

# Capacity of Cooperative Vehicular Networks with Infrastructure Support: Multi-user Case

Jieqiong Chen, *Student Member, IEEE*, Guoqiang Mao, *Senior Member, IEEE*, Changle Li, *Senior Member, IEEE*, Weifa Liang, *Senior Member, IEEE* and De-gan Zhang *Member, IEEE*

**Abstract**—Capacity of vehicular networks with infrastructure support is both an interesting and challenging problem as the capacity is determined by the inter-play of multiple factors including vehicle-to-infrastructure (V2I) communications, vehicle-to-vehicle (V2V) communications, density and mobility of vehicles, and cooperation among vehicles and infrastructure. In this paper, we consider a typical delay-tolerant application scenario with a subset of vehicles, termed Vehicles of Interest (VoIs), having download requests. Each VoI downloads a distinct large-size file from the Internet and other vehicles without download requests assist the delivery of the files to the VoIs. A cooperative communication strategy is proposed that explores the combined use of V2I communications, V2V communications, mobility of vehicles and cooperation among vehicles and infrastructure to improve the capacity of vehicular networks. An analytical framework is developed to model the data dissemination process using this strategy, and a closed form expression of the achievable capacity is obtained, which reveals the relationship between the capacity and its major performance-impacting parameters such as inter-infrastructure distance, radio ranges of infrastructure and vehicles, sensing range of vehicles, transmission rates of V2I and V2V communications, vehicular density and proportion of VoIs. Numerical result shows that the proposed cooperative communication strategy significantly boosts the capacity of vehicular networks, especially when the proportion of VoIs is low. Our results provide guidance on the optimum deployment of vehicular network infrastructure and the design of cooperative communication strategy to improve the capacity.

**Index Terms**—Data dissemination, cooperative communication, capacity, vehicular networks, vehicle-to-infrastructure communications, vehicle-to-vehicle communications.

## I. INTRODUCTION

Interest is surging on vehicular networks and connected vehicle technology due to their increasingly important role in improving road traffic efficiency, enhancing road safety and providing real-time information to drivers and passengers

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org. (Corresponding author: De-gan Zhang.)

J. Chen and G. Mao are with the School of Computing and Communications, The University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: jieqiong.chen@student.uts.edu.au; g.mao@ieee.org).

C. Li is with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China (e-mail: clli@mail.xidian.edu.cn).

W. Liang is with the Research School of Computer Science, Australian National University, Canberra, ACT 2601, Australia (e-mail: wliang@cs.anu.edu.au).

D. Zhang is with the Tianjin Key Lab of Intelligent Computing & Novel software Technology, Tianjin University of Technology, 300384, China (email: gangedange@126.com).

[1], [2]. Two major wireless communication modes: vehicle-to-infrastructure (V2I) communications and vehicle-to-vehicle (V2V) communications, are supported in vehicular networks by deploying wireless communication infrastructure points along the roadside (e.g., road-side units (RSU), or LTE base stations), equipping vehicles with on-board communication facilities (e.g., on-board units (OBU)), and with the assistance of dedicated short-range communication (DSRC) [3] and LTE technology [4], [5].

V2I and V2V communications, on one hand, are both major techniques to disseminate data for vehicular applications, including safety applications like disseminating real-time information about traffic accidents, traffic congestion or obstacles in the road, and non-safety applications such as offering value-added services (e.g., digital maps with real-time traffic status) and in-car entertainment services [1], [2]. On the other hand, as pointed out in the literature, purely relying on V2I communications or V2V communications alone cannot meet the diverse communication requirements of different vehicular applications. For instance, V2V communications may become unreliable when the number of hops in the communication becomes large [6]. They may also not be supported and incur long communication delay when the vehicular density is low [7]. For V2I communications, on one hand, LTE-based V2I communication is not applicable to vehicular communication environment at this stage due to the high mobility of vehicles that leads to high handover and collision issues [4]. On the other hand, DSRC-based V2I communication may have limited availability, especially in rural areas and in the initial deployment phase of vehicular networks due to the high cost of implementation and maintaining of infrastructure. The aforementioned factors may result in frequent interruptions in data transmissions, especially when downloading files of large sizes from the Internet, e.g., in-car entertainment services. Therefore, V2I and V2V communications have to co-exist and complement each other to improve the network performance.

Capacity is one of the most important performance metrics in vehicular networks. Capacity of vehicular networks with infrastructure support is both an interesting and challenging problem as the capacity is determined by the inter-play of multiple factors including V2I communications, V2V communications, density and mobility of vehicles, and cooperation among vehicles and infrastructure. Since the seminal work of Gupta and Kumar [8], extensive research on capacity has been conducted, e.g., [9], [10]. Focusing on the capacity of vehicular networks, Wang et al. [11] have studied urban vehicular networks with uniformly distributed RSUs and have analyzed

the asymptotic uplink throughput scaling law when the total number of vehicles is sufficiently large. In [12], Huang et al. introduce a Euclidean planar graph and use a practical geometric structure to study the asymptotic capacity of urban Vehicular Ad Hoc networks (VANETs). The aforementioned work all assume that the number of vehicles or vehicular density is sufficiently large and utilize asymptotic analysis to study the capacity scaling law, which is only applicable when the number of vehicles or vehicular density is sufficiently large. In our prior work [13], we have considered a vehicular network scenario where there is only one vehicle, termed vehicle-of-interest (VoI), with download request from the Internet and all other vehicles cooperate to assist the communication of the VoI, and analyze the achievable throughput of the VoI assuming a cooperative communication strategy.

In this paper, we consider a typical delay-tolerant application scenario with a subset of vehicles, termed Vehicles of Interest (VoIs), having download requests, e.g., videos, from the Internet. Each VoI downloads a distinct large-size file from the Internet and other vehicles without download requests, termed *helpers*, assist the delivery of the files to the VoIs. A cooperative communication strategy is proposed that explores the combined use of V2I communications, V2V communications, mobility of vehicles and cooperation among vehicles and infrastructure. Investigating the network performance of delay-tolerant application under the coexistence of V2I and V2V communications is important and has been extensively studied, e.g., [14], [15]. In this work, we are interested in investigating the capacity achievable by the VoIs. In addition to capacity, delay is also an important performance metric that has been extensively investigated, e.g., [16]–[18]. In the delay-tolerant network scenario considered in this paper, delay is not a major concern. Therefore, similar to most of the existing work investigating the network capacity [9], [11], [12], delay is not taken into consideration in this work and we focus on studying the long-term data rate achievable by the VoIs. An analytical framework is developed to model the data dissemination process using the proposed strategy, and a closed form expression of the capacity is obtained, which reveals the relationship between the capacity and its major performance-impacting parameters such as inter-infrastructure distance, radio ranges of infrastructure and vehicles, sensing range of vehicles, transmission rates of V2I and V2V communications, vehicular density and the proportion of VoIs. Different from the single-VoI scenario studied in [13], when there are multiple vehicles with download requests, the possible contention and collision among vehicles in vehicular communications become both important and challenging issue to study. Furthermore, the work presented in this paper distinguishes from previous work [11], [12] in that we focus on an accurate analysis (versus asymptotic analysis) of the capacity of vehicular networks with a moderate vehicular density and explore the combined use of V2I communications, V2V communications, mobility of vehicles and cooperation among vehicles and infrastructure to improve the achievable capacity of vehicular network, whereas the results obtained in previous work [11], [12] are only applicable when the number of vehicles or vehicular density is very large.

The following contributions are made in the paper:

- 1) We propose a novel cooperative communication strategy, which utilizes V2I communications, V2V communications, mobility of vehicles, and cooperation among vehicles and infrastructure to boost capacity of vehicular networks;
- 2) We develop an analytical framework to model and investigate the data dissemination process assuming the aforementioned cooperative communication strategy, and a closed-form expression of the capacity achieved by the VoIs in a vehicular network with a finite vehicular density is obtained, which reveals the relationship between the capacity and its major performance-impacting parameters;
- 3) Both simulations and numerical analysis are conducted, which show that the proposed cooperative strategy significantly improves the capacity of vehicular networks, compared with its non-cooperative counterpart, especially when the proportion of VoIs is small.

Our results shed light on the optimum deployment of vehicular network infrastructure in terms of their interval distance, and the optimum design of cooperative communication strategy to improve the capacity of vehicular networks.

The rest of this paper is organized as follows: Section II reviews related work. Section III introduces the system model, the proposed cooperative communication strategy and the problem formation. Theoretical analysis are provided in Section IV. In Section V, we validate the analytical result using simulations and conduct numerical analysis to discuss our result and its insight. Section VI concludes this paper.

## II. RELATED WORK

Extensive work in the literature investigated the performance of vehicular networks, measured by the information propagation speed [7], transmission delay [16]–[18], downloaded data volume [19], packet reception rate [20], etc. Among the major techniques to enhance these performance measures, cooperative communications, including cooperation among vehicles [18]–[21], cooperation among infrastructure points [2], [22]–[24], and cooperation among both vehicles and infrastructure points [15], [25], [26], stands out as a popular and important technique. In the following, we review work closely related to the work in this paper.

The following work investigate cooperative communications among vehicles in vehicular networks. In [18], Zhu et al. have studied using multiple nearby vehicles to collaboratively download data from a RSU and analyzed the average download time. In [19], Zhou et al. have used a cluster of vehicles on the highway to cooperatively download the same file from the infrastructure to enhance the probability of successful download. In [21], Liu et al. have utilized the location information of each vehicle and cooperation among vehicles to maximize the number of vehicles that successfully retrieve their requested data in vehicular networks. In [20], Das et al. have introduced a coalitional graph game to model cooperative message sharing among vehicles in vehicular networks and propose a coalition formation algorithm to improve the packet reception rate and reduce transmission delay.

Cooperation among infrastructure points can be achieved by caching different files into different infrastructure points to help moving vehicles download from the Internet. In [22], Zhang and Yeo have proposed a cooperative content distribution strategy for vehicles by prefetching different data into some selected APs, so that vehicles can obtain the complete data from those selected APs when traveling through their coverage areas. In [2], Li et al. have proposed a heuristic content distribution algorithm that caches data in different infrastructure points, to maximize the downloaded data size. In [23] and [24], the authors utilized cooperative infrastructure points in vehicular networks to maximize the success probability of download, utilizing greedy algorithm and integer linear programming optimization respectively.

Studies considering both vehicular cooperation and infrastructure cooperation are comparatively scarce. By exploring cooperation among vehicles and inter-connected infrastructure points, Mershad et al. [25] have designed an optimum routing algorithm to reduce end-to-end delay for delivering a packet from a source to its destination; and Si et al. [15] have designed an optimum distributed data hopping mechanism to enable delay-tolerant data routing over a vehicular network. In [26], Wang et al. have proposed a scheme that utilizes moving vehicles to serve as relays to assist data dissemination to a target vehicle, and the relay selection is conducted by the cooperative infrastructure points. They focus on reducing the transmission outage of the target vehicle.

In this paper, we propose a cooperative communication strategy that explores the combined use of V2I communications, V2V communications, mobility of vehicles and cooperation among vehicles and infrastructure to improve the capacity of vehicular network, and analyze the data dissemination process and the capacity of the network.

### III. SYSTEM MODEL AND PROBLEM FORMATION

In this section, we introduce the system model and assumptions used in the analysis, and also give a rigorous definition of the problem studied in the paper.

Specially, we consider a scenario where some VoIs (with proportion  $0 < p < 1$ ) want to download large files, e.g., videos, from a remote server and the file to be downloaded by different VoI is different.

