/d) \quad (28)$$

and note from Fig. 10(a) that if $\Theta \geq \theta_0(d, d^*)$ then the cluster does not hit the radius d^* disc, whence $D^* > d^*$, and $\mathbb{P}(D^* \leq d^* | \mathcal{E}_1, D = d) = 0$.

Case 2: For a given d, d^* and $\Theta = \theta < \theta_0(d, d^*)$ note from Fig. 10(b) that a cluster extending a length $L^{*,b} = l$ where

$$l \leq l_0(d, d^*, \theta) \triangleq d \cos(\theta) - \sqrt{d^{*2} - (d \sin(\theta))^2} \quad (29)$$

will not hit the disc of radius d^* , whence $D^* > d^*$. Here $l_0(d, d^*, \theta)$ is determined by studying the triangle of side lengths l_0, d and d^* , knowing θ . Therefore, here again, $\mathbb{P}(D^* \leq d^* | \mathcal{E}_2, D = d) = 0$.

Case 3: The last case corresponds to event \mathcal{E}_3 illustrated in Fig. 10(c). Given $d, d^*, \Theta \leq \theta = \theta_0(d, d^*)$ and $L^{*,b} = l \geq l_0(d, d^*, \theta)$, the vehicle cluster extends into the circle \mathcal{C}_1 of radius d^* . In order for $D^* \leq d^*$, two conditions must be true: (1) at least one vehicle in the cluster must be in the disk of radius d^* , and (2) none of the vehicles within that disk must be closer to another BS than the one at the origin.

The first condition can be shown to be equivalent to $\left\lfloor \frac{l_0}{d_V} \right\rfloor \neq \left\lfloor \frac{l_0 + 2\sqrt{d^{*2} - (d \sin(\theta))^2}}{d_V} \right\rfloor$, i.e., when the two intersections between the cluster fragment and the circle of radius d^* occur between two different pairs of consecutive vehicles. If this condition does, not hold, then $\mathbb{P}(D^* \leq d^* | \mathcal{E}_2, D = d) = 0$.

The second condition is related to the scenario illustrated in Fig. 10(d) where we draw two additional circles. The first is \mathcal{C}_2 whose center the closest vehicle from the one at $(d, 0)$ that lies within \mathcal{C}_1 and whose radius is its distance to the BS at the origin. The second, \mathcal{C}_3 , is centered at the typical vehicle $(d, 0)$ and has radius d , i.e. also crosses the origin. Recalling that the origin is the location of the closest BS to $(d, 0)$, and thus \mathcal{C}_3 contains no other base stations. A necessary and sufficient condition to ensure that at least one vehicle in the cluster fragment within \mathcal{C}_1 is associated with the BS at the origin is that there are no BSs in the shaded region $\mathcal{R}(d, d^*, \theta)$ representing all locations that are closer to the first vehicle in \mathcal{C}_1 than to the origin. This follows because

- if there is a BS in $\mathcal{R}(d, d^*, \theta)$ then not only will the first cluster vehicle in \mathcal{C}_1 not be associated with b , but so will all the others, since the circle centered on each such vehicle and traversing the origin, will contain $\mathcal{R}(d, d^*, \theta)$. We can then conclude that $D^* > d^*$.

- if $\mathcal{R}(d, d^*, \theta)$ is empty, then at least one vehicle in the cluster is less than d^* meters away from b , and will associate with it, i.e. $D^* \leq d^*$.

Using basic algebra and the law of cosines one can show that the distance d_0 between the typical vehicle and the first cluster vehicle in \mathcal{C}_1 is given by

$$d_0(d, d^*, \theta) \triangleq d_V \cdot \left[\frac{l_0(d, d^*, \theta)}{d_V} \right]$$

and the distance r_0 from that vehicle to the origin (i.e., radius of \mathcal{C}_2) is given by

$$r_0(d, d^*, \theta) \triangleq \sqrt{d^2 + d_0^2 - 2d \cdot d_0 \cos(\theta)}. \quad (30)$$

where we have suppressed the arguments of d_0 for conciseness. Using the expression in [48] for the area of $\mathcal{C}_2 \cap \mathcal{C}_3$, as a function of d, d_0 and r_0 , one can find an expression for the area a_0 of $\mathcal{R}(d, d^*, \theta)$:

