Scaling Laws for Vehicular Networks

by

Ning Lu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2015

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Equipping automobiles with wireless communications and networking capabilities is becoming the frontier in the evolution to the next generation intelligent transportation systems (ITS). By means of vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications, information generated by the vehicle-borne computer, vehicle control system, on-board sensors, or roadside infrastructure, can be effectively disseminated among vehicles/infrastructure in proximity or to vehicles/infrastructure multiple hops away, known as vehicular networks (VANETs), to enhance the situational awareness of vehicles and provide motorist/passengers with an information-rich travel environment.

Scaling law for throughput capacity and delay in wireless networks has been considered as one of the most fundamental issues, which characterizes the trend of throughput/delay behavior when the network size increases. The study of scaling laws can lead to a better understanding of intrinsic properties of wireless networks and theoretical guidance on network design and deployment. Moreover, the results could also be applied to predict network performance, especially for the large-scale vehicular networks. However, map-restricted mobility and spatio-temporal dynamics of vehicle density dramatically complicate scaling laws studies for VANETs. As an effort to lay a scientific foundation of vehicular networking, this thesis investigates capacity scaling laws for vehicular networks with and without infrastructure, respectively.

Firstly, the thesis studies scaling law of throughput capacity and end-to-end delay for a social-proximity vehicular network, where each vehicle has a restricted mobility region around a specific social spot and services are delivered in a store-carry-and-forward paradigm. It has been shown that although the throughput and delay may degrade in a high vehicle density area, it is still possible to achieve almost constant scaling for per-vehicle throughput and end-to-end delay. Secondly, in addition to pure ad hoc vehicular networks, the thesis derives the capacity scaling laws for networks with wireless infrastructure, where services are delivered uniformly from infrastructure to all vehicles in the network. The V2V communication is also required to relay the downlink traffic to the vehicles outside the coverage of infrastructure. Three kinds of infrastructures have been considered, i.e., cellular base stations, wireless mesh backbones (a network of mesh nodes, including one mesh gateway), and roadside access points. The downlink capacity scaling is derived for each kind of infrastructure. Considering that the deployment/operation costs of different infrastructure are highly variable, the capacity-cost tradeoffs of different deployments are examined. The results from the thesis demonstrate the feasibility of deploying non-cellular infrastructure for supporting high-bandwidth vehicular applications. Thirdly, the fundamental impact of traffic signals at road intersection on drive-thru Internet access

is particularly studied. The thesis analyzes the time-average throughput capacity of a typical vehicle driving through randomly deployed roadside Wi-Fi networks. Interestingly, we show a significant throughput gain for vehicles stopping at intersections due to red signals. The results provide a quick and efficient way of determining the Wi-Fi deployment scale according to required quality of services.

In summary, the analysis developed and the scaling laws derived in the thesis should be very useful for understanding the fundamental performance of vehicular networks.

Acknowledgements

First and foremost I would like to express my deepest gratitude to my supervisor Professor Xuemin (Sherman) Shen, for his endless support, patient guidance, and enthusiastic encouragement during my program, without which this thesis would not have been possible. I feel so lucky to have Professor Shen on my way in pursuit of a professional career. With the goal of helping every student to be successful, he has put enormous effort in mentoring students, as well as enhancing the visibility of the BBCR research group, all of which eventually make me a confident and independent researcher.

I would like to thank Professor Amiya Nayak, Liang-liang Xie, Sagar Naik, and Xinzhi Liu, for serving my thesis/oral defence committee. Their valuable comments and insightful questions have significantly improved the presentation of my work. I would also like to thank Professor Weihua Zhuang for helping me to build the knowledge base from her courses, which greatly benefits my research in this thesis.

My personal and professional life in Waterloo is truly precious and unforgotten due to friendship, help, support, and collaboration offered by BBCR members in the past four and a half years. I am grateful for time spent with Nan Cheng, Dr. Ning Zhang, Kuan Zhang, Miao Wang, Ran Zhang, Dr. Zhongming Zheng, Yong Zhou, Dr. Haibo Zhou, Dr. Yi Zhou, Dr. Xiugang Wu, and many others. I wish to thank Dr. Tom H. Luan, Dr. Xiaohui Liang, Dr. Xiaodong Lin, Dr. Rongxing Lu, Dr. Lin X. Cai, Dr. Hao Liang, Dr. Yongkang Liu, Dr. Bong Jun Choi, Dr. Sandra Céspedes, Dr. Cailian Chen, Dr. Zhou Su, and Dr. Kan Zheng for their advice and help in the Ph.D. pursuit. Especially, I gratefully acknowledge Dr. Fan Bai for every single inspiring discussion we had and invaluable suggestions he provided on my research. I would also like to thank Dr. Hassan Omar, Sailesh Bharati, Dr. Khadige Abboud, Kamal Malekshan, and Nizar Alsharif for their supports when I coordinated the VANET research group. Special thanks go to Dr. Hassan Omar and Sailesh Bharati for the teamwork on the VeMAC project. I am proud of what we have achieved and still miss good times we spent in Paris.

Finally, I would like to thank my family for all their love, support, and encouragement, without whom it would not be possible to finish my Ph.D. study. I am grateful for my parents who always believe in me and support me in all my pursuits. I dedicate this thesis to my parents, and most of all to my wife Jing Ma for her faithful support and unwavering love. I appreciate the enjoyable time we spent together in Waterloo.

Ning Lu

January 24, 2015

Waterloo, Ontario, Canada

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Vehicular networks play an important role in both promoting the development of next generation intelligent transportation systems and offering mobile data services to vehicle users. The capacity scaling laws of vehicular networks characterize the trend of network capacity when vehicle population grows in the network, which represent the fundamental property of vehicular networks and could be applied to predict the network performance and thereby provide valuable guidance on network design and deployment. Despite extensive research in the field of vehicular networking, the network capacity is not well understood. In this chapter, we first overview the vehicular network, and then briefly introduce the research of capacity scaling laws for general wireless networks. Finally, we present the existing works in the capacity study of vehicular networks.

1.1 Overview of Vehicular Networks

As a key ingredient of transportation system, motor vehicles have continued to evolve since people expect more than just vehicle quality and reliability. With the rapid development of information and communication technologies (ICT), equipping automobiles with wireless communication capabilities is expected to be the next frontier for automotive revolution. Connected vehicles on the go are proactive, cooperative, well-informed, and coordinated, and will pave the way for supporting various applications for road safety (e.g., collision detection, lane change warning, and cooperative merging), smart and green transportation (e.g., traffic signal control, intelligent traffic scheduling, and fleet management), location dependent services (e.g., point of interest and route optimization), and in-vehicle Internet access. The market of connected vehicles is booming, and according to a recent business

report, the global market is expected to reach USD 131.9 billion by 2019 [5]. Academia and the automotive industry are responding promptly by exploring reliable and efficient connectivity solutions.

There are two immediate driving forces of bringing wireless connectivity to vehicles. The first one is the urgent need to improve efficiency and safety of road transportation systems. Growing urbanization yields an increasing population of vehicles in large cities, which is responsible for traffic congestion and the consequences in terms of huge economic cost and environmental problems. It is reported that the cost of extra travel time and fuel due to congestion in 498 U.S. urban areas was already USD 121 billions in 2011, and CO_2 produced during congestion was 56 billions of pounds, compared to USD 24 billions and 10 billions of pounds in 1982, respectively [6]. Connected vehicle solutions are very promising to alleviate traffic congestions via intelligent traffic control and management [7], as well as to improve the road safety via on-board advanced warning and driving assistance systems [8]. The second one is the ever-increasing mobile data demand of users on road. In recent years, the demand for high-speed mobile Internet services has increased dramatically. People in their own cars expect to have the same connectivity as they have at home and at work. Connecting vehicles to the Internet can be envisioned not only to meet the mobile data demand [9], but also enrich safety-related applications, such as online diagnosis [10], and intelligent anti-theft and tracking [11], in which the servers can be on the Internet cloud. Internet-integrated vehicles have hit the road, and it is predicted that the percentage of Internet-integrated vehicle services will jump from 10% today to 90% by 2020 [12]. In addition, government mandate has put the connected vehicle revolution on the fast track. The European Commission proposed to implement a mandatory "eCall" system in cars from 2015, by which cars can automatically establish a telephone link for emergency services in case of a collision [13]. Not surprisingly, the U.S. Department of Transportation's (DOT) National Highway Traffic Safety Administration (NHTSA) recently announced that it will start taking steps to enable communications between light vehicles [14].

VehiculAr NETwork (VANET)[1] refers to the wireless connectivity enabled vehicles that can communicate with their external environments, i.e., supporting the interactions of V2V (vehicle-to-vehicle), V2R (vehicle-to-road infrastructure), and V2I (vehicle-to-Internet), as shown in Fig. 1.1. These interactions, establishing a multiple levels of data pipeline to in-vehicle information systems, enhance the situational awareness of vehicles and provide motorist/passengers with an information-rich travel environment.

Fascinated by the concepts and visions of VANETs, the academia, industry, and gov-

[1]To deemphasize the ad hoc nature of vehicular networks, we redefine the term VANETs, which is traditionally the acronym of vehicular ad hoc networks.

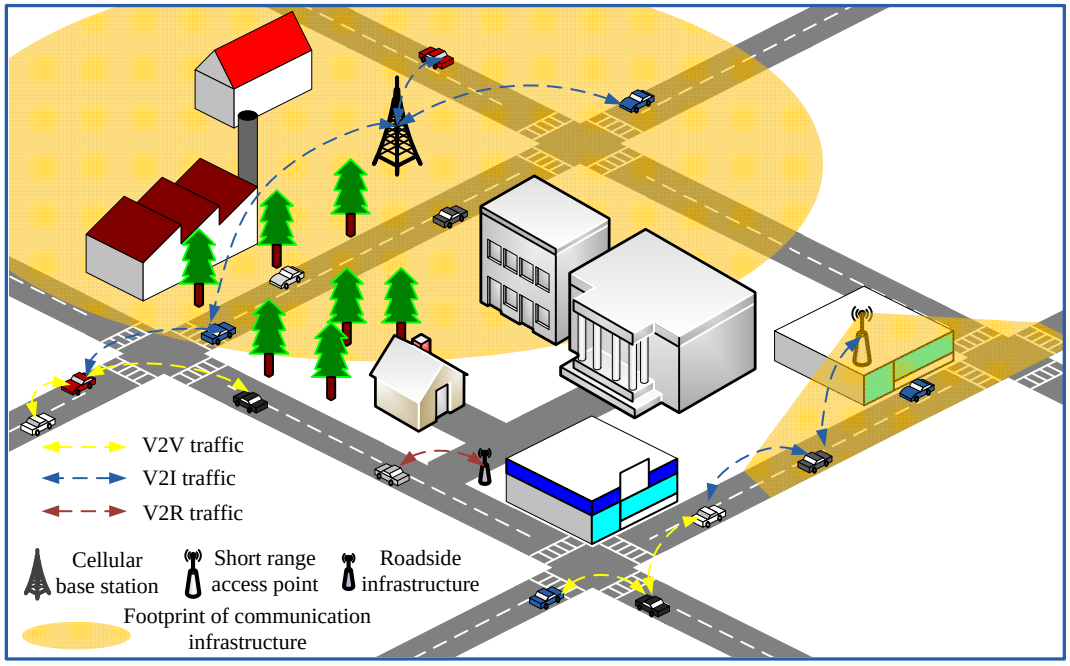


Figure 1.1: An overview of vehicular networks

ernment institutions have initiated numerous activities. An overview of the current and past major programs of the ITS and projects in the USA, Japan, and Europe can be seen in [15]. The standards and standardization process of VANETs are given in [16,17]. There have been tons of research works produced in the past decade to speed up the development of VANETs, including comprehensive surveys (e.g., [15] and [18]).

1.1.1 V2V Connectivity

It is widely believed that the advances of V2V or inter-vehicle communications will reshape the future of road transportation systems, where inter-connected vehicles are no longer information-isolated islands. By means of V2V communications, information generated by the vehicle-borne computer, control system, on-board sensors or passengers can be effectively disseminated among vehicles in proximity, or to vehicles multiple hops away. Without the assistance of any built infrastructure, a variety of active road safety applications (e.g., collision detection, lane changing warning, and cooperative merging) [19] and infotainment applications (e.g., interactive gaming, and file and other valuable information sharing) [20] are enabled by V2V wireless links.

VANETs have attracted extensive research attentions for many years, and how to establish efficient and reliable wireless links between vehicles is a major research focus. The most cumbersome challenge is to combat the harsh communication environment. In urban scenarios, the line-of-sight (LOS) path of V2V communication is often blocked by buildings at intersections. While on a highway, the trucks on a communication path may introduce significant signal attenuation and packet loss [21]. Field tests in [22] demonstrate that multi-path fading, shadowing, and Doppler effects due to high vehicle mobility and the complex urban environment will lead to severe wireless loss, and with a large scale of vehicles transmitting simultaneously, the mutual interference plays an important role as well. Accurate modeling of the propagation environment is premier to design reliable V2V communication systems. [23] presents an overview of the state of the art of the vehicular channel measurements. It is noteworthy that there is a lack of unified channel model that can be applied for all scenarios (e.g., urban, rural, and highway), and the existing channel models, only for a specific scenario, have their own merits and deficiencies. [23] also provides suggestions for V2V communication systems based on the channel characterization. For example, the adoption of multiple antennas would enhance the communication reliability.

From a network perspective, compared to typical low-velocity nomadic mobile communication systems, VANETs also present unique characteristics that have a significant impact on V2V connectivity.

- The network topology changes frequently and very fast due to high vehicle mobility and different movement trajectory of each vehicle;
- Due to the high dynamics of network topology and limited range of V2V communication, frequent network partitioning can occur, resulting in data flow disconnections; and
- Surrounding obstacles (e.g., buildings and trucks) can lead to an intermittent link to a mobile vehicle.

In addition to the technical challenges, the following features can benefit V2V communications: (i) the vehicle mobility is map-restricted and can be predicted in a certain time interval to a certain degree; (ii) there is no power constraint on communications and each vehicle can have relatively powerful processing capability; and (iii) with the aid of Global Positioning System (GPS), vehicles can locate themselves with an error up to a few meters.

DSRC/WAVE

Dedicated Short-Range Communications (DSRC) is a key enabling wireless technology for both V2V and V2R communications. The U.S. Federal Communication Commission (FCC) has allocated 75 MHz bandwidth at 5.9 GHz spectrum band for DSRC. The dedicated bandwidth is further divided into seven channels to support safety and non-safety services simultaneously. The specifications of DSRC are in the IEEE Standard for Wireless Access in Vehicular Environments (WAVE), including the IEEE 802.11p for PHY and MAC layers and the IEEE 1609 family for upper layers. Many automotive and ICT manufacturers, academia, and governments have responded positively and are actively working in collaboration to bring this promising technology to fruition. There have been extensive research efforts from academia to characterize communication properties of DSRC [22, 24–28], and to enhance DSRC performance both in the PHY layer and MAC layer [29–32].

PHY Layer

DSRC PHY layer adopts almost the same Orthogonal Frequency Division Multiplexing (OFDM) modulation as the IEEE 802.11a/g standard and is able to support a data rate of 3–27 Mbps on a 10 MHz channel [16]. Though simulation [24], empirical study [25], and measurement campaigns [22, 26], the performance of DSRC PHY layer has been well understood. Although several research works have demonstrated that DSRC PHY layer is

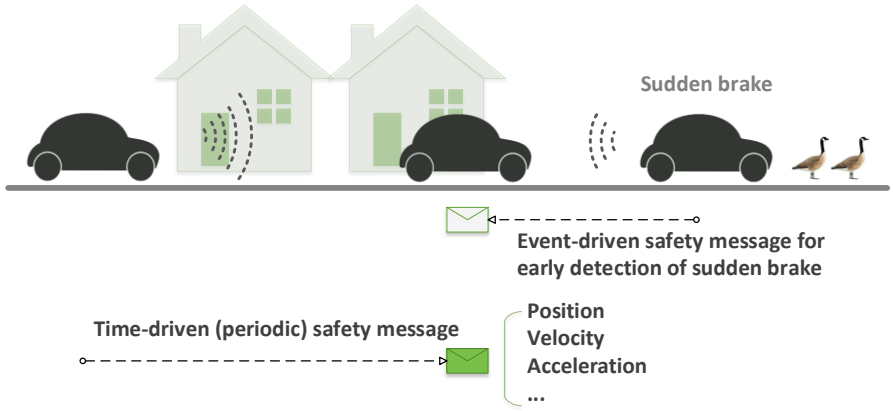


Figure 1.2: Example of safety applications based on V2V communications.

adequate to support safety message delivery, many challenges remain, such as (i) reliable communication is not guaranteed especially when the LOS path is obstructed or the delay spread of wireless channel is too large [27]; (ii) cross-channel interference introduces performance penalty when two adjacent channels are operated simultaneously [16]; and (iii) a gray-zone phenomenon is particularly observed in [26], i.e., the behavior of intermittent loss rate during the transmission. To well fit the vehicular environment, DSRC PHY layer is required to keep evolving. More challenges in this evolution are discussed in [27], such as the difficulty in estimating the channel condition accurately. Some guidelines for OFDM system design in the DSRC PHY layer are also given in [23], such as a suggestion of a modified pilot pattern to reduce receiver complexity.

MAC Layer

Dissemination of safety messages, either time-driven (periodic) or event-driven (as shown in Fig. 1.2), is mostly based on one-hop broadcast, i.e., distributing the same safety message to all the nodes within the communication range, and requires low latency and high reliability, e.g., the dissemination of emergency braking message. However, based on the legacy IEEE 802.11 distributed coordination function (DCF), the current version of DSRC MAC is contention-based and thereby does not support efficient and reliable broadcast services. Specifically, the poor performance of the DSRC MAC in supporting safety applications is mainly due to the high collision probability of the broadcasted packets. For unicast communications using DSRC MAC, the collision probability of two or more transmissions is reduced by the adoption of a two-way handshaking mechanism, i.e., request-

to-send/clear-to-send (RTS/CTS), before the actual data is transmitted. However, the RTS/CTS mechanism and the acknowledgement from data recipients (ACK) are not implemented for broadcast services. As time division multiple access (TDMA) is capable of controlling channel access more precisely, by which the vehicle only needs to listen and broadcast during the acquired time slot, many alternatives have been proposed to guarantee the quality of service (QoS) of safety and other real-time applications in highly densed vehicular scenarios based on TDMA, such as the MAC protocols proposed in [29–32].

1.1.2 V2I Connectivity

Internet connectivity is becoming a must-have feature of modern vehicles. The industry has responded promptly by using off-the-shelf technologies, aiming to build a huge mass market of Internet-connected cars, whereas the academia focuses on the development of optimal solutions to enable connections between vehicles and the Internet. Wireless access technologies play a vital role in delivering the Internet services to vehicle users. Cellular and Wi-Fi are two promising candidates. The cellular networks, such as 3G and 4G-LTE, can provide reliable and ubiquitous access services. The feasibility of using low-cost roadside Wi-Fi access point (AP) for outdoor Internet access at vehicular mobility has also been demonstrated in [33]. We first review the up-to-date status of industrial progress, and then provide an overview of drive-thru Internet.

Industrial Solutions

Earliest concepts of Internet-enabled vehicles were proposed in 1990s in the literature, such as "the Internet multimedia on wheels" [34], "web on wheels" [35], and "the network vehicle" [36]. Nowadays, Internet-integrated vehicle is no longer conceptual due to numerous initiatives in the automotive, telecommunications, and consumer electronics industry. Existing solutions connect vehicles to the Internet through widely deployed cellular network infrastructure, and can be divided into two categories, i.e., brought-in and built-in, advocated by different automobile manufacturers.

The brought-in option caters to 3G/4G mobile users who prefer tethering their own smart phone to the car. The most popular tethering technology, namely *MirrorLink*, is powered by Car Connectivity Consortium (CCC) [37], an organization calling leading automobile (e.g., Volkswagen and Toyota) and ICT manufacturers (e.g., Sony and Nokia) together to create a phone-centric car connectivity solution. By using MirrorLink, a device interoperability standard in essence, the motorist/passengers in a vehicle can connect the

phone to the vehicle infotainment system via wires (e.g., USB) or wirelessly (e.g., Wi-Fi or Bluetooth), so that the vehicle gains immediate access to the Internet and some duplicate functions of smart phones.

Built-in option integrates cellular service in the on-board infotainment system. The Internet connection relies on the built-in cellular module, rather than smart phones of motorist/passengers. For example, through built-in cellular communications, BMW ConnectedDrive [38] combines various elements from online applications, driver assistance, call center services, and solutions to providing Internet connection for in-vehicle mobile devices. The best way to enable Internet connectivity in vehicles is still in debate. Built-in options could provide motorist/passengers with stronger connections and customized services compared to brought-in options. The limitation is that the cellular connectivity cannot evolve once it is embedded.

Drive-Thru Internet

As a popular wireless broadband access technology, Wi-Fi, operating on the unlicensed spectrum, offers the "last-hundred-meter" backhaul connectivity to private or public Internet users. Recent research has demonstrated the feasibility of Wi-Fi for outdoor Internet access at vehicular mobility [33]. The built-in Wi-Fi radio or Wi-Fi-enabled mobile devices in the vehicle can access the Internet when the vehicle is moving in the coverage of Wi-Fi hotspots, which is often referred to as the *drive-thru Internet* [39].

Unlike an indoor Wi-Fi network which only serves stationary or slow-moving users, unique characteristics of drive-thru Internet impose many challenges on reliable and robust Internet access. First of all, high vehicle mobility yields a very short connection time to the Wi-Fi AP, e.g., only several tens of seconds, which greatly limits the amount of data transferred in one connection. For example, the overall connectivity range to a roadside AP is around 500–600 meters, corresponding to a connection time of 18–21 seconds for a vehicle moving at 120 km/h [39]. Moreover, time spent in Wi-Fi association, authentication, and IP configuration before actual data transfer is not negligible. Secondly, V2I communications also suffer from high wireless loss due to the channel fading and shadowing [23]. Thirdly, the Wi-Fi protocol stack is not a specific design for a high mobility environment.

There have been a number of real-world measurements in the literature based on diverse testbed experiments to characterize and evaluate the performance of the drive-thru Internet. In [39] and [40], the drive-thru Internet based on the IEEE 802.11b and 802.11g is evaluated respectively in a planned highway scenario where two APs are closely deployed.

Different vehicle speeds (80, 120, and 180 km/h) and different transport layer protocols (UDP and TCP) are considered. A very important observation is that the drive-thru Internet has a three-phase (*entry*, *production*, and *exit*) characteristic. In the entry and exit phases, vehicles gain less throughput than when in production phase due to such as the weak signal, connection establishment delay, and rate overestimation. To investigate the impact of backhaul capability on drive-thru Internet, a measurement is conducted in [41] on a traffic free road where the interference from other vehicles does not exist. It is evidenced that the performance of the drive-thru Internet suffers a lot from the limitation of the backhaul network. For example, with a 1 Mbps bandwidth backhaul capability, the data volume transferred within a drive-thru reduces from 92 MB to 25 MB. Moreover, a backhaul connection with 100 ms one-way delay significantly degrades the performance of web services due to the time penalty of HTTP requests and responses. The problems that may cause the performance degradation of the drive-thru Internet are thoroughly discussed in [42]. Experiments in [33] and [43] are conducted for large-scale urban scenarios with multiple vehicles. The data sets used in the experiments are collected from the city of Boston with in situ open Wi-Fi APs. It is shown that in [33] with a fixed 1 Mbps data rate, vehicles can gain a median upload throughput of 240 Kbps and a median one-drive-thru uploaded data volume of 216 KB. In addition, the average connection and inter-connection time are shown to be 13 seconds and 75 seconds, respectively. The experiment in [43] shows a long-term throughput of 86 kbps averaged over both connection and inter-connection periods.

Due to high vehicle speeds, intermittent links, and potential severe channel contentions, the throughput of per drive-thru is limited as observed in real-world tests. It fundamentally restricts the quality of service (QoS) of data applications. To improve the performance of drive-thru Internet, solutions in the literature include: (i) reducing connection establishment time [43]; (ii) improving transport protocols to deal with the intermittent connectivity and wireless transmission losses [43]; (iii) enhancing MAC protocols for high mobility [42–44]; (iv) designing efficient handoff schemes to deal with the frequent disruptions [9]; (v) using the multi-hop V2V communication capability for relaying data traffic [45]; (vi) exploiting cooperation among vehicles [46, 47]; and (vii) optimal deployment of Wi-Fi APs [48].

1.1.3 V2R Connectivity

V2R connectivity is critical to avoid or mitigate the effects of road accidents, and to enable the efficient management of ITS. DSRC/WAVE is considered a key technology to enable connections between vehicles and ITS infrastructure, such as traffic lights, street signs,

and roadside sensors. Moreover, roadside infrastructures can also be commercial content providers, such as the roadside unit (RSU) broadcasting flyers of superstores [49]. RSU does not necessarily serve as Internet gateway as wireless infrastructures in V2I communications. Visible light communication (VLC), transmitting data by using light-emitting diodes (LEDs), has also been proposed for road-to-vehicle ITS applications, such as traffic light control at intersections [50–52]. As a key technology specified in the IEEE 802.15.7 standard, VLC can support a data rate up to 96 Mbps through fast modulation of LED light sources [53]. VLC is becoming an intriguing complement to DSRC for light-of-sight communication scenarios. To combat outdoor optical noises, however, advanced receiver is required, such as high-speed camera [50, 51], which may incur a high implementation cost. Compared to the mature IEEE 802.11-based technology, VLC is still in the introductory phase and substantial efforts are needed before it can be widely deployed for short-range ITS applications.

1.2 Capacity Scaling Laws

Wireless networks have received a myriad of research attentions over the past decades, including medium access control, routing, security, cooperation, and energy-efficiency, among others. Despite significant advances in the field of wireless networking, a fundamental question remains open: how much information can a wireless network transfer? To answer this question, we should resort to the study of network capacity which is a central concept in the field of network information theory [54]. Intuitively, if the capacity of a wireless network is known, the network limit of information transfer would be obtained. Moreover, having such knowledge would shed light on what the appropriate architectures and protocols were for operating wireless networks. Although significant efforts have been put on the investigation of network capacity, developing a general theory of such fundamental limit for wireless networks is a long standing open problem [55]. In [56], Claude Shannon successfully determined the maximum achievable rate, called the capacity, for a point-to-point communication channel, below which the reliable communication can be implemented while above which the reliable communication is impossible. However, general wireless networks with sources and destinations sharing channel resources are much more complex, making the quest for fundamental limits of wireless networks a formidable task. For example, even for a simple-looking three-node relay channel [57], the exact capacity still has yet to be determined.

As a retreat when exact fundamental limits are out of reach, capacity scaling laws, first studied by Gupta and Kumar in [58], characterize the trend of node throughput behavior

when the network size increases. The most salient feature of capacity scaling studies is to depict the capacity as a function of the number of nodes in the network, without distractions from minor details of network protocol. This approach is quite different from that of studying network information theory, which is to determine exact capacity region of wireless networks. The seminal work [58] not only provides an alternative and tractable way to study the network capacity, but also obtains insightful capacity results. Great efforts have been made thereafter to derive capacity scaling laws for different paradigms of wireless networks. Scaling laws for network delay and its trade-off with capacity have also been investigated. The study of scaling laws can lead to a better understanding of intrinsic properties of wireless networks and theoretical guidance on network design and deployment [59]. Moreover, the results could also be applied to predict network performance, especially for the large-scale networks [60]. We provide the following illustration. We consider to deploy a large-scale sensor networks for a certain geographic area. Capacity scaling laws show that the network scales poorly when the number of sensors grows, i.e., the throughput of each sensor would decrease. In order to enhance the throughput capacity, we may need to adopt some advanced technologies, such as directional antennas and network coding. However, scaling laws show that exploiting network coding cannot change the trend of throughput capacity; whereas exploiting directional antennas can introduce capacity gains (refer to Chapter 2.2.1). Furthermore, suppose we have deployed a sensor network of 100 sensors with directional antennas. Typically we can obtain the throughput performance (denoted by λ_A) of the network through real measurement. If we need to extend the network to a larger one of 1000 sensors, with the same network settings, by capacity scaling results (denoted by $f(N)$), we are able to have a rough idea that how much throughput (denoted by λ_B) can be supported by the network that we are going to deploy, i.e., $\lambda_B = \lambda_A \cdot f(1000)/f(100)$. We survey the literature in this area in Chapter 2.

1.3 Capacity of Vehicular Networks

The capacity scaling of VANETs is desirable since unlike generic mobile ad hoc networks, VANETs present unique characteristics, which impose distinguished challenges on networking. i) *Large scale*: the VANET is an extremely large-scale mobile network, which is deployed in a large geographic area with a great amount of vehicles and roadside units; ii) *Cars on the road:* the movement of vehicles should follow certain street pattern, different from generic mobile ad hoc networks in which nodes typically move in a free space; iii) *Cars on the wheels:* the vehicle mobility is related to the road traffic environment and the social life of the driver; iv) *Spatio-temporal variations:* there are spatio-temporal variations

of vehicle density and link quality due to vehicle mobility and unstable wireless channels, respectively; and v) *Diversified applications:* applications of VANETs are of a large variety and with different QoS requirements. All these features dramatically complicate studies of scaling laws.

There have been a few efforts to investigate the capacity scaling laws of VANETs. Pishro-Ni *et al.* [61] initiated the study of capacity scaling for vehicular networks with an emphasis on the impact of road geometry on the network capacity. Nekoui *et al.* [62] specially developed a novel notion of capacity for safety applications, which is called *Distance-Limited Capacity*. That is the capacity of VANETs when a pair of vehicles can only communicate if the two vehicles reside in a certain distance of each other. Both [61] and [62] showed that the road geometry has an important role in the capacity of vehicular networks. As the demand of public information dissemination is high in VANETs, multicast flows, in which one source is associated with a set of destinations, may be viable to be deployed for practical applications. In [63], Zhang *et al.* analyzed multicast capacity of hybrid VANETs, in which base stations are deployed to support communications between vehicles. It was assumed that each vehicle is equipped with a directional antenna. By respectively applying the one-dimensional and two-dimensional i.i.d mobility model (refer to Chapter 2.3.1) to vehicles, they derived bounds of the multicast throughput capacity under certain end-to-end delay constraint. In [64], Wang *et al.* studied the uplink capacity of hybrid VANETs, where each vehicle, following random way-point mobility, is required to send packets to regularly placed sink roadside units (RSUs). The basic routing strategy adopted in [64] is to distribute source packets to as many RSUs as possible to increase concurrent uploading opportunities.

One of the limitations of [63] and [64] is that the specific mobility features of vehicles are not fully considered. The i.i.d mobility is not practical for vehicular scenarios. Moreover, the assumption that vehicles are uniformly distributed in the network is also unrealistic. In urban areas, vehicle densities in different regions may be highly diverse. Inhomogeneous vehicle densities were considered in [65], which investigates the throughput capacity of social-proximity VANETs. We will elaborate this work in Chapter 3. Infrastructure, such as Wi-Fi access points and cellular base stations, plays a vital role in providing pervasive Internet access to vehicles. Reference [66] analyzes the downlink capacity of vehicles, i.e., the maximum average downlink rate achieved *uniformly* by vehicles, for each type of access infrastructure considered, and investigates the capacity-cost tradeoffs for access infrastructures since the deployment costs of different access infrastructure are highly variable. We will elaborate the capacity analysis considering access infrastructure in Chapter 4 and Chapter 5.

1.4 Thesis Outline

The rest of this thesis is organized as follows: Chapter 2 presents a comprehensive overview of scaling laws for throughput capacity and delay in wireless networks. Chapter 3 investigates the capacity scaling laws of delay-tolerant vehicular networks in which services/applications are delay-tolerant and can be delivered in a store-carry-and-forward fashion. Chapter 4 analyzes the downlink capacity of vehicles and investigates the capacity-cost tradeoffs for the network in which access infrastructure is deployed to provide a downlink data pipe to all vehicles. Chapter 5 particularly investigates the average throughput capacity of drive-thru Internet access considering interrupted vehicle traffic flow and with a focus on cost-effectiveness of Wi-Fi solution for vehicular Internet access. Finally, Chapter 6 concludes the thesis, and points out our future research directions.

Chapter 2

Background

The capacity scaling law of wireless networks has been considered as one of the most fundamental issues. In this chapter, we aim at providing a comprehensive overview of scaling laws for throughput capacity and delay in wireless networks. We begin with background on the notion of throughput capacity of random networks. Based on the benchmark random network model, we then elaborate the advanced strategies adopted to improve the throughput capacity, and additional factors that affect the scaling laws. We also present the fundamental tradeoffs between throughput capacity, delay, and mobility. Finally, the capacity for hybrid wireless networks are surveyed, in which there are at least two types of nodes functioning differently, e.g., normal nodes and infrastructure nodes.

2.1 Preliminaries: Milestone of Throughput Capacity Scaling

Capacity scaling laws offer fundamental understanding on how per-node capacity scales in an asymptotically large network. The line of investigation began with [58], where Gupta and Kumar introduced two new notions of network capacity: *transport capacity* and *throughput capacity*. In this dissertation, we focus on the throughput capacity. We first introduce the notion of throughput capacity and the capacity result for random networks, as preliminaries for reading the remaining sections in this chapter.