#### A. Network Model

We consider a bi-directional highway segment with length  $L$  where roadside infrastructure, e.g., RSUs, Wi-Fi APs or LTE base stations, are uniformly deployed along the highway and separated by equal distance  $d, d \ll L$ . The width of a lane is typically small compared with the transmission range of vehicles. Therefore, we ignore the road width and model multiple lanes in the same direction as one lane [7], [27]. Even though the analysis is based on a straight linear highway, our obtained result can be readily applied to each separate road in 2D/3D scenarios, e.g., grid 2D network [17] and 3D multi-level network [28], and the network capacity can be readily achieved by adding up the capacity achieved from each separate road. We further assume that all infrastructure points are

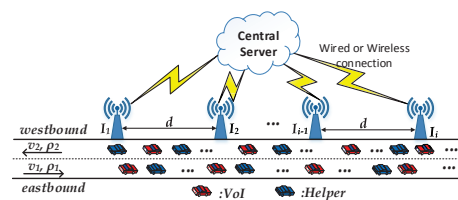


Fig. 1. An illustration of the system model for a bi-directional highway with infrastructure regularly deployed with equal distance  $d$ . The density and speed of vehicles in each direction are  $\rho_1, v_1$  and  $\rho_2, v_2$  respectively.

connected to the Internet through wired or wireless backbone with much larger capacity than the vehicular network.

We adopt a widely used traffic model in highway [7], [27], [29] that the distribution of eastbound and westbound vehicles follows a homogeneous Poisson process with densities  $\rho_1$  and  $\rho_2$  respectively. It follows that the inter-vehicle distances in each direction are exponentially distributed. This exponential inter-vehicle spacing distribution has been supported by some empirical study that it can accurately characterize real traffic distribution when the traffic density is low or medium [27]. For high traffic density however, the Poisson model may no longer be accurate. Furthermore, as a ready consequence of the superposition property of Poisson processes [30], all vehicles on the highway are also Poissonly distributed with density  $\rho = \rho_1 + \rho_2$ . We assume that the proportion of VoIs travel towards each direction is  $p$  ( $0 < p < 1$ ). Therefore, VoIs and helpers respectively have traffic density  $p\rho$  and  $(1-p)\rho$ . Moreover, we assume that eastbound and westbound vehicles travel at a constant speed of  $v_1$  and  $v_2$  respectively. In reality, individual vehicular speed may deviate from the mean speed. We will show later in the analysis that the capacity achieved in this work does not depend on the speed distribution of vehicles, and our analysis also applies to other time-varying speed model, e.g., Gaussian speed model [7], [31], as long as the resulting spatial distribution of vehicles is stationary. The system model is illustrated in Fig. 1.

#### B. Wireless Communication Model

Both V2I and V2V communications are considered. All infrastructure points and vehicles are assumed to have the same radio range, denoted by  $r_I$  and  $r_0$  respectively. A pair of vehicles (or vehicle and infrastructure) can directly communicate with each other if and only if (iff) their Euclidean distance is not larger than the radio range  $r_0$  (or  $r_I$ ) [16], [32]. There are other more realistic and intricate connection models, e.g., the SINR connection model [17], which implies the higher data rate or the larger probability to successfully transmit a packet when the devices are close, the easier occur a wireless connection. Our adopted simplified unit disk model is a special case of the commonly-used random connection model in wireless network [33], [34], where a receiver separated by a Euclidean distance  $x$  from a transmitter receives the message successfully with a probability  $g(x)$ , independent of transmissions by other transmitter-receiver pairs, and  $g(x)$  is a monotonic non-increasing function of  $x$ . This random wireless connection model grossly captures the fact that wireless communications

only occur between nearby devices, namely, the closer two devices become, the easier (with higher probability) it is for them to establish a connection.

We consider that each vehicle has a single antenna and they cannot transmit and receive at the same time [10], [35]. The antenna is omnidirectional so that the signal transmitted by a vehicle is able to reach all the potential receivers within its coverage. It has been shown in [8] that whether adopting Half-duplex that does not split the channel resources, or adopting the Frequency division duplex (FDD) that would break up the channel into two sub-channels to allow simultaneous transmitting and receiving, does not affect capacity calculation. Besides, we consider a unicast scenario and assume that each infrastructure (or vehicle) can only transmit information to one vehicle at a time. Both broadcast (or multicast) and unicast are important in vehicular networks [36]. For some safety-related applications, e.g., disseminating a message about an accident on the road, or when some vehicles request a common content, it is better to use broadcast (or multicast) to inform as many vehicles as possible. In addition to broadcast, unicast is also important and commonly adopted in vehicular networks to transmit data from a single source to a single destination [18], [37]. The scenario being considered in this paper corresponds to a unicast scenario because each VoI requests distinct content from the Internet. In addition, as the major focus of this work is to investigate the impact of the topological aspects of the vehicular network on the achievable capacity, similar as [16], [27], [28], we ignore the packet loss issue. We refer readers to [38], [39] for relevant work on packet loss.

We further assume that V2I and V2V communications are allocated different channels so that there is no mutual interference between them. For V2V communications, CSMA media access control (MAC) protocol is adopted with sensing range  $R_c$ . Moreover, we assume V2I and V2V communicate at a constant data rate  $w_I$  and  $w_V$  respectively [8], [9]. This simplified channel model allows us to omit physical layer details and focus on the topological impact of vehicular networks on the capacity. We will show later in the simulation that for time-varying channels, the values of  $w_I$  and  $w_V$  can be replaced by the respective time-averaged data rate of V2I and V2V communications and our analysis still applies.

We consider a V2I transmitting scheme that infrastructure will transmit its data to VoIs first, i.e., helpers can receive data from infrastructure only when there is no VoI within the coverage of infrastructure. This V2I scheduling scheme makes the VoIs achieve the maximum data rate from infrastructure. For V2V communications, helpers function as transmitters and VoIs as receivers. A transmitter can choose a receiver from either direction within its transmission range. We limit both V2I and V2V communications to one-hop. Allowing multi-hop communications is expected to have marginal improvement on the achievable capacity, which has been verified by our simulation result and the reason is discussed later in the simulation section.

### C. Cooperative Communication Strategy

Now we introduce the cooperative communication strategy considered in this paper. As mentioned previously, we consider

a scenario where some VoIs (with proportion  $0 < p < 1$ ) want to download large files, e.g., videos, from a remote server and the file to be downloaded by different VoI is different. Each requested large file by the VoIs may be first split into multiple pieces and transmitted to different infrastructure points such that each infrastructure point has a different piece of that file, which enables cooperation among infrastructure. Each piece of data delivered to infrastructure may be further split and transmitted either directly to the corresponding VoI requesting it or to helpers when they move into its coverage so that VoIs and helpers have different pieces of data. Each helper may store data for different VoIs. We assume there is a central server that has full knowledge of the network topology and data transmission process to guarantee that the data the helpers receive from infrastructure is the data required by the VoIs they will encounter. This assumption, which may cause large wireless communication overhead, is required to establish the maximum data rate, i.e., capacity, that can be achieved by the VoIs because it assumes a perfect scheduling of data items for both V2I and V2V communications. Some practical issues like out of sequence data delivery and missing packets can be handled by techniques such as network coding (e.g., our previous paper [40]) so that we can focus on the main theme of the paper without the need for considering their impacts. Therefore, when the VoIs are in the coverage of infrastructure, they receive data directly from the infrastructure. In the meantime, the helpers may also receive different pieces of data from the infrastructure when they obtain access to the infrastructure. When the VoIs move outside the coverage of infrastructure, they may continue to receive data from helpers, exploiting the mobility of vehicles and V2V communications. In this way, V2I communications between the VoIs and infrastructure, between helpers and infrastructure, V2V communications between the VoIs and helpers, cooperations among infrastructure and among vehicles, as well as vehicular mobility are coherently combined to maximize the capacity of the VoIs. We do not consider the case that VoIs share their received content with others during V2V communication as it does not improve the capacity considering the fact that different VoIs request different files in our considered scenario, and the fact that a larger number of relay nodes, which leads to a larger number of hops between the source and destination, is detrimental to the achievable capacity [9]. Even though the proposed cooperative communication scheme is simple, all the major topological parameters have been taken into consideration.

### D. Problem Formation

Now we give a formal definition of the capacity considered in this paper. Consider an arbitrarily chosen time interval  $[0, t]$  and denote the amount of data received by all VoIs as  $D^\chi(t)$  during this time interval, which includes data received both directly from infrastructure and indirectly from helpers. The superscript  $\chi \in \Phi$  denotes a scheduling algorithm used to schedule V2I and V2V communications and  $\Phi$  denotes the set of all scheduling algorithms. In this paper, we are interested in finding the maximum average data rate, i.e., capacity, achieved

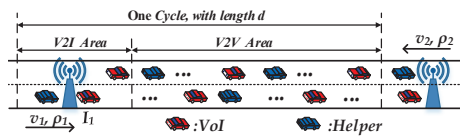


Fig. 2. An illustration of one cycle, which includes V2I Area and V2V Area.

by the VoIs using our cooperative communication strategy, denoted by  $\eta$ , which is mathematically defined as follows:

$$\eta = \max_{\chi \in \Phi} \eta^\chi = \max_{\chi \in \Phi} \lim_{t \rightarrow \infty} \frac{D^\chi(t)}{t}. \quad (1)$$

#### IV. ANALYSIS OF THE CAPACITY

In this section, we will give detailed analysis of the achievable capacity by the VoIs, including analyzing the capacity achieved directly from infrastructure through V2I communications and the capacity achieved indirectly from helpers through V2V communications.

We define the area covered by one infrastructure point (termed *V2I Area*) and the adjacent area between two consecutive infrastructure points but not covered by the infrastructure point (termed *V2V Area*) as a *cycle*, which has length  $d$ . See Fig. 2 for an illustration. It follows from the renewal theory [41] that the long-term achievable capacity by the VoIs from each cycle, denoted by  $\eta_{\text{cycle}}$ , is identical and the total capacity achieved in a given highway segment with length  $L \gg d$  can be readily calculated by (ignoring the trivial fact that  $\frac{L}{d}$  may not be an integer):

$$\eta = \frac{L}{d} \eta_{\text{cycle}}. \quad (2)$$

From (2), to calculate the total capacity achievable by the VoIs from a highway segment with length  $L$ , it suffices to calculate the capacity achieved by the VoIs from one cycle, which includes capacity achieved both from V2I communications and V2V communications, given as follows:

$$\eta_{\text{cycle}} = \lim_{t \rightarrow \infty} \frac{D_{V2I}(t)}{t} + \lim_{t \rightarrow \infty} \frac{D_{V2V}(t)}{t}, \quad (3)$$

where  $D_{V2I}(t)$  and  $D_{V2V}(t)$  are respectively the maximum expected amount of data received by the VoIs from infrastructure in the V2I Area and from helpers in the V2V Area during time period  $t$ . In the following, we will focus on studying one cycle entirely contained within the highway segment of length  $L$ , termed the *cycle of interest*. We will first calculate the two terms on the right hand side of (3) separately, and then combine both terms to obtain the final expression of the achievable capacity.