$$\begin{aligned} a_0(d, d^*, \theta) &\triangleq \pi r_0^2 \\ &- \left[r_0^2 \cos^{-1}\left(\frac{d_0^2 + r_0^2 - d^2}{2d_0 r_0}\right) + d^2 \cos^{-1}\left(\frac{d_0^2 + d^2 - r_0^2}{2d_0 d}\right) - \right. \\ &\quad \left. \frac{\sqrt{(-d_0 + r_0 + d)(d_0 + r_0 - d)(d_0 - r_0 + d)(d_0 + r_0 + d)}}{2} \right] \end{aligned} \quad (31)$$

where we have suppressed the arguments of d_0 and r_0 for conciseness. Since BSs form a PPP, the probability there is no BS in \mathcal{R} is given by

$$P(\Phi_{\text{BS}} \cap \mathcal{R}(d, d^*, \theta) = \emptyset) = e^{-\lambda_{\text{BS}} a_0(d, d^*, \theta)}.$$

Now considering the two necessary and sufficient conditions to ensure that at least one vehicle in the cluster fragment is within a distance d^* to the origin, we get

$$\begin{aligned} e_0(d, d^*, \theta) &\triangleq \mathbb{P}(D^* \leq d^* | \mathcal{E}_3, D = d) \\ &= e^{-\lambda_{\text{BS}} a_0(d, d^*, \theta)} \cdot \mathbb{1}\left\{ \left\lfloor \frac{l_0}{d_V} \right\rfloor \neq \left\lfloor \frac{l_0 + 2\sqrt{d^{*2} - (d \sin(\theta))^2}}{d_V} \right\rfloor \right\} \end{aligned} \quad (32)$$

Therefore, (27) now reduces to

$$\begin{aligned} \mathbb{P}(D^* \leq d^* | D = d) &= \int_0^{\theta_0(d, d^*)} \sum_{i=\lceil \frac{l_0(d, d^*, \theta)}{d_V} \rceil}^{\infty} e_0(d, d^*, \theta) p_{L^{*,b}}(i \cdot d_V) f_{\Theta}(\theta) d\theta \end{aligned} \quad (33)$$

$$= \int_0^{\theta_0(d, d^*)} \mathbb{P}(L^{*,b} \geq d_0(d, d^*, \theta)) \cdot e_0(d, d^*, \theta) f_{\Theta}(\theta) d\theta \quad (34)$$

where $f_{\Theta}(\theta) = \frac{1}{\pi/2}$, for $\theta \in [0, \pi/2]$. The result follows from this expression and Lemma 1. \square

APPENDIX B PROOF OF THEOREM 2

The sequential algorithm is executed in $|\phi_{V,c}|$ steps, where one seeks to associate at step t one additional vehicle to one of the serving BSs that provides the largest marginal improvement

in cluster utility given that $t - 1$ vehicles have already been associated. We shall prove the theorem by contradiction. We shall need the following two definitions.

- Define $\Delta_c^b(\mathbf{n})$ to be the change in cluster utility if an additional vehicle in cluster c were to associate to BS b . More specifically, $\Delta_c^b(\mathbf{n}) = \mathcal{L}_{c,\alpha}(\mathbf{n} + \mathbf{e}_b) - \mathcal{L}_{c,\alpha}(\mathbf{n}) = (n^b + 1) \cdot \mathcal{U}_\alpha\left(\frac{r_c^{b,*}}{(n^b+1)+k_c^b}\right) - n^b \cdot \mathcal{U}_\alpha\left(\frac{r_c^{b,*}}{n^b+k_c^b}\right)$. Note that $\Delta_c^b(\mathbf{n})$ is decreasing in the entries of \mathbf{n} as $\mathcal{L}_{c,\alpha}(\mathbf{n})$ is concave in these entries.
- Similarly, define $\Delta_c^{-b}(\mathbf{n})$ to be the change in cluster utility if a vehicle in cluster c associated to BS b were to be removed. We have, $\Delta_c^{-b}(\mathbf{n}) = \mathcal{L}_{c,\alpha}(\mathbf{n} - \mathbf{e}_b) - \mathcal{L}_{c,\alpha}(\mathbf{n}) = (n^b - 1) \cdot \mathcal{U}_\alpha\left(\frac{r_c^{b,*}}{(n^b-1)+k_c^b}\right) - n^b \cdot \mathcal{U}_\alpha\left(\frac{r_c^{b,*}}{n^b+k_c^b}\right)$. Note that $\Delta_c^{-b}(\mathbf{n}) = -\Delta_c^b(\mathbf{n} - \mathbf{e}_b)$.