2.1.1 Notion of Throughput Capacity

Let N denote the number of nodes in a network. The per-node throughput of the network, denoted by $\lambda(N)$, is the average transmission rate, measured in bits or packets per unit time, that can be supported uniformly for each node to its destination in the network. A per-node throughput of $\lambda(N)$ bits per second is said to be ***feasible*** if there exists a spatial and temporal scheme for scheduling transmissions, such that each node can send $\lambda(N)$ bits per second on average to its destination node. The throughput capacity of the network is said of order $\Theta(f(N))$[1] bits per second if there are deterministic constants $c_1 > 0$ and $c_2 < \infty$ such that

$$\lim_{N\to\infty} \mathbf{Pr}\big(\lambda(N) = c_1 f(N) \text{ is feasible}\big) = 1$$
$$\liminf_{N\to\infty} \mathbf{Pr}\big(\lambda(N) = c_2 f(N) \text{ is feasible}\big) < 1.$$

Therefore, vanishingly small probabilities are allowed for in this definition of "throughput capacity" when considering the randomness involved in the network, such as the location and the destination of each node. Note that the notion of throughput capacity is different from the information-theoretic capacity notion that describes the exact region of simultaneous rates of communications from many senders to many receivers in the presence of interference and noise [67].

2.1.2 Random Networks

A wireless random network consisting of N identical immobile nodes randomly located in a disk of unit area in the plane and operating under a ***multi-hop*** fashion of information transfer, is shown in Fig. 2.1 [58]. Each node having a randomly chosen destination is capable of transmitting at W bits per second over a common wireless channel. The requirements for successful transmission are described as per two interference models: i) the Protocol Model, which is a binary model, i.e., the transmission is successful if there is enough spatial separation from simultaneous transmissions of other nodes otherwise fails;

[1]Since studies of throughput capacity focus on the scaling behavior instead of a specific value, the order notation is involved to describe how capacity scales with the number of nodes N. Specifically, the following Knuth's notation is used throughout all the papers on scaling laws: given nonnegative functions $f_1(n)$ and $f_2(n)$, $f_1(n) = O(f_2(n))$ means $f_1(n)$ is asymptotically upper bounded by $f_2(n)$; $f_1(n) = \Omega(f_2(n))$ means $f_1(n)$ is asymptotically lower bounded by $f_2(n)$; and $f_1(n) = \Theta(f_2(n))$ means $f_1(n)$ is asymptotically tight bounded by $f_2(n)$; $f_1(n) = \omega(f_2(n))$ means $f_1(n)$ is asymptotically dominant with respect to $f_2(n)$; $f_1(n) = o(f_2(n))$ means $f_1(n)$ is asymptotically negligible with respect to $f_2(n)$.

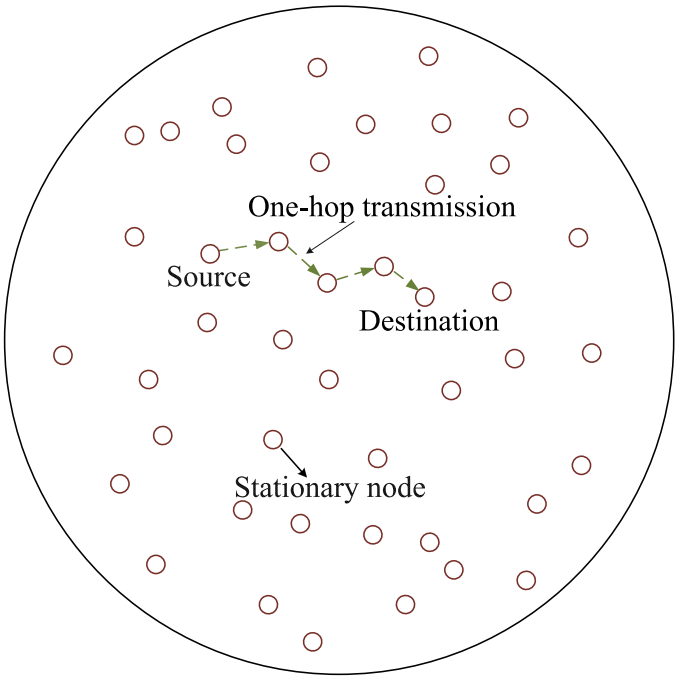


Figure 2.1: A static ad hoc network in a unit disk.

and ii) the Physical Model, based on signal-to-interference ratio requirements. In such a static random ad hoc network, all the nodes are assumed to be homogeneous, i.e., all transmissions employ the same range or power, and wish to transmit at a common rate.

2.1.3 Throughput Capacity of Random Networks

For random networks, the order of the throughput capacity is $\lambda(N) = \Theta(\frac{W}{\sqrt{N \log N}})$ under Protocol Model (see main result 3 in [58]); while under Physical Model, the throughput capacity is given by $\Theta(\frac{W}{\sqrt{N \log N}}) \leq \lambda(N) < \Theta(\frac{W}{\sqrt{N}})$ (see main result 4 in [58]). An explanation of the results is as follows. For Protocol Model, the lower bound and upper bound are of the same order such that there exists a sharp order estimation of the throughput capacity; for Physical Model, a throughput of order $\Theta(\frac{W}{\sqrt{N \log N}})$ is feasible, while $\Theta(\frac{W}{\sqrt{N}})$ is not. Fig. 2.2 gives three examples to show the trend of throughput capacity in the order sense.

The throughput capacity is studied asymptotically, i.e., capacity scaling law results hold with high probability when the population of nodes is larger than some threshold; on

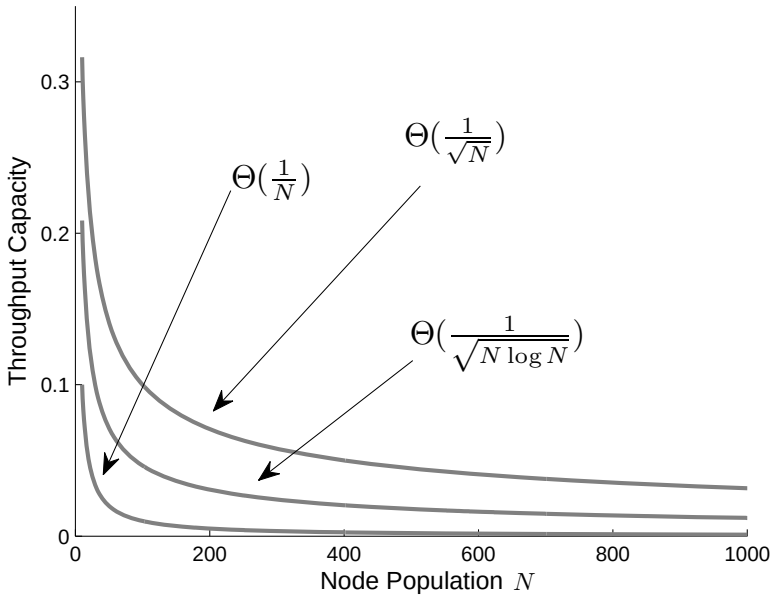


Figure 2.2: Examples of showing throughput capacity trend in the order sense

the other hand, results may not hold, or hold with small probability if the population of nodes is small. The scaling result for random networks is pessimistic because the per-node throughput tends to zero similar to $\frac{1}{\sqrt{N \log N}}$ as the population of nodes goes to infinity, which indicates that static ad hoc networks are not feasible to scale to a large size. What causes such discouraging results? The fundamental reason is that every node in the network needs to share the channel resources or certain geographic area with other nodes in proximity, which constricts the capacity. Specifically, concurrent wireless transmissions in a wireless network limit its throughput capacity, because they create mutual interference so that nodes cannot communicate as that in the wireline network where much less mutual interference exists. This interpretation also demonstrates how desirable it is to mitigate the mutual interference in wireless communications, although it is very challenging.

2.2 Throughput Capacity of Ad Hoc Networks

2.2.1 Strategies to Improve Throughput Capacity

One natural question is if it is possible to improve throughput capacity of random networks by employing any advanced techniques or sophisticated strategies. After significant

progress that has been made to further the investigation on throughput capacity scaling, the answer is positive.

First of all, by allowing both long-distance and short-distance transmissions, the throughput capacity can be improved slightly to $\Theta(\frac{1}{\sqrt{N}})$ [68]. The scheme constructed to achieve this throughput relies on multi-hop transmission, pairwise coding and decoding at each hop, and a time-division multiple access. The gain of throughput capacity can also be achieved by employing directional antennas. Yi *et al.* in [69] considered different beamform patterns, and showed that the throughput capacity can be achieved with a gain of $\frac{4\pi^2}{\alpha\beta}$ using directional transmission and reception, where α and β are antenna parameters. A capacity gain of $\Theta(\log N)$ is proved in [70]. Peraki *et al.* in [71] further revealed that the maximum capacity gain is $\Theta((\log N)^2)$ by using directional antennas at the transmitters and receivers, corresponding to a throughput of $\Theta((\log N)^{3/2}/\sqrt{N})$. If nodes have multi-packet reception (MPR) capabilities, i.e., a receiver is capable of correctly decoding multiple packets transmitted concurrently from different transmitters, the capacity gain can also be achieved. Sadjadpour *et al.* in [72] showed that with MPR, the throughput capacity of random ad hoc networks can be improved at least by an order of $\Theta(\log N)$ and $\Theta((\log N)^{\frac{\alpha-2}{2\alpha}})$ under Protocol Model and Physical Model, respectively, where α is the path loss exponent in the Physical Model. Similar research efforts applying MPR can be found in [73–75].

By means of long-range multiple-input multiple-output (MIMO) communications with local cooperations as proposed in [76], significant improvement of throughput capacity scaling in random networks is attainable [77], i.e., almost constant per-node throughput of $\Theta(n^{-\epsilon})$ on average is achievable, where $\epsilon > 0$ can be arbitrarily small. This yields an aggregate throughput ($N\lambda(N)$) of $\Theta(N^{1-\epsilon})$ for the whole network, indicating almost linear capacity scaling in N. $\epsilon = \Theta(\frac{1}{\sqrt{\log N}})$ was explicitly obtained later in [78] and [79]. However, the capacity gain is at the cost of increased system complexity due to the intelligent hierarchical cooperation among nodes. Regardless the complexity of the constructed strategy, the result in [77] is inspiring but still controversial. Franceschetti *et al.* in [80] claimed that a throughput higher than $O((\log N)^2/\sqrt{N})$ cannot be achieved because of degrees of freedom limitation which is a result of laws of physics. Artificial assumptions and models lead to the impossible linear capacity scaling in [77]. While using Maxwell's equations without any artificial assumptions, Lee *et al.* in [81] established the capacity scaling laws for the line-of-sight (LOS) environments, which show that a linear scaling of the aggregate throughput is indeed possible for static random networks. Thus, the conflict between [77] and [80] is resolved. It is worth noting that even if such physical limits in [80] do exist and sophisticated strategies like the hierarchical cooperation cannot further improve the per-node throughput ($\Theta(1/\sqrt{N})$) in the scaling limit sense,

Table 2.1: Summary: Strategies to Improve Throughput Capacity

Strategies	Throughput Capacity Gain [Compared to $\Theta(\frac{1}{\sqrt{N \log N}})$]
Power Control	$\Theta((\log N)^{1/2})$ [68]
Directional Antenna	$\Theta((\log N)^2)$ [71]
Multi-Packet Reception	$\Theta(\log N)$ for Protocal Model;
	$\Theta((\log N)^{\frac{\alpha-2}{2\alpha}})$ for Physical Model [72], $\alpha \geq 2$
Hierarchical Cooperation	Almost $\Theta((N \log N)^{1/2})$ [77]
Network Coding	Constant gain [82–86]
Ultra-wideband	$\Theta(N^{\alpha/2}(\log N)^{1/2})$ [87], $\alpha \geq 2$
Mobility	$\Theta((N \log N)^{1/2})$ [88]

* α is the path-loss component.

these strategies generally could be considerably beneficial in networks of any finite size. An example is the physical-layer network coding. In [82], it was shown that although the physical-layer network coding scheme does not change the scaling law, it improves throughput performance of the network in the sense by enlarging the constant component of the scaling result. The similar studies applying the network coding schemes can be found in [83–86].

Since the above research works are based on the assumption that the network is bandwidth-constrained, i.e., each node is only capable of transmitting at W bits per second, it is interesting to consider a scenario where each node has power constraint but can utilize unlimited bandwidth. Hence, there have been a few research efforts which focus on the ultra-wideband (UWB) techniques. In [89], Negi and Rajeswaran showed that under the limiting case $W \to \infty$, the throughput capacity is lower bounded by $\Omega\big(P_0\sqrt{N^{\alpha-1}/(\log N)^{\alpha+1}}\big)$ and upper bounded by $O(P_0(\sqrt{N \log N})^{\alpha-1})$, where α is the path loss exponent and P_0 is the maximum transmission power. The gap between the upper bound and the lower bound was closed by Tang and Hua in [90]. They showed that the throughput capacity of a UWB power-constrained ad hoc network is given by $\Theta(P_0(\sqrt{N/\log N})^{\alpha-1})$. A better result was obtained in [87] that the throughput capacity scales as $\Theta(P_0 N^{(\alpha-1)/2})$.

Without leveraging aforementioned advanced techniques in the static random network, what if nodes move? The effect of node mobility on throughput capacity scaling was first investigated by Grossglauser and Tse in [88]. By applying an i.i.d mobility model (see Section 2.3) to each node, they have shown that the per-node throughput of the mobile ad hoc network could remain constant, i.e., $\Theta(1)$, by using a ***two-hop relaying***

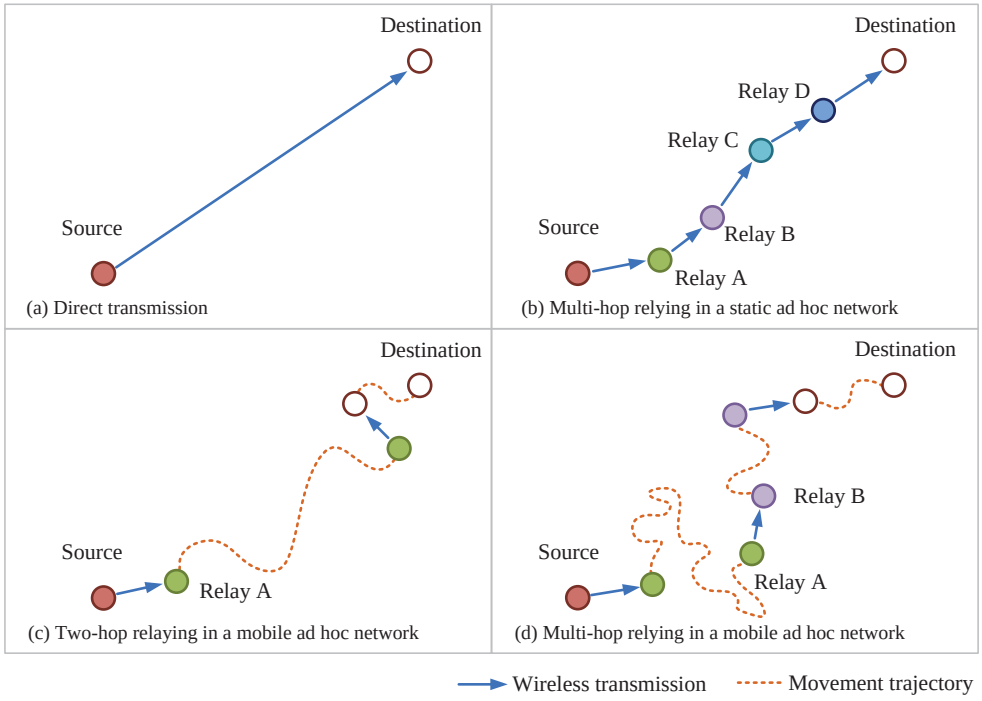


Figure 2.3: An illustration of packet transmission strategies

scheme (see Fig. 2.3(c)) and allowing finite but arbitrary delay. This result provides an interesting implication that dramatic gains in network capacity are possible when mobility is considered so that the nodes can exploit mobile relays to carry packets to distant nodes. Compared with such *store-carry-and-forward* communication paradigm, in the absence of mobility, direct transmission (see Fig. 2.3(a)) between distant nodes causes too much interference, or equivalently requires a large spatial area, so that the number of concurrent transmissions are reduced; on the other hand, if the network only allows the communication between nearest neighbors (see Fig. 2.3(b)), most of the packets will be delivered through multiple hops, resulting in the decrease in throughput capacity as well. Inspired by the promising result in [88], extensive works have been done to investigate capacity scaling in mobile ad hoc networks. In [91], Diggavi *et al.* have shown that even one-dimensional mobility benefits capacity scaling. Restricted to move on a great circle, each node can attain a constant throughput. In [92], Syed Ali Jafa explored the capacity of high mobile ad hoc networks in the presence of channel uncertainty, and has shown that high mobility introduces rapid channel fluctuations and hence limits the capacity of wireless networks. A summary of capacity gains by using each strategy is given in Table 2.1.

2.2.2 Additional Factors Affecting Scaling Laws

The random network considered in [58] is a benchmark network model, in which nodes have basic communication capabilities (i.e., simple coding and decoding strategies implemented on the single radio), and the traffic model (symmetric unicast) and interference model (Protocol Model or Physical Model) are simplified. Besides the strategies mentioned in Section 2.2-A to improve throughput capacity, significant research efforts have been made to study the impact of different modeling factors on capacity scaling laws.

Multi-channel multi-interface: In [58], it has been shown that with a single radio mounted on each vehicle, splitting the total bandwidth W into multiple sub-channels does not change the order of throughput capacity of random networks. However, in practice, a communication device may have multiple radio interface operating on one or different channels. What if each node is equipped with multiple radio interfaces? In [93] and [94], Kyasanur and Vaidya derived the capacity scaling laws for a general multi-channel networks with $l \leq c$, where c is the number of channels, and l is the number of interfaces per node. It was shown that different ratios between c and l yield different capacity bounds, and in general, the network capacity is reduced except when c is upper bounded by $O(\log N)$. Kodialam and Nandagopal [95] also provided capacity trends for multi-channel multi-interface wireless mesh networks by considering channel assignment and scheduling.

Channel model: Most research efforts follow either Protocol Model (governed by geometry) or Physical Model (governed by path loss), which only characterize the deterministic behavior of wireless channel connection. To consider the randomness which is more realistic, several works have been done assuming different channel models. The impact of channel fading on capacity scaling was studied by Toumpis and Goldsmith [96]. They showed that a throughput of $\Theta(\frac{1}{\sqrt{N(\log N)^3}})$ is feasible under a general model of fading for static random networks. The Rayleigh fading was considered by Ebrahimi *et al.* [97] for a single-hop scenario and the lower and upper bounds of throughput capacity were derived. The random connection model was considered in [98] and [99], i.e., the signal strengths of the connections between nodes are independent from each other and follow a common distribution. In [98], a throughput capacity of $\Omega((\log N)^{-d})$ is proved for some $d > 0$. In [99], Cui *et al.* showed that a constant throughput is achievable by relaxing some constraints of the connection model. In [100], Gowaikar and Hassibi considered a *hybrid* channel connection model: for a short distance between transmitter and receiver, the channel strengths are governed by the random connection model; while for a long range, the channel strengths are governed by a Rayleigh distribution. They showed that a throughput of $\Theta\big(\frac{1}{(\log N)^4}\big)$ is achievable. The lower bound on the capacity of wireless erasure networks was reported by Jaber and Andrews [101], in which an erasure channel model

Table 2.2: Summary: Additional Factors Affecting Scaling Laws

Factors	Main Results on Throughput Capacity (λ)
Multi-channel &	No capacity gain when $c/l = O(\log N)$;
Multi-interface*	Capacity loss when $c/l = \Omega(\log N)$ [94]
Channel Model	$\lambda = \Theta(\frac{1}{\log N} \cdot \frac{1}{\sqrt{N \log N}})$ under general fading model [96]
	$\lambda = \Theta(\log N/N)$ under Rayleigh fading (single-hop scenario) [97]
	$\lambda = \Omega((\log N)^{-d})$ under random connection model for some $d > 0$ [98]
	$\lambda = \Theta(1)$ under random connection model for some cases [99]
	$\lambda = \Theta(1/(\log N)^4)$ under hybrid random connection model [100]
	$\lambda = \Omega(\sqrt{\log N/N})$ under independent erasure channel model [101]
	$\lambda = \Omega(\frac{1}{\sqrt{N \log N}})$ under correlated erasure channel model [101]
Network Topology	Symmetric topology yields a high throughput capacity [102]
	$\lambda = \Theta\big(\frac{1}{N^{\frac{1}{3}}(\log N)^{\frac{2}{3}}}\big)$ under Protocol Model for 3-D networks [103]
	$\lambda = \Theta(1/N^{\frac{1}{3}})$ under Physical Model for 3-D networks [103]
Traffic Pattern**	A unified framework for (N, m, k)-casting [105]

* c denotes the number of channels and l denotes the number of interfaces per node.

** m denotes the number of intended recipients of a source packet and k denotes the number of successful recipients.

is considered, i.e., each channel is associated with an erasure probability. Such a channel model incorporates erasure events which may correspond to packet drops or temporary outages when transmission is undergoing. It is proved that the capacity lower bound scales as $\Theta(\sqrt{\log N/N})$ and $\Theta(\frac{1}{\sqrt{N \log N}})$ with independent and correlated erasure channels, respectively.

Network topology: The shape of geographic area where the network is deployed has a significant impact on capacity scaling laws. Hu *et al.* [102] investigated the effect of various geometries, including the strip, triangle, and three-dimensional cube. The main implication from [102] is that the symmetry of the network shape plays an important role. In other words, a high throughput capacity can be achieved if the network is symmetric. In addition to two-dimensional (2-D) networks, several efforts have been put on investigation of three-dimensional (3-D) networks. In [103], a throughput capacity of $\Theta\big(\frac{1}{N^{\frac{1}{3}}(\log N)^{\frac{2}{3}}}\big)$ and $\Theta(1/N^{\frac{1}{3}})$ is reported for 3-D random networks under Protocol Model and Physical Model, respectively. In [104], Li *et al.* respectively derived the capacity bound for the 3-D network with regularly and heterogeneously deployed nodes.

Traffic pattern: Besides symmetric unicast, i.e., each node is only the source of one

unicast flow and the destination of another, dissemination of information in other fashions has been extensively studied in the literature. The broadcast capacity is reported in [106–108], which is the maximum per-node throughput of successfully delivered broadcast packets. For each broadcast packet, it is successfully delivered if all nodes in the network other than the source receive the packet correctly in a finite time. The multicast capacity has been widely investigated [109–115] considering different network settings. By employing multicast, each packet is disseminated to a subset of $N-1$ nodes which are interested in the common information from the source. Nie [116] reported a short survey on multicast capacity scaling. A unifying study was provided by Wang *et al.* [105], in which how information is disseminated is generally modeled by the (N, m, k)-casting. In this particular context, m and k denote the number of intended recipients of a source packet and the number of successful recipients, respectively. For unicast, $m = k = 1$; for multicast, $k \leq m < N$; and for broadcast, $k \leq m = N - 1$. The capacity bounds were established in [105] for each type of traffic pattern. A summary for this subsection is given in Table 2.2.

2.3 Fundamental Tradeoffs: Capacity, Delay, and Mobility

Capacity is not the only metric to evaluate network performance. From applications point of view, network delay (its average, maximum, or distribution) is also an important design aspect [59]. In [88], it has been shown that striking performance gains in throughput capacity are achievable in mobile ad hoc networks, however at the expense of enlarged delay. With the same time scale of node mobility, the delay is incurred by the ***movements*** of the relay (transmitter) and the destination (receiver) since they have to be geographically close enough for transmission, as shown in Fig. 2.3(c). Basically, there are two ways to transfer an information packet from the source to the destination: wireless transmission and node movement. Since wireless transmission is typically at a much smaller time scale, the time spent on the relay movements towards the destination contributes to the major component of the delay. There is a tradeoff between capacity and delay: if an increase in throughput is desired, we should reduce the distance of wireless transmission to allow more concurrent transmissions in the network; while if a decrease in delay is desired, we should reduce the distance of relay movement towards the destination. However, it is impossible to reduce both distances simultaneously given a fixed distance between the source and the destination. Furthermore, intuitively, different mobility models may incur different delays, because the node movement pattern determines the time spent on the relay movements.

For example, if a node always wanders around (see Relay A in Fig. 2.3(d)), it is very difficult for the node to move a long distance in one direction. To understand the tradeoffs between capacity and delay of wireless networks, a large body of research studies have been done under a variety of mobility models.

2.3.1 Mobility Models

The type of node mobility studied includes the i.i.d mobility, random walk model, random way-point model, Brownian motion, and Lévy mobility. Besides, there are two more general mobility models defined in [117] to study the relationship between delay and throughput capacity from a global perspective. We introduce these mobility models in the following and give a brief summary in Table 2.3 as well.

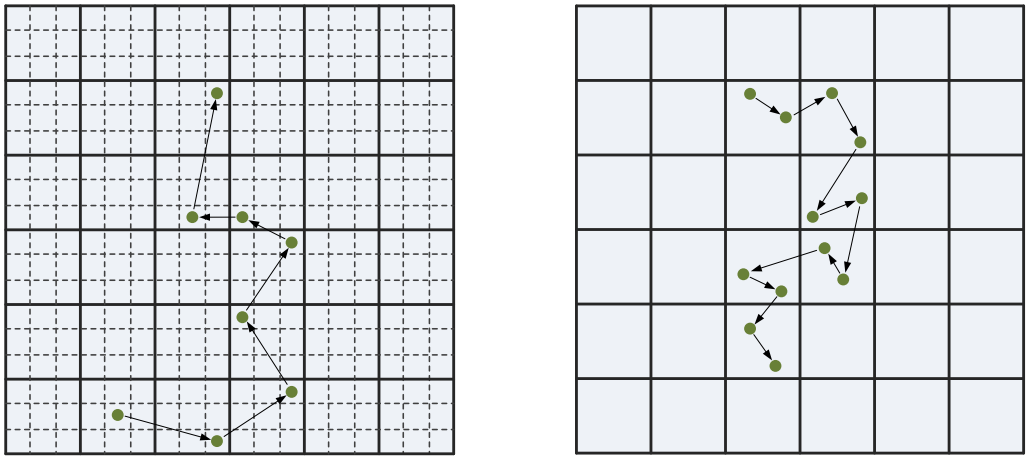
- i.i.d Mobility Model: In time-slotted system, at each time slot, each node selects a new position independently and identically distributed over all positions in the network. The position distributions of the nodes are independent between time slots. The i.i.d mobility is also referred to as the reshuffling model [118]. Depicting an extreme mobility, the i.i.d mobility model is unrealistic but analytically tractable.

- Random Walk Model: Random walk can be described by Markovian dynamics from i.i.d mobility and is often considered symmetric, i.e., nodes select new positions for next time slot equally likely from the set of current neighboring positions.

- Random Way-Point Model: In random way-point model, at each time slot, the mobile node chooses a random destination in the network and moves toward it at a random speed. The node pauses for some random time after reaching the destination, and then repeats this process.

- Brownian Motion: Brownian motion is like the motion conducted by a small particle totally immersed in a liquid or gas. Brownian mobility has a strong connection with random walk model and is a limiting case when taking smaller and smaller steps in smaller and smaller time intervals in symmetric random walk [119].

- Lévy Walk and Lévy Flight: Lévy mobility is a special type of random walk in which the distribution of flights, i.e., step-lengths, is heavy tailed. In other words, the trajectory of Lévy mobility contains many short flights and an exponentially small number of long flights. The difference between Lévy walk and Lévy flight is that the former has constant flight speed and the latter has constant flight time [120]. It is



(a) Hybrid random walk (b) Random direction model

Figure 2.4: An example of trajectories of hybrid random walk (a) and random direction model (b), respectively.

reported that Lévy mobility has certain statistical similarity to human mobility and some animals' hunting patterns [121].

- Hybrid Random Walk Models: A family of hybrid random walk models is considered in [117] and characterized by a single parameter $\beta \in [0, \frac{1}{2}]$. The unit area of the network is divided into $N^{2\beta}$ equal-sized squares, each of which is further divided into $N^{1-2\beta}$ equal-sized sub-squares. At the beginning of each time slot, each node jumps from its current sub-square to a random sub-square of one uniformly selected neighboring square, as shown in Fig. 2.4(a). It can be seen that the model turns to the i.i.d mobility model and the random walk model when $\beta = 0$ and $\beta = 1/2$, respectively.

- Discrete Random Direction Models: A family of discrete random direction models is also considered in [117] and characterized by a single parameter $\alpha \in [0, \frac{1}{2}]$. The unit area of the network is divided into $N^{2\alpha}$ squares with equal area. The movement of each node is of the following pattern: at the beginning of each time slot, the node jumps from its current square to a uniformly selected neighboring square; and during the time slot, the node moves from a start position to an end position at a certain velocity, as shown in Fig. 2.4(b). The two positions are uniformly selected from all the positions in the square. It can be seen that the above mobility model turns to the random way-point model and the discrete Brownian motion when $\alpha = 0$ and

Table 2.3: Summary of mobility models

Mobility models	Key features
i.i.d Mobility	No motion constraints
Random walk	Next position is chosen from current neighboring positions
Randomly way-point	Randomly selected destination, speed, and pause duration
Brownian motion	The limit form of random walk
Lévy walk	Heavy-tailed distribution of flights; constant flight speed
Lévy flight	Heavy-tailed distribution of flights; constant flight time
Hybrid random walk	Hybrid of i.i.d mobility and random walk
Discrete random direction	Hybrid of randomly way-point and discrete Brownian motion

$\alpha = 1/2$, respectively.

2.3.2 Tradeoffs Between Throughput Capacity and Delay

The throughput capacity and delay under the i.i.d. mobility model were reported by Neely and Modiano [122] for a cell-partitioned ad hoc network. They found that a general delay-throughput tradeoff can be established: the ratio of delay and throughput is at least $O(N)$ under different scheduling policies (i.e., two-hop or multi-hop relaying) with or without packet redundancy[2]. The optimality of delay-capacity tradeoffs under i.i.d. mobility model was studied in [123]. Different time scales of node mobility are taken into consideration: fast mobility, only allowing one-hop transmissions during a time slot after which the topology changes; and slow mobility, allowing packets to be delivered through multiple hops during a time slot since the mobility of nodes is much slower than packet delivery time. It was shown that under i.i.d. fast mobility, a per-node capacity is $O(\sqrt{\mathcal{D}/N})$ given a delay constraint $\mathcal{D}$; while a per-node capacity is $O(\sqrt[3]{\mathcal{D}/N})$ under i.i.d. slow mobility, which is a tighter bound than $O(\sqrt[3]{\mathcal{D}/N}\log N)$ obtained in [124].

In [125], El Gamal *et al.* studied the throughput and delay under random walk model. It was shown that the ratio of delay and throughput is $\Theta(N)$ for throughput of $O(\frac{1}{\sqrt{N\log N}})$, while the delay remains $\Theta(N\log N)$ for almost any throughput of a higher order, indicating an unsmooth tradeoff under random walk model. Similar insights can be obtained for Brownian motion. In [127], Lin *et al.* first derived a lower bound of $\Omega(\log N/\sigma^2)$ for average delay associated with capacity of $\Theta(1)$ by using the two-hop relaying scheme proposed

[2]Redundancy of the packet means extra copies of the original packet, which are issued by the source node.

Table 2.4: Summary of capacity-delay tradeoffs for random ad hoc networks

	Two-hop delay	Critical delay	Any tradeoff?
i.i.d Mobility	$\Theta(N \log N)$ [125]	$\Theta(1)$ [124]	Yes
Random walk	$\Theta(N \log N)$ [125]	$\Theta(N \log N)$ [125]	No
Random way-point	$\Theta(N)$* [126]	$\approx \Theta(\sqrt{N})$ [117]	Yes
Brownian motion	$\approx \Theta(N)$ [117]	$\Theta(N)$ [120]	No
Discrete random direction (α)	$\Theta(N)$ [117]	$\approx \Theta(N^{\alpha+0.5})$ [117]	Yes
Hybrid random walk (β)	$\Theta(N)$ [117]	$\Theta(N^{2\beta} \log\log N)$ [117]	Yes
Lévy walk (γ)	$\approx \Theta(N)$ [120]	$\Theta(N^{\frac{1}{2}})$ for $0 < \gamma < 1$; $\Theta(N^{\frac{\gamma}{2}})$ for $1 \leq \gamma < 2$ [120]	Yes
Lévy flight (γ)	$\approx \Theta(N)$ [120]	$\Theta(N^{\frac{\gamma}{2}})$ [120]	Yes

* The result is for the case in which the velocity does not scale with the network size.
** $\alpha \in (0, 0.5)$, $\beta \in (0, 0.5)$ and $\gamma \in (0, 2]$.

in [88], where σ^2 is related to the Brownian mobility model. More importantly, they demonstrated that it is impossible to reduce a large amount of delay without dropping the throughput to $O(\frac{1}{\sqrt{N}})$. From [125] and [127], significant increase in delay cannot be circumvented if a larger throughput than $\Theta(\frac{1}{\sqrt{N}})$[3] is desired by using random walk mobility or Brownian motion. Without showing any tradeoff, Sharma and Mazumda [126] analyzed the average delay of the two-hop relaying scheme in a network of N nodes following random way-point mobility.