##### A. Capacity achieved by VoIs from V2I communications

Without loss of generality, we call the infrastructure point located in our cycle of interest  $I_1$ . We assume that time is divided into time slots of equal length  $\Delta t$ , and  $\Delta t$  is sufficiently small that we can approximately regard vehicles as stationary during each time slot. Denote by  $q_1(i)$ ,  $i = 1, 2, \dots$  a discrete random variable representing the fraction of time

that VoIs' V2I communication with  $I_1$  happens during the  $i$ th time slot  $[(i-1)\Delta t, i\Delta t)$ . Recall that to make the VoIs achieve the maximum data rate from the infrastructure, we adopt a V2I communication scheme that infrastructure delivers its data directly to the VoIs as long as there are VoIs within its coverage. Therefore,  $q_1(i)$  is equal to 1 when there exist at least one VoIs within the coverage of  $I_1$  during the  $i$ th time slot; otherwise it is equal to 0. This follows that the maximum expected amount of data the VoIs can obtain from one cycle through V2I communications during time period  $[0, t]$  can be calculated by (ignoring the trivial fact that  $\frac{t}{\Delta t}$  may not be an integer):

$$D_{V2I}(t) = \lim_{\Delta t \rightarrow 0} E \left( w_I \sum_{i=1}^{t/\Delta t} q_1(i) \Delta t \right) \quad (4)$$

According to the ergodicity and stationarity properties of homogeneous Poisson point process [42], the time average of  $q_1(i)$  is equal to the probability that there is at least one VoI within the coverage of  $I_1$  at a randomly chosen time slot, denoted by  $\bar{q}_1$ . Considering the Poisson distribution of vehicles, it can be readily shown that  $\bar{q}_1 = 1 - e^{-\rho \rho_1 2r_I}$ . Therefore, we have

$$\lim_{t \rightarrow \infty} \frac{\lim_{\Delta t \rightarrow 0} E \left( \sum_{i=1}^{t/\Delta t} q_1(i) \Delta t \right)}{t} = \bar{q}_1 = 1 - e^{-\rho \rho_1 2r_I}. \quad (5)$$

Combing (4)-(5), we have the long-term capacity achieved by the VoIs from one cycle through their V2I communications:

$$\lim_{t \rightarrow \infty} \frac{D_{V2I}(t)}{t} = w_I (1 - e^{-\rho \rho_1 2r_I}). \quad (6)$$

##### B. Capacity achieved by VoIs from V2V communications

Note that the data received by the VoIs from helpers through V2V communications eventually comes from the data received by the helpers from infrastructure during their V2I communications. Therefore, the amount of data the VoIs can receive from V2V communications during time period  $[0, t]$ , on one hand, is constrained by how much data the helpers can receive via their V2I communications during time period  $[0, t]$ ; on the other hand, is limited by how much data the helpers can transmit to the VoIs through V2V communications during time period  $t$ . Taking the above two constraints into account, we have the following results:

**Theorem 1.** *The capacity the VoIs can achieve through V2V communications from one cycle is given by:*

$$\lim_{t \rightarrow \infty} \frac{D_{V2V}(t)}{t} = \min \left\{ \lim_{t \rightarrow \infty} \frac{D_{I-H}(t)}{t}, \lim_{t \rightarrow \infty} \frac{D_V(t)}{t} \right\}, \quad (7)$$

where  $D_{I-H}(t)$  is the expected amount of data received by helpers from one cycle through their V2I communications during time period  $[0, t]$  under the optimum scheme, and  $D_V(t)$  is the maximum expected amount of data the helpers can deliver to the VoIs through V2V communications in the V2V Area during time period  $[0, t]$  without considering the limitation of the amount of data received by helpers from the infrastructure.



Since the bottleneck is either in the V2V communications between VoIs and helpers, represented by  $\lim_{t \rightarrow \infty} \frac{D_V(t)}{t}$ , or in the V2I communications between helpers and infrastructure, represented by  $\lim_{t \rightarrow \infty} \frac{D_{I-H}(t)}{t}$ , the proof of Theorem 1 follows readily. More specifically, imagine the V2V communication process between helpers and VoIs as a single-queue queuing system. The rate the helpers receive data from infrastructure,  $\lim_{t \rightarrow \infty} \frac{D_{I-H}(t)}{t}$ , is equivalent to the incoming rate of the queue. The rate helpers deliver data to the VoIs,  $\lim_{t \rightarrow \infty} \frac{D_V(t)}{t}$ , is equivalent to the processing speed of the queue. The outgoing rate of the queue,  $\lim_{t \rightarrow \infty} \frac{D_{V2V}(t)}{t}$ , is equal to either the incoming rate or the processing speed.

From Theorem 1, to obtain the capacity achieved by the VoIs through V2V communications with the helpers from one cycle, it remains to calculate the long-term data rate achieved by the helpers from one cycle through their V2I communications,  $\lim_{t \rightarrow \infty} \frac{D_{I-H}(t)}{t}$ , and the maximum long-term data rate the VoIs can achieve through V2V communications from one cycle without considering the limitation of the amount of data the helpers received,  $\lim_{t \rightarrow \infty} \frac{D_V(t)}{t}$ . In the following, we will calculate these two terms separately.

1) *Calculation of  $\lim_{t \rightarrow \infty} \frac{D_{I-H}(t)}{t}$* : Denote by  $q_2(i)$  a discrete random variable, which is equal to 1 when helpers' V2I communication happens during the  $i$ -th time slot  $[(i-1)\Delta t, i\Delta t)$ ,  $i = 1, 2, \dots$ , otherwise it is equal to 0. Similar to the analysis in section IV-A, we have

$$\lim_{t \rightarrow \infty} \frac{D_{I-H}(t)}{t} = \lim_{t \rightarrow \infty} \frac{w_I \lim_{\Delta t \rightarrow 0} E \left( \sum_{i=1}^{t/\Delta t} q_2(i) \Delta t \right)}{t} = w_I \bar{q}_2, \quad (8)$$

where  $\bar{q}_2$  is the probability that helpers' V2I communication happens at a randomly chosen time slot. Note that an infrastructure point only delivers its data to helpers when both of the following conditions are met: (i) there is no VoI within its coverage, and (ii) there is at least one helpers within its coverage. Thus,  $\bar{q}_2$  can be readily calculated by using the Poisson distribution of the VoIs and the helpers:

$$\bar{q}_2 = e^{-p\rho 2r_I} \left( 1 - e^{-(1-p)\rho 2r_I} \right) = e^{-p\rho 2r_I} - e^{-\rho 2r_I}. \quad (9)$$

Combing (8) and (9), we have:

$$\lim_{t \rightarrow \infty} \frac{D_{I-H}(t)}{t} = w_I \bar{q}_2 = w_I \left( e^{-p\rho 2r_I} - e^{-\rho 2r_I} \right). \quad (10)$$

2) *Calculation of  $\lim_{t \rightarrow \infty} \frac{D_V(t)}{t}$* : In this subsection, we analyze the maximum data rate achieved by the VoIs through V2V communications from one cycle area without considering the amount of data each helper has.

Recall that for V2V communications, we adopt CSMA multiple access protocol with sensing range  $R_c$ . Therefore, a helper within the V2V Area can potentially be chosen as one of the simultaneously transmitters when there is no other helper transmitting within its sensing range and there is at least one VoIs within its transmission range. We call a helper, together with the VoI that the helper transmits to, an *active helper-VoI pair* iff this helper is chosen as a transmitter and chooses this VoI within its transmission range as its receiver.

Denote by  $N_p^\chi(i)$  the number of simultaneous helper-VoI pairs in the V2V Area during the  $i$ -th time slot  $[(i-1)\Delta t, i\Delta t)$ , where the superscript  $\chi \in \Phi$  denotes the scheduling algorithm that selects the simultaneously active helper-VoI pairs and  $\Phi$  denotes the set of all scheduling algorithms. Using the analysis in [9], it can be shown that:

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{D_V(t)}{t} &= \max_{\chi \in \Phi} \lim_{t \rightarrow \infty} \frac{D_V^\chi(t)}{t} \\ &= \max_{\chi \in \Phi} \lim_{t \rightarrow \infty} \frac{w_V \lim_{\Delta t \rightarrow 0} E \left( \sum_{i=1}^{t/\Delta t} N_p^\chi(i) \Delta t \right)}{t} \\ &= \max_{\chi \in \Phi} w_V E [N_p^\chi], \end{aligned} \quad (11)$$

where  $E [N_p^\chi]$  is the expected number of simultaneously active helper-VoI pairs in the V2V Area at a randomly chosen time slot assuming scheduling algorithm  $\chi$ .

From (11), the maximum value of  $\lim_{t \rightarrow \infty} \frac{D_V(t)}{t}$  is achieved when using an optimum V2V scheduling algorithm that schedules as many active helper-VoI pairs as possible. Therefore, in the following analysis, we shall establish the optimum V2V scheduling algorithm and the maximum  $E [N_p^\chi]$  that can be achieved by the algorithm. Specifically, we will first find an optimum V2V scheduling scheme, denoted by  $\chi_{opt}$ , that leads to the maximum number of simultaneously active helper-VoI pairs, and then calculate  $E [N_p^{\chi_{opt}}]$  under this optimum algorithm. Without loss of generality, we designate the left boundary point of the V2V Area, i.e., the point to the right of infrastructure point  $I_1$  and at a distance  $r_I$  to  $I_1$ , as the origin of the coordinate system, and the east (right) direction as positive ( $+x$ ) direction. The following theorem summarizes the optimum scheduling scheme.

**Theorem 2.** *An optimum scheduling scheme  $\chi_{opt}$ , which leads to the maximum number of simultaneously active helper-VoI pairs in V2V Area is as follows: select active helper-VoI pairs in order from left to the right. First, choose the first helper to the right of the origin that has at least one VoIs within its coverage as the first transmitter, and the left-most VoI within the coverage of that helper as its receiver. The next transmitter is the nearest helper to the current transmitter, and satisfies the following conditions: 1) the distance between this helper and the current transmitter is no smaller than  $R_c$ ; 2) it can find at least one VoIs within its coverage, which is different from the receiver of the current transmitter. If there are multiple VoIs, always chooses the leftmost VoI. Repeat the above process until the rightmost border of the V2V Area is reached.*

*Proof:* See Appendix A. ■

**Remark 3.** Note that the optimum scheduling algorithm that achieves the maximum number of active helper-VoI pairs may not be unique.

Now we calculate the maximum expected number of helper-VoI pairs,  $E [N_p^{\chi_{opt}}]$ , and the corresponding value of  $\lim_{t \rightarrow \infty} \frac{D_V(t)}{t}$  under the optimum scheme  $\chi_{opt}$ .

Denote by  $S_k \in [0, d - 2r_I]$ ,  $k = 1, 2, \dots$  the position of the  $k$ -th transmitter (helper in the active helper-VoI pair) under the optimum scheduling scheme  $\chi_{opt}$ . Denote by  $L_k$ ,  $k = 1, 2, \dots$  the distance between the  $k$ -th and the  $(k+1)$ -th transmitter,

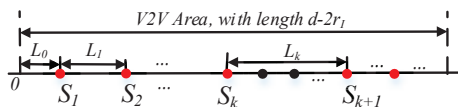


Fig. 3. An illustration of the distribution of distances between two consecutive simultaneous transmitters.

and  $L_0$  the distance between the first transmitter and the origin. See Fig. 3 for an illustration. It is straightforward that  $L_k, k = 1, 2, \dots$  are identically and independently distributed (i.i.d.) and are also independent of  $L_0$ . Note that the distribution of  $L_0$  is not the same as that of  $L_k, k = 1, 2, \dots$ . This is due to the fact that when selecting the first transmitter assuming  $\chi_{opt}$ , we directly choose the first (leftmost) helper within  $[0, d-2r_I]$  that has at least one VoI within its coverage. In contrast, when  $k > 1$ , we select the  $k$ -th transmitter from helpers located within  $[S_{k-1} + R_c, d - 2r_I]$ , which imposes an additional condition that  $L_k > R_c$ . It can be shown that the expected number of helper-VoI pairs,  $E[N_p^{\chi_{opt}}]$ , is exactly the expected number of renewals of a renewal counting process in the V2V Area, with a delay length of  $E[L_0]$  (as  $L_0$  has a different distribution from  $L_k, k = 1, 2, \dots$ ) and each renewal has an average length  $E[L_k], k = 1, 2, \dots$ . Using the renewal theory [41],  $E[N_p^{\chi_{opt}}]$  can be calculated as follows:

$$E[N_p^{\chi_{opt}}] = \sum_{n=1}^{\infty} Pr(\sum_{i=0}^{n-1} L_i \leq d - 2r_I), \quad (12)$$

An alternative way of obtaining (12) is by noting that the inner term of (12), i.e.,  $Pr(\sum_{i=0}^{n-1} L_i \leq d - 2r_I)$ , gives the cumulative distribution function (cdf) of the maximum number of simultaneously active helper-VoI pairs assuming the optimum scheduling  $\chi_{opt}$ . It then follows that the summation of the cdf gives the expected value of  $N_p^{\chi_{opt}}$ .