Suppose $\mathcal{L}_{c,\alpha}(\mathbf{n}_c^*) > \mathcal{L}_{c,\alpha}(\tilde{\mathbf{n}}_c^{(\phi_{v,c})})$. Then $\exists b_+, b_- \in \phi_{BS,c}$ such that $n_c^{b_+,*} > \tilde{n}_c^{b_+,(|\phi_{v,c}|)}$ and $n_c^{b_-,*} < \tilde{n}_c^{b_-,(|\phi_{v,c}|)}$. Let t_0 be the index of the last iteration of the sequential algorithm that associated an additional vehicle to BS b_- . We seek to show that $\Delta_c^b(\mathbf{n}_c^*) + \Delta_c^{-b_+}(\mathbf{n}_c^*) \geq 0$. We have successively:

$$\Delta_c^b(\mathbf{n}_c^*) \geq \Delta_c^b(\tilde{\mathbf{n}}_c^{(t_0-1)}) \quad (35)$$

$$\geq \Delta_c^{b_+}(\tilde{\mathbf{n}}_c^{(t_0-1)}) \quad (36)$$

$$\geq \Delta_c^{b_+}(\mathbf{n}_c^* - \mathbf{e}_{b_+}) \quad (37)$$

$$= -\Delta_c^{-b_+}(\mathbf{n}_c^*) \quad (38)$$

Where inequality 35 follows from the facts that $n_c^{b_-,*} \leq \tilde{n}_c^{b_-,(t_0-1)}$ and Δ_c^b is decreasing. Inequality 36 is a necessary condition for the sequential algorithm to associate the vehicle to BS b_- at step t_0 , and inequality 37 follows from the facts that $n_c^{b_+,*} \geq 1 + \tilde{n}_c^{b_+,(|\phi_{v,c}|)} \geq 1 + \tilde{n}_c^{b_+,(t_0-1)}$ and $\Delta_c^{b_+}$ is decreasing in $n_c^{b_+}$. Finally equality 38 follows from the definition of $\Delta_c^b(\mathbf{n})$. Therefore, we get $\Delta_c^b(\mathbf{n}_c^*) + \Delta_c^{-b_+}(\mathbf{n}_c^*) \geq 0$, hence \mathbf{n}_c^* is not optimal. We conclude from this contradiction that the sequential algorithm finds an optimal association vector.