To further investigate the impact of node mobility on throughput capacity and delay, Sharma *et al.* [117] proposed two general classes of mobility models, i.e., hybrid random walk models and discrete random direction models, incorporating mobility models aforementioned in [122, 125, 127]. The objective of this systematical study is to inquiry how much delay the network has to bear to achieve a per-node capacity better than $\Theta(\frac{1}{\sqrt{N}})$ under different mobility models, resulting in the notion of *critical delay*. Considering that the worst performance in network delay is incurred by the two-hop relaying scheme (*two-hop delay*), however, with an optimal throughput, the room left for tradeoff is actually determined by these two important delays. In [117], it was shown that tradeoffs are negligible under random walk model and Brownian motion, as also shown in [125] and [127], respectively; However, the tradeoff between delay and capacity is quite smooth under i.i.d. mobility and random way-point model. In [120], Lee *et al.* studied the delay-capacity tradeoffs under Lévy mobility. By using the limiting features of the joint spatio-temporal

[3]This throughput scaling is achievable in static random ad hoc networks.

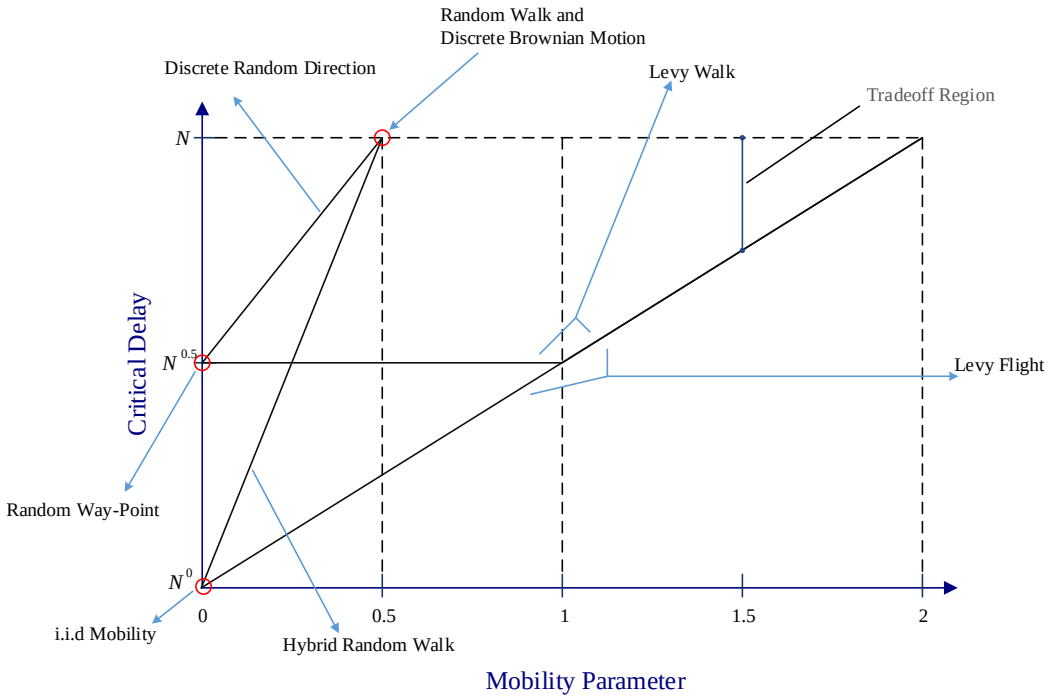


Figure 2.5: Tradeoff regions for a particular mobility parameter under different mobility models

probability density functions of Lévy models, they derived the critical delay under Lévy walk and Lévy fight, respectively. It was shown that smooth tradeoffs can be obtained and are determined by the distribution parameter related to Lévy mobility. A summary of delay-capacity tradeoffs for random ad hoc networks is given in Table 2.4. Fig. 2.5 also shows delay-capacity tradeoff regions under different mobility models.

2.3.3 Impact of Restricted and Correlated Mobility

The mobility models considered in aforementioned delay-capacity studies rely on the following assumptions: i) the mobility pattern of each node is identical; ii) following certain ergodic mobility process, each node can visit the entire network area equally likely; and iii) the movements of different nodes are independent. There have been several efforts made by

follow-up investigations to relax these assumptions and then find the impact of restricted and correlated mobility on delay and throughput performance in ad hoc networks.

Restricted Mobility: By noticing that nodes often spend most of the time in proximity of a few preferred places within a localized area, some researchers have studied the throughput and delay under the restricted node mobility, which is more realistic to characterize mobility traces of humans, animals, and vehicles. Li *et al.* [128] investigated the impact of a restricted mobility model on throughput and delay of a cell-partitioned network. They found that smooth throughput-delay tradeoffs in mobile ad hoc networks can be obtained by controlling the mobility pattern of nodes. Unlike the network in [129] showing homogeneous node density, Garetto *et al.* have done a series of research [130–134] on the network with heterogeneous node density under restricted mobility model. The capacity scaling of a class of mobile ad hoc networks which show spatial inhomogeneities by considering a cluster mobility model was analyzed in [132] and [133]. In [134], Garetto and Leonardi demonstrated that the delay-throughput tradeoffs can be improved by restricting the node mobility. They considered a restricted mobility that the node moves around a fixed home-point according to a Markov process, and the stationary distribution of the node location decays as a power law of exponent δ with the distance from the home-point. They showed that it is possible to exploit node heterogeneity under a restricted mobility model to achieve $\Theta(1/\log^2(N))$ throughput capacity and $O(\log^4(N))$ delay by using a sophisticated bisection routing scheme.

Correlated Mobility: Instead of exploring the full range of possible capacity-delay tradeoffs, Ciullo *et al.* [118] studied the impact of correlated mobility on performance of delay and throughput capacity. They considered a mobility model in which nodes in the network are grouped and each group, occupying a disc area, moves following i.i.d mobility. Although each node visits uniformly the entire network, movements of different nodes belonging to the same group are not independent. It was shown that the correlated mobility pattern has a significant impact on asymptotic network performance and it is possible to achieve better delay and throughput performance than that shown in [123].

2.3.4 Delay and Capacity Scaling without Exploiting Mobility

In [125], El Gamal *et al.* established delay-capacity tradeoffs for static ad hoc networks. It was shown that the tradeoff when applying multi-hop schemes is given by $\mathcal{D} = \Theta(N\lambda)$, where λ and $\mathcal{D}$ are respectively the throughput and delay. Following [77], throughput and delay tradeoff by means of hierarchical cooperation has been studied in [135], showing that $\mathcal{D} = N\log^2(N)\lambda$ for λ between $\Theta(\frac{1}{\sqrt{N}\log N})$ and $\Theta(\frac{1}{\log N})$. To serve delay sensitive traffic,

Comaniciu and Poor [136] reported the delay-constrained capacity scaling of mobile ad hoc networks. Without taking advantage of mobility, they exploited multiuser detection among other signal processing techniques to enhance user capacity.

2.4 Infrastructure Matters: Capacity and Delay of Hybrid Wireless Networks

Unlike pure ad hoc networks of homogeneous nodes operating in the same manner, hybrid wireless networks consist of at least two types of nodes functioning differently. After [58], significant efforts have been made to investigate capacity and delay scaling considering node heterogeneity, i.e., for hybrid networks, including wireless ad hoc networks with infrastructure aiding nodes, ad hoc networks with wireless helping nodes, multihop acess networks, and cognitive radio networks, among others.

2.4.1 Ad Hoc Networks with Supportive Infrastructure

It has been shown that adding wired infrastructure nodes, such as base stations, to ad hoc networks can render significant benefits in terms of both throughput capacity and delay. In the context of related investigations, the fixed infrastructure supports the underlying ad hoc networks by relaying their packets, rather than access points to the Internet. The advantage of infrastructure nodes is to overcome geographic limitations since the packet can be relayed over a long distance through high-bandwidth wired links, as a complement of local ad hoc delivery.

Liu *et al.* [137] initiated the study on capacity scaling of hybrid wireless networks. By placing N stationary nodes and M base stations in the network, they found that the throughput capacity increases linearly with M if $M = \omega(\sqrt{N})$, otherwise the improvement is negligible. Different from the hexagonal cell structure of base station in [137], access points in [138] are randomly distributed in the network and the results show that it is possible to achieve a throughput of $\Theta(\frac{1}{\log N})$ under the condition that the number of ad hoc nodes associated with each access point is upper bounded. Allowing power control, a constant throughput of $\Theta(1)$ is reported in [139]. In [140], Toumpis derived capacity bounds of hybrid wireless networks assuming randomly located access points and a general fading channel model and reported very similar results to those in [137]. In [1], Zemlianov *et al.* provided upper bounds of per-node throughput capacity for the network of randomly distributed ad hoc nodes and base stations placed in any deterministic fashion. By allowing

Table 2.5: Scaling regimes shown in [1]

Regime	Number of infrastructure nodes	Per-node throughput capacity
i)	$M \lesssim \sqrt{N/\log N}$	$\Theta(1/\sqrt{N \log N})$
ii)	$\sqrt{N/\log N} \lesssim M \lesssim N/\log N$	$\Theta(M/N)$
iii)	$M \gtrsim N/\log N$	$\Theta(1/\log N)$

power control of base stations, they determined three scaling regimes as shown in Table 2.5. It can be seen that there is no need to deploy any infrastructure for regime i), since the throughput is achievable by only leveraging ad hoc communications; and for regime iii), adding more infrastructure nodes does not make any improvement in throughput, at least in the order sense.

Table 2.6: Impact of network geometry [2]

Network geometry	Number of BS	Throughput capacity	Average delay
1-D network &	$M \log M = O(N)$	$\Omega(M/N)$	$O(N/M \log N)$
2-D strip with	$M \log M = \omega(N)$	$\Omega(1/\log M)$	$O(N/M \log N)$
strip width of $o(\log N)$			
2-D square &	$M = O(\sqrt{N})$	$\Omega(1/\sqrt{N})$	$O(\sqrt{N})$
2-D strip with	$M = \omega(\sqrt{N})$	$\Omega(\min\{M/N, 1/\log M\})$	$O(\sqrt{N/M \log N})$
strip width of $\Omega(\log N)$			

By noting that previous studies usually consider a two-dimensional square or disk network area, Liu *et al.* [2] investigated the impact of network geometry on capacity scaling by exploring one-dimensional networks and two-dimensional strip networks with regularly placed base stations. The main implications of theirs results (shown in Table 2.6) are: i) for the one-dimensional network, even a small number of supportive base stations can significantly increase the per-node throughput capacity; and ii) for a two-dimensional strip network, depending on the width of the strip, the behavior of capacity scaling is the same as that of either the one-dimensional network or the two-dimensional square network. The upper bound of average packet delay for each type of network was also derived, as shown in Table 2.6. Impacts of both network topology and traffic pattern were considered in [141]. Traffic patterns differ from each other in number of destination nodes in the network. The capacity scaling is determined by the number of base stations, the shape of network area, and the traffic pattern. Moreover, the impact of base station placement, i.e., regular or random placement, was also considered in [141].

Table 2.7: Scaling regimes shown in [3]

Regime	Number of base stations	Throughput capacity	Average delay
i)	$M = O(N/\log N)$	$\Omega(\sqrt{\frac{M}{N \log N}})$	$\Omega(\sqrt{\frac{N}{M \log N}})$
ii)	$M = \Omega(N/\log N)$	$\Omega(M/N)$	O(1)

An important implication of results shown in [1, 2, 137, 138, 140, 141] is that capacity gain will be insignificant if the number of infrastructure nodes placed in a square or disk network area grows asymptotically slower than certain threshold. By pointing out that such a "threshold" comes from the underutilization of the capability of base stations, Shila *et al.* [3] provided a better capacity and delay scaling, as shown in Table 2.7. The basic strategy they adopted is to deliver a packet to the nearest base station through multiple hops, in contrast to the one-hop transmission from the node to the associated base station assumed in previous studies, which yields a sublinear capacity scaling with the number of base stations.

Li *et al.* [142] revisited capacity and delay scaling in hybrid wireless networks by exploiting an *L-maximum-hop routing strategy*. Specifically, if the destination can be reached within L hops, packets from the source are delivered without relying on any infrastructure node. More importantly, it was shown that without degrading throughput, network delay can be improved substantially, however, at the expense of built infrastructure. It is possible to achieve both constant throughput and delay in this type of networks. By using the L-maximum-hop routing strategy as well, Zhang *et al.* [143] studied the throughput capacity for a network of N randomly distributed nodes, each of which is equipped with a directional antenna, and M regularly placed base stations. By analyzing the relationship between L, M, and directional antenna beamwidth θ, they showed a "threshold" result on impacts of directional antenna, i.e., throughput gain can be achieved by implementing directional antenna only when the number of base stations grows slower than certain threshold. Multiantenna systems were also considered. In [144], Shin *et al.* investigated the capacity scaling in the network with supportive base stations, at each of which the number of antennas scales at arbitrary rates relative to N. It is beneficial to exploit the spatial dimension of infrastructure by deploying multiple antennas, which enable simultaneous uplinks, at each base station. Wang *et al.* in [145] considered the impact of fading impairments when operating hybrid wireless networks where base stations are deployed to support long-range communications between ad hoc nodes. The throughput capacity of mobile hybrid networks was reported in [146], in which the mobility model considered is similar to that in [134].

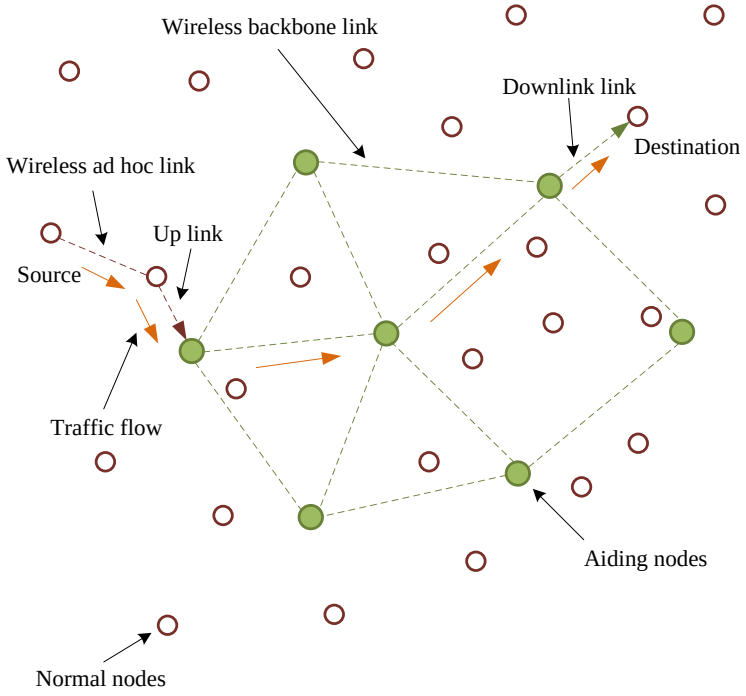


Figure 2.6: Ad hoc network supported by wirelessly connected aiding nodes

2.4.2 Ad Hoc Networks with Wireless Aiding Nodes

Deploying wired infrastructure to support ad hoc networks may incur a prohibitive cost which is always an important concern of building real-world communication networks. Moreover, under some emergency (e.g., earthquake) or extreme (e.g., underwater) circumstances, infrastructure is typically unavailable. Therefore, a potential substitute is to deploy a set of aiding nodes which are wirelessly connected and more powerful than normal nodes, as shown in Fig. 2.6. A natural question arises in the context: how much capacity gain can be achieved? To answer this question, Li *et al.* [147] studied the throughput capacity of ad hoc networks with the deployment of wireless helping nodes. Other specific network features considered in [147] are rectangular network area, both regular and random placement of helping nodes, and asymmetric traffic in which the number of destination nodes can scale at a lower rate than $\Theta(N)$, all of which have large impacts on throughput capacity. The main result of [147] illustrates that it is possible to achieve higher per-node throughput than that of pure ad hoc networks when the allocated bandwidth of helping

nodes scales at a much higher rate than $\Theta(1)$. In [148], Zhou *et al.* provided another promising solution of wireless mesh structures. In such a hierarchical wireless mesh network, mesh clients (normal nodes) are uniformly distributed, and mesh routers (aiding nodes) constitute a wireless mesh backbone, some of which can function as infrastructure gateways. Asymptotic throughput was derived and represented by the number of mesh clients, the number of mesh routers, and the number of mesh gateways. Relying on only a small number of mesh gateways, it was shown that such a mesh network can achieve the same throughput capacity as that of a hybrid infrastructure-based network, however, with a much lower cost. Literature [149] investigates a special scenario in which there exists only one active source-destination communication pair, and all remaining nodes act as aiding nodes. A constant capacity scaling is proved for that particular case.

2.4.3 Multihop Access Networks

Unlike ad hoc networks with supportive infrastructure nodes which do not generate or consume any data traffic, multihop access networks consist of infrastructure gateways bring/routing data traffic from/to the outside, such as Internet. Moreover, ad hoc transmissions between normal nodes are enabled and expected to enhance performance of such access networks, including capacity, coverage, and connectivity. To justify the benefit of augmenting access networks with multihop wireless links, Law *et al.* [150] investigated the downlink capacity of multihop cellular networks with regular placement of normal nodes and base stations. Due to poor spatial reuse, it was shown that one-dimensional multihop cellular networks yield almost no capacity gain compared to pure cellular networks. However, it is possible to significantly improve capacity of hexagonal hybrid network by exploiting multihop wireless links. By analyzing mathematically, they also found that capacity scaling in this type of networks mainly depends on the coverage of the base station, the transmission range of ad hoc links, and bandwidth allocation between different types of links. As a follow-up effort, Li *et al.* in [151] investigated capacity scaling for multihop cellular networks of randomly placed base stations and normal nodes distributed following a general inhomogeneous poisson process (IPP). In addition, throughput capacity was analyzed under different fairness constraints: i) throughput-fairness, making throughput equal over all the nodes; and ii) bandwidth-fairness, which guarantees that each node has equal allocated bandwidth. A "$\log_2 N$" result was shown in [151], i.e., multihop cellular networks with regular placement of nodes and base stations achieve higher per-node throughput than pure cellular networks by a scaling factor of $\log_2 N$, regardless the underlying fairness constraint. For the network with heterogeneous node distribution, it is possible to obtain the "$\log_2 N$" result under certain conditions.

2.4.4 Cognitive Radio Networks

Nowadays, the demand on the frequency spectrum is increasingly difficult to meet due to scarce and underutilized spectrum resources. Cognitive radio is a paradigm created in an attempt to enhance spectrum utilization, by enabling unlicensed users to opportunistically utilize the spectrum bands owned by licensed users [152]. In cognitive radio networks, licensed users and unlicensed users are referred to as primary users (PUs) and secondary users (SUs), respectively. With overlapping primary and secondary networks operating simultaneously, capacity and delay scaling laws of cognitive radio networks need to be investigated carefully.

By only allowing single-hop communication between a pair of SUs, Vu and Tarok [153] showed that the aggregate throughput of SUs can scale linearly with the number of SUs in the presence of a single or multiple pairs of primary transmitter (TX) and receiver (RX). In [154], Jeon *et al.* considered an ad hoc primary network of N randomly distributed PUs overlapped with an ad hoc secondary network of M randomly distributed SUs. Assuming M is much larger than N, they showed that an aggregate throughput of $\Theta(\sqrt{N})$ is achievable for the primary network, and in the meantime, the aggregate throughput of the secondary network is $\Theta(M^{\frac{1}{2}-\delta})$, for any arbitrarily small fraction of outage δ. The main implication of their result is that both two networks have almost the same capacity scaling as if each were a single network, given that one is much denser than the other. Another assumption made in [154] is that SUs know the locations of primary RXs. However, such prior knowledge is typically unavailable in practical scenarios. Instead, Yin *et al.* [155] studied capacity scaling of cognitive radio networks on the assumption that the locations of primary TXs are available to SUs and obtained very similar results to those in [154]. Huang and Wang [156] considered a more general model of cognitive radio networks, where the primary network can be different types, including classic static network, network with random walk mobility, and hybrid network, among others. Within this scope, they showed that the secondary network can attain the same asymptotic capacity and delay as standalone networks. The literature [157] is different from previous works in twofold. First, SUs are mobile and follow a specific heterogeneous speed-restricted mobility model. Second, cooperative communications are enabled so that SUs are allowed to relay packets for PUs. By exploiting the mobility heterogeneity of SUs, it was shown that almost constant capacity and delay scalings (except for poly-logarithmic factors) are possible in such a kind of cognitive radio networks.

2.5 Summary

This chapter has surveyed the existing literature for scaling laws of throughput capacity for both ad hoc wireless networks and wireless networks with communication infrastructure. It has also presented a comprehensive overview of capacity-delay tradeoffs under a variety of mobility models. Extensive comparisons of existing results have been done to reach a better understanding.

Chapter 3

Capacity of Delay-Tolerant Vehicular Networks with Socialized Mobility

Mobility of vehicles facilitates the store-carry-and-forward delivery of information in vehicular networks. In this chapter, we investigate the capacity of delay-tolerant vehicular networks in which services/applications are delay-tolerant and can be delivered in a store-carry-and-forward fashion. In addition, we specifically consider the social-proximity feature of the network which consists of N vehicles moving and communicating on a scalable grid-like street layout following the social-proximity model: each vehicle has a restricted mobility region around a specific social spot, and transmits via a unicast flow to a destination vehicle which is associated with the same social spot. Furthermore, the spatial distribution of the vehicle decays following a power-law distribution from the central social spot towards the border of the mobility region. With vehicles communicating using a variant of the two-hop relay scheme, the asymptotic bounds of throughput capacity and the average end-to-end delay are derived in terms of the number of social spots, the size of the mobility region, and the decay factor of the power-law distribution. By identifying these key impact factors of performance mathematically, we find three possible regimes for the throughput capacity and the end-to-end delay. It is shown that although the throughput and delay may degrade in a high density area, it is still possible to achieve almost constant scaling for per-vehicle throughput and end-to-end delay, i.e., such a network is truly scalable. Moreover, it is shown that inherent mobility patterns of vehicles have considerable impact on network performance.

3.1 Introduction

We consider the following challenging features of VANETs in this chapter. *Firstly*, in VANETs, vehicles have ***map-restricted*** and ***localized*** mobility with specific social features. Notably, for most of the time, a vehicle only moves within a bounded region related to the social life of the driver. For example, a vehicle often moves within a small area daily close to the driver's home, the work place or the city center. Such a mobility feature has also been reported in [158] based on the analysis of the real-world mobility trace of taxis in the city of Warsaw, Poland. It is observed that the mobility of taxis is typically around certain social spots. *Secondly*, VANETs show high ***spatial variations*** of vehicle density [19]. The analysis of the Warsaw trace data in [158] also reveals that the density of vehicles within the proximity area of social spots is much higher than on average and follows the empirical heavy-tailed distribution. *Thirdly*, we consider delay-tolerant applications running on top of the VANET, many of which are ***proximity-related***, such as traffic information publishing and localized social content sharing, since it is neither practical nor necessary to maintain a long-lasting unicast communication flow among vehicles over a long-distance. Although VANETs have received extensive attentions, the in-depth investigations on scaling laws on throughput capacity and end-to-end delay are very limited. Such scaling laws are critical to predict network performance in face of the large-scale networks of connected vehicles [60]. Thus, it is desirable to know the fundamental capability of VANETs for supporting delay-tolerant applications especially with the specific features aforementioned, which motivates our work. (Refer to Chapter 1 for related works).

In this chapter, we investigate the throughput capacity and average packet delay of the social-proximity urban VANET. Specifically, we model the urban area as a scalable grid with equal-length road segments and a set of social spots. Considering the localized and social features of vehicle's mobility, we apply a restricted mobility model to each vehicle surrounding a fixed social spot with the spatial stationary distribution of the vehicle following a power-law decay from the social spot to the border of the mobility region. Over this network model, we consider the delay-tolerant proximity applications such that the data traffic is delivered through unicast flows; and for each unicast flow, its source and destination vehicles belong to the same social spot. With a variant of the two-hop relay scheme [122] applied, we derive the bounds of throughput capacity and average packet delay, and show how the scaling laws depend on the inherent mobility pattern of the network which is characterized by the number of social spots, size of the mobility region, and the decay factor of the spatial distribution. The main contributions are three-fold:

- Our work represents the first theoretical study on the social-proximity vehicular networks. As vehicular communications are intensively affected by the social behaviors

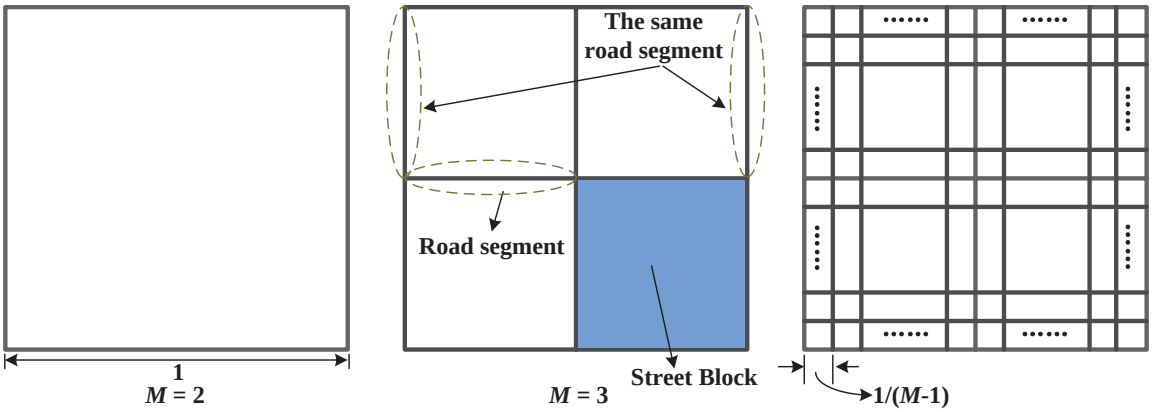


Figure 3.1: A grid-like street layout

of drivers, we argue that to accurately model the social features of vehicle mobility is crucial for the study of vehicular communications.

- We provide a generic modeling framework to unveil the asymptotic performance limits of the social-proximity vehicular networks. We obtain the bounds on per-vehicle throughput, average per-vehicle throughput, and average packet delay.
- The attained asymptotic property of capacity and delay can be used to predict network performance and provide guidance on design and analysis for different application scenarios of VANETs.

The remainder of this chapter is organized as follows: In Section 3.2, we introduce the system models. Section 3.3 summarizes the main results of this work. We analyze the asymptotic throughput capacity and delay with the proposed two-hop relay scheme in Section 3.4. Section 3.5 gives a brief summary.

3.2 System Model

Table 3.1: Summary of Notations for Chapter 3

Notations	Description
N	Number of vehicles in the network
M	Number of parallel roads in the grid
G	Number of road segments in the grid
C	Number of street blocks in the grid
ψ	Network density
V	Number of social spots
$\mathbf{S}$	Set of social spots, $\mathbf{S} = \{S_1, \ldots, S_V\}$
H_k	Vehicle k's social spot, $H_k \in \mathbf{S}$
ν	Exponent of V
κ	Exponent of $\mathcal{A}$
$\mathcal{A}$	Outermost tier of vehicle's mobility region
π'_α	Steady-state location probability of each vehicle on a segment of $Tier(\alpha)$
γ	Decay factor
r	Communication radius of the vehicle
Δ	Guard factor
$\mathbf{\Phi}$	Family of social-proximity VANETs
$\lambda(\Phi)$	Per-vehicle throughput
$\widetilde{\lambda}(\Phi)$	Average per-vehicle throughput
p_{ac}	Probability of a randomly selected road segment being active at a time slot
d_i^N	Vehicle density of road segment i
$\underline{d}_i^N$	Lower bound of d_i^N
$\overline{d}_i^N$	Upper bound of d_i^N
F_i	Rectangular area of $2\mathcal{A}(2\mathcal{A}-1)$ street blocks centered at road segment i
F_i^s	Number of social spots in F_i
S_j^v	Number of $\mathbb{S}$-$\mathbb{D}$ pairs associated with social spot S_j
$\mathcal{N}_i$	Number of vehicles on road segment i during a time slot
$\mathcal{N}_i^{\mathbb{SD}}$	Number of $\mathbb{S}$-$\mathbb{D}$ pairs on road segment i during a time slot
$P(\Phi)$	Average number of road segments where there are at least two vehicles during a time slot
$Q(\Phi)$	Average number of road segments where there is at least one $\mathbb{S}$-$\mathbb{D}$ pair during a time slot

3.2.1 Street Pattern

The geographic area where the network is deployed is modeled as a grid-like street layout, which consists of a set of M vertical roads intersected with a set of M horizontal roads, as shown in Fig. 3.1. Each line segment of equal length represents a road segment with bi-directional vehicle traffic. The grid street pattern is very common in many cities, such as Houston and Portland [159]. In the model, M is used to characterize the scale of the city grid. For example, M is roughly 100 for the downtown area of Toronto [160]. In addition, the city grid is considered as a torus of unit area to eliminate the border effects, which is a common practice to avoid tedious technicalities [134].

Let C denote the number of street blocks in the grid. The total number of road segments (the road section between any two neighboring intersections) is therefore $G = 2C = 2(M-1)^2$. We define the *network density* $\psi = \frac{N}{G} = \frac{N}{2(M-1)^2}$, where N is the total number of vehicles on the roads. Since N would tend to infinity in the asymptotic study, the city size, determined by M, cannot be fixed and should be scalable as well. Let $\Theta(1) \leq \psi \leq o(N)$ to avoid two extreme cases which are not practical in real-world scenarios: 1) when $\psi = o(1)$, the city size increases faster than the population of vehicles; and 2) when $\psi = \Theta(N)$, the city size is fixed such that the network density will become extremely high when more and more vehicles appear in the city. Note that ψ can represent the average vehicle density on each road segment. However, as each vehicle moves following the mobility model with social features, the spatial distribution of vehicles is inhomogeneous, as examples shown in Fig. 3.3(b) and 3.3(c). It can be seen that a network with a very large M and a relatively large ψ can represent metropolitan areas like New York City; whereas for a small town, M and ψ are relatively small. Therefore, from a macroscopic view, the grid street pattern with different values of M and ψ can model urban scenarios of different scales. A summary of the mathematical notations used in this chapter is given in Table 3.1.

3.2.2 Socialized Mobility Model

Markovian Mobility Pattern

We consider a time-slotted communication system, where time is slotted with equal duration. The road segments are indexed from 1 to G and vehicle nodes are indexed from 1 to N. Vehicles move independently from each other in the city. The mobility of a vehicle k follows a discrete time Markovian process, denoted by $\mathcal{C}_k, k \in \{1, 2, \ldots, N\}$, which is uniquely represented by a one-dimensional G-state ergodic Markov Chain. $\mathcal{C}_k(t) = i$ if

vehicle k appears on road segment i, $i \in \{1, 2, \ldots, G\}$, at time slot t, $t \in \{1, 2, \ldots, T\}$. Let P_k^{ij} denote the transition probability that vehicle k moves from road segment i to the next road segment j, $j \in \{1, 2, \ldots, G\}$. Let $\mathbf{P}_k = \{P_k^{ij}\}_{G \times G}$ denote the transition probability matrix of $\mathcal{C}_k$; the element P_k^{ij} in $\mathbf{P}_k$ is non-zero only if j is a neighboring road segment of i. The steady-state location distribution of vehicle k is $\boldsymbol{\pi}_k = \{\pi_k(i)\}_{1 \times G}$, where $\pi_k(i)$ denotes the long-term proportion of time that vehicle k stays on road segment i. In our analysis, we focus on the steady-state location distribution of the vehicles, since the capacity region only depends on how the node location distributes in the steady state [161], and the Markovian mobility model converges to its steady-state location distribution at an exponential rate [162].

Restricted Mobility Region with Social Spot

The mobility region of each vehicle is restricted and associated with a fixed social spot. Geographically, the social spot is the center of a certain street block, as shown in Fig. 3.2. Let V denote the number of social spots in the grid. We assume that all the social spots in the grid are uniformly distributed and thereby do not consider the inhomogeneous distribution of social spots in this study. Indexing all the street blocks from 1 to C, we denote by $\mathbf{S} = \{S_1, S_2, \ldots, S_V\} \subseteq \{1, 2, \ldots, C\}$ the set of social spots. Since we are interested in the capacity order, let $V = |\mathbf{S}| = \lceil C^\nu \rceil$, where $\nu \in (0, 1]$. Notably, $V = \Theta((N/\psi)^\nu)$, which is represented by a power function of N. When $\nu = 1$, all the street blocks in the network contain a social spot. In addition, we consider all the intermediate cases of ν between 0 and 1 in this research[1]. Each vehicle uniformly and independently selects one social spot out of all the social spots. Let $\mathbf{H}_N = (H_1, H_2, \ldots, H_N)$ denote the vector which collects the locations of all the vehicles' social spots, with each element $H_k \in \mathbf{S}$, denoting the index of the street block where vehicle k's social spot is located. The set $\mathbf{S}$ is fixed once the network is defined.

The mobility region of each vehicle consists of multiple tiers co-centered at its social spot, as shown in Fig. 3.2. $Tier(1)$ of the mobility region is collocated with the social spot and contains four road segments. The adjacent street blocks surrounding $Tier(1)$ form $Tier(2)$, and so on. We denote by $Tier(\mathcal{A})$ the outermost tier of the mobility region, where $\mathcal{A} = \Theta(M^\kappa) = \Theta((N/\psi)^{\frac{\kappa}{2}}) \leq \lfloor \frac{M}{2} \rfloor$, $\kappa \in [0, 1)$. When $\kappa = 0$, the size of the mobility region is fixed and does not scale with the city grid. It can be easily derived that $Tier(\alpha)$, $\alpha \in \{1, 2, \ldots, \mathcal{A}\}$, contains $16\alpha - 12$ road segments. Thus, the mobility of each vehicle is constrained in $\mathcal{A}$ tiers with a total number of $\sum_{\alpha=1}^{\mathcal{A}} 16\alpha - 12 = 4\mathcal{A}(2\mathcal{A} - 1)$ road

[1]We do not consider the extreme case in which $\nu = 0$. When $\nu = 0$, there is only one social spot in the network.