Equation (12) shows that to calculate  $E[N_p^{\chi_{opt}}]$ , we first need to calculate the distribution of each  $L_k, k = 0, 1, 2, \dots$ . The following two theorems characterize the probability density function (pdf) of  $L_k$ , denoted by  $f_{L_k}(x), k = 1, 2, \dots$  and the pdf of  $L_0$ , denoted by  $f_{L_0}(x)$  respectively.

**Theorem 4.** Consider a bi-directional vehicular network with vehicular densities  $\rho_1$  and  $\rho_2$ , with  $p$  percentage of vehicles being the VoIs and the remaining  $1 - p$  percentage of vehicles being the helpers. Furthermore, each vehicles' radio range is  $r_0$  and sensing range is  $R_c$ . The distance between two consecutive transmitters (helpers in two consecutive active helper-VoI pairs),  $L_k, k \geq 1$ , under the optimum scheduling scheme  $\chi_{opt}$ , has the pdf as follows:

$$f_{L_k}(x) = \begin{cases} \sum_{m=1}^{\infty} f(x - R_c; m, (1-p)\rho) Pr(m_k = m), & x \geq R_c \\ 0 & x < R_c \end{cases}, \quad (13)$$

where  $f(x; k, \alpha) = \frac{\alpha^k x^{k-1} e^{-\alpha x}}{(k-1)!}$  is the pdf of Erlang distribution with shape parameter  $k$  and rate parameter  $\alpha$ , and  $m_k$  is the (random) number of helpers within

$[S_k + R_c, \min\{S_{k+1}, d - 2r_I\}]$ . The probability mass function of  $m_k$  is given by:

$$Pr(m_k = 1) \approx 1 - e^{-p\rho 2r_0}, \quad (14)$$

and for  $m \geq 2$ ,

$$Pr(m_k = m) \approx e^{-p\rho 2r_0} (p - pe^{-\rho 2r_0}) (1 - p + pe^{-\rho 2r_0})^{m-2}. \quad (15)$$

Moreover, when  $R_c \geq 2r_0$ , the approximation in (14) and (15) becomes accurate and can be replaced by equality.

*Proof:* See Appendix B. ■

**Theorem 5.** Under the same setting as that described in Theorem 4, the distance between the first transmitter and the origin under the optimum scheduling scheme  $\chi_{opt}$ ,  $L_0$ , has the pdf as follows:

$$f_{L_0}(x) = \sum_{m=1}^{\infty} f(x; m, (1-p)\rho) Pr(m_0 = m), \quad (16)$$

where  $m_0$  is the (random) number of helpers within  $(0, S_1)$  with distribution:  $Pr(m_0 = 1) = 1 - (1 - p + pe^{-\rho r_0}) e^{-p\rho r_0}$ ; and for  $m \geq 2$ ,

$$Pr(m_0 = m) = e^{-p\rho r_0} (1 - c_1) c_1^{m-2} (1 - p + pe^{-\rho r_0}),$$

where  $c_1 = 1 - p + pe^{-\rho 2r_0}$ .

*Proof:* The cdf of  $L_0$  can be derived using the same method as that used in the proof of Theorem 4. Particularly, we have  $L_0 = \sum_{i=0}^{m_0-1} l_{0,i}$ , and in this case, define  $h_{0,i} = \min\{l_{0,i}, 2r_0\}, i = 1, \dots, m_0 - 1$ ,  $h_{0,0} = \min\{l_{0,0}, r_0\} + r_0$ , and define  $H_{0,m} = \sum_{i=1}^{m-2} h_{0,i} + h_{0,0}$ . See proof of Theorem 4 for definitions of these parameters. The proof follows. ■

From (13) and (16), we can see that the pdf of  $L_0$  and  $L_k, k = 1, 2, \dots$  are both in the forms of rather complicated expressions. Accordingly, the computation of the distribution of  $\sum_{i=0}^{n-1} L_i, n = 1, 2, \dots$ , which is required for computing  $E[N_p^{\chi_{opt}}]$  and relies on the joint the distribution of  $L_k, n = 1, 2, \dots$ , can become even more intricate. In our case, assuming that  $d$  is much larger compared with the distance between two consecutive simultaneous transmitters, i.e., we have  $d \gg E[L_k]$ . It is then reasonably accurate to calculate the value of  $E[N_p^{\chi_{opt}}]$  approximately using the Elementary Delayed Renewal Theorem [41, Theorem 5.8.4], shown as follows:

$$E[N_p^{\chi_{opt}}] \approx \frac{d - 2r_I}{E[L_k]}, \quad (17)$$

where only the expected value of  $L_k, k = 1, 2, \dots$  is needed. In the following, we first calculate the expected value of  $L_k, k = 1, 2, \dots$ , and then use the obtained result of  $E[L_k]$  to calculate  $E[N_p^{\chi_{opt}}]$ .

According to the pdf of  $L_k$  provided in (13), the expectation of  $L_k, k = 1, 2, \dots$  can be readily calculated as follows:

$$\begin{aligned} E[L_k] &= E\left[R_c + \sum_{i=0}^{m_k-1} l_{k,i}\right] \\ &= R_c + \sum_{m=1}^{\infty} E\left[\sum_{i=0}^{m-1} l_{k,i}\right] Pr(m_k = m) \end{aligned}$$

$$\begin{aligned}
 &= R_c + \sum_{m=1}^{\infty} \frac{m}{(1-p)\rho} Pr(m_k = m) \\
 &= R_c + \frac{p - pe^{-\rho 2r_0} + e^{-p\rho 2r_0}}{(1-p)p\rho(1 - e^{-\rho 2r_0})}, \quad (18)
 \end{aligned}$$

where the first step results by using (24) and the second step is obtained by using the total probability theorem; the third step is obtained due to fact that  $\sum_{i=0}^{m-1} l_{k,i}$  is the sum of  $m$  i.i.d. exponential random variable with mean  $\frac{1}{(1-p)\rho}$ , which is independent of  $m_k$ ; and the last step results by plugging in  $Pr(m_k = m)$  shown as (35).

Putting (18) into (17) and simplifying it, we have

$$\begin{aligned}
 &E[N_p^{\chi_{opt}}] \\
 &= \frac{(1-p)p\rho(1 - e^{-\rho 2r_0})(d - 2r_I)}{(1-p)p\rho(1 - e^{-\rho 2r_0})R_c + p - pe^{-\rho 2r_0} + e^{-p\rho 2r_0}}. \quad (19)
 \end{aligned}$$

Combining (11) and (19), we have:

$$\begin{aligned}
 &\lim_{t \rightarrow \infty} \frac{D_V(t)}{t} \\
 &= \frac{w_V(1-p)p\rho(1 - e^{-\rho 2r_0})(d - 2r_I)}{(1-p)p\rho(1 - e^{-\rho 2r_0})R_c + p - pe^{-\rho 2r_0} + e^{-p\rho 2r_0}}. \quad (20)
 \end{aligned}$$

### C. Achievable capacity

In this subsection, we first give the final result of the capacity achieved by the VoIs by combing the results from Section IV-A and IV-B; then we analyze the capacity achieved by eastbound and westbound VoIs separately to demonstrate the relationship between capacity achieved by eastbound and westbound VoIs and their vehicular density.

1) *Total Achievable Capacity*: Combining the capacity achieved by the VoIs from V2I communications and V2V communications from one cycle with length  $d$  shown in Section IV-A and IV-B, the total capacity achieved by the VoIs from a highway segment with length  $L$  can be readily obtained as follows:

$$\begin{aligned}
 \eta &= \frac{L}{d} \lim_{t \rightarrow \infty} \frac{D_{V2I}(t) + D_{V2V}(t)}{t} \\
 &= \frac{L}{d} \min \left\{ w_I (1 - e^{-\rho 2r_I}), \right. \\
 &\quad \left. w_I (1 - e^{-p\rho 2r_I}) + \frac{w_V c_2 (d - 2r_I)}{c_2 R_c + p - pe^{-\rho 2r_0} + e^{-p\rho 2r_0}} \right\}, \quad (21)
 \end{aligned}$$

where  $c_2 = (1-p)p\rho(1 - e^{-\rho 2r_0})$ .

*Remark 6.* It is interesting to note from (21) that the achievable capacity does not depend on the speed of vehicles, which appears to be counter-intuitive at the first sight. This can be explained from the data dissemination process. As (6) and (11) show, both the capacity achieved by the VoIs from infrastructure and from helpers only depend on the spatial distribution of vehicles. In our system, the vehicles' arrival follows a Poisson process and the vehicles move at a constant speed. Therefore, the spatial distribution of the vehicles are both stationary and ergodic [42] (ignoring the finite length of the road segment  $L$ ). It follows that the capacity that can be achieved by the VoIs is independent of vehicular speed. This observation implies that when vehicles arrive following

a Poisson process, our analysis assuming the constant speed model is also applicable to other time-varying speed model, e.g., Gaussian speed model [7], as long as the resulting spatial distribution of vehicles is time-invariant, i.e., stationary.

*Remark 7.* In the extreme case when all vehicles have download requests, i.e., when  $p = 1$ , the achievable capacity from one cycle is  $\eta_{cycle} = \eta_{max} = w_I (1 - e^{-\rho 2r_I})$ , which is exactly the achievable capacity when all vehicles directly receive data from the infrastructure without cooperative V2V communications. It can be further established when  $p$  is greater than a certain threshold, cooperative V2V communications between the helpers and the VoIs are of little use in boosting the capacity. This can be explained by that in the particular scenario considered in the paper, all new data comes from outside the vehicular network. V2V communications between the helpers and the VoIs only help to extend the communication range of the VoIs when there is no VoI in the infrastructure's coverage and balance data among the VoIs, but cannot increase the net amount of data available in the vehicular network. Therefore, when the density of the VoIs is high, the probability that there is no VoI in the infrastructure's coverage is negligible and thus it is more beneficial for the VoIs to retrieve data directly from the infrastructure. In this situation, V2V cooperative communications offer little benefit in boosting the vehicular network capacity. Unsurprisingly, following the argument outlined earlier, when the vehicular density is sufficiently large, even for a small value of  $p$ , the benefit of V2V cooperative communications vanishes very quickly. This can be also validated using (21).

2) *Capacity achieved by eastbound and westbound VoIs*: In this subsection, we analyze the capacity achieved from one cycle by all eastbound and westbound VoIs separately, denoted by  $\eta_e$  and  $\eta_w$  respectively. The following theorem summarizes the results.

**Theorem 8.** *The capacities achieved from one cycle by the eastbound and westbound VoIs, are proportional to the traffic density of eastbound vehicles and westbound vehicles respectively, which are given by:*

$$\eta_e = \eta_{cycle} \cdot \frac{\rho_1}{\rho_1 + \rho_2}, \quad (22)$$

and

$$\eta_w = \eta_{cycle} \cdot \frac{\rho_2}{\rho_1 + \rho_2}. \quad (23)$$

*Proof:* See Appendix C. ■

*Remark 9.* It is interesting to note that the capacities achieved by VoIs traveling in each direction are strictly proportional to the vehicular densities of that direction respectively. This can be explained by that when the vehicular density in a direction increases, there is a higher chance for the VoIs travel in that direction to communicate with the infrastructure and to receive more data indirectly from the helpers via V2V communication. As a VoI travels in a direction is statistically indistinguishable from another VoI travels in the same direction, this result suggests that the achievable throughput by a (any) VoI, no matter which direction the VoI is traveling in, will be statistically the same.



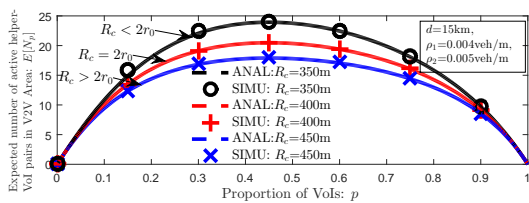


Fig. 4. A comparison of the expected number of simultaneously active helper-VoI pairs in one V2V Area with respect to the proportion of VoIs  $p$  between simulation and analysis, for different sensing ranges  $R_c$ .