REFERENCES

- [1] S. Kassir, P. C. Garces, G. de Veciana, N. Wang, X. Wang, and P. Palacharla, "An analytical model and performance evaluation of multihomed multilane VANETs," *IEEE/ACM Trans. Netw.*, vol. 29, no. 1, pp. 346–359, Feb. 2021.
- [2] J. Wang, C. Jiang, K. Zhang, T. Q. S. Quek, Y. Ren, and L. Hanzo, "Vehicular sensing networks in a smart city: Principles, technologies and applications," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 122–132, Feb. 2018.
- [3] J. Wang *et al.*, "Dynamic clustering and cooperative scheduling for vehicle-to-vehicle communication in bidirectional road scenarios," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 6, pp. 1913–1924, Jun. 2018.
- [4] W. Sun, D. Yuan, E. G. Ström, and F. Brännström, "Cluster-based radio resource management for D2D-Supported safety-critical V2X communications," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2756–2769, Apr. 2016.
- [5] J. Wang, C. Jiang, Z. Han, Y. Ren, and L. Hanzo, "Internet of vehicles: Sensing-aided transportation information collection and diffusion," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 3813–3825, May 2018.
- [6] S. Kassir, G. de Veciana, N. Wang, X. Wang, and P. Palacharla, "Enhancing cellular performance via vehicular-based opportunistic relaying and load balancing," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, 2019, pp. 91–99.
- [7] H. Wu, Chunming Qiao, S. De, and O. Tonguz, "Integrated cellular and ad hoc relaying systems: ICAR," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 2105–2115, Oct. 2001.
- [8] J. Cho and Z. Haas, "On the throughput enhancement of the downstream channel in cellular radio networks through multihop relaying," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 7, pp. 1206–1219, Sep. 2004.
- [9] C. Yu and O. Tirkkonen, "Opportunistic multiple relay selection with diverse mean channel gains," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 885–891, Mar. 2012.
- [10] P. Bahl *et al.*, "Opportunistic use of client repeaters to improve performance of WLANs," *IEEE/ACM Trans. Netw.*, vol. 17, no. 4, pp. 1160–1171, Aug. 2009.
- [11] J. Yoo, B. Choi, and M. Gerla, "An opportunistic relay protocol for vehicular road-side access with fading channels," in *Proc. 18th IEEE Int. Conf. Netw. Protoc.*, 2010, pp. 233–242.
- [12] J. Lee and J. H. Lee, "Performance analysis and resource allocation for cooperative D2D communication in cellular networks with multiple D2D pairs," *IEEE Commun. Lett.*, vol. 23, no. 5, pp. 909–912, May 2019.
- [13] F. Zabini, A. Bazzi, B. M. Masini, and R. Verdone, "Optimal performance versus fairness tradeoff for resource allocation in wireless systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2587–2600, Apr. 2017.
- [14] C. Karakus and S. Diggavi, "Enhancing multiuser MIMO through opportunistic D2D cooperation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5616–5629, Sep. 2017.
- [15] S. Das, S. Sen, and R. Jayaram, "A dynamic load balancing strategy for channel assignment using selective borrowing in cellular mobile environment," *Wireless Netw.*, vol. 3, no. 5, pp. 333–347, Oct. 1997.
- [16] A. Sang *et al.*, "A load-aware handoff and cell-site selection scheme in multi-cell packet data systems," in *Proc. IEEE Glob. Telecommun. Conf.*, 2004, pp. 3931–3936.
- [17] Y. Bejerano and S. Han, "Cell breathing techniques for load balancing in wireless LANs," *IEEE Trans. Mobile Comput.*, vol. 8, no. 6, pp. 735–749, Jun. 2009.
- [18] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706–2716, Jun. 2013.
- [19] A. Sharma, A. Trivedi, and N. Roberts, "Efficient load balancing using D2D communication and biasing in LTE-advance Het-Nets," in *Proc. IEEE 6th Int. Conf. Comput. Commun. Technol.*, 2015, pp. 456–460.
- [20] K. Son, S. Chong, and G. de Veciana, "Dynamic association for load balancing and interference avoidance in multi-cell networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3566–3576, Jul. 2009.
- [21] H. Kim, G. de Veciana, X. Yang, and M. Venkatchalam, "Distributed α -optimal user association and cell load balancing in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 1, pp. 177–190, Feb. 2012.
- [22] L. Deng *et al.*, "Device-to-device load balancing for cellular networks," *IEEE Trans. Commun.*, vol. 67, no. 4, pp. 3040–3054, Apr. 2019.
- [23] H. Zhang, L. Song, and Y. Zhang, "Load balancing for 5G ultra-dense networks using device-to-device communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4039–4050, Jun. 2018.
- [24] F. Baccelli and B. Błaszczyszyn, *Stochastic Geometry and Wireless Networks*. vol. 1, Boston, MA, USA: Now Publishers Inc, 2009.
- [25] L. Le and E. Hossain, "Multihop cellular networks: Potential gains, research challenges, and a resource allocation framework," *IEEE Commun. Mag.*, vol. 45, no. 9, pp. 66–73, Sep. 2007.
- [26] T. Van, "Location-aware and load-balanced data delivery at road-side units in vehicular ad hoc networks," in *Proc. IEEE Int. Symp. Consum. Electron.*, 2010, pp. 1–5.
- [27] D. Wu, J. Luo, R. Li, and A. Regan, "Geographic load balancing routing in hybrid vehicular ad hoc networks," in *Proc. 14th Int. IEEE Conf. Intell. Transp. Syst.*, 2011, pp. 2057–2062.
- [28] B. Sliwa, R. Falkenberg, and C. Wietfeld, "A simple scheme for distributed passive load balancing in mobile ad-hoc networks," Feb. 2017, *arXiv:1702.05235*.
- [29] C. Saha and H. Dhillon, "D2D underlaid cellular networks with user clusters: Load balancing and downlink rate analysis," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2017, pp. 1–6.
- [30] S. Das, H. Viswanathan, and G. Rittenhouse, "Dynamic load balancing through coordinated scheduling in packet data systems," in *Proc. IEEE 22nd Annu. Joint Conf. IEEE Comput. Commun. Societies*, 2003, pp. 786–796.
- [31] J. Liu, Y. Kawamoto, H. Nishiyama, N. Kato, and N. Kadowaki, "Device-to-device communications achieve efficient load balancing in LTE-advanced networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 57–65, Apr. 2014.