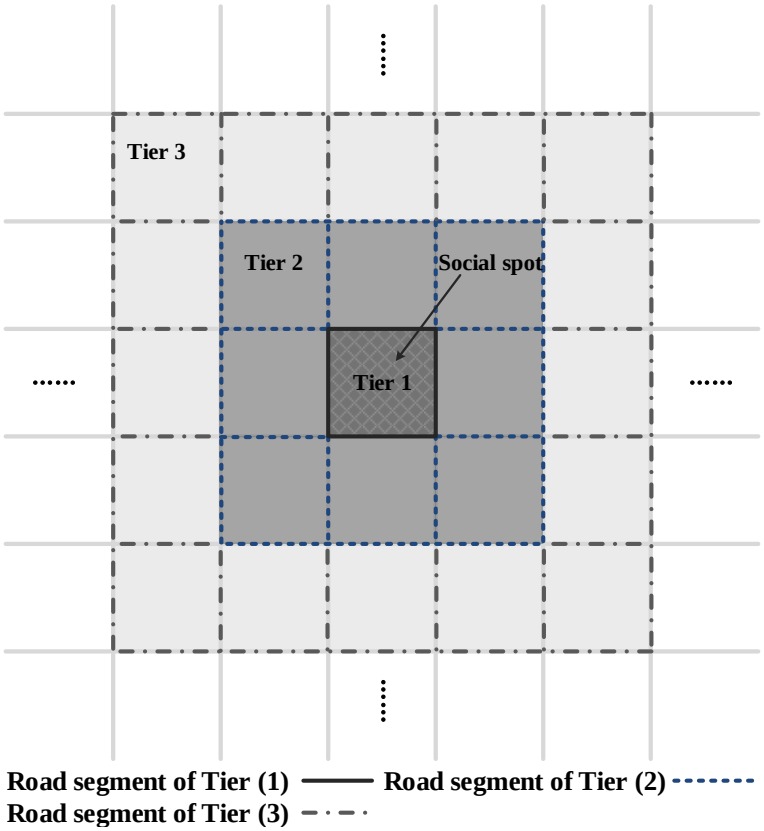


Figure 3.2: Restricted and socialized mobility with different tiers centered at a social spot for a given vehicle.

segments, and further the mobility region covers an area of $(\mathcal{A}+1)^2/C = \Theta((N/\psi)^{\kappa-1})$. For a randomly selected $Tier(\alpha)$, a vehicle has equal steady-state probability to appear on each road segment. Let π'_α denote the steady-state location probability of each vehicle on one of the road segments of $Tier(\alpha)$. From $Tier(1)$ to $Tier(\mathcal{A})$, the steady-state location probability of vehicles is modeled to exponentially decay as a power-law function with exponent $\gamma > 0$. Therefore, we have $\pi'_\alpha = \alpha^{-\gamma}\pi'_1$ which indicates that a vehicle is more likely to stay in the area near its social spot. The same model has been used in [134] and its accuracy is validated in [158] through real-world measurements. As the summation of steady-state probability on road segments equals to 1, i.e.,

$$\sum_{\alpha=1}^{\mathcal{A}}(16\alpha-12)\pi'_\alpha = \sum_{\alpha=1}^{\mathcal{A}}(16\alpha-12)\alpha^{-\gamma}\pi'_1 = 1,$$

we have,

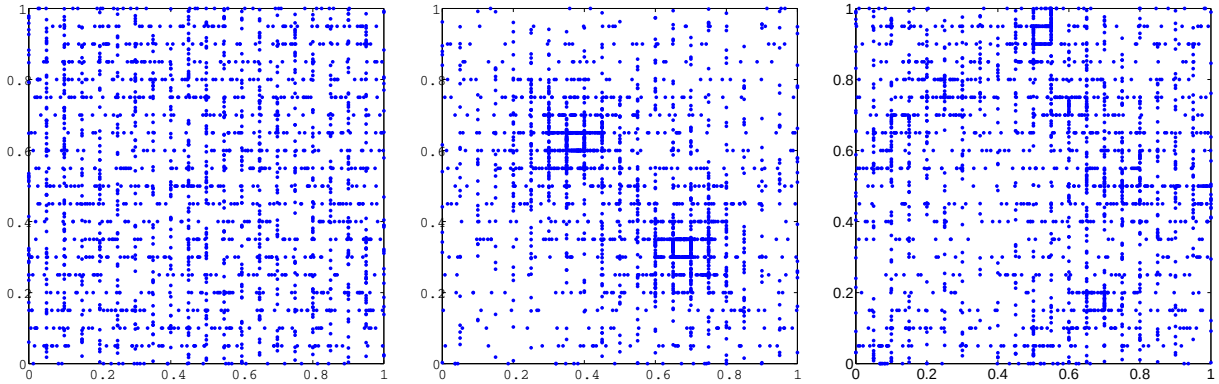
$$\pi'_1 = \frac{1}{\sum_{\alpha=1}^{\mathcal{A}}(16\alpha-12)\alpha^{-\gamma}}. \tag{3.1}$$

Lemma 1 *Given that $\kappa > 0$, as $N \to \infty$, $\pi'_1 = \Theta((N/\psi)^{-\kappa(1-\frac{1}{2}\gamma)})$, for $0 < \gamma < 2$; $\pi'_1 = \Theta(\frac{1}{\log(N/\psi)})$, for $\gamma = 2$; π'_1 converges to a constant value, for any $\gamma > 2$.*

This lemma can be proved by applying results of partial sums of p-series [163]. Note that when $\kappa = 0$, π'_1 is constant for all γ. Under the socialized mobility model, the network presents inhomogeneous vehicle densities. Fig. 3.3 illustrates the vehicle density when vehicles are uniformly distributed and follow the socialized mobility model, respectively.

3.2.3 Traffic Model

We consider that there exist N unicast flows concurrently in the network. Each vehicle is the source of one unicast flow and the destination of another unicast flow. We consider the case in which the source and destination vehicles of each unicast flow have the same social spot. This is motivated by the dominant proximity applications in vehicular communications. By doing so, the source and destination vehicles of each unicast flow are spatially close to each other. Without loss of generality, N is considered to be even. We sort the index of vehicles such that vehicle k communicates with vehicle $k+1$, $k \in \{1, 3, 5, \ldots, N-1\}$, and each communication pair independently and uniformly chooses a social spot from $\mathbf{S}$. The packet arrives in each unicast flow at an average rate η.



(a) Uniform distribution

(b) Distribution with social spots ($V = 2$)

(c) Distribution with social spots ($V = 10$)

Figure 3.3: Examples of homogeneous (a) and inhomogeneous (b) and (c) distributions of vehicles in the network, in the case of $N = 2000$, $M = 21$, $\mathcal{A} = 10$ and $\gamma = 2$.

3.2.4 Communication Model

Since the communication range of a vehicle is geographically limited in practice, the communication radius should scale with M. Let $r = \frac{1}{M-1}$ denote the communication radius of each vehicle which can always cover the entire road segment, as shown in Fig. 3.4. Without loss of generality, a pair of vehicles can communicate only when they are on the same road segment at the same time slot, and the transmission spans the whole time slot. Although the communication model has been simplified, such simplification does not affect the order of throughput capacity and average packet delay derived in this chapter. The success or failure of a transmission is determined by the protocol model defined in [58] as follows. The transmission from vehicle i to vehicle j can be successful during time slot t if and only if the following condition holds: $d_{kj}(t) \geq (1 + \Delta)r$, for every other vehicle k transmitting simultaneously, where $d_{kj}(t)$ denotes the Euclidean distance between vehicle k and j at time slot t, and $\Delta > 0$ is the guard factor.

3.2.5 Definitions

We denote by $\mathbf{\Phi}(N, \psi, \gamma, \mathcal{A}(\kappa), \mathbf{S}(\nu), \mathbf{H}_N)$ the family of social-proximity VANETs. Let $L_k(T)$ be the number of packets received by the destination of flow k, $k \in \{1, 2, \ldots, N\}$, up to time T. An asymptotic per-vehicle throughput $\lambda(\Phi)$ of $\mathbf{\Phi}$ is said feasible if there

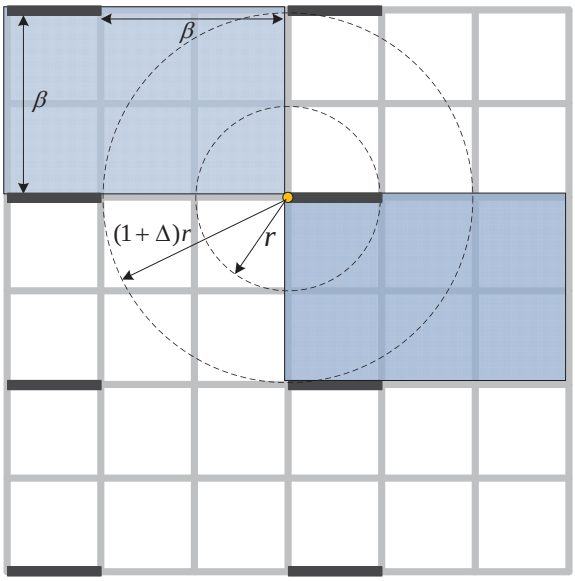


Figure 3.4: An example of non-interfering transmission group of road segments.

exist a scheduling policy and an N_0 such that for any $N > N_0$, we have,

$$\lim_{T\to\infty} \mathbf{Pr}\left(\frac{L_k(T)}{T} \geq \lambda(\Phi), \forall k\right) = 1. \tag{3.2}$$

Furthermore, an average per-vehicle throughput $\widetilde{\lambda}(\Phi)$ of $\mathbf{\Phi}$ is said feasible if there exist a scheduling policy and an N_0, such that for any $N > N_0$, the following holds

$$\lim_{T\to\infty} \mathbf{Pr}\left(\frac{\sum_{k=1}^{N} L_k(T)}{NT} \geq \widetilde{\lambda}(\Phi)\right) = 1. \tag{3.3}$$

3.3 Summary of Main Results

This section presents the summary of our main results. The formal statement of the results (Theorems 1, 2, and 3) and the derivations are given in Section 3.4.

The results of capacity and delay obtained in the analysis demonstrate three possible regimes depending on different values of κ and ν, as shown in Fig. 3.5. Recall that the size of mobility region and the number of social spots scale as $\Theta((N/\psi)^{\kappa-1})$ and $\Theta((N/\psi)^{\nu})$, respectively. i) ***Dense regime:*** when $\kappa+\nu > 1$, the sum of all mobility regions associated

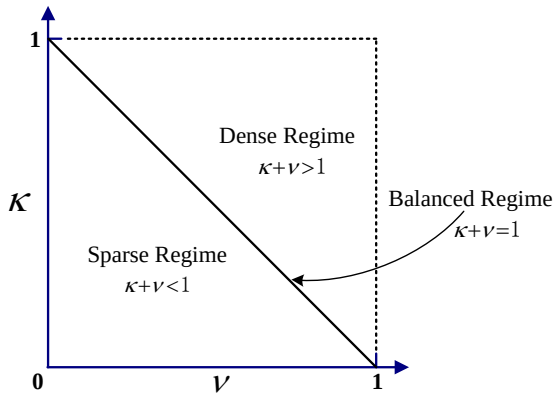
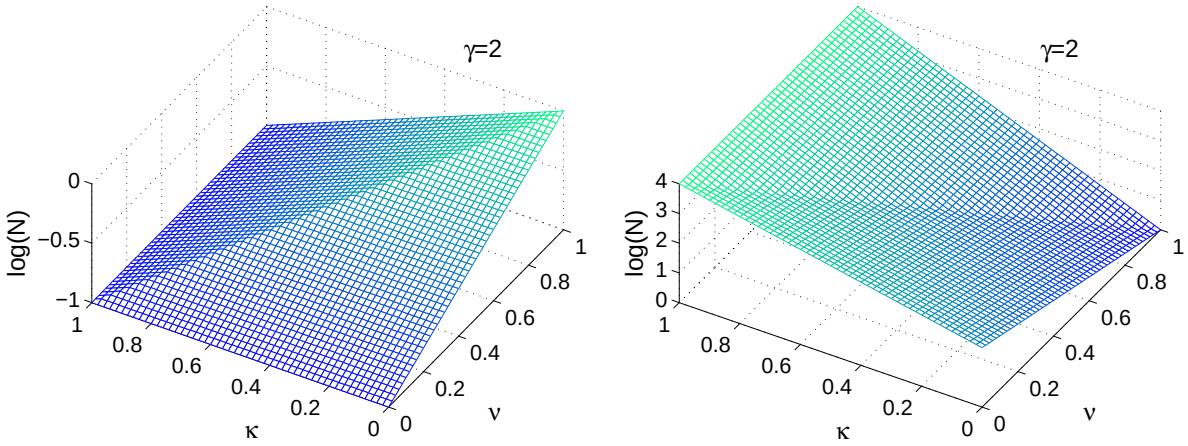


Figure 3.5: Regimes for asymptotic performance limits with respect to κ and ν.

with different social spots is $\Theta((N/\psi)^{\kappa+\nu-1}) = \omega(1)$, which indicates that V different mobility regions are overlapped and fully cover the city grid; ii) ***Sparse regime:*** when $\kappa + \nu < 1$, the sum of all different mobility regions is $o(1)$, which results in V typically isolated mobility regions sparsely distributed in the city grid; and iii) ***Balanced regime:*** when $\kappa + \nu = 1$, the sum of all different mobility regions has the same scale with the grid area, making the mobility area of vehicles perfectly fit the grid area in the order sense.

A graphical representation of our results is reported in Fig. 3.6 and Fig. 3.7. The results are shown in log-scale in terms of κ and ν, with $\psi = \Theta(1)$. For example, "-0.5" corresponds to a throughput of $\Theta(\frac{1}{\sqrt{N}})$. Fig. 3.6(a) shows the lower bound of the per-vehicle throughput capacity for $\gamma = 2$. In the dense regime, bounds of per-vehicle throughput capacity are dominated by κ given γ and ψ. It is observed that a large κ indicates a large size of mobility region, which results in decrease in per-vehicle throughput because 1) the contact probability of a pair of vehicles is reduced; and 2) different mobility regions are largely overlapped so that potentially increase the vehicle density. In the sparse regime, the performance is mainly dominated by ν. When ν tends to 1, the number of vehicles associated with each social spot is significantly reduced, avoiding a high vehicle density in the proximity of social spots. Therefore, the throughput performance is enhanced with a large ν. The performance decreases in the sparse regime when $\kappa + \nu$ tends to zero, due to increasing empty area in the city grid where there is no any packet transmission occurs. When $\kappa + \nu = 1$, the network achieves optimal bounds of per-vehicle throughput capacity, since the geographic area of the city grid, i.e., the spatial resource of the network, is just fully utilized for packet transmissions. From Fig. 3.6(b), the same insight on average packet delay can be obtained. Therefore, it is possible to achieve almost constant (except for the

polylogarithmic factor) per-vehicle throughput and average packet delay, i.e., in the case of $\kappa = 0$ and $\nu = 1$.



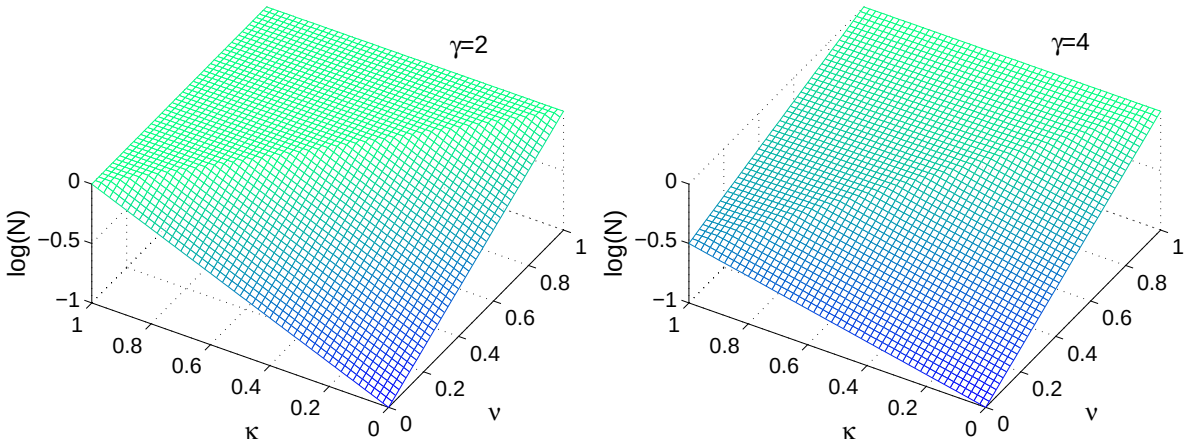
(a) Lower bound of per-vehicle throughput

(b) Upper bound of average packet delay

Figure 3.6: Per-vehicle throughput and average packet delay

The average per-vehicle throughput represents a global performance metric of the network with inhomogeneous vehicle densities. Fig. 3.7(a) and Fig. 3.7(b) demonstrate the average per-vehicle throughput for $\gamma = 2$ and $\gamma = 4$, respectively. For $\gamma = 2$, almost constant average per-vehicle throughput is achievable with high probability in the dense regime. However, in this case, the per-vehicle throughput may degrade dramatically, as shown in Fig. 3.6(a), in some hot area where is covered by a large number of different overlapped mobility regions. With a larger value of γ, e.g., $\gamma = 4$, the vehicles usually move in a very limited area centered at the social spot. Due to the limited spatial resource, the total number of concurrent transmissions is reduced. Therefore, the average per-vehicle throughput decreases since the geographic area of the city grid is not fully used for packet transmissions.

It has been shown that the performance metrics of interest depends on inherent mobility patterns of the network. Notice that the parameters of the socialized mobility are not easy to obtain, although they can be extracted from real-world mobility traces of vehicles. Once the mobility pattern of the real-world scenario is determined, our results can be applied to predict network performance, at least in the order sense. We provide an example in the following. Consider a network of 10^4 vehicles with parameters of mobility model



(a) Lower bound for $\gamma = 2$

(b) Lower bound for $\gamma = 4$

Figure 3.7: Average per-vehicle throughput

$\kappa = 0.5$, $\nu = 0.7$, and $\gamma = 2$. The bandwidth of point-to-point link is 1 Mbps. From the results obtained in the analysis, neglecting polylogarithmic factors and constant factor ψ, a per-vehicle throughput of around 10 kbps is achievable. If we consider a duration of 1 ms for each time slot, the time to deliver a packet could be from seconds to days. We notice that the delay performance may not satisfy requirements of many applications. It is important to note that there exists a throughput-delay tradeoff for a given mobility pattern. The throughput and delay achieved in the example is under the proposed two-hop relay scheme which is elaborated in Section 3.4. Better delay performance can be obtained by using other forwarding schemes, such as multi-hop scheme with or without packet redundancies, however, with a lower throughput. Considering another mobility pattern where $\kappa = 0.2$ and $\nu = 0.8$, the per-vehicle throughput is around 150 kbps and the average packet delay could be at most several seconds. It can be seen that when κ and ν tend to 0 and 1, respectively, the delay is small enough for many applications by using the two-hop relay scheme, indicating that there is not necessary to sacrifice throughput to improve the network delay. Therefore, another implication from our results is that it is beneficial to design suitable forwarding schemes according to different mobility patterns of the network.

3.4 Asymptotic Capacity and Delay Analysis

In this section, a two-hop relay scheme is first proposed to delivery the packets from the source to the destination. We then derive the bounds of per-vehicle throughput capacity and average per-vehicle throughput, which are stated in Theorem 1 and 2, respectively.

3.4.1 Two-Hop Relay Scheme

All packets are transmitted by using a two-hop relay scheme $\mathcal{X}$: a packet is either transmitted directly from the source to the destination, or relayed through one intermediate vehicle from the source to the destination. The packet transmission consists of two phases:

$\mathcal{X}$-I: Each road segment in the network becomes "active" in every $1/p_{ac}$ time slots[2].

$\mathcal{X}$-II: For each active road segment where there are at least two vehicles,

1) if there exists at least one source-destination ($\mathbb{S}$-$\mathbb{D}$) pair on the road segment, one pair is uniformly selected. If the source has a buffering packet for the destination, it transmits the packet and deletes it from the buffer after the transmission; otherwise, the source stays idle.

2) if there is no any $\mathbb{S}$-$\mathbb{D}$ pair on the road segment, a vehicle, e.g., υ_A, is uniformly selected out of all vehicles on this road segment to be the source or the destination equally likely, and in the meantime another vehicle, e.g., υ_B, is independently and uniformly selected over the rest of vehicles to be the relay.

- If υ_A is the source, a source-to-relay transmission from υ_A to υ_B is scheduled. If υ_A has a buffering packet to transmit, υ_A transmits the packet to υ_B and deletes the packet from the buffer; otherwise, υ_A remains idle.
- If υ_A is the destination, a relay-to-destination transmission from υ_B to υ_A is scheduled. If υ_B has a buffering packet destined for υ_A, υ_B transmits the packet to υ_A and deletes the packet from the buffer; otherwise, υ_B remains idle.

We calculate the value of p_{ac} in the following, which is the probability of a randomly selected road segment being active at a time slot. As shown in Fig. 3.4, we partition the network into equal-size sub-areas. Each sub-area consists of $\beta(\beta+1)$ street blocks where β

[2]A road segment is active when vehicles on the road segment can transmit successfully without any interference of transmissions from other road segments. The value of p_{ac} is discussed later in the section.

is an integer number. The road segments highlighted in each sub-area in Fig. 3.4 constitute a non-interfering transmission group, such that simultaneous transmissions within one non-interfering group do not interfere with each other. Totally, there are $2\beta(\beta+1)$ road segments within one sub-area, and collectively $2\beta(\beta+1)$ non-interfering groups in the network. With non-interfering groups transmitting alternately, each non-interfering group becomes active every $1/p_{ac} = 2\beta(\beta+1)$ time slots. This indicates that the vehicles on one specific road segment obtain a transmission opportunity with probability p_{ac} at a randomly selected time slot. With the grid scale of M, the minimum distance between any two neighboring road segments of a non-interfering group is $\frac{\beta}{M-1}$. By using the protocol model, we have

$$\beta/(M-1) \geq (1+\Delta)r.$$

With $r = 1/(M-1)$, we have $\beta \geq 1+\Delta$. We set $\beta = \lceil 1+\Delta \rceil$. By substituting it into $1/[2\beta(\beta+1)]$, we have

$$p_{ac} = 1/(2\lceil 1+\Delta \rceil \lceil 2+\Delta \rceil).$$

3.4.2 Bounds of Per-Vehicle Throughput Capacity

In the following, we derive the bounds of the per-vehicle throughput capacity by using the two-hop relay scheme $\mathcal{X}$, which are formally stated in Theorem 1. We first obtain an important result of vehicle density of a generic road segment (Lemma 3) by applying Chernoff bounds (Lemma 2) and the Vapnik-Chervonenkis Theorem which gives the uniform convergence in the weak law of large numbers.

To characterize the vehicle spatial inhomogeneities of the network, inspired by [132], we define the vehicle density (*vehicles/road segment*) of a generic road segment i by

$$d_i^N = \sum_{k=1}^{N} E[\mathbb{I}_{\mathcal{C}_k(t)=i} | \mathbf{H}_N]. \tag{3.4}$$

where $\mathbb{I}_{\mathcal{C}_k(t)=i}$ is the indicator variable that takes value 1 if $\mathcal{C}_k(t) = i$ and 0 otherwise.

Lemma 2 *(Chernoff bounds [164]) Let X be a sum of n independent random variables $\{X_i\}$, with $X_i \in \{0,1\}$ for all $i \leq n$. Write $\mu' = E[X] = E[X_1] + \cdots + E[X_n]$. Then for*

any $0 < \varepsilon \leq 1$,

$$\mathbf{Pr}(X > (1+\varepsilon)\mu') \leq e^{-\frac{\varepsilon^2}{2+\varepsilon}\mu'}, \quad \textit{and}$$
$$\mathbf{Pr}(X < (1-\varepsilon)\mu') \leq e^{-\frac{\varepsilon^2}{2}\mu'}.$$

Lemma 2 is a well-known result and used to prove the following important lemma which presents a bound of d_i^N.

Lemma 3 *The following bounds of the vehicle density* d_i^N *hold* w.h.p.[3], $\forall i \in \{1, 2, \ldots, G\}$*:*

i. when $\kappa + \nu > 1$,

$$d_i^N = \begin{cases} \Omega(\psi), \quad O(\psi(N/\psi)^{\frac{1}{2}\kappa\gamma}) & 0 < \gamma < 2 \\ \Omega(\frac{\psi}{\log(N/\psi)}), \quad O(\frac{\psi(N/\psi)^{\kappa}}{\log(N/\psi)}) & \gamma = 2 \\ \Omega(\frac{\psi}{(N/\psi)^{\kappa(\frac{1}{2}\gamma-1)}}), \quad O(\psi(N/\psi)^{\kappa}) & \gamma > 2. \end{cases}$$

ii. when $\kappa + \nu \leq 1$,

(a) $\kappa = 0$, $d_i^N = O(\psi(N/\psi)^{1-\nu}\log(N/\psi))$*;*

(b) $\kappa \neq 0$,

$$d_i^N = \begin{cases} O\big(\frac{\psi\log(N/\psi)}{(N/\psi)^{\nu+\kappa(1-\frac{1}{2}\gamma)-1}}\big) & 0 < \gamma < 2 \\ O(\psi(N/\psi)^{1-\nu}) & \gamma = 2 \\ O(\psi(N/\psi)^{1-\nu}\log(N/\psi)) & \gamma > 2. \end{cases}$$

Proof. The proof of Lemma 3 consists of three parts in terms of different values of ν and κ.

(i) $0 < \nu < 1$ and $\kappa + \nu > 1$:

We first show that Lemma 3 holds for $0 < \nu < 1$ and $\kappa + \nu > 1$. Let F_i denote the rectangular area of $2\mathcal{A}(2\mathcal{A}-1)$ street blocks centered at road segment i, as shown in Fig. 3.8. If the social spot of vehicle k is not located in F_i, $E[\mathbb{I}_{\mathcal{C}_k(t)=i}] = 0$. In other words, vehicles whose social spots are located in F_i contribute d_i^N. We denote by F_i^s the number of social spots contained in F_i. Intuitively, a social spot represents a mobility region of

[3]As $N \to \infty$, the probability of the event approaches 1.

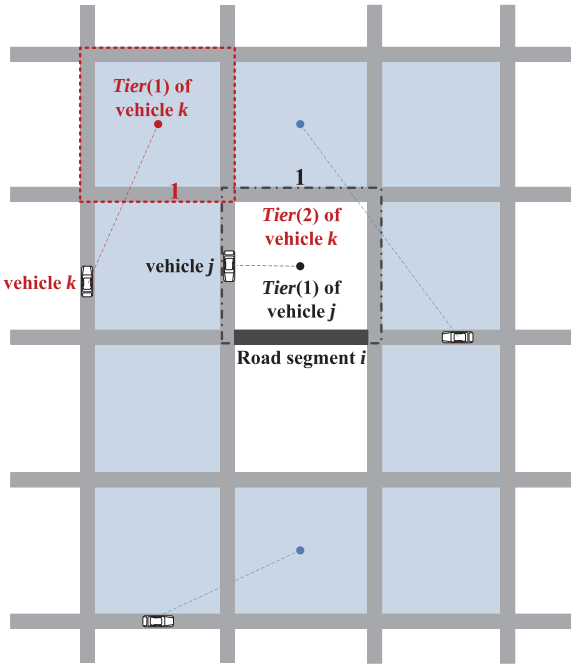


Figure 3.8: An example of one given road segment contained by different vehicles' mobility regions.

vehicles that are associated with this social spot. The vehicle density of a road segment depends on how many mobility regions contain this road segment. Thus, the number of social spot in an area plays an important role in determining the vehicle density of a road segment, and further the throughput capacity. We first bound F_i^s, for all $i \in \{1, 2, \ldots, G\}$. By definition, we have,

$$F_i^s = \sum_{j=1}^{V} \mathbb{I}_{S_j \in F_i} \tag{3.5}$$

where $\mathbb{I}_{S_j \in F_i}$, $\forall j \in \{1, 2, \ldots, V\}$, are i.i.d Bernoulli random variables with expectation $\mu = \frac{2\mathcal{A}(2\mathcal{A}-1)}{C}$. Inspired by [132], from Lemma 2, we have

$$\mathbf{Pr}\Big\{F_i^s < \frac{1}{2}E[F_i^s] = \frac{1}{2}V\mu\Big\} < e^{-\frac{1}{8}V\mu}, \quad \text{and}$$
$$\mathbf{Pr}\Big\{F_i^s > 2E[F_i^s] = 2V\mu\Big\} < e^{-\frac{1}{3}V\mu} < e^{-\frac{1}{8}V\mu}.$$

Notably, $\mu = \Theta((N/\psi)^{\kappa-1})$. By applying the union bound, we have

$$\begin{aligned}
&\mathbf{Pr}\big\{\bigcap_i \{\frac{1}{2}E[F_i^s] < F_i^s < 2E[F_i^s]\}\big\} \\
&\quad = 1 - \mathbf{Pr}\big\{\bigcup_i \{F_i^s < \frac{1}{2}E[F_i^s] \cup F_i^s > 2E[F_i^s]\}\big\} \\
&\quad \geq 1 - \sum_{i=1}^{G} \mathbf{Pr}\big\{F_i^s < \frac{1}{2}E[F_i^s] \cup F_i^s > 2E[F_i^s]\big\} \\
&\quad \geq 1 - G2e^{-\frac{1}{8}V\mu}.
\end{aligned}$$

For $\kappa + \nu > 1$, $G2e^{-\frac{1}{8}V\mu} \to 0$, as $N \to \infty$. Hence, *w.h.p.*, we get

$$\frac{1}{2}E[F_i^s] < F_i^s < 2E[F_i^s],\ \forall i \in \{1, 2, \ldots, G\}. \tag{3.6}$$

Let $S_j^v = \sum_{k=1}^{N-1} \mathbb{I}_{H_k=S_j}$ denote the number of $\mathbb{S}$-$\mathbb{D}$ pairs associated with social spot S_j, where $\mathbb{I}_{H_k=S_j}$, $\forall k \in \{1, 3, \ldots, N-1\}$, are i.i.d Bernoulli random variables with expectation $1/V$. From Lemma 2, for $0 < \varepsilon \leq 1$, we have

$$\mathbf{Pr}\big\{\frac{(1-\varepsilon)N}{2V} < S_j^v < \frac{(1+\varepsilon)N}{2V}\big\} \geq 1 - 2e^{-\frac{\varepsilon^2 N}{2(2+\varepsilon)V}}.$$

By applying union bound in the same manner,

$$\begin{aligned}
&\mathbf{Pr}\big\{\bigcap_j \{\frac{(1-\varepsilon)N}{2V} < S_j^v < \frac{(1+\varepsilon)N}{2V}\}\big\} \\
&\quad = 1 - \mathbf{Pr}\big\{\bigcup_j \{S_j^v < \frac{(1-\varepsilon)N}{2V} \cup S_j^v > \frac{(1+\varepsilon)N}{2V}\}\big\} \\
&\quad \geq 1 - \sum_{j=1}^{V} \mathbf{Pr}\big\{S_j^v < \frac{(1-\varepsilon)N}{2V} \cup S_j^v > \frac{(1+\varepsilon)N}{2V}\big\} \\
&\quad \geq 1 - V2e^{-\frac{\varepsilon^2 N}{2(2+\varepsilon)V}}.
\end{aligned}$$

Since $V = \Theta((N/\psi)^{\nu})$ and $\nu \neq 1$, $V2e^{-\frac{\varepsilon^2 N}{(2+\varepsilon)V}} \to 0$, as $N \to \infty$. Hence, *w.h.p.*, we have

$$\frac{(1-\varepsilon)N}{2V} < S_j^v < \frac{(1+\varepsilon)N}{2V}, \ \forall j \in \{1, 2, \ldots, V\}. \tag{3.7}$$

By definition (3.4) and from (3.6) and (3.7), *w.h.p.*, we obtain, $\forall i$,

$$\begin{aligned} \frac{1}{2}V\mu \cdot \frac{(1-\varepsilon)N}{V} \cdot \mathcal{A}^{-\gamma}\pi_1' = \underline{d}_i^N < d_i^N < \overline{d}_i^N \\ = 2V\mu \cdot \frac{(1+\varepsilon)N}{V} \cdot \pi_1' \end{aligned}$$

denoting by $\underline{d}_i^N$ and $\overline{d}_i^N$ the lower bound and upper bound of d_i^N, respectively. Letting $\varepsilon \to \infty$, we have $\underline{d}_i^N = \Theta(\pi_1'\psi(N/\psi)^{\kappa(1-\frac{1}{2}\gamma)})$ and $\overline{d}_i^N = \Theta(\pi_1'\psi(N/\psi)^{\kappa})$. Then (i) follows according to Lemma 1.