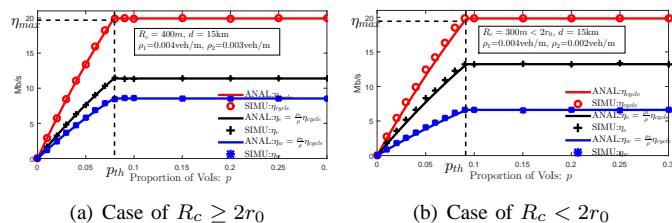


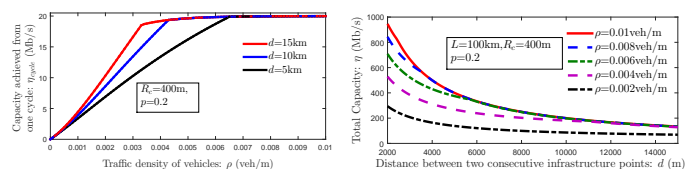
Fig. 5. A comparison of the capacities achieved from one cycle by all VoIs, by VoIs in the eastbound direction and westbound direction.

## V. SIMULATION AND DISCUSSION

In this section we conduct Monte-Carlo simulations to establish the accuracy of the theoretical analysis and discuss its insights. Specifically, we set the length of a highway segment  $L=100\text{km}$ . Eastbound and westbound vehicles move at constant speeds of  $v_1=20\text{m/s}$  and  $v_2=25\text{m/s}$  respectively. The radio ranges of infrastructure points and vehicles are  $400\text{m}$  and  $200\text{m}$  (typical radio ranges using DSRC [19]) respectively. The data rates of V2I and V2V communications are  $w_I=20\text{Mb/s}$  and  $w_V=2\text{Mb/s}$ . Each simulation is repeated 2000 times and the average value is shown in the plot.

Fig. 4 shows a comparison between the expected number of simultaneous active helper-VoI pairs in a V2V Area from analysis and simulation, under three different sensing ranges  $R_c$ . It is shown that the analytical result match perfectly with the simulation result when  $R_c \geq 2r_0$ . When  $R_c < 2r_0$ , there is a marginal gap between the simulation and analytical result, and the gap reduces with an increase of  $p$ . This is due to the fact that when  $R_c < 2r_0$ , other things being equal, with an increase of  $p$ , the density of helpers becomes smaller. Therefore, it is less likely to occur the scenario discussed in the proof of Theorem 4 that the VoI of the  $k$ -th ( $k \geq 1$ ) helper-VoI pair is located within the coverage of the helper  $V_{k,1}$ , helper  $V_{k,2}, \dots$ . Consequently, the approximations used in (14) and (15) become more accurate. Furthermore, as expected, a higher value of sensing range  $R_c$  results in a lower expected number of simultaneous helper-VoI pairs.

Fig. 5 shows a comparison of the capacities achieved from one cycle by all VoIs, by all eastbound VoIs, and by all westbound VoIs. Both Fig. 5(a) and Fig. 5(b) show that the analytical results match very well with simulations. Furthermore, it can be seen that the capacity achieved by the VoIs traveling towards each direction ( $\eta_e$  and  $\eta_w$  respectively) is exactly proportional to the traffic density in that direction, as predicted by Theorem 8.



(a) Achievable capacity from one cycle under different  $d$ . (b) Total achievable capacity under different traffic density.

Fig. 6. Relationship between capacity, distance between adjacent infrastructure points and vehicular density.

Fig. 5 also reveals the relationship between the capacity and the proportion of VoIs  $p$ , and shows that the capacity increases to its maximum value when the proportion of VoIs is larger than a certain threshold. Beyond that threshold, a further increase in  $p$  has little impact on the capacity. Specifically, as shown in Fig. 5(a), when  $p$  is small, the capacity increases sharply with an increase of  $p$ ; however, when  $p$  increases beyond a certain threshold, e.g.,  $p_{th}=0.08$  in this case, a further increase in  $p$  has no impact on the capacity. This can be explained by that when  $p < p_{th}$ , the number of VoIs is insufficient to retrieve all the data received by the helpers from their V2I communications. That is, vehicular networks offer more data (capacity) than that can be retrieved by the VoIs and the capacity is limited by the V2V communications between the VoIs and the helpers. Therefore, an increase in  $p$  would significantly increase the number of simultaneous active helper-VoI pairs and consequently boost the capacity. However, when the proportion of VoIs reaches a certain threshold, VoIs can retrieve almost all the data received by the helpers from their V2I communications. In this case, the capacity achieved by the VoIs approaches its maximum  $\eta_{max} = w_I(1 - e^{-\rho 2r_0 t})$ , which is equal to the average data rate the infrastructure point delivers its data to all vehicles, including both VoIs and helpers.

Fig. 6 demonstrates the relationship between the capacity and the inter-infrastructure distance, and gives insight into the optimum vehicular network infrastructure deployment in terms of their inter-distances under different vehicular density. It is shown in Fig. 6(a) that when the vehicular density  $\rho$  is small, a larger  $d$  will lead to a larger capacity achieved from one cycle because a large  $d$  will increase the capacity achieved by the VoIs from V2V communications with helpers. However with an increase of  $\rho$ , the capacity achieved by the VoIs from one cycle under different values of  $d$  differ marginally and converge to the same maximum value. This can be explained by that when  $\rho$  is large, most of the VoIs can receive data directly from the infrastructure, and the contribution from V2V communications with helpers becomes less significant. Even though an increase in  $d$  would help to boost the capacity achieved from one cycle when the vehicular density is small, Fig. 6(b) shows that the total achievable capacity decreases with an increase of  $d$ . This is due to the fact that an increase in  $d$  on one hand brings marginal improvement on the capacity achieved from one cycle, on the other hand, it reduces the number of cycles, which consequently leads to a reduction in the total capacity. Furthermore, it can be

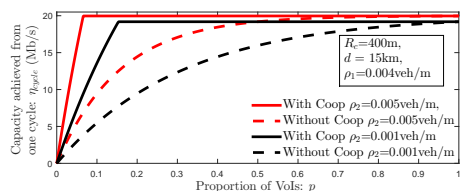


Fig. 7. A comparison between the capacity achieved from one cycle, with and without cooperative communications.

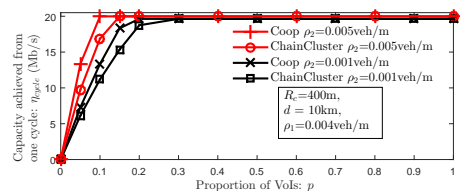


Fig. 8. A comparison between the capacity achieved from one cycle assuming our proposed strategy and that assuming the strategy proposed in [26].

seen that to achieve the same capacity, when the vehicular density is larger, the inter-infrastructure distance needs to be higher. Therefore, when determining the optimum deployment of vehicular network infrastructure, it is important to take the vehicular density into account, e.g., in areas where the vehicular density is usually large, by utilizing a cooperative communication strategy, the number of infrastructure points can be reduced.

Fig. 7 compares the capacity from one cycle using our cooperative communication strategy (labeled as With Coop) with its non-cooperative counterpart (labeled as Without Coop), and shows that our cooperative communication strategy can improve the capacity, even when there is only a small proportion of vehicles with download requests, i.e., a small  $p$ . The result for the non-cooperative counterpart is obtained by letting the VoIs only receive data from infrastructure. It is shown that with an increase in  $p$ , the proposed cooperative communication strategy becomes less effective in improving the capacity. This is due to the fact that a larger  $p$  leads to a smaller number of helpers, which results in a reduction in the amount of data the helpers can help to retrieve from the infrastructure. Thus, the contribution to the capacity from the proposed cooperative communication strategy becomes less significant. Furthermore, we can see that under the same network setting, without using the cooperative communication strategy, only when all vehicles have download requests, i.e.,  $p = 1$ , the maximum capacity  $\eta_{max} = w_I(1 - e^{-\rho^2 r_I})$  can be achieved. In contrast, with the cooperative communication strategy, this maximum capacity  $\eta_{max}$  can be achieved even when a small proportion of vehicles have download requests. This validates the effectiveness of cooperative communications to boost network performance.

Fig. 8 compares the achievable capacity assuming our proposed cooperative communication strategy (labeled as Coop) with that assuming the strategy proposed in [19] (labeled as ChainCluster). Specifically, the strategy proposed in [19] utilized vehicles moving in the same direction as the target vehicles (VoIs) to form clusters to help the VoIs' download.

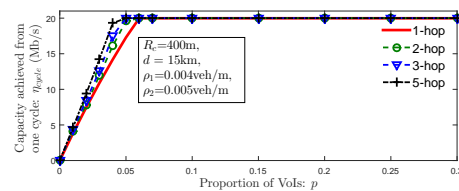


Fig. 9. A comparison between capacity achieved from one cycle when allowing one-hop communication and multi-hop communications.

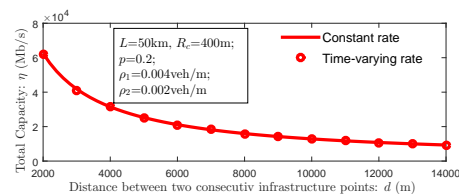


Fig. 10. A comparison between capacity achieved from constant channel and time-varying channel which considering Rayleigh fading and path loss.

A vehicle can be chosen into a cluster if and only if it can connect to a VoI via a multi-hop path. It can be seen that our scheme achieves better performance in terms of the achievable capacity than that proposed in [19]. This is due to the fact that in [19], the authors only used the cooperation among vehicles moving in the same direction and within the same cluster of the VoIs, while in our strategy, both cooperation among infrastructure and cooperation of vehicles traveling in each direction are utilized to help the VoIs' download, which improves the capacity achieved by the VoIs.

Fig. 9 compares the capacity achieved by allowing only one-hop communication and allowing both  $k$ -hop ( $k = 2, 3, 5$ ) V2I communications between the VoIs and infrastructure and  $k$ -hop V2V communication between the VoIs and helpers. It is shown that allowing multi-hop communications beyond one hop has little impact on the capacity. This can be explained by the fact that in the specific scenario being considered, there are only a subset of vehicles with download request (VoIs), all other vehicles (helpers) assist the VoIs to receive more data. Any new data in the vehicular network must come from the infrastructure. Therefore, allowing multi-hop V2V communications only helps to balance the distribution of information among vehicles but does not increase the net amount of information available in the network. The marginal increase in the achievable capacity comes from multi-hop V2I communications between the VoIs and infrastructure, because it allows the VoIs to have longer connection time (via some intermediate vehicles) with the infrastructure. This increase only occurs when the proportion of Vols,  $p$ , is smaller than a threshold, e.g.,  $p_{th} = 0.06$  in the considered scenario.

Fig. 10 compares capacity achieved from the constant channel model with that from the time-varying channel model, and shows that our analysis under the constant channel model is applicable to a more realistic time-varying channel model. Specifically, for the time-varying channel model, we adopt the model used in [35] that considers Rayleigh fading and path loss, from which the transmission rate is given by  $w'_I = B_I \log_2(1 + P_I |\beta d_i^{-2}|^2)$  and  $w'_V =$

$B_V \log_2(1 + P_V |\beta d_{ij}^{-2}|^2)$ , with the bandwidth and transmit power of each infrastructure and vehicle being  $B_I=40\text{MHz}$ ,  $P_I=52\text{dBm}$  and  $B_V=5\text{MHz}$ ,  $P_V=20\text{dBm}$  [43] respectively. Parameter  $\beta$  is a Gaussian random variable with mean 0 and variance 1 and  $d_i, d_{ij}$  are the distances between a vehicle and its associated infrastructure point, between vehicle and vehicle when conducting V2I and V2V communications respectively. By dividing the total coverage length of the transmitter (infrastructure or vehicle) into  $K$  (here we set  $K=1000$ ) small segments, the average channel throughput  $w_I$  and  $w_V$  in the time-varying channel model can be obtained by averaging the transmission rates of all segments. This obtained average throughput  $w_I$  and  $w_V$  are then used in our constant channel model. It is obvious from Fig. 10 that the achievable capacity from the above two channel models match each other. This phenomenon can be explained by equation (21) which shows that the achievable capacity is a linear function of  $w_I$  and  $w_V$ . Then, it follows that  $E[\eta(w_I, w_V)] = \eta(E[w_I], E[w_V])$ , which implies that for time-varying channels, the time-varying values of  $w_I$  and  $w_V$  can be replaced by the respective time-averaged throughput of V2I and V2V communications and our analysis still applies.