- [32] Z. Li, C. Wang, and C. Jiang, "User association for load balancing in vehicular networks: An online reinforcement learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 8, pp. 2217–2228, Aug. 2017.
- [33] F. Castro, A. Martins, N. Capela, and S. Sargento, "Multihoming for uplink communications in vehicular networks," in *Proc. Wireless Days*, 2017, pp. 230–237.
- [34] J. Andrews, A. Gupta, and H. Dhillon, "A primer on cellular network analysis using stochastic geometry," Oct. 2016, *arXiv:1604.03183*.
- [35] C. Choi and F. Baccelli, "Poisson Cox point processes for vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 10160–10165, Oct. 2018.
- [36] V. Chetlur and H. Dhillon, "Coverage analysis of a vehicular network modeled as cox process driven by poisson line process," Sep. 2017, *arXiv:1709.08577*.
- [37] M. Gramaglia, P. Serrano, J. A. Hern, M. Calderon, and C. J. Bernardos, "New insights from the analysis of free flow vehicular traffic in highways," in *Proc. IEEE Int. Symp. World Wireless, Mobile Multimedia Netw.*, 2011, pp. 1–9.
- [38] 5G; Study on channel model for frequencies from 0.5 to 100GHz (3GPP TR 38.901 v14.3.0 Release 14). Tech. Rep., ETSI, 2018. [Online]. Available: https://www.etsi.org/deliver/etsi_tr/138900_138999/138901/14.00.00_60/tr_138901v140000p.pdf
- [39] N. Jindal, J. Andrews, and S. Weber, "Optimizing the sinr operating point of spatial networks," Feb. 2007, *arXiv:cs/0702030*.
- [40] M. H. C. Garcia *et al.*, "A tutorial on 5G NR V2X communications," *IEEE Commun. Surv. Tut.*, vol. 23, no. 3, pp. 1972–2026, Jul.–Sep. 2021.
- [41] T. Lan, D. Kao, M. Chiang, and A. Sabharwal, "An axiomatic theory of fairness in network resource allocation," in *Proc. IEEE INFOCOM*, 2010, pp. 1–9.
- [42] B. Hajek, "Balanced loads in infinite networks," *Ann. Appl. Probability*, vol. 6, no. 1, pp. 48–75, 1996.
- [43] L. Vigneri, G. Paschos, and P. Mertikopoulos, "Large-scale network utility maximization: Countering exponential growth with exponentiated gradients," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, 2019, pp. 1630–1638.
- [44] E. Reingold, J. Nievergelt, and N. Deo, *Combinatorial Algorithms: Theory and Practice*. Englewood Cliffs, NJ, USA: Prentice Hall, 1977.
- [45] A. Samuylov *et al.*, "Characterizing spatial correlation of blockage statistics in urban mmwave systems," in *Proc. IEEE GLOBECOM Workshops*, 2016, pp. 1–7.
- [46] S. Zhong, J. Chen, and Y. Yang, "Sprite: A simple, cheat-proof, credit-based system for mobile ad-hoc networks," in *Proc. IEEE INFOCOM22nd Annu. Joint Conf. IEEE Comput. Commun. Societies*, 2003, pp. 1987–1997.
- [47] P. Mach, T. Spyropoulos, and Z. Becvar, "Incentive-based D2D relaying in cellular networks," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1775–1788, Mar. 2021.
- [48] E. Weisstein, "Circle-circle intersection," MathWorld. [Online]. Available: <https://mathworld.wolfram.com/Circle-CircleIntersection.html>