(ii) $\nu = 1$:

To prove Lemma 3 for the case in which $\nu = 1$, we recall the Vapnik-Chervonenkis Theorem [165]. Some relevant definitions are first provided. A Range Space is a pair $(X, \mathcal{F})$, where X is a set and $\mathcal{F}$ is a family of subsets of X. For any $A \subseteq X$, we define $P_{\mathcal{F}}(A)$, the projection of $\mathcal{F}$ on A, as $\{F \cap A : F \in \mathcal{F}\}$. We say that A is *shattered* by $\mathcal{F}$ if $P_{\mathcal{F}}(A) = 2^A$, i.e., if the projection of $\mathcal{F}$ on A includes all possible subsets of A. The VC-dimension of $\mathcal{F}$, denoted by VC-d($\mathcal{F}$) is the cardinality of the largest set A that $\mathcal{F}$ shatters. If arbitrarily large finite sets are shattered, the VC dimension of $\mathcal{F}$ is infinite.

The Vapnik-Chervonenkis Theorem: If $\mathcal{F}$ is a set of finite VC-dimension and $\{Y_j\}$ is a sequence of N i.i.d. random variables with common probability distribution P, then for every ϵ, $\delta > 0$

$$\mathbf{Pr}\left\{\sup_{F\in\mathcal{F}}\left|\frac{1}{N}\sum_{j=1}^{N}\mathbb{I}_{Y_j\in F} - P(F)\right| \leq \epsilon\right\} > 1-\delta \tag{3.8}$$

if

$$N > \max\left\{\frac{8\text{VC-d}(\mathcal{F})}{\epsilon}\log\frac{16e}{\epsilon}, \frac{4}{\epsilon}\log\frac{2}{\delta}\right\}. \tag{3.9}$$

We use the Vapnik-Chervonenkis Theorem to show that Lemma 3 holds for $\nu = 1$. Recall that F_i denotes the rectangular area of $2\mathcal{A}(2\mathcal{A}-1)$ street blocks centered at road

segment i. $\sum_{k=1}^{N-1} \mathbb{I}_{H_k \in F_i}$, $k \in \{1, 3, \ldots, N-1\}$, is the number of $\mathbb{S}$-$\mathbb{D}$ pairs whose social spot falls into the region F_i. $\mathbf{Pr}(\mathbb{I}_{H_k \in F_i} = 1) = \frac{2\mathcal{A}(2\mathcal{A}-1)}{C} = \Theta((N/\psi)^{\kappa-1})$, $\forall k$. Let $\mathcal{F}$ be the class of all such F_i rectangular areas. It is easy to show that the VC-dimension of $\mathcal{F}$ is at most 4 [166]. Thus, $\forall F_i$,

$$\mathbf{Pr}\left\{\sup_{F_i \in \mathcal{F}} \left| \frac{\sum_{k=1}^{N-1} \mathbb{I}_{H_k \in F_i}}{N/2} - \frac{2\mathcal{A}(2\mathcal{A}-1)}{C} \right| \leq \epsilon \right\} > 1 - \delta.$$

The condition (3.9) holds when $\epsilon = \delta = \frac{\Delta_\epsilon \log(N/\psi)}{N/\psi}$, where $\Delta_\epsilon := \max\{8\text{VC-d}(\mathcal{F}), 16e\}$. Thus, the Vapnik-Chervonenkis Theorem states that

$$\mathbf{Pr}\left\{\sup_{F_i \in \mathcal{F}} \left| \sum_{k=1}^{N-1} \mathbb{I}_{H_k \in F_i} - \Theta(\psi(\frac{N}{\psi})^\kappa) \right| \leq \Theta(\psi \log \frac{N}{\psi}) \right\} > 1 - \frac{\Delta_\epsilon \log(N/\psi)}{N/\psi}.$$

We conclude that *w.h.p.*, for $\kappa = 0$,

$$\overline{d}_i^N = 2\max\{\sum_{k=1}^{N-1} \mathbb{I}_{H_k \in F_i}\}\pi_1' = \Theta(\psi \log(N/\psi)),\ \forall i;$$

for $0 < \kappa < 1$, $\underline{d}_i^N = \Theta(\pi_1' \psi (N/\psi)^{\kappa(1-\frac{1}{2}\gamma)})$ and $\overline{d}_i^N = \Theta(\pi_1' \psi (N/\psi)^\kappa)$, $\forall i$. Hence, (ii) follows according to Lemma 1.

(iii) $0 < \nu < 1$ and $\kappa + \nu \leq 1$:

We apply the Vapnik-Chervonenkis Theorem in the same manner. From (3.5), we have,

$$\mathbf{Pr}\left\{\sup_{F_i \in \mathcal{F}} \left| \frac{\sum_{j=1}^{V} \mathbb{I}_{S_j \in F_i}}{V} - \frac{2\mathcal{A}(2\mathcal{A}-1)}{C} \right| \leq \epsilon \right\} > 1 - \delta.$$

The condition (3.9) holds when $\epsilon = \delta = \frac{\Delta_\epsilon \log(V)}{V} = \frac{\Delta_\epsilon \log(N/\psi)}{(N/\psi)^\nu}$. Thus,

$$\mathbf{Pr}\left\{\sup_{F_i\in\mathcal{F}}\left|\sum_{j=1}^{V}\mathbb{I}_{S_j\in F_i} - \Theta((\frac{N}{\psi})^{\kappa+\nu-1})\right| \le \Delta_\epsilon \log(\frac{N}{\psi})\right\} > 1 - \frac{\Delta_\epsilon \log(N/\psi)}{(N/\psi)^\nu}. \tag{3.10}$$

Since $\kappa + \nu \le 1$, we conclude that *w.h.p.*, $F_i^s = O(\log(N/\psi))$, $\forall i$. From (3.7) and (3.10), the upper bound of vehicle density $\overline{d}_i^N = \Theta(\pi_1'\psi(N/\psi)^{1-\nu}\log(N/\psi))$, $\forall i$, *w.h.p.*. Then (iii) follows according to Lemma 1. □

Theorem 1 *For the social-proximity grid-like VANETs, with the two-hop relay scheme $\mathcal{X}$, the per-vehicle throughput $\lambda(\Phi)$ cannot be better than $\frac{1}{2\psi\lceil 1+\Delta\rceil\lceil 2+\Delta\rceil}$ and* w.h.p.*, we obtain*

i. when $\kappa + \nu > 1$,

$$\lambda(\Phi) = \begin{cases} \Omega(\frac{1}{\psi(N/\psi)^{\frac{1}{2}\kappa\gamma}}) & 0 < \gamma < 2 \\ \Omega(\frac{1}{\psi(N/\psi)^{\kappa}\log(N/\psi)}) & \gamma = 2, \psi = \Theta(1) \\ \Omega(\frac{\log(N/\psi)}{\psi(N/\psi)^{\kappa}}) & \gamma = 2, \psi = \omega(1) \\ \Omega(\frac{1}{\psi(N/\psi)^{\kappa}}) & \gamma > 2. \end{cases}$$

ii. when $\kappa + \nu \le 1$,

(a) $\kappa = 0$, $\lambda(\Phi) = \Omega(\frac{(N/\psi)^{\nu-1}}{\psi\log(N/\psi)})$;

(b) $\kappa \ne 0$,

$$\lambda(\Phi) = \begin{cases} \Omega(\frac{(N/\psi)^{\nu-\frac{1}{2}\kappa\gamma-1}}{\psi\log(N/\psi)}) & 0 < \gamma < 1 \\ \Omega(\frac{(N/\psi)^{\nu-\frac{1}{2}\kappa-1}}{\psi}) & \gamma = 1 \\ \Omega(\frac{(N/\psi)^{\nu-\kappa(1-\frac{1}{2}\gamma)-1}}{\psi\log(N/\psi)}) & 1 < \gamma < 2 \\ \Omega(\frac{(N/\psi)^{\nu-1}}{\psi\log^2(N/\psi)}) & \gamma = 2 \\ \Omega(\frac{(N/\psi)^{\nu-1}}{\psi\log(N/\psi)}) & \gamma > 2. \end{cases}$$

Proof. The proof consists of two parts. First, we apply Lemma 1 and Lemma 3 to derive the lower bound of the per-vehicle throughput. Following the two-hop relay scheme $\mathcal{X}$, the

long-term throughput of flow k (denoting the source and destination of flow k by $\mathbb{S}$ and $\mathbb{D}$, respectively) is given by

$$\begin{aligned}\lambda_k(\Phi) &= \lim_{T\to\infty} L_k(T)/T \\ &= \frac{1}{2}p_{ac}\sum_{i=1}^{G}\mathbf{Pr}(\mathcal{N}_i \geq 2, \mathcal{N}_i^{\mathbb{SD}} = 0|\mathcal{C}_{\mathbb{D}} = i)\pi_{\mathbb{D}}(i)\frac{1}{\mathcal{N}_i} \\ &+ p_{ac}\sum_{i=1}^{G}\mathbf{Pr}(\mathcal{C}_{\mathbb{S}} = i|\mathcal{C}_{\mathbb{D}} = i)\pi_{\mathbb{D}}(i)\frac{1}{\mathcal{N}_i^{\mathbb{SD}}},\end{aligned} \tag{3.11}$$

where $\mathcal{N}_i$ and $\mathcal{N}_i^{\mathbb{SD}}$ denote the number of vehicles and the number of $\mathbb{S}$-$\mathbb{D}$ pairs on road segment i in a time slot, respectively. Recall that $\pi_{\mathbb{D}}(i)$ is the steady-state probability that $\mathbb{D}$ stays on road segment i.

Let $\mathfrak{N}_i$ denote the number of $\mathbb{S}$-$\mathbb{D}$ pairs whose mobility region contains road segment i. The probability of finding at least two vehicles and no any $\mathbb{S}$-$\mathbb{D}$ pair on road segment i given that $\mathbb{D}$ is on road segment i is given by, *w.h.p.*, $\forall i$,

$$\begin{aligned}&\mathbf{Pr}(\mathcal{N}_i \geq 2, \mathcal{N}_i^{\mathbb{SD}} = 0|\mathcal{C}_{\mathbb{D}} = i) \\ &\geq (1-\pi_1')\bigg(1-\bigg(1-\frac{2\pi_1'}{\mathcal{A}^\gamma}+2\big(\frac{\pi_1'}{\mathcal{A}^\gamma}\big)^2\bigg)^{\mathfrak{N}_i-1}\bigg) \\ &= (1-\pi_1')\bigg(1-\bigg(1+\big(-\frac{\mathcal{A}^\gamma}{2\pi_1'}\big)^{-1} \\ &\qquad + \frac{1}{2}\big(-\frac{\mathcal{A}^\gamma}{2\pi_1'}\big)^{-2}\bigg)^{-\frac{\mathcal{A}^\gamma}{2\pi_1'}(-2\mathfrak{N}_i\mathcal{A}^{-\gamma}\pi_1'+2\mathcal{A}^{-\gamma}\pi_1')}\bigg) \\ &\geq (1-\pi_1')\bigg(1-\bigg(\big(1+\frac{1}{\mathscr{Z}}+\frac{1}{2\mathscr{Z}^2}\big)^{\mathscr{Z}}\bigg)^{-\underline{d}_i^N-1/\mathscr{Z}}\bigg),\end{aligned}$$

where we denote $-\mathcal{A}^\gamma/(2\pi_1')$ by $\mathscr{Z}$. If $\underline{d}_i^N = \omega(1)$, $((1+1/\mathscr{Z}+1/(2\mathscr{Z}^2))^{\mathscr{Z}})^{-\underline{d}_i^N-1/\mathscr{Z}} \to 0$, as $N\to\infty$. Hence, the event "$\mathcal{N}_i \geq 2, \mathcal{N}_i^{\mathbb{SD}} = 0|\mathcal{C}_{\mathbb{D}} = i$" holds *w.h.p.* when $\gamma \leq 2$, and at least with a constant probability $1-\pi_1'$ when $\gamma > 2$, according to Lemma 1. If $\underline{d}_i^N = \Theta(1)$, $\mathbf{Pr}(\mathcal{N}_i \geq 2, \mathcal{N}_i^{\mathbb{SD}} = 0|\mathcal{C}_{\mathbb{D}} = i)$ is lower bounded by $(1-\pi_1')(1-e^{-\underline{d}_i^N})$. If $\underline{d}_i^N = o(1)$, $\mathbf{Pr}(\mathcal{N}_i \geq 2, \mathcal{N}_i^{\mathbb{SD}} = 0|\mathcal{C}_{\mathbb{D}} = i) \to 0$, as $N\to\infty$.

According to the results of partial sum of p-series, the probability of finding $\mathbb{S}$ and $\mathbb{D}$

on the same road segment during a slot is asymptotically given by

$$\begin{aligned}\sum_{i=1}^{G}\mathbf{Pr}(\mathcal{C}_{\mathbb{S}}=i|\mathcal{C}_{\mathbb{D}}=i)\pi_{\mathbb{D}}(i) &= \sum_{i=1}^{G}\pi_{\mathbb{S}}(i)\pi_{\mathbb{D}}(i)\\ &= \sum_{\alpha=1}^{\mathcal{A}}(16\alpha-12)\alpha^{-2\gamma}\pi_1'^2 = \pi_1'^2\sum_{\alpha=1}^{\mathcal{A}}\big(\frac{16}{\alpha^{2\gamma-1}}-\frac{12}{\alpha^{2\gamma}}\big)\\ &= \begin{cases}\Theta(\pi_1'^2(N/\psi)^{\kappa(1-\gamma)}) & 0<\gamma<1\\ \Theta(\pi_1'^2\log(N/\psi)) & \gamma=1\\ \Theta(\pi_1'^2) & \gamma>1.\end{cases}\end{aligned} \tag{3.12}$$

Base on the analysis above, when $\underline{d}_i^N=\Omega(1)$, $\forall k$, *w.h.p.* we have,

$$\begin{aligned}\lambda_k(\Phi) &\geq \frac{c'p_{ac}}{2\overline{d}_i^N}\sum_{i=1}^{G}\pi_{\mathbb{D}}(i)+\frac{p_{ac}}{\overline{d}_i^N}\sum_{i=1}^{G}\pi_{\mathbb{S}}(i)\pi_{\mathbb{D}}(i)\\ &= c'p_{ac}/\overline{d}_i^N+O(1)p_{ac}/\overline{d}_i^N,\end{aligned}$$

where c' is constant. Therefore, $\lambda(\Phi)=\Omega(p_{ac}/\overline{d}_i^N)$, *w.h.p.*. When $\underline{d}_i^N=o(1)$, $\forall k$,

$$\lambda_k(\Phi)\geq p_{ac}(\sum_{i=1}^{G}\pi_{\mathbb{S}}(i)\pi_{\mathbb{D}}(i))/\overline{d}_i^N.$$

Thus, *w.h.p.*, we have

$$\lambda(\Phi)=\Omega(p_{ac}(\sum_{i=1}^{G}\pi_{\mathbb{S}}(i)\pi_{\mathbb{D}}(i))/\overline{d}_i^N).$$

From Lemma 1 and 3, the assert follows.

We then derive an upper bound of per-vehicle throughput considering any possible stabilizing scheduling policies under $\mathcal{X}$-I. Let $\mathscr{X}_d(T)$ denote the number of packets delivered in the network through direct transmissions from the source to destination, and $\mathscr{X}_r(T)$ denote the number of packets delivered to the destination via relaying, during the interval $[0,T]$. Thus, provided the arbitrary and fixed $\epsilon>0$, there must exist arbitrarily large

values of T such that the per-vehicle throughput $\lambda(\Phi)$ satisfies

$$\frac{\mathscr{X}_d(T) + \mathscr{X}_r(T)}{T} \geq N\lambda(\Phi) - \epsilon. \tag{3.13}$$

Let $\mathscr{Y}(T)$ denote the total number of transmission opportunities during the interval $[0, T]$. From (3.13), we have

$$\begin{aligned}\frac{1}{T}\mathscr{Y}(T) &\geq \frac{1}{T}\mathscr{X}_d(T) + \frac{2}{T}\mathscr{X}_r(T) \\ &\geq \frac{1}{T}\mathscr{X}_d(T) + 2\big((N\lambda(\Phi) - \epsilon) - \frac{1}{T}\mathscr{X}_d(T)\big).\end{aligned}$$

The first inequality holds since the relayed packet reaches to the destination through at least two hops. Hence,

$$\lambda(\Phi) \leq \frac{\frac{1}{T}\mathscr{Y}(T) + \frac{1}{T}\mathscr{X}_d(T) + 2\epsilon}{2N},$$

i.e.,

$$\lambda(\Phi) \leq \lim_{T\to\infty} \frac{\frac{1}{T}\mathscr{Y}(T) + \frac{1}{T}\mathscr{X}_d(T)}{2N}. \tag{3.14}$$

Due to the transmission interference, the total number of transmission opportunities is no larger than the maximum number of concurrent transmissions during $[0, T]$. We have $\lim_{T\to\infty} \frac{1}{T}\mathscr{Y}(T) \leq Gp_{ac}$. Similarly, we have $\lim_{T\to\infty} \frac{1}{T}\mathscr{X}_d(T) \leq Gp_{ac}$, where the equality holds when there is always an $\mathbb{S}$-$\mathbb{D}$ transmission on each road segment of a non-interference group during each time slot. By plugging the inequalities into (3.14), we have

$$\lambda(\Phi) \leq \frac{Gp_{ac} + Gp_{ac}}{2N} = \frac{p_{ac}}{\psi} = \Theta(\frac{1}{\psi}). \tag{3.15}$$

From (3.15), the per-vehicle throughput capacity of $\mathbf{\Phi}$ cannot be better than $\Theta(\frac{1}{\psi})$. □

3.4.3 Average Per-Vehicle Throughput

We next derive a lower bound of the average per-vehicle throughput $\widetilde{\lambda}(\Phi)$, stated in Theorem 2, based on the proposed two-hop relay scheme for $\gamma \geq 2$, where the network shows dramatic social features. To simplify the analysis, we let $\psi = \Theta(1)$ in this subsection, i.e.,

the network density keeps constant and does not scale up with the population of vehicles. Considering all possible functions of ψ with the order of $o(N)$ makes the derivation very complex. The following lemmas (Lemma 4 to Lemma 8) are first presented to prove Theorem 2.

Lemma 4 *Let $\mathscr{T}$ be a regular tessellation of the network, whose elements $\mathscr{T}_i$ contains $\lceil 100N^{1-\nu}\log(N)\rceil$ street blocks.* w.h.p.*, every element of $\mathscr{T}$ contains at least one social spot.*

Proof. Recall that each element $\mathscr{T}_i$ of $\mathscr{T}$ is a regular area containing $\lceil 100N^{1-\nu}\log(N)\rceil$ street blocks. Let $\mathscr{T}_i^s$ denote the number of social spots contained in $\mathscr{T}_i$. Applying the Vapnik-Chervonenkis Theorem, $\forall \mathscr{T}_i$,

$$\mathbf{Pr}\left\{\sup_{\mathscr{T}_i\in\mathscr{T}}\left|\frac{\mathscr{T}_i^s}{V}-\frac{100N^{1-\nu}\log(N)}{C}\right|\leq\epsilon\right\}>1-\delta.$$

Note that VC-d($\mathscr{T}$) is at most 4. The condition (3.9) is satisfied when $\epsilon=\delta=\frac{50\log(N)}{N^{\nu}}$. Thus, $\forall\mathscr{T}_i$,

$$\mathbf{Pr}\big\{\mathscr{T}_i^s\geq 50\log(N)\big\}>1-\frac{50\log(N)}{N^{\nu}}.$$

The lemma follows as $N\to\infty$. □

We denote by $P(\Phi)=E[\sum_{i=1}^{G}\mathbb{I}_{\mathcal{N}_i\geq 2}]$ the average number of road segments where there are at least two vehicles during a time slot. Similarly, let $Q(\Phi)=E[\sum_{i=1}^{G}\mathbb{I}_{\mathcal{N}_i^{\mathbb{SD}}\geq 1}]$ denote the average number of road segments where there is at least one $\mathbb{S}$-$\mathbb{D}$ pair during a time slot. Lemma 5 and 8 present a lower bound of $P(\Phi)$ respectively when $\nu\neq 1$ and $\nu=1$.

Lemma 5 *When $\nu\neq 1$,* w.h.p.*, we have,*

$$P(\Phi)=\begin{cases}\Omega(N^{\frac{2}{\gamma}+\nu(1-\frac{2}{\gamma})}/\log^3(N)) & \kappa+\nu>1\\ \Omega(N^{\nu+\frac{2\kappa}{\gamma}}/\log(N)) & \kappa+\nu<1\\ \Omega(N^{\nu+\frac{2\kappa}{\gamma+\vartheta}}/\log(N)) & \kappa+\nu=1\end{cases}$$

where ϑ is a positive and arbitrarily small value.

Proof. We consider a single social spot S_j in an area. Recall that S_j^v is the number of vehicles associated with S_j. For road segment i in the mobility region of the vehicles, from (3.7), we have,

$$\begin{aligned}
&\mathbf{Pr}(\mathcal{N}_i \geq 2) = 1 - \mathbf{Pr}(\mathcal{N}_i \leq 1) \\
&\quad \geq 1 - (1-\pi'_{\mathcal{B}})^{\frac{(1-\varepsilon)N}{V}} - \frac{(1-\varepsilon)N}{V}\pi'_{\mathcal{B}}(1-\pi'_{\mathcal{B}})^{\frac{(1-\varepsilon)N}{V}-1} \\
&\quad \geq 1 - e^{-\pi'_{\mathcal{B}}\frac{(1-\varepsilon)N}{V}} - \pi'_{\mathcal{B}}\frac{(1-\varepsilon)N}{V}e^{-\pi'_{\mathcal{B}}\frac{(1-\varepsilon)N}{V}+\pi'_{\mathcal{B}}} \\
&\quad = 1 - e^{-\pi'_{\mathcal{B}}\frac{(1-\varepsilon)N}{V}}(1 + \pi'_{\mathcal{B}}\frac{(1-\varepsilon)N}{V}e^{\pi'_{\mathcal{B}}})
\end{aligned}$$

where $\pi'_{\mathcal{B}} = \mathcal{B}^{-\gamma}\pi'_1$ and $\mathcal{B} \leq \min\{\mathcal{A}, \lceil 10\sqrt{N^{1-\nu}\log(N)}\rceil\}$ from Lemma 4. Further, it is satisfied that $\pi'_{\mathcal{B}}\frac{N}{V} = \omega(1)$. Thus, letting $\varepsilon \to 0$, as $\to \infty$, the event "$\mathcal{N}_i \geq 2$" holds *w.h.p.*. Considering the regular tessellation $\mathscr{T}$ of the network, according to Lemma 4, *w.h.p.*, we have,

$$\begin{aligned}
P(\Phi) &= \sum_{i=1}^{G} E[\mathbb{I}_{\mathcal{N}_i \geq 2}] = \sum_{i=1}^{G} \mathbf{Pr}(\mathcal{N}_i \geq 2) \\
&\geq \frac{C}{100N^{1-\nu}\log(N)} \cdot \frac{1}{4} \cdot 4\mathcal{B}(2\mathcal{B}-1) = \Theta(\frac{\mathcal{B}^2 N^{\nu}}{\log(N)}).
\end{aligned} \tag{3.16}$$

For $\kappa + \nu > 1$, we choose $\mathcal{B} = \Theta(N^{\frac{1-\nu}{\gamma}}/\log(N))$. $\mathcal{B}$ can scale as $\Theta(N^{\frac{\kappa}{\gamma+\vartheta}})$ for $\kappa + \nu = 1$. When $\kappa + \nu < 1$, $\mathcal{B}$ can be $\Theta(N^{\frac{\kappa}{\gamma}})$. The lemma follows. □

Lemma 6 *(Chebyshev's Inequality) If X is a random variable with mean $E[X]$ and variance $Var(X)$, then, for any value $k > 0$,*

$$\mathbf{Pr}(\mid X - E[X] \mid \geq k) \leq \frac{Var(X)}{k^2}.$$

Lemma 6 is well known and we use it to prove Lemma 7.

Lemma 7 *When $\nu = 1$, at least $(1 - e^{-\psi})C$ social spots associate with at least one* $\mathbb{S}$-$\mathbb{D}$ *pair* w.h.p..

Proof. We denote by $\mathbb{I}_C = \sum_{i=1}^{C} \mathbb{I}_{S_i^v=0}$ the number of social spots that are not selected by any $\mathbb{S}$-$\mathbb{D}$ pair in the network. $\mathbf{Pr}(\mathbb{I}_{S_i^v=0} = 1) = (1 - \frac{1}{C})^{\frac{N}{2}}$. Thus, the expectation and

variance of $\mathbb{I}_{S_i^v=0}$ are $E[\mathbb{I}_{S_i^v=0}] = (1-\frac{1}{C})^{\frac{N}{2}}$ and $\mathrm{Var}(\mathbb{I}_{S_i^v=0}) = (1-\frac{1}{C})^{\frac{N}{2}} - (1-\frac{1}{C})^N$, respectively. Next we need to determine the variance of $\mathbb{I}_C$. For any $i \neq j$, $j \in \{1, 2, \ldots, C\}$, $\mathrm{Cov}(\mathbb{I}_{S_i^v=0}, \mathbb{I}_{S_j^v=0}) = E[\mathbb{I}_{S_i^v=0}\mathbb{I}_{S_j^v=0}] - E[\mathbb{I}_{S_i^v=0}]E[\mathbb{I}_{S_j^v=0}]$, where $\mathrm{Cov}(\mathbb{I}_{S_i^v=0}, \mathbb{I}_{S_j^v=0})$ is the covariance of variable $\mathbb{I}_{S_i^v=0}$ and $\mathbb{I}_{S_j^v=0}$. It is easy to get that $E[\mathbb{I}_{S_i^v=0}\mathbb{I}_{S_j^v=0}] = (1-\frac{2}{C})^{\frac{N}{2}}$. Since $\mathrm{Cov}(\mathbb{I}_{S_i^v=0}, \mathbb{I}_{S_i^v=0}) = \mathrm{Var}(\mathbb{I}_{S_i^v=0})$, we have

$$\begin{aligned}
\mathrm{Var}(\mathbb{I}_C) &= \mathrm{Var}\left(\sum_{i=1}^{C} \mathbb{I}_{S_i^v=0}\right) = \sum_{i=1}^{C}\sum_{j=1}^{C} \mathrm{Cov}(\mathbb{I}_{S_i^v=0}, \mathbb{I}_{S_j^v=0}) \\
&= \sum_{i=1}^{C} \mathrm{Cov}(\mathbb{I}_{S_i^v=0}, \mathbb{I}_{S_i^v=0}) + 2\sum_{i=1}^{C}\sum_{j<i} \mathrm{Cov}(\mathbb{I}_{S_i^v=0}, \mathbb{I}_{S_j^v=0}) \\
&= C\left((1-\frac{1}{C})^{\frac{N}{2}} - (1-\frac{1}{C})^N\right) \\
&\qquad + C(C-1)\left((1-\frac{2}{C})^{\frac{N}{2}} - (1-\frac{1}{C})^N\right) \\
&\leq C\left((1-\frac{1}{C})^{\frac{N}{2}} - (1-\frac{1}{C})^N\right).
\end{aligned}$$

The inequality holds because $(1-\frac{2}{C})^{\frac{N}{2}} - (1-\frac{1}{C})^N = (1-\frac{2}{C})^{\frac{N}{2}} - (1-\frac{2}{C}+\frac{1}{C^2})^{\frac{N}{2}} \leq 0$. From Lemma 6, choosing $k = \epsilon C$, we have

$$\mathbf{Pr}(\mathbb{I}_C - E[\mathbb{I}_C] \geq \epsilon C) \leq \frac{C[(1-\frac{1}{C})^{\frac{N}{2}} - (1-\frac{1}{C})^N]}{\epsilon^2 C^2}.$$

Note that $E[\mathbb{I}_C] = C(1-\frac{1}{C})^{\frac{N}{2}}$. Thus,

$$\mathbf{Pr}\left(\frac{\mathbb{I}_C}{C} \geq (\rho + \epsilon)\right) \leq \frac{\rho - \rho^2}{\epsilon^2} \cdot \frac{1}{C},$$

where $\rho = (1-\frac{1}{C})^{\frac{N}{2}}$. Since $N = 2\psi C$, as $N \to \infty$, $\rho \to e^{-\psi}$. Therefore, we have

$$\lim_{N\to\infty} \mathbf{Pr}(\mathbb{I}_C/C \geq e^{-\psi}) = 0,$$

i.e., the probability of $\mathbb{I}_C$ being over a constant proportion of C tends to zero as $N \to \infty$. The lemma follows. □

Lemma 8 *When $\nu = 1$,* w.h.p., *$P(\Phi) = \Omega(N/\log^2(N))$ for $\gamma = 2$ and $\kappa \neq 0$; $P(\Phi) =$*

$\Theta(N)$ *for* $\gamma = 2$ *and* $\kappa = 0$*;* $P(\Phi) = \Theta(N)$ *for* $\gamma > 2$*.*

Proof. According to Lemma 7, we can obtain that *w.h.p.*, there are at least $2(1 - e^{-\psi})C$ road segments, each of which belongs to a $Tier(1)$ of vehicles' mobility region. Let $\mathcal{S}$ denote the set of road segments that are not contained in the mobility region of any vehicle. $\bar{\mathcal{S}}$ is the complementary set of $\mathcal{S}$ in $\{1, 2, \ldots, G\}$. Note that $\mathbf{Pr}(\mathcal{N}_i \geq 2) = 0$, $\forall i \in \mathcal{S}$. Thus,

$$\begin{aligned} P(\Phi) &= \sum_{i=1}^{G} E[\mathbb{I}_{\mathcal{N}_i \geq 2}] = \sum_{i=1}^{G} \mathbf{Pr}(\mathcal{N}_i \geq 2) \\ &= \sum_{i \in \mathcal{S}} \mathbf{Pr}(\mathcal{N}_i \geq 2) + \sum_{j \in \bar{\mathcal{S}}} \mathbf{Pr}(\mathcal{N}_j \geq 2) \\ &= |\mathcal{S}| \cdot 0 + \sum_{j \in \bar{\mathcal{S}}} \mathbf{Pr}(\mathcal{N}_j \geq 2) \geq \sum_{j \in \bar{\mathcal{S}}} \pi_1'^2, \end{aligned}$$

since for any $j \in \bar{\mathcal{S}}$, $\mathbf{Pr}(\mathcal{N}_j \geq 2) \geq \pi_1'^2$. The Lemma follows from Lemma 7. □

Theorem 2 *For the social-proximity grid-like VANETs, with the two-hop relay scheme* $\mathcal{X}$*, a bound of average per-vehicle throughput capacity* $\widetilde{\lambda}(\Phi)$ *is given by* w.h.p.*, i. when* $\nu \neq 1$ *and* $\gamma \geq 2$*,*

$$\widetilde{\lambda}(\Phi) = \begin{cases} \Omega(N^{\frac{2}{\gamma} + \nu(1 - \frac{2}{\gamma}) - 1} / \log^3(N)) & \kappa + \nu > 1 \\ \Omega(N^{\nu + \frac{2\kappa}{\gamma + \vartheta} - 1} / \log(N)) & \kappa + \nu = 1 \\ \Omega(N^{\nu + \frac{2\kappa}{\gamma} - 1} / \log(N)) & \kappa + \nu < 1 \end{cases}$$

and ii. when $\nu = 1$*,* $\widetilde{\lambda}(\Phi) = \Omega(\frac{1}{\log^2(N)})$*, for* $\gamma = 2$ *and* $\kappa \neq 0$*;* $\widetilde{\lambda}(\Phi) = \Theta(1)$ *for* $\gamma = 2$ *and* $\kappa = 0$*;* $\widetilde{\lambda}(\Phi) = \Theta(1)$ *for* $\gamma > 2$*.*

Proof. Under the two-hop relay scheme $\mathcal{X}$, we can use a decoupling queue structure, similar to that in [122], to model each unicast flow, as shown in Fig. 3.9. Without loss of generality, we consider that the packet arrival rate η follows the Bernoulli process. In other words, in each unicast flow, one packet arrives with the probability η at the current slot, and with the rest probability if there is no packet arrival. Hence, the source vehicle, e.g., v_k, can be represented as a Bernoulli/Bernoulli queue with packet arrival rate η_k and service rate ζ_k. The buffering packet in the source is transmitted (served) to either its destination directly or one of the relays within the mobility region of the source. The transmission

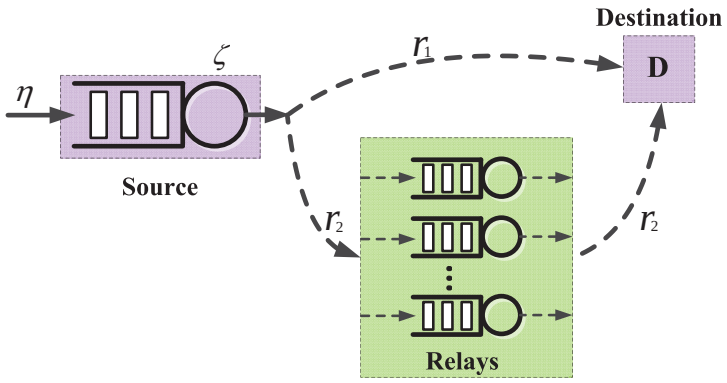


Figure 3.9: A decoupling queue structure.

opportunity arises with probability ζ_k. Let $r_1^k(N)$ denote the long term average rate at which a direct transmission to the destination is scheduled to source v_k, and $r_2^k(N)$ denote the long term average rate at which a source-to-relay transmission is scheduled to source v_k. The transmission opportunity arises at the rate $\zeta_k(N) = r_1^k(N) + r_2^k(N)$. As per the definition, $\widetilde{\lambda}(\Phi) = \frac{\sum_{k=1}^N \zeta_k(N)}{N}$. Since the two-hop relay scheme $\mathcal{X}$ schedules a source-to-relay transmission and a relay-to-destination transmission equally likely, the rate into the relays is equal to the rate out of the relays. During each time slot, the total number of transmission opportunities over the network is $\sum_{k=1}^N (r_1^k(N) + 2r_2^k(N))$. Given that the transmission opportunity arises on a road segment when it is active and at least two vehicles are on it, we have,

$$p_{ac} P(\Phi) = \sum_{k=1}^{N} (r_1^k(N) + 2r_2^k(N)). \tag{3.17}$$

Since the two-hop relay scheme $\mathcal{X}$ schedules the source-to-destination transmission whenever possible, we have,

$$p_{ac} Q(\Phi) = \sum_{k=1}^{N} r_1^k(N). \tag{3.18}$$

From (3.17) and (3.18), we obtain $\sum_{k=1}^N r_2^k(N) = \frac{p_{ac}(P(\Phi)-Q(\Phi))}{2}$ and therefore

$$\widetilde{\lambda}(\Phi) = \frac{\sum_{k=1}^N (r_1^k(N) + r_2^k(N))}{N} = \frac{p_{ac}(P(\Phi) + Q(\Phi))}{2N}.$$

Since $P(\Phi) \geq Q(\Phi)$ and from Lemma 5 and 8, the theorem follows. $\square$ *Remark:* The average per-vehicle throughput is analyzed as a global metric to evaluate the network performance with inhomogeneous vehicle density. For example, from Theorem 2, we can attain that the constant per-vehicle throughput is feasible *w.h.p.* for N_f $\mathbb{S}$-$\mathbb{D}$ pairs, where $N_f = \Theta(N) \leq \frac{N}{2}$, when $\nu = 1$ and $\gamma > 2$. Due to the socialized mobility of vehicles and the randomness of the locations of vehicle's social spots, the network shows spatial variations of vehicle density. Hence, the throughput performance of vehicles in different areas of the city grid may be different. For example, in a hot area where is covered by a large number of different overlapped mobility regions, the throughput of an $\mathbb{S}$-$\mathbb{D}$ pair in that area may drop significantly.