## VI. CONCLUSIONS

In this paper, we analyzed the capacity of vehicular networks with a finite traffic density adopting a cooperative communication strategy, which utilizes V2I communications, V2V communications, mobility of vehicles, and cooperation among vehicles and infrastructure to facilitate the transmission. A closed-form expression of the achievable capacity was obtained. Our result showed that the proposed cooperative strategy can improve the capacity of vehicular networks, and the improvement is more pronounced when the proportion of vehicles with download request is low. Moreover, our result sheds insight into the optimum deployment of vehicular network infrastructure and the design of cooperative communication strategy to maximize the capacity.

### APPENDIX A: PROOF OF THEOREM 2

Recall that we set the point to the right of infrastructure point  $I_1$  and at a distance  $r_I$  to  $I_1$  (the left boundary point of the V2V Area) as the origin of the coordinate system, and the right direction as the positive ( $+x$ ) direction. Denote by  $X_k, k = 1, 2, \dots$  the location of the  $i$ -th transmitter helper of the active helper-VoI pair, numbered from left to the right, under the optimum scheduling scheme  $\chi_{opt}$ . Denote by  $Y_k, k = 1, 2, \dots$  the location of the  $i$ -th transmitter under an arbitrary scheduling scheme  $\chi'$ . It follows that  $X_1 < X_2 < \dots < X_k < X_{k+1} < \dots$  and  $Y_1 < Y_2 < \dots < Y_k < Y_{k+1} < \dots$ . See Fig. 11 for an illustration. In the following, we prove that  $\chi_{opt}$  described in Theorem 2 is an optimum scheduling scheme that would lead to the maximum number of active helper-VoI pairs by recursion that  $X_k \leq Y_k$  holds for any  $k = 1, 2, \dots$

For  $k = 1$ , noting that according to the scheduling scheme  $\chi_{opt}$ , the first transmitter is the leftmost helper in the V2V

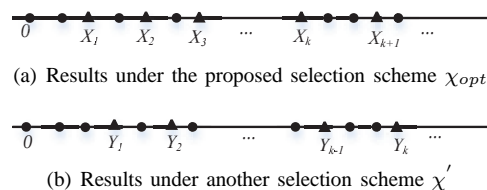


Fig. 11. An illustration of the distribution of simultaneous transmitters, where the triangular point represent the helpers that are chosen as simultaneous transmitters and the dots represent the helpers that are not chosen as transmitters.

Area that has at least one VoI within its coverage. Therefore, it follows readily that  $X_1 \leq Y_1$ .

Assuming that  $X_k \leq Y_k$  when  $k = n, n \geq 1$ , we will show that  $X_{n+1} \leq Y_{n+1}$ . We consider two different cases:  $X_{n+1} \leq Y_n$  and  $X_{n+1} > Y_n$ :

(i) Case  $X_{n+1} \leq Y_n$ : in this case, it can be readily shown that  $X_{n+1} \leq Y_n < Y_{n+1}$ .

(ii) Case  $X_{n+1} > Y_n$ : in this case, under the scheduling scheme  $\chi_{opt}$ , the  $(n+1)$ -th transmitter is the *nearest helper* to the right of the  $n$ -th transmitter satisfying simultaneous transmission conditions: it is outside the sensing range of the  $n$ -th transmitter who are located at  $X_n$  and has at least one VoI within its transmission range that is different from the VoI that the  $n$ -th transmitter transmits to. Therefore, there is no helper within road segment  $(X_n, X_{n+1})$  that can transmit simultaneously. If  $Y_{n+1} < X_{n+1}$ , a contradiction must occur. Thus,  $X_{n+1} \leq Y_{n+1}$ .

Therefore,  $X_k \leq Y_k$  holds for any  $k = 1, 2, \dots$

It readily follows that the number of simultaneous active helper-VoI pairs under  $\chi'$  must be less than or equal to that under  $\chi_{opt}$ .

### APPENDIX B: PROOF OF THEOREM 4

Recall that we denote by  $S_k \in [0, d - 2r_I], k = 1, 2, \dots$  the position of the  $k$ -th transmitter in the V2V Area and  $L_k, k = 1, 2, \dots$  the distance between the  $k$ -th and the  $(k+1)$ -th transmitter. Note that each  $L_k, k = 1, 2, \dots$  are i.i.d.

Denote by  $V_{k,1}$  the first helper located in the road segment  $[S_k + R_c, d - 2r_I]$ , by  $V_{k,2}$  the second helper, and so on. Denote by  $l_{k,i}, i = 1, 2, \dots$  the distance between helpers  $V_{k,i}$  and  $V_{k,i+1}$  and by  $l_{k,0}$  the distance between helper  $V_{k,1}$  and the point  $S_k + R_c$ . See Fig. 3 for an illustration. As an easy consequence of the Poisson distribution of helpers,  $l_{k,i}, i = 1, 2, \dots$  follows identical and independent exponential distribution with a mean value  $\frac{1}{(1-p)\rho}$ , and further due to the memoryless property of exponential distribution [44],  $l_{k,0}$  also has the same distribution as  $l_{k,i}, i = 1, 2, \dots$

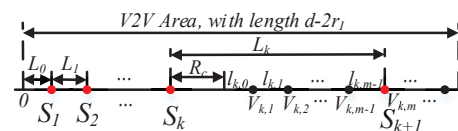


Fig. 12. An illustration of the distribution of distances between two consecutive simultaneous transmitters.

Consider the  $k$ -th transmitter and suppose the  $(k+1)$ -th transmitter is exactly the  $m_k$ -th helper located within the road

segment  $[S_k + R_c, d - 2r_I]$ , where  $m_k$  is a random integer. Note that the distribution of  $m_k$  is independent of  $l_{k,i}$ ,  $i = 1, 2, \dots$ , but is determined by the distribution of VoIs within the coverage of each helper  $V_{k,1}, V_{k,2}, \dots$ . Therefore,

$$L_k = R_c + \sum_{i=0}^{m_k-1} l_{k,i}. \quad (24)$$

From (24), to calculate the distribution of  $L_k$ , it remains to calculate the distribution of  $m_k$ ,  $Pr(m_k = m)$ . We compute the distribution of  $m_k$  in the following paragraphs.

- 1) when  $m = 1$ , it means that the first helper  $V_{k,1}$  located within road segment  $[S_k + R_c, d - 2r_I]$  is chosen as the  $(k + 1)$ -th transmitter. This implies that there should be at least one VoIs different from the VoI of the  $k$ th active helper-VoI pair within the coverage of helper  $V_{k,1}$ . When  $R_c \geq 2r_0$ , the VoI of the  $k$ th active helper-VoI pair cannot be possibly located within the coverage of helper  $V_{k,1}$ . Therefore, when  $R_c \geq 2r_0$ , the condition that the helper  $V_{k,1}$  is chosen as the  $(k + 1)$ th transmitter is that there exists at least one VoI within the coverage of helper  $V_{k,1}$ , which has a length  $r_x = 2r_0$  (see Fig. 13(a) for an illustration). In contrast, when  $R_c < 2r_0$ , it may happen that the VoI of the  $k$ th active helper-VoI pair is located within the coverage of helper  $V_{k,1}$  (see Fig. 13(b) for an illustration). In this situation, helper  $V_{k,1}$  may be chosen as the  $(k + 1)$ th transmitter iff there exists at least one VoI within the road segment with length  $r_x$ , which starts from the position of the VoI of  $k$ th active helper-VoI pair and ends at the right boundary of the coverage of helper  $V_{k,1}$ , and  $r_x < 2r_0$ .

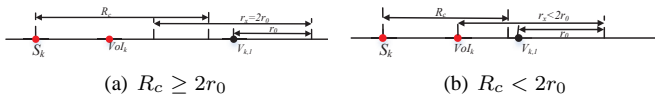


Fig. 13. An illustration of the case that the helper  $V_{k,1}$  is chosen as the  $(k + 1)$ -th transmitter

Here we approximately omit the scenario that the VoI of  $k$ th active helper-VoI may be located within the coverage of helper  $V_{k,1}$ . That is, we approximately consider  $r_x$  to be equal to the length of the coverage area of helper  $V_{k,1}$ ,  $2r_0$ . We will use simulation later to validate the accuracy of this approximation and its impact on our result. This approximation allows us to write:

$$\begin{aligned} Pr(m_k = 1) &= Pr(\exists \text{VoI within length } r_x) \\ &\approx Pr(\exists \text{VoI within coverage of } V_{k,1}) \\ &= 1 - e^{-\rho \rho 2r_0}. \end{aligned} \quad (25)$$

- 2) when  $m \geq 2$ , it means that the  $m$ -th helper located within road segment  $[S_k + R_c, d - 2r_I]$ ,  $V_{k,m}$ , is chosen as the  $(k + 1)$ -th transmitter. This implies that a) none of the helper  $V_{k,1}, V_{k,2}, \dots, V_{k,m-1}$  satisfies the condition to become the  $(k + 1)$ -th transmitter, i.e., none of the helpers  $V_{k,i}$ ,  $i = 1, \dots, m - 1$  can find a VoI that is not the same as the VoI of the  $k$ -th active helper-VoI pair within their coverage; and b) there exist at least one VoIs that is

not the same as the VoI of the  $k$ -th active helper-VoI pair within the coverage  $V_{k,m}$ . Using the same *approximation* as that used previously, it can be obtained that:

$$\begin{aligned} &Pr(m_k = m) \\ &\approx Pr(\text{no VoI within the coverage of } V_{k,i}, i = 1, \dots, m - 1 \\ &\quad \cap \exists \text{VoI within the coverage of } V_{k,m}) \\ &= Pr\left(\text{no VoI in } \left[ \sum_{i=1}^{m-2} \min\{l_{k,i}, 2r_0\} + 2r_0 \right] \right. \\ &\quad \left. \cap \exists \text{VoI in } \min\{l_{k,m-1}, 2r_0\} \right) \\ &= Pr\left(\text{no VoI in } \left[ \sum_{i=1}^{m-2} \min\{l_{k,i}, 2r_0\} + 2r_0 \right] \right) \times \\ &\quad Pr(\exists \text{VoI in } \min\{l_{k,m-1}, 2r_0\}), \end{aligned} \quad (26)$$

where the last step results due to the property that VoIs have a Poisson distribution and therefore the numbers of VoIs in non-overlapping intervals are independent. The summation ends at  $m - 2$  is due to that the total length of the coverage area of helper  $V_{k,i}$ ,  $i = 1, \dots, m - 1$  is exactly  $\sum_{i=1}^{m-2} \min\{l_{k,i}, 2r_0\} + 2r_0$ . Define two parameters  $h_{k,i}$  and  $H_{k,m}$  as follows:

$$h_{k,i} = \min\{l_{k,i}, 2r_0\}, i = 1, 2, \dots, m - 1 \quad (27)$$

and

$$H_{k,m} = \sum_{i=1}^{m-2} h_{k,i} + 2r_0, m = 2, 3, \dots, \quad (28)$$

with  $H_{k,2} = 2r_0$ . Because  $l_{k,i}$ ,  $i = 1, 2, \dots, m - 1$  are i.i.d. random variables, it follows that  $h_{k,i}$ ,  $i = 1, 2, \dots, m - 1$  are i.i.d. random variables. Denote by  $\bar{f}_{h_{k,i}}(x)$  and  $\bar{f}_{H_{k,m}}(x)$  the pdf of random variables  $h_{k,i}$  and  $H_{k,m}$  respectively. As an easy consequence of the total probability theorem, (26) can be calculated by:

$$\begin{aligned} &Pr(m_k = m) \\ &\approx Pr(\text{no VoI in } H_{k,m}) \times Pr(\exists \text{VoI in } h_{k,m-1}) \\ &= \int_0^\infty e^{-\rho \rho x} \bar{f}_{H_{k,m}}(x) dx \times \\ &\quad \int_0^\infty (1 - e^{-\rho \rho y}) \bar{f}_{h_{k,m-1}}(y) dy \\ &= E[e^{-\rho \rho H_{k,m}}] \times (1 - E[e^{-\rho \rho h_{k,m-1}}]). \end{aligned} \quad (29)$$