3.4.4 Average Packet Delay

We first analyze the average packet delay of a given unicast flow. The packet delay is accounted starting from the time slot when the packet arrives at the source until the time slot when the packet is delivered to its destination (including the queueing delay at the source or relay vehicle).

Recall that the source v_k can be represented as a Bernoulli/Bernoulli queue with arrival rate η_k and service rate ζ_k. The expected number of packets buffered at the source is

$$E_k[n_s] = \frac{\eta_k(1-\eta_k)}{\zeta_k - \eta_k}. \tag{3.19}$$

It has been shown in [122] that packets depart from the source at the rate of η_k when the buffer of the source is stable. For a packet from the source, it is delivered to a relay vehicle, e.g., v_i, with the probability $\frac{r_2^k}{\zeta_k} \cdot P_{ki}$, where P_{ki} is the contact probability between v_k and v_i. Therefore, the packet arrival rate to the relay v_i is $\eta_{ki} = \frac{\eta_k r_2^k}{\zeta_k} P_{ki}$. The packets depart to the destination from the relay v_i at the rate $\zeta_{ki} = r_2^k P_{ki}$. This is because that the source and the destination have the equal contact probability with the relay vehicles, and moreover the packet injection rate from the source to the relays equals to that from the relays to the destination, as shown in Fig. 3.9. With the packet arrivals and departures at the relay v_i following the Bernoulli process with mean rates η_{ki} and ζ_{ki}, respectively, the average number of packets held by v_i is

$$E_{ki}[n_r] = \frac{\eta_{ki}}{\zeta_{ki} - \eta_{ki}} = \frac{\eta_k}{\zeta_k - \eta_k}. \tag{3.20}$$

Note that (3.20) holds for every relay. From Little's law, the average packet delay of the flow from v_k is

$$\mathcal{D}_k(N) = \frac{E_k[n_s] + R_k(N)E_{ki}[n_r]}{\eta_k} = \frac{R_k(N) + 1 - \eta_k}{\zeta_k - \eta_k}, \tag{3.21}$$

where $R_k(N)$ is the total number of relay vehicles which have an overlapped mobility region with source v_k. As indicated by (3.21), the average packet delay is dependent of vehicle density in the proximity region of a unicast flow.

We proceed to derive the lower bound and upper bound of the average packet delay of the entire network. We neglect the queueing delay at the source vehicle and the propagation delay in the calculation, as we are only interested in the packet delay caused by vehicles' mobility.

Theorem 3 *For the social-proximity grid-like VANETs, with the two-hop relay scheme $\mathcal{X}$,* w.h.p.*, we obtain the following bounds of the average packet delay $\mathcal{D}(N)$:*

i. when $\kappa + \nu > 1$,

$$\mathcal{D}(\Phi) = O(\psi^2(N/\psi)^{\kappa(2+\gamma)});$$

ii. when $\kappa + \nu \leq 1$,

$$\mathcal{D}(\Phi) = O\big(\frac{\psi^2 \log^2(N/\psi)}{(N/\psi)^{2(\nu-1)-\kappa\gamma}}\big), \text{ and}$$

iii. $\mathcal{D}(\Phi) = \Omega((N/\psi)^{\kappa})$, for $0 < \gamma < 1$; $\mathcal{D}(\Phi) = \Omega((N/\psi)^{\kappa}/\log(N/\psi))$, for $\gamma = 1$; $\mathcal{D}(\Phi) = \Omega((N/\psi)^{\kappa(2-\gamma)})$, for $1 < \gamma < 2$; $\mathcal{D}(\Phi) = \Omega(\log^2(N/\psi))$, for $\gamma = 2$; $\mathcal{D}(\Phi) = \Omega(1)$, for $\gamma > 2$.

Proof. The minimal delay of a flow is achieved when the source delivers the flow packets to its destination with the highest transmission priority. Moreover, the direct packet transmission from the source to the destination has lower average delay compared to the relay transmissions, with the condition that the contact probability between the source and one of its relay vehicles is no larger than the contact probability between the source and its destination. The source encounters the destination on the same road segment with the probability $P_{\mathbb{SD}} = \sum_{i=1}^{G} \pi_{\mathbb{S}}(i)\pi_{\mathbb{D}}(i)$. Therefore, the minimum packet delay is geometric distributed with mean $1/P_{\mathbb{SD}}$. According to (3.12), we obtain a lower bound of $\mathcal{D}(\Phi)$.

Next we derive an upper bound of the average packet delay. Let υ_A and υ_B be a given pair of vehicles whose mobility regions are overlapped. υ_A intends to transmit a packet to υ_B. The transmission between υ_A and υ_B can be scheduled during a time slot only when the following three events occur at the same time: i) υ_A and υ_B are located on the same road segment during the time slot; ii) that road segment is active in the slot; and iii) υ_A and υ_B are both selected for a transmission from υ_A to υ_B. These three events occur with probability ϕ_1, p_{ac} and ϕ_2 respectively. Thus, the distribution of the packet delay between υ_A and υ_B can be treated as geometric with mean

$$\frac{1}{\phi_1 \phi_2 p_{ac}} = O((\frac{\mathcal{A}^{\gamma}}{\pi_1'} \overline{d}_i^N)^2),$$

where ϕ_1 cannot be lower than $(\mathcal{A}^{-\gamma}\pi_1')^2$ and ϕ_2 is $\Omega(1/(\overline{d}_i^N)^2)$. From Lemma 3, we can attain an upper bound of $\mathcal{D}(\Phi)$. The theorem follows. □

3.5 Summary

In this chapter, we have investigated the asymptotic capacity and delay for social-proximity VANETs. We adopt a scalable city grid to deploy the VANET and consider a socialized mobility model for each vehicle. The user applications have proximity nature, i.e., the source and the destination of each flow have the same social spot. Under the proposed two-hop relay scheme, the bounds of the per-vehicle throughput capacity and average per-vehicle throughput have been derived with respect to different network parameters. We have shown that the throughput capacity and delay of the network highly depend on the inherent parameters of mobility patterns. Results in this chapter can be applied to predict the network performance and provide guidance on the design and implementations for large-scale VANETs.

Chapter 4

Downlink Capacity of Vehicular Networks with Access Infrastructure

Wireless access infrastructure, such as Wi-Fi access points and cellular base stations, plays a vital role in offering pervasive Internet services to vehicles. However, the deployment costs of different access infrastructure are highly variable. In this chapter, we analyze the downlink capacity of vehicles and investigate the capacity-cost tradeoffs for the network in which access infrastructure is deployed to provide a downlink data pipe to all vehicles. Three alternatives of wireless access infrastructure are considered, i.e., cellular base stations (BSs), wireless mesh backbones (WMBs), and roadside access points (RAPs). We first derive a lower bound of downlink capacity for each type of access infrastructure. We then present a case study based on an ideal city grid of 400 km^2 with 0.4 million vehicles, in which we examine the capacity-cost tradeoffs for different deployment solutions in terms of both capital expenditures (CAPEX) and operational expenditures (OPEX). Rich implications from the results provide fundamental guidance on the choice of cost-effective wireless access infrastructure for the emerging vehicular networking.

4.1 Introduction

With growing awareness of road safety and the ever-increasing demand for high-speed mobile Internet services, Internet connectivity is becoming a must-have feature of modern vehicles. The telecommunication industry has responded promptly by using off-the-shelf wireless technologies to establish a huge mass market of Internet-connected cars, which is

expected to reach USD 131.9 billion by 2019 [5]. Not surprisingly, cellular technology, such as 3G (UMTS, HSPA, HSPA+) and 4G (LTE), is the top choice for delivering Internet service to cars due to its prominent role in providing reliable and ubiquitous mobile Internet access. Very recently, via AT&T's LTE network, General Motors's LTE-connected car has begun hitting the roads, powering many automotive telematics applications (e.g., emergency services and online diagnostics) [167]. However, the cellular network nowadays faces an uphill battle against the explosive growth of mobile data traffic which has been reportedly doubling each year in the last few years. Exploiting complementary spectrum for vehicular Internet access is thereby an immediate need, which is also part of the solution to the so-called 1000x data challenge [168].

Operating in unlicensed frequency bands, Wi-Fi is astonishingly popular with millions of hotspots deployed all over the world for public Internet access. Due to its low per-bit cost and the feasibility of serving outdoor users at vehicular mobility [33], Wi-Fi is expected to be an attractive and complementary tool to deliver broadband services to moving cars — the built-in Wi-Fi radio or Wi-Fi-enabled mobile devices on board can access the Internet when vehicles *drive-thru* the coverage of Wi-Fi hotspots. Recent advances in Passpoint/Hotspot 2.0, powered by Wi-Fi Alliance, make Wi-Fi more capable of providing secure connectivity than before. Moreover, with subscriber identity module (SIM)-based authentication, seamless roaming between Wi-Fi and cellular networks is also enabled.

In this chapter, we derive the capacity scaling laws for networks with wireless access infrastructure, where Internet services are delivered uniformly from infrastructure to all vehicles in the network. Three candidate solutions are considered to provide Internet connectivity to vehicles, i.e., off-the-shelf 3G or 4G cellular networks, drive-thru or roadside Wi-Fi access points, and particularly a fixed wireless mesh backbone [169], which consists of wirelessly connected mesh nodes (MNs) including one gateway to the Internet. The difference between Wi-Fi access point and wireless mesh is that the latter uses wireless mesh-to-mesh links as backhaul, while the former fully relies on external wired connectivity. Since VANETs have yet to become reality, there remains great uncertainty as to the feasibility of each type of wireless access infrastructure in terms of both network performance and deployment cost. To better understand the capacity-cost issue in vehicular access networks, we consider a scalable urban area where vehicles access Internet through deployed infrastructure nodes. We first analyze the downlink capacity of vehicles to show how it scales with the number of deployed infrastructure nodes. The downlink capacity is defined as the maximum average downlink throughput achieved *uniformly* by all the vehicles from the access infrastructure. Two operation modes of the network are considered to provide pervasive Internet access: ***infrastructure mode***, in which the network is fully covered by infrastructure nodes, i.e., all the vehicles are within the coverage of the infras-

tructure, and thereby only the infrastructure-to-vehicle (I2V) communication is utilized to deliver the downlink traffic; and ***hybrid mode***, in which the network is not fully covered and the downlink flow is relayed to the vehicles in the area without coverage by means of *multi-hop* V2V communications. A lower bound of the downlink capacity is derived for the network with deployment of cellular base stations (BSs), wireless mesh backbones (WMBs), and roadside access points (RAPs) (Wi-Fi/DSRC-like access), respectively. To understand the effect of key factors, such as the deployment scale and the coverage of infrastructure nodes, we present a case study based on a ideal city grid of 400 km^2 with 0.4 million vehicles. Furthermore, we examine the capacity-cost tradeoffs for different deployments. We show that in the hybrid mode, to achieve the same downlink throughput, the network roughly needs X BSs, or $6X$ MNs, or $25X$ RAPs[1]; whereas in the infrastructure mode, if it is desired to improve the downlink throughput by the same amount for each deployment, we roughly need to additionally deploy X BSs, or $5X$ MNs, or $1.5X$ RAPs. By explicitly taking capital expenditures (CAPEX) and operational expenditures (OPEX) of access infrastructures into consideration, the deployment of BSs or WMBs is cost-effective to offer a low-speed downlink rate to vehicles; nonetheless, when providing a high-speed Internet access, the deployment of RAPs outperforms the other two alternatives in terms of deployment costs. This implication can provide valuable guidance on the choice of access infrastructures for the automobile and telecommunication industry. In particular, as automotive industry gears for supporting high-bandwidth applications, non-cellular access infrastructure will play an increasingly important role in offering a cost-effective data pipe for vehicles.

To the best of our knowledge, this research represents the first theoretical study on the capacity-cost tradeoffs when providing pervasive Internet access to vehicles. [170] is the most relevant literature, in which Banerjee *et al.* first examined the performance-cost tradeoffs for VANETs by considering three infrastructure enhancement alternatives: BSs, meshes, and relays. They showed that if the average packet delay can be reduced by a factor of two by adding X BSs, the same reduction needs $2X$ MNs or $5X$ relays. They argued that relays or meshes can be a more cost-effective enhancement due to the high cost of deploying BSs. The objective of their work is to improve network delay by augmenting mobile ad hoc networks with infrastructure, which is different from ours. In addition, our methodology is also different from that adopted in [170]. Notably, quite a few research works [43, 171, 172] focus on content downloading in VANETs. Although we consider a downlink scenario as well, our focus is to unveil capacity-cost tradeoffs for deployment of vehicular access networks.

The remainder of this chapter is organized as follows: Section 4.2 introduces the system

[1] X is used to represent a ratio relationship rather that a specific value.

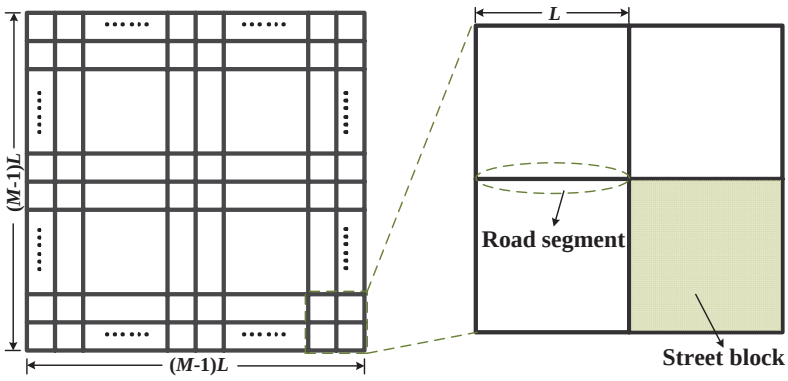


Figure 4.1: A grid-like urban street pattern.

model. In Section 4.3, we analyze the downlink capacity for each type of infrastructure deployment. We present the case study and examine the capacity-cost tradeoffs in Section 4.4. Section 4.5 summarizes this chapter.

4.2 System Model

4.2.1 Urban Street Pattern

Similar to Chapter 3, the street layout of the urban area is modeled by a perfect grid $\mathbb{G}(M, L)$, which consists of a set of M vertical roads intersected with a set of M horizontal roads. Each line segment of length L represents a road segment, as shown in Fig. 4.1. Let $\mathbb{G}$ be a torus to eliminate the border effects. We denote the total number of road segments in $\mathbb{G}$ by $\mathscr{G} = 2(M-1)^2$. The scale of the urban grid is therefore determined by M and L. For example, M is roughly 100 and L is generally from 80 m to 200 m for the downtown area of Toronto [160]. A summary of the mathematical notations used in the chapter is given in Table 4.1.

4.2.2 Spatial Distribution of Vehicles

Taking a snapshot of the city grid where vehicles are moving, it is considered that vehicles are distributed according to a Poisson Point Process (p.p.) Φ with intensity measure Ξ on

Table 4.1: Summary of Notations for Chapter 4

Notations	Description
N	The average number of vehicles in the grid
M	The number of parallel roads in the grid
L	The length of road segment
$\mathscr{G}$	The total number of road segments
$\mathbb{G}$	The urban grid
β	Path-loss exponent
ξ	Vehicle density
W	Communication bandwidth
θ	Ratio between the number of MGs and N_M
N_B	The number of deployed BSs
N_M	The number of deployed MNs
N_R	The number of deployed RAPs
R_V	Transmission radius of V2V communications
R_M	Transmission radius of M2M communications
$\tau_{\mathcal{B}}$	The number of tiers in BS service square
$\tau_{\mathcal{C}}$	The number of tiers in the coverage of BS
λ_B	Downlink capacity for deployment of BSs
λ_B^P	Downlink capacity of B2V transmissions
λ_B^A	Downlink capacity of V2V transmissions (BS)
τ_M	The number of tiers in WMB service square
τ_{MR}	The number of tiers in the coverage of MN
τ_W	The number of tiers in the coverage of WMB
λ_M	Downlink capacity for deployment of WMBs
λ_M^M	Downlink capacity of M2M transmissions
λ_M^P	Downlink capacity of M2V transmissions
λ_M^A	Downlink capacity of V2V transmissions (WMB)
L_R	Service region of an RAP
R_C	Transmission radius of RAP
λ_R	Downlink capacity for deployment of RAPs
λ_R^P	Downlink capacity of R2V transmissions
λ_R^A	Downlink capacity of V2V transmissions (RAP)

$\mathbb{G}(M, L)$. Further, $\Xi(\mathrm{d}x) = \xi \mathrm{d}x$, where $\xi \in (0, +\infty)$, indicating that the average number of vehicles on the road of length $\mathrm{d}x$ is $\xi \mathrm{d}x$. We denote by N the average number of vehicles in the grid. Thus,

$$N = \Xi(\mathbb{G}) = \int_{\mathbb{G}} \Xi(\mathrm{d}x) = \mathscr{G} L \xi. \tag{4.1}$$

And then $\xi = \frac{N}{\mathscr{G}L} = \frac{N}{2L(M-1)^2}$. Notably, $M = \Theta(\sqrt{N})$, since ξ should be positive and bounded. Particularly, ξL is typically much larger than 1 for urban areas. The assumption of p.p. for vehicle distribution on the road can be found in many studies such as [61] and [173].

4.2.3 Propagation and Channel Capacity

The received signal power P_{ij} at receiver j from transmitter i follows the propagation model described in the following: $P_{ij} = KP_i/l(d_{ij})$, where P_i is the transmission power of transmitter i, d_{ij} is the Euclidean distance between i and j, and K is a parameter related to the hardware of communication systems. The path-loss function is given by $l(d_{ij}) = (d_{ij})^{\beta}$, where β is positive and called the path-loss exponent. Typically, we have $\beta = 4$ for urban environments [174]. Note that the phenomenon of channel fluctuations is not considered since a macroscopic description of power attenuation shown above is sufficient for throughput analysis of a long-term average.

The channel capacity of transmitter i and its receiver j is described by Shannon capacity:

$$\mathcal{T}_{ij} = W_{ij} \log_2(1 + SINR_{ij}), \tag{4.2}$$

where W_{ij} is the spectrum bandwidth for the transmission and $SINR_{ij}$ is the *signal-to-interference-plus-noise ratio* (SINR) at receiver j. The interference experienced by receiver j is the aggregation of the signal powers received from all simultaneous transmitters, except its own transmitter i. For ease of comparison, the same path-loss exponent and total bandwidth, which is denoted by W, are applied for each type of deployment of wireless access infrastructure.

4.3 Analysis of Downlink Capacity

We derive a lower bound of downlink capacity in this section for each type of infrastructure deployment, i.e., BSs, WMBs, and RAPs. Asymptotic results are also provided, describing how the downlink capacity scales with the number of deployed infrastructure nodes (directly related to the deployment cost). The derivations in this chapter are mainly based on geometric considerations about interference patterns under certain bandwidth planning. Note that the coverage of the infrastructure node is treated independently from the transmission power in the analysis. It is not necessary to explicitly show the relationship between these two parameters, since our results only depend on the coverage of infrastructure node. In addition, it is noteworthy that the difference between WMB and RAP is that WMBs use wireless mesh-to-mesh links as backhaul, whereas RAPs fully rely on external wired connectivity.

4.3.1 Network with Deployment of BSs

Let N_B denote the number of BSs deployed in the city grid $\mathbb{G}(M, L)$. The grid is thereby divided into N_B squares of equal area, which is denoted by $\mathcal{B}$ and hence $|\mathcal{B}| = (M-1)^2L^2/N_B$. Each square is associated with one BS, which is deployed in the central street block of the square, as shown in Fig. 4.2. It is required that $N_B < (M-1)^2$, i.e., the number of deployed BSs should be less than the total number of street blocks of $\mathbb{G}$. Furthermore, each square consists of multiple tiers co-centered at the BS. $Tier(1)$ of the square is the street block where the BS is located and contains four road segments. The adjacent street blocks surrounding $Tier(1)$ form $Tier(2)$, and so on. It is easy to find that $Tier(\tau)$ contains $16\tau - 12$ road segments. We denote by $\tau_{\mathcal{B}}$ the number of tiers of each square. Hence,

$$\tau_{\mathcal{B}} \leq \lceil \frac{1}{2}\sqrt{\frac{|\mathcal{B}|}{L^2}} + 1 \rceil = \lceil \frac{M-1}{2\sqrt{N_B}} + 1 \rceil, \tag{4.3}$$

where $\lceil \cdot \rceil$ is the ceiling function.

The coverage of the BS is simply considered as a square area of $\tau_{\mathcal{C}}$ tiers, although it is often assumed that the cellular BS covers a hexagon region. A similar approximation can be found in [175]. When $\tau_{\mathcal{C}} \geq \tau_{\mathcal{B}}$, we let $\tau_{\mathcal{C}} = \tau_{\mathcal{B}}$. In this case, the network is fully covered by BSs and thereby operates in the infrastructure mode. When $\tau_{\mathcal{C}} < \tau_{\mathcal{B}}$, the network is partially covered by BSs and operates in the hybrid mode, i.e., BS-to-vehicle (B2V) transmissions and vehicle-to-vehicle (V2V) transmissions coexist. Let $\lambda_B(N, N_B)$ denote the downlink capacity for the deployment of BSs. Furthermore, we denote by λ_B^P

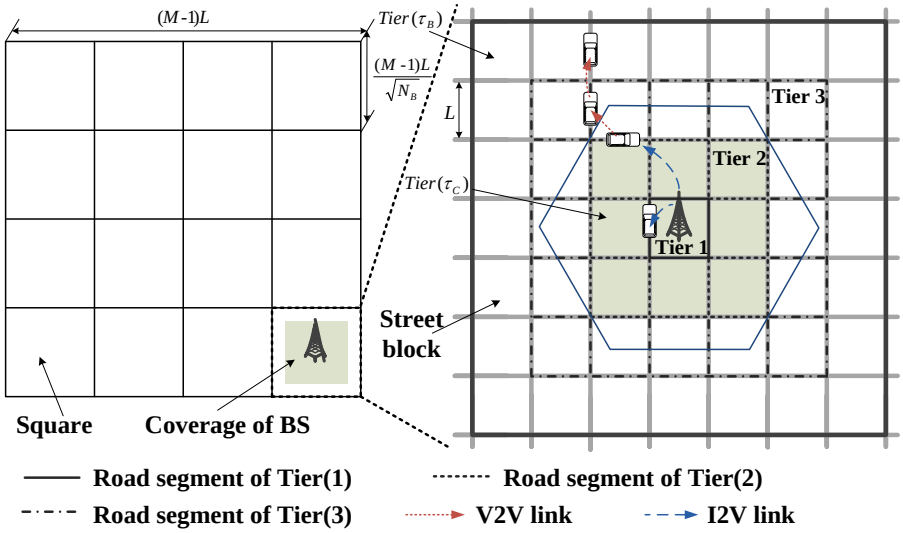


Figure 4.2: Grid-like VANETs with deployment of cellular BSs.

and λ_B^A the downlink capacity of B2V and V2V transmissions, respectively. The downlink capacity in the hybrid mode is determined in the following.

$$\lambda_B(N, N_B) = \min\{\lambda_B^P, \lambda_B^A\}. \tag{4.4}$$

We first study the downlink throughput λ_B^P for B2V transmissions in the hybrid mode. The total bandwidth W is further divided into αW and $(1-\alpha)W$ respectively for B2V and V2V transmissions. A simple spectrum reuse scheme is adopted to mitigate the interference from neighboring squares in B2V transmissions: a square and its eight neighboring squares use different channels for B2V transmissions, each of which is of bandwidth $\alpha W/9$.

Let P_r^τ denote the received signal power of vehicle $\mathcal{V}_0$ on a road segment of $Tier(\tau)$ from its own BS in the square $\mathcal{S}_0$, where $\tau \leq \tau_C$. From the propagation model, we have

$$P_r^\tau \geq \frac{KP_B}{\left[\sqrt{2}L(\tau - \frac{1}{2})\right]^\beta}, \tag{4.5}$$

where P_B is the transmission power of BSs. The interference experienced by $\mathcal{V}_0$, denoted by I_B, is the aggregated signal power of all the other BSs transmitting on the same channel.

Thus

$$I_B \leq \sum_{q=1}^{\infty} 8q \cdot \frac{KP_B}{\left[(3q-\frac{1}{2})\sqrt{|\mathcal{B}|}\right]^{\beta}} = \sum_{q=1}^{\infty} \frac{8qKP_B}{\left[(3q-\frac{1}{2})\frac{(M-1)L}{\sqrt{N_B}}\right]^{\beta}}$$
$$\leq \frac{8KP_B N_B^{\frac{\beta}{2}}}{L^{\beta}(M-1)^{\beta}}\left[(\frac{2}{5})^{\beta} + \int_1^{\infty} \frac{1}{(3q-\frac{1}{2})^{\beta-1}}\mathrm{d}q\right]$$
$$\leq \frac{2^{\beta+1}KP_B N_B^{\frac{\beta}{2}}}{5^{\beta}L^{\beta}(M-1)^{\beta}} \cdot \frac{12\beta+1}{3\beta-6}.$$

Given that $\mathcal{V}_0$ is on a road segment of $Tier(\tau)$, the SINR of the received signal from the BS at $\mathcal{V}_0$ is given by

$$SINR_{\tau} \geq \frac{5^{\beta}(3\beta-6)}{(12\beta+1)2^{\frac{3}{2}\beta+1}}\left[\frac{M-1}{(\tau-\frac{1}{2})\sqrt{N_B}}\right]^{\beta}. \tag{4.6}$$

Throughout this chapter, we neglect the noise as did in previous works like [150] and [151], since we focus on an interference-dominated vehicular environment.

For $\mathcal{V}_0$ on a road segment of $Tier(\tau)$, where $\tau \leq \tau_{\mathcal{C}} - 1$, from (4.2), we have

$$\lambda_B^P = W_{\tau} \log_2(1 + SINR_{\tau}), \tag{4.7}$$

where W_{τ} out of $\alpha W/9$ is the bandwidth allocated to a single vehicle on a road segment of $Tier(\tau)$. Since vehicles on road segments of $Tier(\tau_{\mathcal{C}})$ are required to relay the downlink traffic to vehicles in the area without the BS coverage (see Fig. 4.2), we have

$$\lambda_B^P = \frac{W_{\tau_{\mathcal{C}}} \log_2(1 + SINR_{\tau_{\mathcal{C}}})}{(\sum_{\tau=\tau_{\mathcal{C}}}^{\tau_{\mathcal{B}}} 16\tau - 12)/(16\tau_{\mathcal{C}} - 12)}. \tag{4.8}$$

From (5.13) and (5.14), we have

$$\sum_{\tau=1}^{\tau_{\mathcal{C}}-1} \frac{(16\tau-12)\xi L\lambda_B^P}{\log_2(1+SINR_{\tau})} + \frac{(\sum_{\tau=\tau_{\mathcal{C}}}^{\tau_{\mathcal{B}}} 16\tau - 12)\xi L\lambda_B^P}{\log_2(1+SINR_{\tau_{\mathcal{C}}})} = \frac{\alpha W}{9}.$$

Hence, $\lambda_B^P = \frac{\alpha W/9}{\xi L \mathcal{U}_1}$, where

$$\begin{aligned}\mathcal{U}_1 &= \sum_{\tau=1}^{\tau_{\mathcal{C}}-1} \frac{16\tau - 12}{\log_2(1+SINR_\tau)} + \frac{\sum_{\tau=\tau_{\mathcal{C}}}^{\tau_{\mathcal{B}}} 16\tau - 12}{\log_2(1+SINR_{\tau_{\mathcal{C}}})} \\ &\leq \frac{\sum_{\tau=1}^{\tau_{\mathcal{B}}} 16\tau - 12}{\log_2(1+SINR_{\tau_{\mathcal{C}}})} = \frac{4\tau_{\mathcal{B}}(2\tau_{\mathcal{B}}-1)}{\log_2\left(1+\mathcal{U}_2\left[\frac{M-1}{(\tau_{\mathcal{C}}-\frac{1}{2})\sqrt{N_B}}\right]^\beta\right)} \\ &\leq \frac{2(\frac{M-1}{\sqrt{N_B}}+4)^2}{\log_2\left(1+\mathcal{U}_2\left[\frac{M-1}{(\tau_{\mathcal{C}}-\frac{1}{2})\sqrt{N_B}}\right]^\beta\right)}.\end{aligned}$$

The inequalities hold according to (4.3) and (4.6). We denote $\frac{5^\beta(3\beta-6)}{(12\beta+1)2^{\frac{3}{2}\beta+1}}$ by $\mathcal{U}_2$. A lower bound of λ_B^P is given by

$$\lambda_B^P \geq \frac{\alpha W/(9\xi L)}{2(\frac{M-1}{\sqrt{N_B}}+4)^2} \log_2\left(1+\mathcal{U}_2\left[\frac{M-1}{(\tau_{\mathcal{C}}-\frac{1}{2})\sqrt{N_B}}\right]^\beta\right). \tag{4.9}$$

We denote $\tau_{\mathcal{C}} = \tau_{\mathcal{B}}^\kappa$, $0<\kappa<1$ and $N_B = N^\nu$, $0<\nu<1$. Asymptotically, it is clear that $\lambda_B^P = \Omega(\frac{N_B}{N}\log_2(\frac{N}{N_B})) = \Omega(N^{\nu-1}\log_2 N)$. Notably, $\lambda_B^P = \Omega(\frac{N_B}{N}) = \Omega(N^{\nu-1})$ when $\kappa = 1$, i.e., the network operates in the infrastructure mode.

We then study the downlink capacity λ_B^A for V2V transmissions. Let P_V and $R_V(\geq L)$ denote the transmission power and the transmission radius of V2V communications, respectively. The Carrier Sensing Multiple Access (CSMA) with a carrier sensing radius of $2R_V$ is applied by vehicles to access the channel of bandwidth $(1-\alpha)W$. Due to that simultaneous transmitters cannot be within a distance of $2R_V$ as per the stipulation of CSMA, the distribution of transmitting vehicles in the area without the BS coverage follows a Matérn-like hard core (MHC) p.p. [176]. The MHC p.p. is a dependent marked p.p. of original Poisson p.p. Φ of vehicles. Following [177], an average medium access probability over all the vehicles of Φ is given by

$$P_{ac} = (1-e^{-\bar{\mathcal{N}}})/\bar{\mathcal{N}},$$

where $\bar{\mathcal{N}}$ is the average number of neighbors of a generic vehicle within the carrier sensing

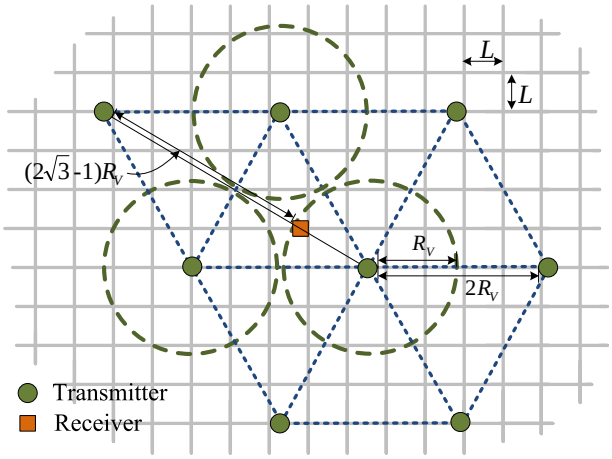


Figure 4.3: A triangular lattice of simultaneous transmitting vehicles.

range. We have

$$\begin{aligned}\bar{\mathcal{N}} &\leq \xi L \cdot 2\lceil\frac{4R_V}{L}\rceil(\lceil\frac{4R_V}{L}\rceil+1)\\ &\leq 8\xi L\big(\frac{2R_V}{L}+1\big)^2.\end{aligned}$$

Therefore,

$$P_{ac} \geq \frac{1-\exp\left(-8\xi L(2R_V/L+1)^2\right)}{8\xi L(2R_V/L+1)^2}. \tag{4.10}$$

Since $\exp\left(-8\xi L(2R_V/L+1)^2\right)$ decays to 0 very fast, we can ignore this exponential term in (4.10).