From (29), we can see that when  $m \geq 2$ , the value of  $Pr(m_k = m)$  is the product of two factors,  $E[e^{-\rho \rho H_{k,m}}]$ , and  $(1 - E[e^{-\rho \rho h_{k,m-1}}])$ . Further note that  $E[e^{-\rho \rho H_{k,m}}]$  and  $E[e^{-\rho \rho h_{k,m-1}}]$  are in the form of the moment generating functions (MGF) of the random variables  $H_{k,m}$  and  $h_{k,m-1}$  respectively. For a random variable  $X$ , its MGF is defined as follows [30]:

$$M_X(t) \triangleq E[e^{tX}], \quad t \in \mathbb{R}. \quad (30)$$

Let  $M_{H_{k,m}}(t)$  and  $M_{h_{k,m-1}}(t)$  be the MFG of  $H_{k,m}$  and  $h_{k,m-1}$  respectively. It follows that

$$Pr(m_k = m)$$



$$\begin{aligned}
 &\approx (M_{H_{k,m}}(t) \cdot (1 - M_{h_{k,m-1}}(t)))|_{t=-p\rho} \\
 &= \left( \prod_{i=1}^{m-2} M_{h_{k,i}}(t) \cdot M_{2r_0}(t) \cdot (1 - M_{h_{k,m-1}}(t)) \right)|_{t=-p\rho} \\
 &= \left( (M_{h_{k,i}}(t))^{m-2} \cdot e^{2r_0 t} \cdot (1 - M_{h_{k,i}}(t)) \right)|_{t=-p\rho} \quad (31)
 \end{aligned}$$

where the second step results because of the fact that the random variable  $H_{k,m}$  is the sum of  $m-2$  i.i.d. random variables  $h_{k,i}$  and a constant  $2r_0$ .

It remains to calculate the MGF of  $h_{k,i}$ . Noting that  $l_{k,i}, i = 1, 2, \dots, m-1$  has an exponential distribution:  $f_{l_{k,i}}(x) = (1-p)\rho e^{-(1-p)\rho x}$ . The pdf of  $h_{k,i} = \min\{l_{k,i}, 2r_0\}$ ,  $\bar{f}_{h_{k,i}}(x)$ , can be calculated as follows:

$$\bar{f}_{h_{k,i}}(x) = \begin{cases} (1-p)\rho e^{-(1-p)\rho x} & x < 2r_0 \\ e^{-(1-p)\rho 2r_0} \delta(x - 2r_0) & x \geq 2r_0 \end{cases} \quad (32)$$

where  $\delta(x)$  is a delta function:

$$\delta(x) = \begin{cases} 1, & x = 0 \\ 0, & x \neq 0 \end{cases} \quad (33)$$

Using (32), the MGF of  $h_{k,i}$  can be obtained as follows:

$$\begin{aligned}
 M_{h_{k,i}}(t) &= E[e^{t \cdot h_{k,i}}] \\
 &= \int_0^\infty e^{tx} \bar{f}_{h_{k,i}}(x) dx \\
 &= \int_0^{2r_0} e^{tx} (1-p)\rho e^{-(1-p)\rho x} dx \\
 &\quad + e^{t2r_0} e^{-(1-p)\rho 2r_0} \\
 &= \frac{te^{(t-(1-p)\rho)2r_0} - (1-p)\rho}{t - (1-p)\rho}. \quad (34)
 \end{aligned}$$

Combing (31) and (34) and simplifying it, for  $m \geq 2$ , we have

$$Pr(m_k = m) \approx e^{-p\rho 2r_0} \cdot c_1^{m-2} (1 - c_1) \quad (35)$$

where  $c_1 = M_{h_{k,i}}(t)|_{t=-p\rho} = 1 - p + pe^{-\rho 2r_0}$ .

Combing the above result, we have the cdf of  $L_k, k \geq 1$ , denoted by  $F_{L_k}(x)$ , as follows:

$$\begin{aligned}
 &F_{L_k}(x) \\
 &= Pr(L_k \leq x) \\
 &= \sum_{m=1}^\infty Pr(R_c + \sum_{i=0}^{m_k-1} l_{k,i} \leq x | m_k = m) Pr(m_k = m) \\
 &= \sum_{m=1}^\infty Pr(\sum_{i=0}^{m-1} l_{k,i} \leq x - R_c) Pr(m_k = m) \\
 &= \sum_{m=1}^\infty \sum_{n=m}^\infty \frac{e^{-(1-p)\rho(x-R_c)} [(1-p)\rho(x-R_c)]^n}{n!} Pr(m_k = m) \quad (36)
 \end{aligned}$$

where the second step is obtained by putting (24) into  $L_k$  and using the total probability theorem and the last step is obtained by that the distribution of  $\sum_{i=0}^{m-1} l_{k,i}$  is independent of the distribution of  $m_k$ .

From (36) and with  $f_{L_k}(x) = \frac{dF_{L_k}(x)}{dx}$ , we have the pdf of  $L_k$  shown as (13), which completes the proof.

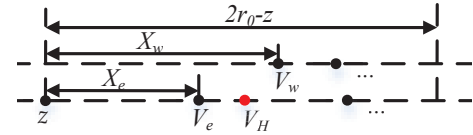


Fig. 14. An illustration of the coordinate system, the location of the randomly chosen transmitter, and the left-most VoI from each direction that are located at the right of origin.

### APPENDIX C: PROOF OF THEOREM 8

To calculate the capacity achieved by the eastbound and westbound VoIs, we will analyze the V2V communications and V2I communications in one cycle area separately.

#### 1). V2V communications:

Recall that under our optimum helper-VoI scheduling algorithm  $\chi_{opt}$  for the V2V communications proposed in Theorem 2, for any two randomly chosen helper-VoI pairs at a randomly chosen time slot, the travel direction of the VoI in one pair is independent of the travel direction of the helper in the same pair, and is also independent of the travel direction of the VoI in the other pair. Therefore, at any randomly chosen time slot, the proportion of the helper-VoI pairs whose VoI is eastbound (westbound) is equal to the probability that the VoI of a randomly chosen helper-VoI pair is eastbound (westbound), denoted by  $P_{V_e}$  ( $P_{V_w}$ ). Obviously,  $P_{V_e} + P_{V_w} = 1$ . It follows that the maximum expected amount of data the eastbound and westbound VoIs can receive through V2V communications during time period  $[0, t]$  from one cycle area, denote by  $D_{V2V_e}(t)$  and  $D_{V2V_w}(t)$  respectively, can be calculated by  $D_{V2V_e}(t) = P_{V_e} \cdot D_{V2V}(t)$  and  $D_{V2V_w}(t) = P_{V_w} \cdot D_{V2V}(t)$ , where  $D_{V2V}(t)$  is the expected amount of data received by all the VoIs given in (3). Therefore, the capacity achieved by the eastbound and westbound VoIs through V2V communications from one cycle area are respectively:

$$\lim_{t \rightarrow \infty} \frac{D_{V2V_e}(t)}{t} = P_{V_e} \lim_{t \rightarrow \infty} \frac{D_{V2V}(t)}{t}, \quad (37)$$

and

$$\lim_{t \rightarrow \infty} \frac{D_{V2V_w}(t)}{t} = P_{V_w} \lim_{t \rightarrow \infty} \frac{D_{V2V}(t)}{t}. \quad (38)$$

In the following, we will calculate  $P_{V_e}$  and  $P_{V_w}$ . Suppose helper  $V_H$  is one of the simultaneous transmitters at a randomly chosen time slot. Noting that for a randomly chosen helper-VoI pair, the travel direction of its VoI is irrelevant to the travel direction of its helper, but only dependent on the original distribution of VoIs in each direction. Therefore, without loss of generality, we assume that helper  $V_H$  travels eastbound. Recall that we designate the east (right) direction as  $+x$  direction. Here, designate the point to the  $-x$  direction of helper  $V_H$  and at a distance  $r_0$  to  $V_H$  as the origin of the coordinate system. Denote by  $z$  the location of the point from which the helper  $V_H$  starts to choose its receiver (VoI), i.e., the helper  $V_H$  chooses its receiver within road segment  $[z, 2r_0]$ . Therefore, according to the scheduling scheme  $\chi_{opt}$ ,  $z$  is equal to 0 if the VoI of the previous helper-VoI pair is not located within the coverage of helper  $V_H$ , otherwise  $z > 0$ . See Fig. 14 for an illustration.



Denote the left-most eastbound VoI located at the right side of  $z$  by  $V_e$ , and its location by  $X_e$ . Further denote the left-most westbound VoI located at the right of  $z$  by  $V_w$ , and its location by  $X_w$ . Noting that eastbound and westbound VoIs follow Poisson distributions with densities  $p\rho_1$  and  $p\rho_2$  respectively, it follows that  $X_e$  and  $X_w$  are exponentially distributed:  $f_{X_e}(x) = p\rho_1 e^{-p\rho_1 x}$  and  $f_{X_w}(x) = p\rho_2 e^{-p\rho_2 x}$ . Given that there is at least one VoI within road segment  $[z, 2r_0]$  (otherwise helper  $V_H$  can not be one of the simultaneous transmitters), VoI  $V_e$  can be chosen as the receiver of helper  $V_H$  iff when  $X_e \leq X_w$ . Therefore, we have the probability that the receiver of the helper  $V_H$  travels towards east as follows (conditioned on  $z$  is fixed):

$$\begin{aligned} P_{V_e} &= Pr(X_e \leq X_w | \text{there exists VoI in } [z, 2r_0]) \\ &= \frac{Pr(X_e \leq X_w, \text{there exists VoI in } [z, 2r_0])}{Pr(\text{there exists VoI in } [z, 2r_0])} \\ &= \frac{\int_0^{2r_0-z} Pr(X_e \leq x) f_{X_w}(x) dx}{1 - e^{-p\rho(2r_0-z)}} + \\ &\quad \frac{\int_{2r_0-z}^{\infty} Pr(X_e \leq 2r_0 - z) f_{X_w}(x) dx}{1 - e^{-p\rho(2r_0-z)}} \\ &= \frac{\rho_1}{\rho_1 + \rho_2}, \end{aligned} \quad (39)$$

where the second step results by using Bayes' theorem; the third step is obtained by using total probability theorem. In addition, the result of (39) is irrelevant to the random variable  $z$ . Straightforwardly, we have:

$$P_{V_w} = 1 - P_{V_e} = \frac{\rho_2}{\rho_1 + \rho_2}. \quad (40)$$

Plugging (39) and (40) into (37) and (38) respectively, we can conclude that the expected amount of data received by the eastbound and westbound VoIs through V2V communications are proportional to their respective traffic densities.