In V2V transmissions, the received signal power at destination $\mathcal{V}_0$ from its transmitter is given by $P_r \geq KP_V/R_V^{\beta}$. Let $I_{\mathcal{V}_0}$ denote the aggregate interference power experienced by $\mathcal{V}_0$. A close-form expression of $I_{\mathcal{V}_0}$ is difficult to determine. Therefore, we derive an upper bound of $I_{\mathcal{V}_0}$ in the following. Since we consider a high density urban environment, simultaneous V2V transmitters under CSMA scheme with carrier sensing radius $2R_V$ cannot be denser than a triangular lattice [178]. As shown in Fig. 4.3, the six nearest interferers in the first layer are at distance $2R_V$. The next twelve form the second layer, and so on. The distance between the receiver marked and interferers in the first layer is at least R_V, and

at least $(\sqrt{3}q-1)R_V$ in the qth layer, $q>1$. Thus,

$$\begin{aligned} I_{\mathcal{V}_0} &\leq \frac{6KP_V}{R_V^\beta} + \sum_{q=2}^{\infty} 6q \cdot \frac{KP_V}{\left[(\sqrt{3}q-1)R_V\right]^\beta} \\ &\leq \frac{6KP_V}{R_V^\beta}\left[1+\int_1^\infty \frac{1}{(\sqrt{3}q-1)^{\beta-1}}\mathrm{d}q\right] \\ &= \frac{6KP_V}{R_V^\beta}\left(1+\frac{1}{\sqrt{3}(\beta-2)(\sqrt{3}-1)^{\beta-2}}\right). \end{aligned}$$

We denote by $SINR_V$ the SINR of received signal at $\mathcal{V}_0$ from its V2V transmitter. It follows that

$$SINR_V \geq \frac{(\beta-2)(\sqrt{3}-1)^{\beta-2}}{2\sqrt{3}+(\beta-2)(\sqrt{3}-1)^{\beta-2}} = \mathcal{U}_3(\beta). \tag{4.11}$$

Notably, $SINR_V$ is lower bounded by $\mathcal{U}_3(\beta)$, which merely depends on β.

Note that vehicles on road segments of $Tier(\tau_{\mathcal{C}})$ are required to relay the downlink traffic to vehicles from $Tier(\tau_{\mathcal{C}}+1)$ to $Tier(\tau_{\mathcal{B}})$. On the average, every vehicle on road segments of $Tier(\tau_{\mathcal{C}})$ needs to relay the traffic for $\bar{\eta_1}$ vehicles. We have,

$$\begin{aligned} \bar{\eta_1} &= \frac{(\sum_{\tau=\tau_{\mathcal{C}}+1}^{\tau_{\mathcal{B}}} 16\tau-12)\xi L}{(16\tau_{\mathcal{C}}-12)\xi L} \\ &= \frac{(2\tau_{\mathcal{B}}+2\tau_{\mathcal{C}}-1)(\tau_{\mathcal{B}}-\tau_{\mathcal{C}})}{4\tau_{\mathcal{C}}-3} \sim \frac{\tau_{\mathcal{B}}^{2-\kappa}-\tau_{\mathcal{B}}^{\kappa}}{2}. \end{aligned} \tag{4.12}$$

Recall that $\tau_{\mathcal{C}} = \tau_{\mathcal{B}}^{\kappa}$, $0<\kappa<1$. Hence, from (4.10), (4.11) and (4.12), the downlink capacity λ_B^A can be lower bounded as follows.

$$\begin{aligned} \lambda_B^A &\geq \frac{(1-\alpha)W\log_2(1+SINR_V)P_{ac}}{\bar{\eta_1}} \\ &\geq \frac{(1-\alpha)W\log_2(1+\mathcal{U}_3(\beta))}{8\xi L(2R_V/L+1)^2\bar{\eta_1}} \\ &\sim \frac{(1-\alpha)W\log_2(1+\mathcal{U}_3(\beta))}{4\xi L(2R_V/L+1)^2\cdot(\frac{M-1}{2\sqrt{N_B}}+2)^{2-\kappa}}. \end{aligned} \tag{4.13}$$

Let $(R_V/L)=\tau_{\mathcal{B}}^{\mu}$ establish a relationship between the transmission range of vehicles and the number of tiers of $\mathcal{B}$, where $0<\mu<1$. Further, it is required that $\mu<\kappa$, due to that

the transmission range of vehicles should be smaller than that of BSs. Then, we obtain an asymptotic lower bound of λ_B^A from (4.13), i.e., $\lambda_B^A = \Omega((\frac{N_B}{N})^{1-\frac{\kappa}{2}+\mu})$. Recall that $N_B = N^\nu$, $0 < \nu < 1$. Hence, $\lambda_B^A = \Omega(N^{(\nu-1)(1-\frac{\kappa}{2}+\mu)})$.

As per (4.9) and (4.13), we obtain a feasible downlink throughput $\lambda_B(N, N_B)$ when related network parameters are given. In the following, we show an asymptotic lower bound of λ_B. Since $\lambda_B^P = \Omega(\frac{N_B}{N}\log_2(\frac{N}{N_B}))$ and $\lambda_B^A = \Omega((\frac{N_B}{N})^{1-\frac{\kappa}{2}+\mu})$, we have

i). when $\mu < \frac{\kappa}{2}$, $\lambda_B(N, N_B) = \Omega\big(\frac{N_B}{N}\log_2(\frac{N}{N_B})\big)$;

ii). when $\frac{\kappa}{2} \leq \mu < \kappa$, $\lambda_B(N, N_B) = \Omega\big((\frac{N_B}{N})^{1-\frac{\kappa}{2}+\mu}\big)$.

Notably, the downlink throughput of the network mostly depends on the number of deployed BSs, the coverage of the BS, and the transmission radius of the vehicle. For the case in which the transmission range of vehicles is relatively small compared with the coverage of BSs, the downlink throughput of B2V transmissions is lower than that of V2V transmissions and thereby determines the network throughput; with a relatively large vehicular transmission range, V2V communications limit the network throughput due to that medium access probability of vehicles is quite small and thereby degrades the per-vehicle throughput in V2V transmissions.

4.3.2 Network with Deployment of WMBs

Fig. 4.4 shows the network with deployment of WMBs. There are N_M MNs in the network, θN_M of which are functioned as mesh gateways (MGs) connecting to the Internet through the wireline, where $0 < \theta < 1$. Similar to BSs, MGs are regularly deployed in the city grid, each of which is deployed at the center of a square of area $\frac{(M-1)^2L^2}{\theta N_M}$. Let τ_M denote the number of tiers of each square. We have,

$$\tau_M \leq \Big\lceil \frac{M-1}{2\sqrt{\theta N_M}} + 1 \Big\rceil. \tag{4.14}$$

There are $\frac{(1-\theta)N_M}{\theta N_M}$ mesh routers (MRs) deployed in each square, each of which can be reached wirelessly by the MG through one hop or multiple hops. As such, $\frac{1-\theta}{\theta}$ MRs and one MG constitute a WMB in each square. Let R_M denote the transmission radius of mesh-to-mesh (M2M) communications. We consider a regular lattice deployment of MRs with nearest nodal distance of $\frac{\sqrt{2}}{2}R_M$, as shown in Fig. 4.4, so that the Internet traffic is

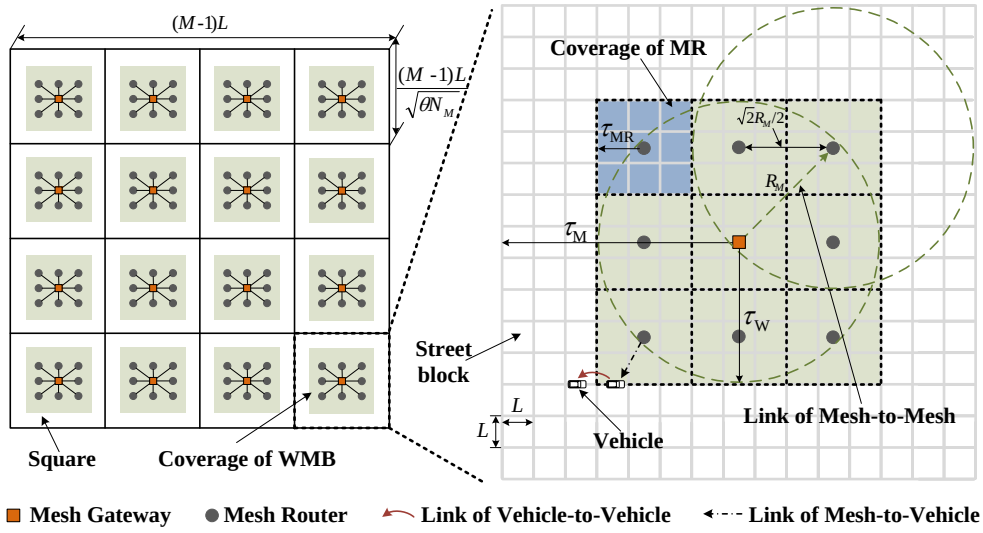


Figure 4.4: Grid-like VANETs with deployment of WMBs.

delivered from the MG to MRs of the first layer through one hop and to MRs of other layers through multiple hops. In addition, each MN covers an area of $\frac{\sqrt{2}}{2}R_M \times \frac{\sqrt{2}}{2}R_M$ with τ_{MR} tiers, where

$$\tau_{MR} \leq \lceil \sqrt{2}R_M/(4L) + 1 \rceil. \tag{4.15}$$

Vehicles within the coverage of the MN receive the downlink traffic through mesh-to-vehicle (M2V) communications. We denote by Q and τ_W the number of layers of MRs and the number of tiers of the coverage region of each WMB, respectively. It follows that $\sum_{q=1}^{Q-1} 8q \leq (1-\theta)/\theta$. Thus, $Q \leq \frac{1}{2}\sqrt{(1-\theta)/\theta} + 1$. We have

$$\tau_W \leq \lceil \frac{\sqrt{2}R_M(3 + \sqrt{(1-\theta)/\theta})}{4L} \rceil. \tag{4.16}$$

When $\tau_W > \tau_M$, let $\tau_W = \tau_M$. The network is completely covered by WMBs if $\tau_W = \tau_M$, otherwise not. In the case where $\tau_W < \tau_M$, vehicles in the area without the WMB coverage receive the downlink traffic through V2V transmissions and require the assistance of vehicles on road segments of $Tier(\tau_W)$. We denote the downlink capacity for the deployment of WMBs by $\lambda_M(N, N_M)$. Furthermore, we denote by λ_M^M, λ_M^P, and λ_M^A the downlink capacity of M2M, M2V, and V2V transmissions in the hybrid mode, respectively.

We first study λ_M^M for delivering Internet traffic from the MG to MRs. All the MNs

apply the same transmission power P_M for M2M transmissions. The total bandwidth W is divided into W_1, W_2, and W_3 respectively for M2M, M2V, and V2V transmissions. It follows that $W = W_1 + W_2 + W_3$. We consider that M2M communications are under the coordination of CSMA scheme with carrier sensing radius $2R_M$. Let I_M denote the interference experienced by a receiver in M2M transmissions. Similar to the derivation of $I_{\mathcal{V}_0}$, I_M can be upper bounded as follows.

$$I_M \leq \frac{6KP_M}{R_M^\beta}\left(1 + \frac{1}{\sqrt{3}(\beta-2)(\sqrt{3}-1)^{\beta-2}}\right).$$

Hence, the SINR of the M2M transmission is given by $SINR_M \geq \mathcal{U}_3(\beta)$. Note that on the average, every MG needs to deliver the downlink traffic for $\frac{1-\theta}{\theta}$ MRs. Given a carrier sensing radius of $2R_M$, an average medium access probability over all MNs, denoted by P'_{ac}, is at least $P'_{ac} = 1/\sum_{q=1}^{2} 8q$. Especially, $P'_{ac} = 1$ for $Q = 1$ and $P'_{ac} \geq 1/9$ for $Q = 2$. Hence, λ_M^M can be lower bounded as follows.

$$\begin{aligned}\lambda_M^M &\geq \frac{W_1 \log_2(1+SINR_M)P'_{ac}}{(1-\theta)/\theta} \\ &\geq \frac{W_1 \log_2(1+\mathcal{U}_3(\beta))P'_{ac}}{(1-\theta)/\theta}.\end{aligned} \tag{4.17}$$

We then study λ_M^P for delivering traffic from the MN to vehicles within its coverage. Similarly, to mitigate the interference from neighboring MNs in M2V transmissions, an MN and its neighbors (at most eight) use different channels for M2V transmissions, each of which has bandwidth $W_2/9$. Let P_{MV} denote the transmission power for M2V communications. The interference experienced by vehicles in M2V communications, denoted by I_{MV}, is given by

$$I_{MV} \leq \sum_{q=1}^{\infty} \frac{8qKP_{MV}}{\left[(3q-\frac{1}{2})\frac{\sqrt{2}}{2}R_M\right]^\beta} \leq \frac{2^{\frac{3}{2}\beta+1}KP_{MV}}{5^\beta R_M^\beta} \cdot \frac{12\beta+1}{3\beta-6}.$$

Let P_{MV}^τ denote the received power of a vehicle on the road segment of $Tier(\tau)$ from its own MN, where $\tau \leq \tau_{MR}$. Since $P_{MV}^\tau \geq KP_{MV}/(\sqrt{2}L(\tau-\frac{1}{2}))^\beta$, we have

$$SINR'_\tau \geq \frac{5^\beta(3\beta-6)}{(12\beta+1)2^{2\beta+1}}\left[\frac{R_M}{(\tau-\frac{1}{2})L}\right]^\beta, \tag{4.18}$$

where $SINR'_\tau$ is the SINR of the received signal from the MN for vehicles on road segments

of $Tier(\tau)$.

Similar to the deployment of BSs, W_τ out of $W_2/9$ is the bandwidth allocated to a single vehicle on the road segment of $Tier(\tau)$ for each coverage of MNs. Since vehicles on road segments of $Tier(\tau_W)$ of the WMB are required to relay the downlink traffic, additional bandwidth needs to be allocated to vehicles on the road segments of $Tier(\tau_{MR})$ for MNs located in the outmost layer Q of the WMB, as shown in Fig. 4.4. In the following, we consider an MN on the boundary of the WMB and derive a lower bound of λ_M^P. For vehicles of $Tier(\tau)$, where $\tau \leq \tau_{MR} - 1$, we have

$$\lambda_M^P = W_\tau \log_2(1 + SINR'_\tau). \tag{4.19}$$

We denote by $\bar{\eta_2}$ the average number of vehicles that need a vehicle of $Tier(\tau_W)$ to relay the downlink traffic. We have,

$$\bar{\eta_2} = \frac{\sum_{\tau=\tau_W+1}^{\tau_M} 16\tau - 12}{16\tau_W - 12} \leq \frac{\tau_M^2 - \tau_W^2}{\tau_W - 1}. \tag{4.20}$$

Hence,

$$\lambda_M^P = \frac{W_{\tau_{MR}} \log_2(1 + SINR'_{\tau_{MR}})}{1 + \bar{\eta_2}}. \tag{4.21}$$

From (4.19), (4.20), and (4.21), it holds that $\lambda_M^P = \frac{W_2/9}{\xi L \mathcal{U}_4}$, where

$$\begin{aligned} \mathcal{U}_4 &= \sum_{\tau=1}^{\tau_{MR}-1} \frac{(16\tau - 12)}{\log_2(1 + SINR'_\tau)} + \frac{(16\tau_{MR} - 12)(1 + \bar{\eta_2})}{\log_2(1 + SINR'_{\tau_{MR}})} \\ &\leq \frac{4\tau_{MR}(2\tau_{MR} - 1) + \bar{\eta_2}(16\tau_{MR} - 12)}{\log_2(1 + SINR'_{\tau_{MR}})}. \end{aligned}$$

Let $\mathcal{U}_5$ denote the numerator of the last fraction, which is an upper bound of the average number of vehicles served by a single MN. From (4.14), (4.15), and (4.16), we can attain a lower bound of λ_M^P, i.e.,

$$\begin{aligned} \lambda_M^P &\geq \frac{W_2 \log_2(1 + SINR'_{\tau_{MR}})}{9\xi L \mathcal{U}_5} \\ &\sim \frac{W_2}{9\xi L \mathcal{U}_5} \log_2 \left(1 + \frac{5^\beta (3\beta - 6)}{(12\beta + 1) 2^{\frac{1}{2}\beta + 1}}\right). \end{aligned} \tag{4.22}$$

Furthermore, let $N_M = N^{\gamma}$, where $0 < \gamma < 1$. We have $\lambda_M^P = \Omega(\frac{N_M}{N}) = \Omega(N^{\gamma-1})$ asymptotically.

We follow the derivation of (4.13) to obatin λ_M^A, since V2V communications are with the same configurations in both BSs and WMBs deployments. Hence,

$$\begin{aligned} \lambda_M^A &\geq \frac{W_3 \log_2(1 + SINR_V) P_{ac}}{\bar{\eta_2}} \\ &\geq \frac{W_3 \log_2(1 + \mathcal{U}_3(\beta))(\tau_W - 1)}{8\xi L(2R_V/L + 1)^2(\tau_M^2 - \tau_W^2)}. \end{aligned} \tag{4.23}$$

Asymptotically,

$$\lambda_M^A = \Omega\bigg(\frac{N_M(R_M/L)}{N(R_V/L)^2}\bigg).$$

We let $(R_M/L) = \tau_M^{\sigma_1}$ to establish a relationship between the transmission range of MNs and the size of the mesh square, where $0 < \sigma_1 < 1$. Similarly, $R_V/L = \tau_M^{\sigma_2}$, where $0 < \sigma_2 < 1$ and $\sigma_2 < \sigma_1$. Therefore, $\lambda_M^A = \Omega(N^{(\gamma-1)(1+\sigma_2-\frac{1}{2}\sigma_1)})$. From (4.17), (4.22), and (4.23), we can attain a lower bound of $\lambda_M(N, N_M)$ as follows.

$$\lambda_M(N, N_M) = \min\left(\frac{\lambda_M^M}{\mathcal{U}_5}, \min\left(\lambda_M^P, \lambda_M^A\right)\right). \tag{4.24}$$

Notably, $\lambda_M^M/\mathcal{U}_5 = \Omega(N^{\gamma-1})$. Then, we obtain the following asymptotic bound of λ_M^M in the hybrid mode:

i). when $\sigma_2 < \frac{1}{2}\sigma_1$,

$$\lambda_M(N, N_M) = \Omega\big(\frac{N_M}{N}\big);$$

ii). when $\frac{1}{2}\sigma_1 \leq \sigma_2 < \sigma_1$,

$$\lambda_M(N, N_M) = \Omega\big((\frac{N_M}{N})^{1-\frac{1}{2}\sigma_1+\sigma_2}\big).$$

When the network is fully covered by deployed WMBs, each MN covers an area of

$(M-1)^2L^2/N_M$. Hence, $R_M \geq \sqrt{2}(M-1)L/\sqrt{N_M}$. We then have

$$\begin{aligned}\lambda_M^P &\geq \frac{(W-W_1)\log_2(1+SINR'_{\tau_{MR}})}{9N/N_M} \\ &\sim \frac{(W-W_1)N_M}{9N}\log_2\left(1+\frac{5^\beta(3\beta-6)}{(12\beta+1)2^{\frac{1}{2}\beta+1}}\right).\end{aligned}$$

It follows that $\lambda_M(N,N_M) = \min\left(N_M\lambda_M^M/N, \lambda_M^P\right)$ in the infrastructure mode. Asymptotically, $\lambda_M(N,N_M) = \Omega(N_M/N) = \Omega(N^{\gamma-1})$.

4.3.3 Network with Deployment of RAPs

We consider that the coverage of the RAP is one-dimensional along the road, as shown in Fig. 4.5. There are N_R RAPs regularly deployed in the network and each RAP provides Internet access service to vehicles on the road of length L_R, which is called the RAP cell. It can be seen that $L_R = \frac{2(M-1)^2L}{N_R}$. The coverage radius of RAP is denoted by R_C. When $R_C > \frac{1}{2}L_R$, let $R_C = \frac{1}{2}L_R$. The network is fully covered by RAPs if $R_C = \frac{1}{2}L_R$. To provide pervasive Internet access, the network operates in the hybrid mode when $R_V < R_C < \frac{1}{2}L_R$: vehicles within the coverage of the RAP receive the downlink traffic through RAP-to-vehicle (R2V) communications; vehicles at distance $(R_C - R_V, R_C]$ from the RAP are required to relay the downlink traffic for vehicles in the area without the RAP coverage, given the transmission radius of V2V communications R_V. The downlink capacity for the deployment of RAPs is denoted by $\lambda_R(N,N_R)$. Moreover, the downlink capacity of R2V and V2V transmissions are denoted by λ_R^P and λ_R^A, respectively. Similarly, in the hybrid mode,

$$\lambda_R(N,N_R) = \min\left\{\lambda_R^P, \lambda_R^A\right\}. \tag{4.25}$$

We first study the downlink throughput λ_R^P in the hybrid mode. A spectrum reuse scheme is adopted to mitigate the inter-RAP interference: i) RAPs deployed along the same road operate on one common channel; ii) RAPs on any two adjacent parallel roads use different channels; and iii) RAPs on the horizontal roads and on the vertical roads use different channels. To that end, four different communication channels, each of which has bandwidth $\frac{1}{4}\phi W$, are allocated. The remaining bandwidth of $(1-\phi)W$ is allocated for V2V communications. The interference I_d experienced by a vehicle at distance d from the RAP, where $d \leq R_C$, in R2V communications is the aggregated signal power from all the other RAPs operating on the same channel, as shown in the Fig. 4.6. We have

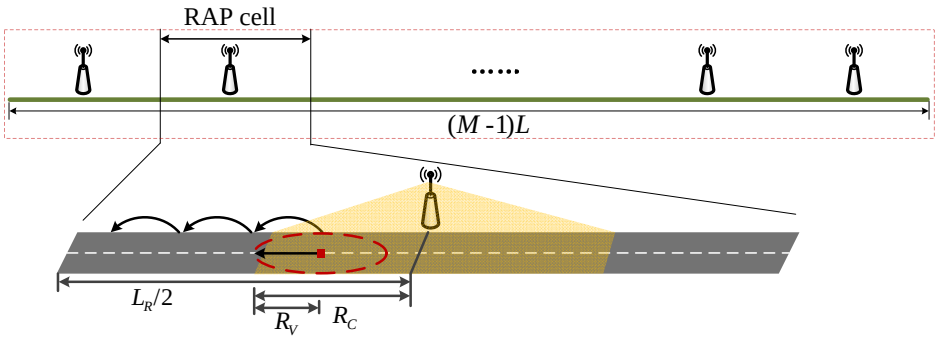


Figure 4.5: Grid-like VANETs with deployment of RAPs.

$$
\begin{aligned}
I_d &\leq \sum_{q=1}^{\infty}\left[\frac{KP_R}{(qL_R-d)^{\beta}}+\frac{KP_R}{(qL_R+d)^{\beta}}\right] \\
&\quad+\sum_{q=1}^{\infty}\frac{2KP_R}{(2qL)^{\beta}}+\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{4KP_R}{(i^2(2L)^2+j^2L_R^2)^{\frac{\beta}{2}}} \\
&\leq 2KP_R\left[\frac{1}{(L_R-d)^{\beta}}+\int_1^{\infty}\frac{1}{(qL_R-d)^{\beta}}\mathrm{d}q\right] \\
&\quad+\frac{2^{1-\beta}\beta KP_R}{(\beta-1)L^{\beta}}+\frac{2^{2-\beta}KP_R}{(LL_R)^{\frac{\beta}{2}}}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{(ij)^{\frac{\beta}{2}}} \\
&\leq \frac{2KP_R}{\beta-1}\left(\frac{\beta L_R-d}{L_R(L_R-d)^{\beta}}+\frac{\beta}{(2L)^{\beta}}\right)+\frac{2^{2-\beta}\beta^2KP_R}{(\beta-2)^2(LL_R)^{\frac{\beta}{2}}},
\end{aligned}
$$

where P_R is the transmission power of RAPs. The SINR of received signal from the RAP is thereby given as follows.

$$
SINR_d \geq \frac{(\beta-1)/(2d^{\beta})}{\frac{\beta L_R-d}{L_R(L_R-d)^{\beta}}+\frac{\beta}{(2L)^{\beta}}+\frac{2^{1-\beta}(\beta-1)\beta^2}{(\beta-2)^2(LL_R)^{\frac{\beta}{2}}}}=\mathcal{U}_6(d).
$$

For vehicle $\mathcal{V}_d$ at distance d from the RAP, where $d \leq R_C$, it holds that

$$
\lambda_R^P = W_d \log_2(1+SINR_d)
$$

where W_d out of $\frac{1}{4}\phi W$ is the bandwidth allocated to $\mathcal{V}_d$. As discussed before, vehicles at distance $(R_C - R_V, R_C]$ from the RAP are required to relay the downlink traffic to the vehicles at distance $(R_C, \frac{1}{2}L_R]$, which yields an average relaying traffic load of that $\bar{\eta_3} = (\frac{1}{2}L_R - R_C)/R_V$. Therefore, for vehicles at distance $d \in (R_C - R_V, R_C]$ from the RAP,

$$\lambda_R^P = \frac{W_d \log_2(1 + SINR_d)}{1 + \bar{\eta_3}}.$$

Given the constraint of the total bandwidth, we have

$$\begin{aligned} \lambda_R^P &\geq \frac{\frac{1}{4}\phi W}{\frac{2\xi(R_C - R_V)}{\log_2(1+SINR_{R_C-R_V})} + \frac{2\xi(1+\bar{\eta_3})R_V}{\log_2(1+SINR_{R_C})}} \\ &\geq \frac{\frac{1}{8}\phi W/\xi}{\frac{R_C - R_V}{\log_2(1+\mathcal{U}_6(R_C-R_V))} + \frac{R_V + \frac{1}{2}L_R - R_C}{\log_2(1+\mathcal{U}_6(R_C))}}. \end{aligned} \tag{4.26}$$

Further, we let $R_C = (\frac{1}{2}L_R)^{\rho_1}$ and $R_V = (\frac{1}{2}L_R)^{\rho_2}$, where $0 < \rho_2 < \rho_1 < 1$. Denoting $N_R = N^{\varphi}$, where $0 < \varphi < 1$, it can be obtained that $\lambda_R^P = \Omega(\frac{N_R}{N}\log_2(\frac{N}{N_R})) = \Omega(N^{\varphi-1}\log_2 N)$ asymptotically when $\rho_1 < \frac{1}{2}$; $\lambda_R^P = \Omega(\frac{N_R}{N}) = \Omega(N^{\varphi-1})$ when $\rho_1 = \frac{1}{2}$; $\lambda_R^P = \Omega(\frac{N_R}{N}\log_2(1 + (\frac{N_R}{N})^{\beta(\rho_1 - \frac{1}{2})})) = \Omega(N^{(\varphi-1)[1+\beta(\rho_1-\frac{1}{2})]})$ when $\rho_1 > \frac{1}{2}$.

The derivation of λ_R^A is in the same manner.

$$\begin{aligned} \lambda_R^A &\geq \frac{(1-\phi)W\log_2(1 + SINR_V)P_{ac}}{\bar{\eta_3}} \\ &\geq \frac{(1-\phi)W\log_2(1 + \mathcal{U}_3(\beta))R_V}{8\xi L(2R_V/L + 1)^2(\frac{1}{2}L_R - R_C)}. \end{aligned} \tag{4.27}$$

Asymptotically, $\lambda_R^A = \Omega((N_R/N)^{1+\rho_2}) = \Omega(N^{(\varphi-1)(1+\rho_2)})$.

As per (4.26) and (4.27), $\lambda_R(N, N_R)$ can be obtained from (4.25) when values of all the impact factors are determined. In addition, the asymptotic bound of $\lambda_R(N, N_R)$ is given by

i). when $\rho_1 \leq \frac{1}{2}$,

$$\lambda_R(N, N_R) = \Omega((N_R/N)^{1+\rho_2});$$

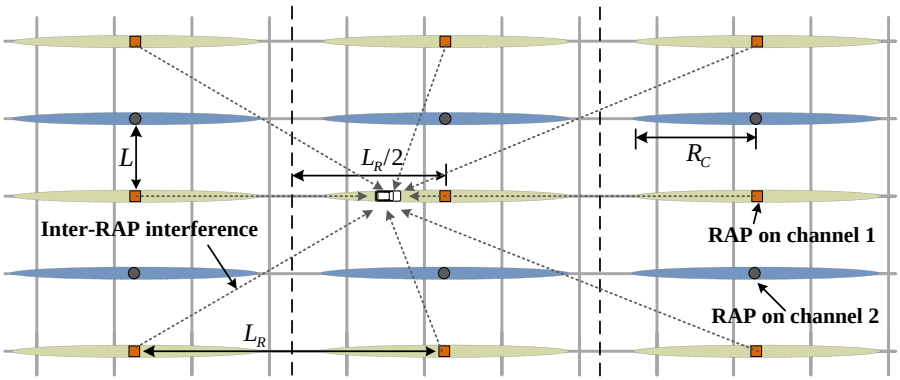


Figure 4.6: An illustration of inter-RAP interference for horizontal roads.

ii). when $\frac{1}{2} < \rho_1 < 1$,

$$\lambda_M(N, N_M) = \Omega\big((N_R/N)^{\max[1+\rho_2, 1+\beta(\rho_1-\frac{1}{2})]}\big).$$

Especially, when the network is completely covered by RAPs, $\lambda_R(N, N_R) = \lambda_R^P \geq WN_R \log_2(1+\mathcal{U}_6(R_C))/(4N)$. The asymptotic result of $\lambda_R(N, N_R)$ in the infrastructure mode is the same as that of λ_R^P in the hybrid mode.

4.4 Case Study

We present a case study of downlink capacity of vehicles based on the analytical results from Section 5.5. The objective is to evaluate the impact of key factors, i.e., the number of infrastructure nodes deployed and the coverage of infrastructure nodes, on capacity performance and compare the three types of infrastructures in terms of the deployment cost. The values of parameters for this study are given in Table 4.2.

Table 4.2: Values of Parameters

Parameter	Value	Parameter	Value
M	201	L	100 m
ξ	0.05 veh/m	N	4×10^5
W	10 MHz	β	4
R_V	100 m	θ	0.25

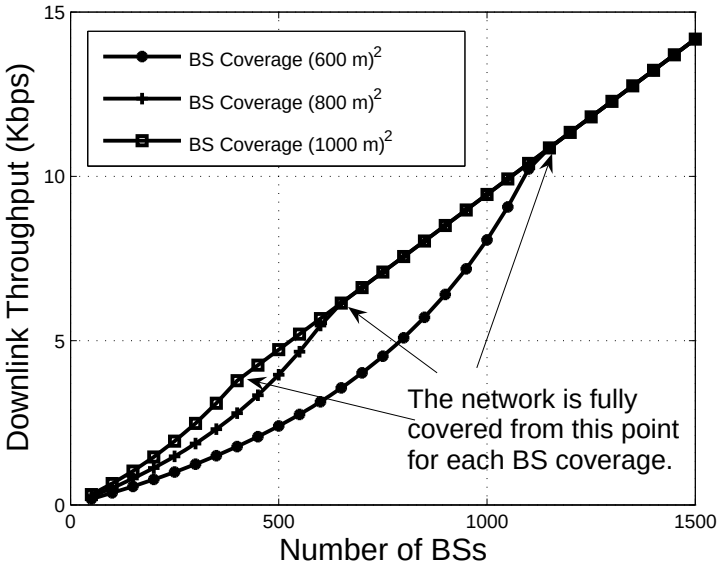


Figure 4.7: Network with deployment of BSs.

4.4.1 Impact of Coverage of Infrastructure Nodes

We consider a ideal city grid of 20 km×20 km with an average vehicle density of 0.05 vehicle per meter (veh/m). The total bandwidth of 10 MHz is applied in all three types of infrastructure deployment. The bandwidth allocation is done to maximize the downlink throughput for each case. The downlink capacity is plotted with respect to the number of deployed infrastructure nodes, as shown in Fig. 4.7, Fig. 4.8, and Fig. 4.9. With the increase of the number of deployed infrastructure nodes, the network transits from being partially covered to being fully covered, and accordingly the downlink throughput increases gradually. The impact of coverage of infrastructure nodes on downlink throughput is also investigated. Three different sizes of BS footprint are considered in Fig. 4.7. We show that for each BS coverage, the achievable downlink throughput increases faster than a linear increase with N_B in the hybrid mode. The reason is that the relaying traffic load of relay vehicles decreases very fast when the network gradually becomes fully covered and thereby the capacity of V2V communications increases. When the network is fully covered by BSs, the downlink throughput increases almost linearly with N_B. In addition, it is very intuitive that the network needs more BSs to be fully covered with a smaller size of BS coverage. The similar insights for the other two deployments are obtained from Fig. 4.8 and Fig. 4.9.