## 2). V2I communications:

Denote by  $P_{I_e}$  and  $P_{I_w}$  respectively the probability that at a randomly chosen time slot, the receiver of a VoIs' V2I communication travels towards east and west. Using the same method above for analyzing the V2V communications (the infrastructure point in this case corresponds to the randomly chosen transmitter  $V_H$ , and coverage area of infrastructure  $2r_I$  corresponds the VoI choosing area  $2r_0 - z$ ), it is ready to have  $P_{I_e} = \frac{\rho_1}{\rho_1 + \rho_2}$ , and  $P_{I_w} = \frac{\rho_2}{\rho_1 + \rho_2}$ . Therefore, the capacity achieved by the eastbound and westbound VoIs respectively through V2I communications in one cycle area, can be obtained by

$$\lim_{t \rightarrow \infty} \frac{D_{V2I_e}(t)}{t} = \frac{\rho_1}{\rho_1 + \rho_2} \cdot \lim_{t \rightarrow \infty} \frac{D_{V2I}(t)}{t}, \quad (41)$$

and

$$\lim_{t \rightarrow \infty} \frac{D_{V2I_w}(t)}{t} = \frac{\rho_2}{\rho_1 + \rho_2} \cdot \lim_{t \rightarrow \infty} \frac{D_{V2I}(t)}{t}. \quad (42)$$

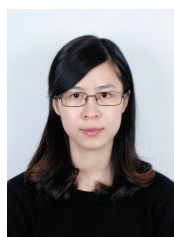
where  $D_{V2I_e}(t)$  and  $D_{V2I_w}(t)$  are respectively the maximum expected amount of data the eastbound and westbound VoIs can receive through V2I communications during time period  $t$  from one cycle area.

Noting that both the analysis in 1) and 2) show that the maximum amount of data received from V2V communications and V2I communications by the eastbound and westbound VoIs are proportional to their respective traffic densities. Therefore, the capacity achieved from one cycle area by the eastbound and westbound VoIs, are also proportional to their traffic densities respectively, which finalizes the proof.

## REFERENCES

- [1] S. Ilarri, T. Delot, R. Trillo-Lado, "A Data Management Perspective on Vehicular Networks," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2420-2460, Fourth Quarter, 2015.
- [2] Y. Li, X. Zhu, D. Jin and D. Wu, "Multiple Content Dissemination in Roadside-Unit-Aided Vehicular Opportunistic Networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 8, pp. 3947-3956, Oct. 2014.
- [3] J. B. Kenney, "Dedicated Short-Range Communications (DSRC) Standards in the United States," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1162-1182, Jun. 2011.
- [4] H. Seo, K-D. Lee, et al., "LTE Evolution for Vehicle-to-Everything Services," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 22-28, Jun. 2016.
- [5] A. Bazzi, B. M. Masini, and A. Zanella, "Performance Analysis of V2V Beaconing Using LTE in Direct Mode With Full Duplex Radios," *IEEE Wireless Commun. Lett.*, vol. 4, no. 6, pp. 685-688, Dec. 2015.
- [6] M. Khabazian, S. Aissa, and M. Mehmet-Ali, "Performance Modeling of Safety Messages Broadcast in Vehicular Ad Hoc Networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 380-387, Mar. 2013.
- [7] Z. Zhang, G. Mao, and B. D. O. Anderson, "Stochastic Characterization of Information Propagation Process in Vehicular Ad hoc Networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 122-135, Feb. 2014.
- [8] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388-404, Mar. 2000.
- [9] G. Mao, Z. Lin, X. Ge, and Y. Yang, "Towards a Simple Relationship to Estimate the Capacity of Static and Mobile Wireless Networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 3883-3895, Aug. 2013.
- [10] H. N. Dai, R. C. W. Wong, and H. Wang, "On Capacity and Delay of Multi-channel Wireless Networks with Infrastructure Support," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1589 - 1604, Feb. 2017.
- [11] M. Wang, H. Shan, et al., "Asymptotic Throughput Capacity Analysis of VANETs Exploiting Mobility Diversity," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4187 - 4202, Sep. 2015.
- [12] Y. Huang, M. Chen, et al., "Graph Theory Based Capacity Analysis for Vehicular Ad Hoc Networks," in *Proc. IEEE Globecom, Dec. 2015*, pp. 1-5.
- [13] J. Chen, G. Mao et al., "Throughput of Infrastructure-Based Cooperative Vehicular Networks", to appear in *IEEE Trans. Intell. Transp. Syst.* 2017.
- [14] L. Huang, H. Jiang, et al., "Optimal traffic scheduling between roadside units in vehicular delay-tolerant networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 3, pp. 1079-1094, Mar. 2015.
- [15] P. Si, Y. He, et al., "DaVe: Offloading Delay-Tolerant Data Traffic to Connected Vehicle Networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 3941 - 3953, Jun. 2016.
- [16] Y. Wang, J. Zheng, and N. Mitton, "Delivery Delay Analysis for Roadside Unit Deployment in Vehicular Ad Hoc Networks with Intermittent Connectivity", *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8591 - 8602, Oct. 2016.
- [17] N. Lu, N. Zhang, et al., "Vehicles Meet Infrastructure: Toward Capacity-Cost Tradeoffs for Vehicular Access Networks", *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1266-1277, Sep. 2013.
- [18] W. Zhu, D. Li, and W. Saad, "Multiple Vehicles Collaborative Data Download Protocol via Network Coding," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1607-1619, Apr. 2015.
- [19] H. Zhou et.al, "ChainCluster: Engineering a Cooperative Content Distribution Framework for Highway Vehicular Communications," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 6, pp. 2644-2657, Dec. 2014.
- [20] B. Das, S. Misra, and U. Roy, "Coalition Formation for Cooperative Service-Based Message Sharing in Vehicular Ad Hoc Networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 1, pp. 144-156, Jan. 2016.
- [21] K. Liu, et al., "Cooperative Data Scheduling in Hybrid Vehicular Ad Hoc Networks: VANET as a Software Defined Network," *IEEE/ACM Trans. Netw.*, vol. 24, no. 3, pp. 1759 - 1773, Jun. 2016.
- [22] D. Zhang and C. Yeo, "Enabling efficient WiFi-based vehicular content distribution," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 3, pp. 479-492, Mar. 2013.

- [23] R. Kim, H. Lim, and B. Krishnamachari, "Prefetching-Based Data Dissemination in Vehicular Cloud Systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 1, pp. 292 - 306, Jan. 2016.
- [24] G. Mauri, M. Gerla, et al., "Optimal Content Prefetching in NDN Vehicle-to-Infrastructure Scenario," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2513 - 2525, Mar. 2017.
- [25] K. Mereshad, H. Artail, and M. Gerla, "We Can Deliver Messages to Far Vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1099-1115, Sep. 2012.
- [26] Y. Wang, Y. Liu, et al., "Cooperative Store-Carry-Forward Scheme for Intermittently Connected Vehicular Networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 777 - 784, Jan. 2017.
- [27] N. Wisitpongphan, B. Fan, et al., "Routing in Sparse Vehicular Ad Hoc Wireless Networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 8, pp. 1538-1556, Oct. 2007.
- [28] L. Zhu, C. Li, et al., "On Stochastic Analysis of Greedy Routing in Vehicular Networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3353 - 3366, Dec. 2015.
- [29] R. Mao and G. Mao, "Road traffic density estimation in vehicular networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2013, pp. 1-6.
- [30] R. Nelson, *Probability, Stochastic Processes, and Queueing Theory: The Mathematics of Computer Performance Modeling*, New York: Springer-Verlag, 1995.
- [31] X. Ge, S. Tu, et al., "Energy Efficiency of Small Cell Backhaul Networks Based on Gauss-Markov Mobile Models," *IET Networks*, Vol. 4, No. 2, pp. 158-167, Mar. 2015.
- [32] G. Mao, and B. D. O. Anderson, "Graph Theoretic Models and Tools for the Analysis of Dynamic Wireless Multihop Networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2009, pp. 1-6.
- [33] G. Mao, *Connectivity of Communication Networks*, Springer, 2017.
- [34] G. Mao and B. D. O. Anderson, "Towards a Better Understanding of Large Scale Network Models," *IEEE/ACM Trans. Netw.*, Vol. 20, No. 2, pp. 408 - 421, Apr. 2012.
- [35] T. Wang, L. Song and Z. Han, "Coalitional Graph Games for Popular Content Distribution in Cognitive Radio VANETs," *IEEE Trans. Veh. Technol.*, vol. 62, no. 8, pp. 4010-4019, Oct. 2013.
- [36] J. Cheng, J. Cheng, et al., "Routing in Internet of Vehicles: A Review," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2339-2352, Oct. 2015.
- [37] N. Lu, T. H. Luan, et al., "Bounds of Asymptotic Performance Limits of Social-Proximity Vehicular Networks," *IEEE/ACM Trans. Netw.* vol. 22, no. 3, pp. 812-825, Jun. 2014.
- [38] J. Ning, S. Singh, et al., "Forensic Analysis of Packet Losses in Wireless Networks," *IEEE/ACM Trans. Netw.*, vol. 24, no. 4, pp. 1975-1988, Aug. 2016.
- [39] A. Le, A. S. Tehrani, et al., "Recovery of Packet Losses in Wireless Broadcast for Real-Time Applications," *IEEE/ACM Trans. Netw.*, vol. 25, no. 2, pp. 676-689, Apr. 2017.
- [40] P. Wang, G. Mao, et al., "Network Coding based Wireless Broadcast with Performance Guarantee," *IEEE Trans. Wireless Commun.*, Vol. 14, No. 1, pp. 532 - 544, Jan. 2015.
- [41] R. G. Gallager, *Stochastic Processes: Theory for Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2013.
- [42] F. Baccelli and B. Błaszczyszyn, *Stochastic Geometry and Wireless Networks Volume II Applications*, NOW: Foundations and Trends in Networking, 2010.
- [43] K. Zheng, F. Liu, et al., "A Graph-Based Cooperative Scheduling Scheme for Vehicular Networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1450-1458, May. 2013.
- [44] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 2. New York, NY, USA: Wiley, 1971.



**Jieqiong Chen** (S'16) received the Bachelor's degree in Engineering from Zhejiang University, Zhejiang, China, in 2012, and she is currently working toward the Ph.D. degree in engineering at the University of Technology Sydney, Australia. Her research interests include wireless communications and vehicular networks.



**Guoqiang Mao** (S'98-M'02-SM'08) joined the University of Technology Sydney in February 2014 as Professor of Wireless Networking and Director of Center for Real-time Information Networks. He has published about 200 papers in international conferences and journals, which have been cited more than 5000 times. He is a fellow of IET. His research interest include intelligent transport systems, applied graph theory and its applications in telecommunications, Internet of Things, wireless sensor networks, wireless localization techniques and network performance analysis.



**Changle Li** (M'09--SM'16) received the Ph.D. degrees in communication and information system from Xidian University, China, in 2005. Since then, he conducted his postdoctoral research in Canada and the National Institute of information and Communications Technology (NICT), Japan, respectively. He is currently a Professor with the State Key Laboratory of Integrated Services Networks, Xidian University. He is an IEEE Senior Member and his research interests include intelligent transportation systems, vehicular networks, mobile ad hoc networks, and wireless sensor networks.



**Weifa Liang** (M'99--SM'01) received the PhD degree from the Australian National University in 1998, the ME degree from the University of Science and Technology of China in 1989, and the BSc degree from Wuhan University, China in 1984, all in Computer Science. He is currently a Full Professor in the Research School of Computer Science at the Australian National University. His research interests include design and analysis of energy efficient routing protocols for wireless ad hoc and sensor networks, cloud computing, Software-Defined Networking, design and analysis of parallel and distributed algorithms, approximation algorithms, combinatorial optimization, and graph theory. He is a senior member of the IEEE.



**De-gan Zhang** (M'01) Born in 1970, Ph.D. Graduated from Northeastern University, China. Now he is professor of Tianjin Key Lab of Intelligent Computing and Novel software Technology, Key Lab of Computer Vision and System, Ministry of Education, Tianjin University of Technology, Tianjin, 300384, China. His research interest includes IOT, WSN, IOV, etc. His E-mail: gandegande@126.com. He is the corresponding author of this paper.