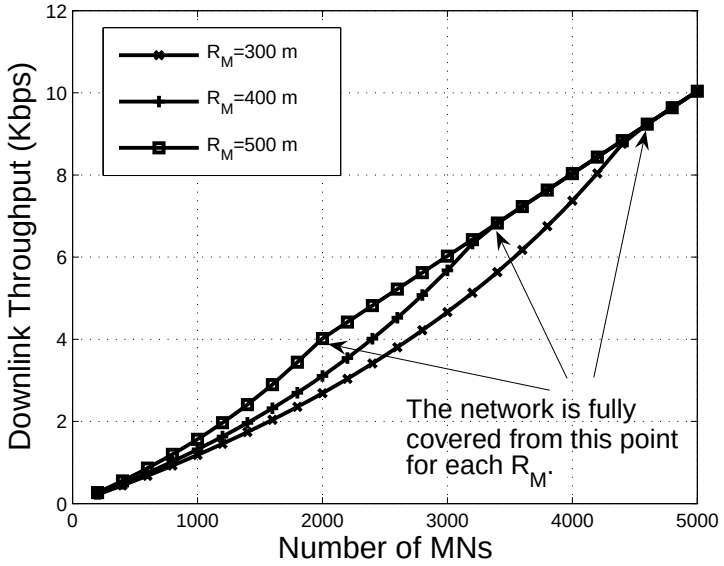


Figure 4.8: Network with deployment of WMBs.

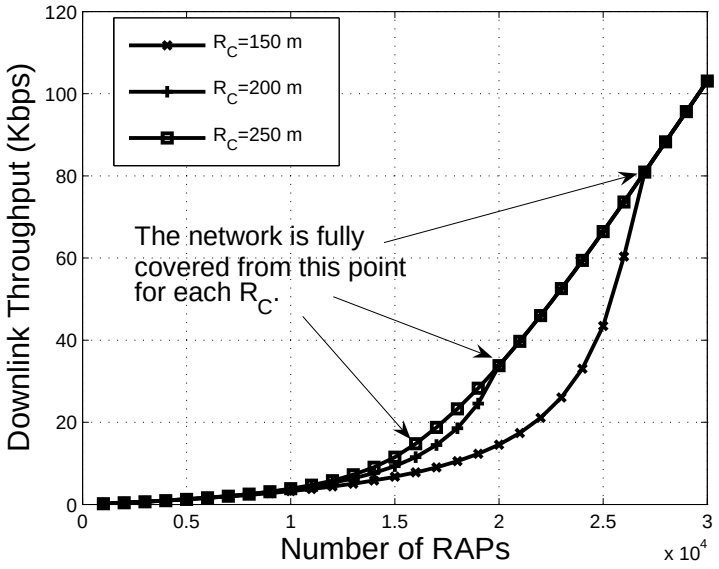


Figure 4.9: Network with deployment of RAPs.

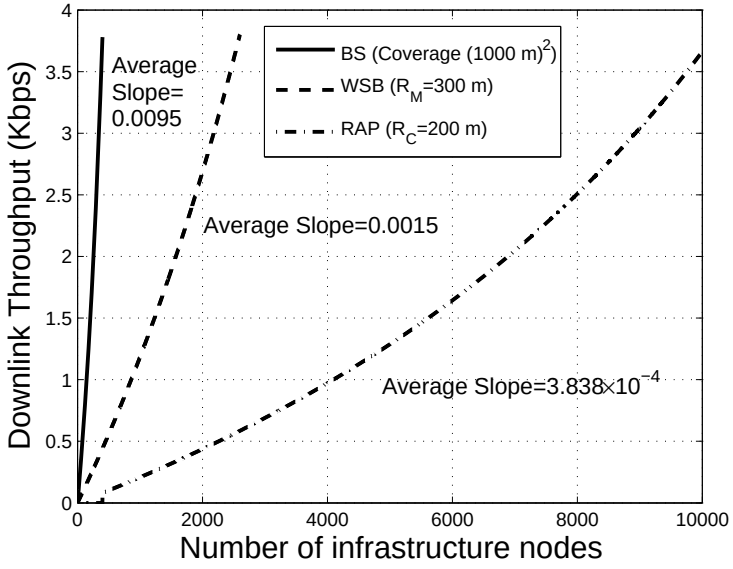


Figure 4.10: Comparison of number of deployed infrastructure nodes in the hybrid mode.

4.4.2 Comparison of Deployment Scales

Different trends of downlink throughput are shown in Fig. 4.10 when the network is not fully covered by any type of infrastructure. From the average slope of each curve, an important observation is attained that the network roughly needs X BSs, or $6X$ MNs, or $25X$ RAPs to achieve a certain downlink throughput in the hybrid mode. A whole picture of the comparison is shown in Fig. 4.11. Regardless of the operation mode (hybrid or infrastructure), on the average, the network requires X BSs, or $5X$ MNs, or $15X$ RAPs to achieve a downlink throughput less than 15 Kbps with our settings. Furthermore, it is observed that more MNs are needed than RAPs to achieve the same throughput after the Point A shown in Fig. 4.11. The reason is that in the infrastructure mode, the relaying traffic load from the MG to MRs limits the downlink throughput, and there is almost no benefit from better coverage of MNs due to that the network is fully covered by either RAPs or MNs. The downlink throughput decreases severely, as shown in Fig. 4.12, with a very small value of θ, which reflects the backhaul capability of wireless mesh networks. Another observation from Fig. 4.11 is that we roughly need to additionally deploy X BSs, or $5X$ MNs, or $1.5X$ RAPs to improve the downlink throughput by the same amount, given that the network operates in the infrastructure mode.

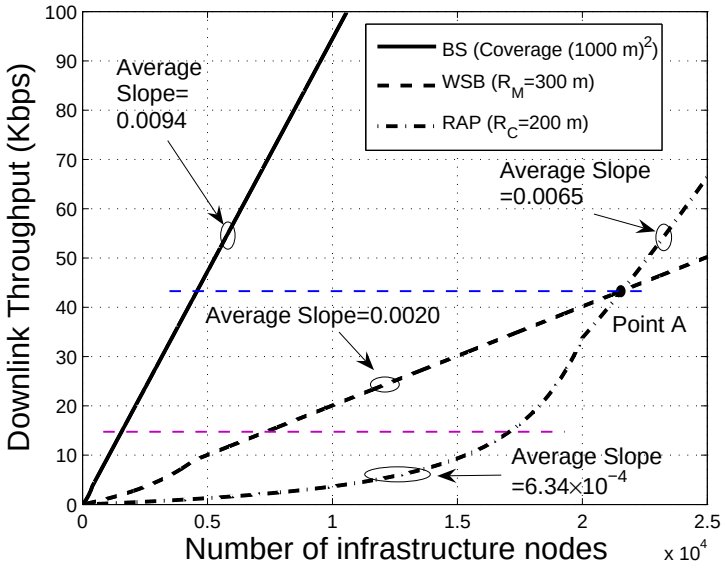


Figure 4.11: Comparison of number of deployed infrastructure nodes in the infrastructure mode.

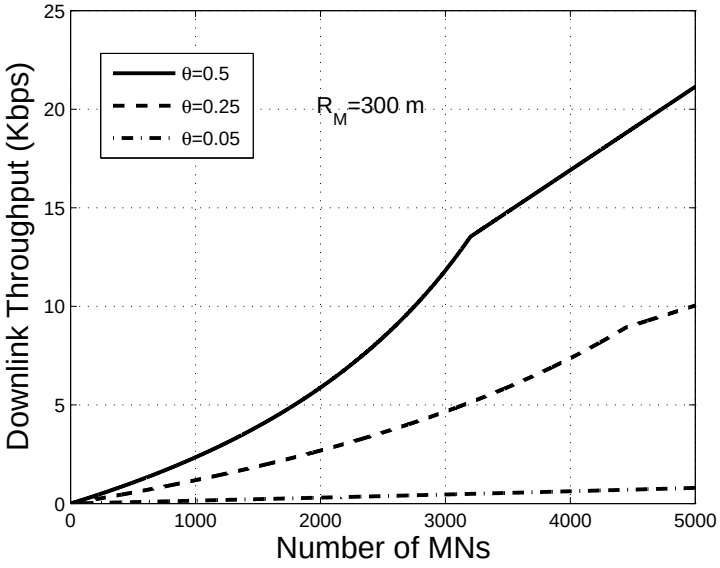


Figure 4.12: Impact of θ on the downlink throughput for the deployment of WMBs.

4.4.3 Capacity-Cost Tradeoffs

Deployment cost plays an important role in determining the cost-effective access infrastructure. CAPEX and OPEX contribute to the major component of the deployment cost [179]. As per the cost models in [179], the estimated deployment cost of each type of access infrastructure is described in Table 4.3. It can be seen that when the network operates in the hybrid mode (low-capacity regime), the deployment of BSs or WMBs is cost-effective for a five-year operation period (the cost is roughly $120X$ K€ to deploy X BSs, or $6X$ MNs). On the other hand, when the network operates in the infrastructure mode (high-capacity regime), the deployment of RAPs outperforms the other two alternatives in terms of deployment costs for a given downlink throughput requirement. For instance, to provide a downlink throughput of 40 Kbps to all the vehicles, roughly we need to pay 530 M€ for the deployment of 4200 BSs, or 210 M€ for the deployment of 2.1×10^4 RAPs for a five-year period. From Fig. 4.11, the choice of the cost-effective access infrastructure can be made as per the data demand of vehicles. Notably, non-cellular infrastructure like RAPs is a good choice to offer a cost-effective high-speed data pipe for vehicles.

Table 4.3: Estimated Deployment Cost(K€)

Deployment Cost	BS	MG (MR)	RAP
CAPEX	58.9	10.9 (7.0)	3.0
OPEX (per year)	13.4	2.9 (2.0)	1.4
5-Year Cost	125.9	25.4 (17.0)	10.0

4.5 Summary

In this chapter, we have investigated the capacity-cost tradeoffs for different wireless access infrastructures in vehicular access networks. The considered alternatives of wireless access infrastructure include BSs, WMBs, and RAPs, which are respectively deployed to provide downlink Internet service to all the vehicles uniformly in the network. The downlink capacity of vehicles for each kind of deployment has been lower-bounded under the same set of benchmark models by considering an ideal city grid with vehicles randomly distributed on the roads. A case study has been presented to examine the capacity-cost tradeoffs of different solutions in terms of both CAPEX and OPEX. Offering fundamental guidance, results in this chapter imply that it is necessary to choose a cost-effective access infrastructure according to the data demand of vehicle users.

Chapter 5

Wi-Fi Hotspot at Signalized Intersection: Capacity of Drive-Thru Internet Access

In this chapter, we particularly investigate the average throughput capacity of drive-thru Internet access considering interrupted vehicle traffic flow and with a focus on cost-effectiveness of Wi-Fi solution for vehicular Internet access. Specifically, we define the cost-effectiveness as the cost saving by deploying and operating a low-cost Wi-Fi infrastructure instead of a costly benchmark cellular network. To characterize the service quality of Wi-Fi deployment, we also define the normalized service delay to fulfill data application via Wi-Fi normalized by that via the cellular. To derive the service delay, we analyze the average throughput capacity of a generic vehicle in the Wi-Fi network and the average downlink capacity in the cellular network. Especially, we propose deploying Wi-Fi access point at signalized intersection and study the fundamental influence of traffic signals (which yield an interrupted vehicle traffic) on Wi-Fi access. Then, we examine the tradeoff between cost-effectiveness and normalized service delay by identifying interplays between controllable (e.g., the density of Wi-Fi deployment and user's satisfaction) and uncontrollable parameters (e.g., vehicle traffic statistics). Our results reveal the fundamental relation between the service quality and cost-effectiveness of Wi-Fi solution, and provide a quick and efficient way of determining the Wi-Fi deployment strategy and the corresponding cost savings.

5.1 Introduction

Compared to cellular access with wide availability, the Wi-Fi infrastructure has limited utility of the service offering for vehicles over intermittent network connectivity, as observed in real-world tests, e.g., [39]. Typically, vehicle users on the road have to experience a number of drive-thrus/connections to fulfill a mobile application (e.g., buffering a video clip of 100 MBytes from the Internet), implying a large service delay that degrades the user's satisfaction. While the Wi-Fi footprint can be enlarged by deploying more Wi-Fi access points (APs), the network cost or TCO[1] would be increased as well. Especially, a solution that tries to achieve a ubiquitous coverage as cellular networks is prohibitive and not practical [48]. Therefore, great uncertainty remains as to whether it makes economic sense to deploy Wi-Fi networks for highly mobile vehicle users. Thanks to the new generation Wi-Fi hotspot, many mobile network operators (MNOs), such as AT&T and China Mobile, have shown strong interest in the Wi-Fi solution. However, the real benefit of Wi-Fi solution should be validated through cost-effectiveness analysis considering user's satisfaction so that MNOs would be fully convinced to turn the strong interest to strong commitment to deploy large-scale outdoor Wi-Fi networks in favor of vehicle users.

As an effort to that end, in this chapter, we study cost-effectiveness of vehicular Internet access through Wi-Fi hotspots. The *cost-effectiveness* is defined as the TCO saving by deploying and operating drive-thru Wi-Fi networks instead of the cellular network for vehicular Internet access. To establish the relationship between cost-effectiveness and network performance, we analyze the maximum average throughput of individual vehicle under any given density of deployed Wi-Fi APs (different deployment yields different cost-effectiveness), which can be used to determine the service delay once the total throughput required for fulfilling a data application is given. For apple-to-apple comparison, the network performance of benchmark cellular network is also analyzed so that the service delay experienced in Wi-Fi network can be normalized by that in cellular network. This *normalized service delay* is then able to reflect the service quality degradation because of using Wi-Fi networks characterized by the corresponding cost-effectiveness. We examine tradeoff between cost-effectiveness and normalized service delay, and then demonstrate the benefit of Wi-Fi solution quantitatively.

The main contributions of this research are as follows:

- We deploy Wi-Fi APs at signalized intersections and study the fundamental influence of traffic signals on drive-thru Internet access, which has attracted significant research

[1]Total cost of ownership, including one-time cost component (CAPEX, capital expenditures) and recurring cost component (OPEX, operational expenditures).

interest. Deploying AP at intersection can reduce the need for non-line-of-sight (N-LOS) transmissions along the road so as to provide better service coverage [180]. One example of real-world deployment is Wickedly Fast Wi-Fi network which covers the downtown area of San Jose, CA, USA [181]. Interestingly, we show a significant throughput gain for vehicles stopping at intersections due to red signals. Our modeling and analysis are validated via SUMO and NS-3 simulations.

- We propose a framework for cost-effectiveness analysis of Wi-Fi network, in which the cost-effectiveness (how much TCO can be saved for the MNO) and normalized service delay (how much service degradation the vehicle user will tolerate) are mathematically defined and the explicit relation between these two metrics are established. Quantitatively, we show that the TCO of Wi-Fi can be traded with user's satisfaction, which could aid MNOs in strategic decision-making for Wi-Fi deployment. Our framework also lays a foundation for helping in understanding cost-effectiveness of other complementary wireless technologies for vehicle users, such as small cells and super Wi-Fi operating on TV white space.

The remainder of the chapter is organized as follows. Section 5.2 presents the problem formulation. We analyze the time-average throughput capacity of drive-thru Wi-Fi and the average downlink capacity of cellular network in Section 5.3 and Section 5.4, respectively. Section 5.5 presents the cost-effectiveness analysis. Further discussions are given in Section 5.6. Section 5.7 provides concluding remarks.

5.2 Problem Formulation

We study the cost-effectiveness of drive-thru Wi-Fi access in a city area Ω with a set of moving vehicles. Two different networking scenarios are considered. In the first scenario (`Wi-Fi`), the Internet gateways are sparsely deployed Wi-Fi APs (at intersections) and vehicles have only opportunistic drive-thru access. We use the second scenario (`Cellular`) with cellular macrocell BSs providing full service coverage as a benchmark for performance comparison.

We denote by N_W the number of Wi-Fi APs deployed in `Wi-Fi` scenario, and N_C the number of macrocell BSs deployed in `Cellular` scenario. In both scenarios, vehicles consume Internet data services as long as Internet connectivity is available. The maximum average data throughput achieved by individual vehicles are denoted by F_W and F_C, respectively for `Wi-Fi` and `Cellular`. To achieve a target aggregate throughput G (considering a mobile application of downloading a file of size G), the average time required

(service delay) in Wi-Fi scenario and in Cellular scenario are denoted by D_W and D_C, respectively.

Definition 1 *The normalized service delay (NSD) of a Wi-Fi deployment is defined as*

$$\alpha = D_W/D_C.$$

Straightforwardly, the NSD characterizes the service quality degradation if the vehicle uses the Wi-Fi network instead of the cellular network. For a fixed deployment of cellular network (N_C), α merely depends on the deployment of Wi-Fi APs (N_W). From users' standpoint, a lower NSD is desirable, as the delay incurred to achieve a target throughput by using drive-thru Internet access would be easier to be tolerant instead of using fast but costly cellular services.

We denote by e_W the TCO (including CAPEX and OPEX) of deploying and operating one Wi-Fi AP. The total cost of a Wi-Fi deployment is thereby $\mathcal{E}_W = e_W N_W$. Similarly, we have e_C for one macrocell BS and $\mathcal{E}_C = e_C N_C$.

Definition 2 *Cost-effectiveness of a Wi-Fi deployment is defined as*

$$\eta = 1 - \mathcal{E}_W/\mathcal{E}_C \tag{5.1}$$

The cost-effectiveness is used to characterize the cost saving for deploying and operating a Wi-Fi infrastructure in a model city Ω. Intuitively, a lower NSD yields a lower cost-effectiveness. The NSD cannot be reduced while increasing the cost-effectiveness. The main objective of this research is to study the tradeoff between α and η. A summary of the mathematical notations used in this chapter is given in Table 5.1.

5.3 Time-Average Throughput Capacity of Wi-Fi

To obtain service delay D_W, we first derive the time-average throughput capacity F_W of Wi-Fi network. In an urban environment, vehicle mobility is regulated by traffic signals at intersection, which imposes a significant impact on drive-thru Wi-Fi access: vehicle stopping at the intersection prolongs the connection time with the Wi-Fi AP so as to potentially increase the data volume downloaded/uploaded. To facilitate the throughput analysis, a simple yet effective modeling of vehicle flow regulated by traffic signals is developed.

Table 5.1: Summary of Notations for Chapter 5

Symbol	Description
Ω	Area of the target city
N_W	Number of Wi-Fi APs in the `WiFi` scenario
N_C	Number of BSs in the `Cellular` scenario
F_W	Time-average throughput capacity of `WiFi`
F_C	Time-average downlink capacity of `Cellular`
D_W	Average service delay of `WiFi`
D_C	Average service delay of `Cellular`
G	Target aggregate throughput
α	Normalized service delay
e_W	Per infrastructure cost of `WiFi`
e_C	Per infrastructure cost of `cellular`
$\mathcal{E}_W$	Total infrastructure cost of `WiFi`
$\mathcal{E}_C$	Total infrastructure cost of `Cellular`
η	Cost-effectiveness of a `WiFi` deployment
L	Distance between two adjacent intersections
R	Radius of Wi-Fi AP coverage
λ	Arrival rate of the vehicle flow to the `AreaOI`
τ	Length of traffic signal cycle
τ_g	Effective green period (`EGP`)
τ_r	Effective red period (`ERP`)
v	Constant vehicle speed unless it stops
Δ	Time loss due to vehicle acceleration
ρ	Density with which vehicles flow into the `IArea`
ρ_{max}	Jam density
$N(t)$	Number of vehicles in an `AreaOI` in a TSC
τ_a	Time instant when the tagged vehicle arrives at the `IArea`
$S(\tau_a)$	Sojourn time during which the tagged vehicle stays in an `AreaOI`
$\mu(t)$	Vehicle depart rate from the `AreaOI`, $t \in [0, \tau)$
$F(\tau_a)$	Throughput capacity per drive-thru achieved by the tagged vehicle
$\Upsilon(t)$	Maximum spectrum efficiency of cellular network (bits/s/Hz)

5.3.1 Modeling Vehicle Flow with Fixed Signals

Appropriate modeling of vehicle flow regulated by traffic signals at intersections is the prerequisite to analyze the network performance of urban drive-thru Wi-Fi networks. However, it turns out to be a challenging task as it is difficult to determine how many details of vehicle mobility and road network should be incorporated into the modeling. Traffic models developed in transportation engineering, such as car following models, depend on many details/factors which increase the accuracy but could make the network analysis intractable or tedious. Hence, we develop a simple yet effective modeling with adequate details, which can capture the main characteristics of vehicle traffic at signalized intersections.

Road Network: The road network of the target city area Ω is considered as a regular grid, which is a common street pattern in many cities, such as Houston and Portland in U.S. [159]. In specific, we consider a two-way traffic on each road and one single lane for each way. The intersection of any two roads is signalized, i.e., having traffic signal control. We denote by L the distance between any two neighboring intersections. Further, to facilitate our analysis, we define the intersection area (`IArea`) for each intersection. It is a square centering at one intersection and consisting four lanes (eastbound, westbound, northbound, and southbound) each of which is of length L, as shown in Fig. 5.1. By doing so, the city area is partitioned into distinct `IArea`s.

Wi-Fi Deployment: In `Wi-Fi` scenario, Wi-Fi APs are deployed in the target city area for vehicular Internet access. We consider the following random deployment strategy[2]: Wi-Fi APs are deployed only at intersections and an intersection has an AP deployed with probability p_{ap}. Deploying APs at intersections can reduce the need for non-line-of-sight (NLOS) transmissions along the road so as to provide better service coverage [180], which is also considered in literature, such as the theoretical study [48]. Moreover, placing APs at intersections mathematically facilitates the investigation on the impact of traffic signals due to the introduced symmetry. We denote by R the radius of AP coverage and define the area of interest (`AreaOI`) for each intersection, which is a disk centering at the intersection with radius R. The `AreaOI` thereby consists four lanes (eastbound, westbound, northbound, and southbound) each of which is of length $2R$, as shown in Fig. 5.1. It is obvious that if one intersection has an AP deployed, the `AreaOI` will be the Wi-Fi coverage region[3].

Stop-and-Go Flow: The objective of this section is to analyze the throughput of

[2]Although Wi-Fi APs may be regularly deployed at intersections, the encounter of next AP for a vehicle is still random due to the randomness in movement when we observe the vehicle. Therefore, we consider a random Wi-Fi deployment.

[3]We only consider the case in which $R < L/2$ so that there is no overlapped Wi-Fi coverage regions, which simplifies the analysis and also makes economic sense.

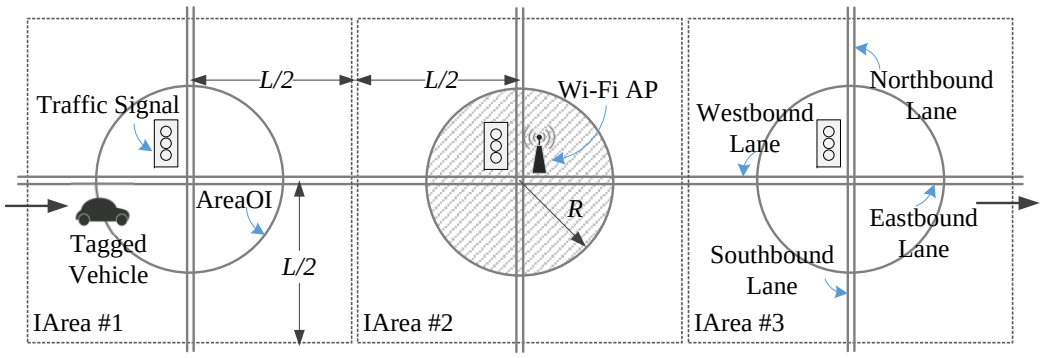


Figure 5.1: Wi-Fi hotspot at signalized intersections.

drive-thru Wi-Fi network, considering the impact of road intersections controlled by traffic signals. To this end, we theoretically derive the vehicle connection time with AP (depending on the time the vehicle arrives at the intersection) and how many vehicles share the wireless resources within the AP coverage (changing over time during the connection period). Under the control of traffic signals, vehicles arriving at the intersection during the red period stop and form a queue, whereas vehicles arriving at the intersection during the green period keep going without delay if the vehicle queue is completely dissolved. In transportation engineering, many modeling approaches for signalized intersection focus on the development of delay and queue models [182]. The main objective is to analyze the signal delay a vehicle experiences at the intersection (i.e., the extra waiting time due to signal operation and the vehicle queue), and the dynamics and the stochastic behavior of the overflow queue (i.e., the vehicle queue at the end of a green period). For example, one of the best-studied models is the fixed-cycle traffic light (FCTL) queue, where the traffic signal alternates between fixed green and red periods, and vehicles queued at the intersection are assumed to depart during the green period at equal time intervals [183]. However, the existing modeling approaches from traffic engineering cannot be directly applied to solve our problem. The reasons are two-fold: (i) the existing approaches focus on the steady-state or time-dependent analysis of delay and overflow queue length (the mean and the distribution), which are unable to characterize the dynamics of vehicle flow in the `AreaOI`; and (ii) the existing approaches are for the analysis of one-way traffic interrupted by traffic signals (the case of multiple lanes may be considered), and are too complex to apply to the scenario where the whole intersection area (including four lanes of different directions) is considered. In wireless networking research area, a stochastic traffic model is proposed for VANETs in signalized urban road systems [173], which can describe the average vehicle density and the random interactions among vehicles. The difficulty of applying

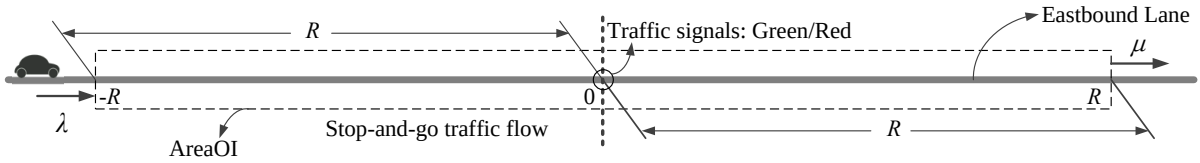


Figure 5.2: Single-lane stop-and-go vehicle flow.

this model to our scenario is also the complexity. To describe the main behaviors of vehicle flow in the `AreaOI`, we have to ignore the minor details, such as the random behavior of individual vehicles. Thus, we develop a *stop-and-go* vehicle flow model controlled by fixed traffic signals. The proposed model is described as follows.

As a common practice, we simplify the three signal periods (i.e., green, amber, and red) into two periods, effective green period (`EGP`) and effective red period (`ERP`). We define the traffic signal cycle (TSC) ($[0, \tau)$). For the eastbound and westbound lanes, an `EGP` ($[0, \tau_g)$) is followed by an `ERP` ($[\tau_g, \tau)$) during a TSC, whereas for the northbound and southbound lanes, an `ERP` ($[0, \tau_g)$) is followed by an `EGP` ($[\tau_g, \tau)$). The length of one TSC is hence denoted by $\tau = \tau_g + \tau_r$. We consider a deterministic vehicle flow, i.e, vehicles arrive at the `IArea` on each lane with arrival rate $\lambda(t) = \lambda$. Every vehicle keeps moving at the same and constant speed v unless it stops and joins a vehicle queue due to red signals. During the `EGP`[4], each stopped vehicle departs the intersection at the speed v after a short delay Δ of being head-of-line. We introduce Δ to consider the time loss due to vehicle acceleration. We also consider that vehicles do not change the lane at the intersection (i.e., no left, right, or U turns) for simplicity. Thus, vehicles flow into the `IArea` with a density ρ vehicle/m, and according to [184],

$$\rho = \lambda / v. \tag{5.2}$$

Further, we denote the jam density (maximum density) by ρ_{max} vehicle/m, which is the density of vehicles stopping and queueing at the intersection due to signal operations (typical range of ρ_{max} is $0.116 - 0.156$ vehicle/m [185]).

We analyze the performance of the urban drive-thru Wi-Fi network by observing a tagged vehicle. The tagged vehicle moves along a fixed lane (the eastbound lane in this study) and traverses `IArea`s in sequence, as shown in Fig. 5.1. To simplify our analysis, we have the following two assumptions[5].

- *Unsaturation*: We consider the case that the length of vehicle queue is less than

[4]We make statement for the eastbound and westbound lanes unless otherwise specified.
[5]Relaxation of these two assumptions would be considered for future works.

R, so that if the tagged vehicle stops at the intersection, it will always be in the `AreaOI`, i.e., the `AreaOI` is always unsaturated. In addition, we assume that there is no overflow queue, i.e., all queued vehicles can pass through the intersection during the `EGP`. This assumption is often valid under a regular traffic load condition and for a typical value of R.

- *Independence*: The `IArea`s are treated independently, i.e., traffic signals at different intersections are not coordinated and the vehicle arrival process for one `IArea` does not depend on the upstream traffic. We further assume that the tagged vehicle arrives at the `IArea` equally likely for any instant τ_a during the cycle $[0, \tau)$, i.e, τ_a is uniformly distributed over the interval $[0, \tau)$.

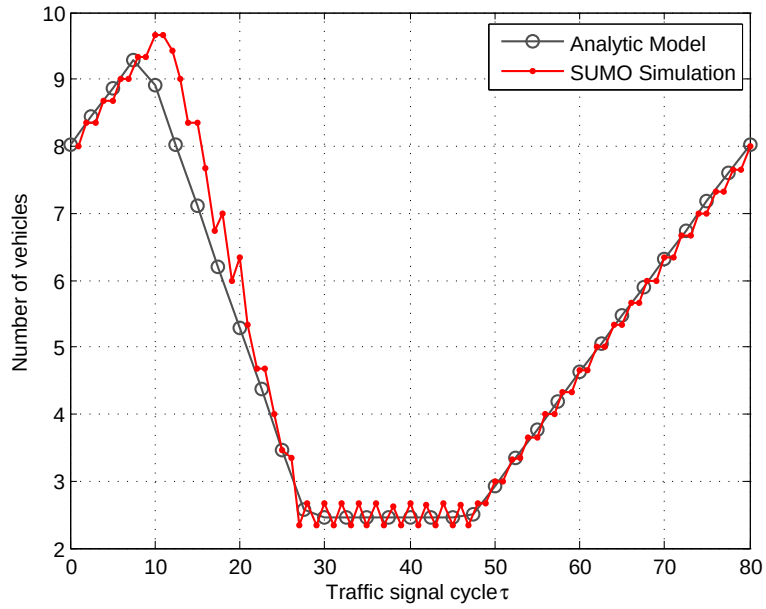
We denote by $N(t)$ the number of vehicles in an `AreaOI` in a TSC, where $t \in [0, \tau)$. Also, we denote by $S(\tau_a)$ the sojourn time the tagged vehicle stays in an `AreaOI`, where the arrival time $\tau_a \in [0, \tau)$. The sojourn time is equal to the connection time if there is an AP deployed in the intersection. Next, we derive the analytic expressions of $N(t)$ and $S(\tau_a)$.

5.3.2 Vehicle Dynamics

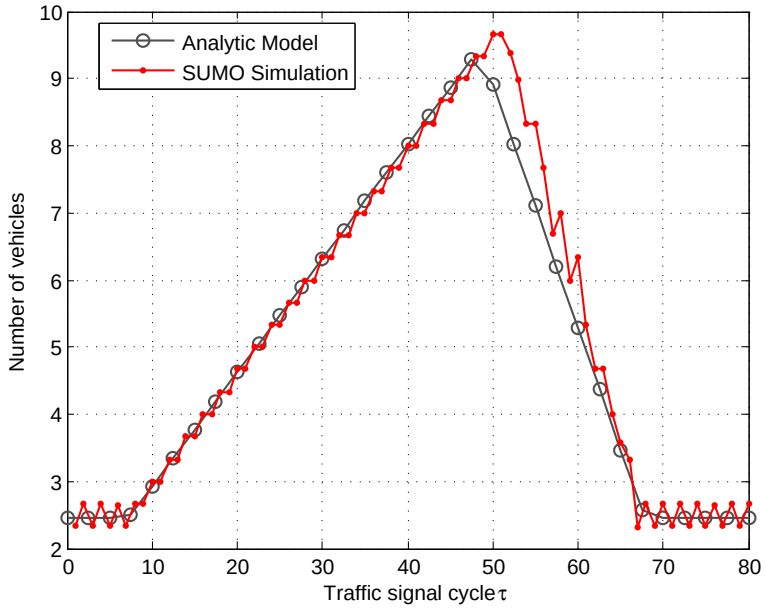
As shown in Fig. 5.2, we first focus on the vehicle flow regulated by traffic signals on the eastbound lane in a generic `IArea`, and denote the number of vehicles at time t in the `AreaOI` (only the eastbound lane is considered) by $N_e(t)$, where $t \in [0, \tau)$. We immediately have $N_e(0) = \lambda(\tau_r + R/v)$, which is the number of vehicles that arrive in the `AreaOI` but do not pass through the intersection during the previous TSC. Note that all these vehicles are located in $[-R, 0]$, i.e., $N_e(0)$ should be less than $\rho_{max}R$, according to the unsaturation assumption. Hence, vehicles arrive in the `AreaOI` at the rate $\lambda(t) = \lambda$ during $[0, \tau)$. As the arrival traffic would be regulated by the traffic signals, the vehicle departure rate from the `AreaOI` is the key to characterize $N_e(t)$. Let $\mu(t)$ denote this departure rate. $N_e(t)$ can be characterized as follows.

$$N_e(t) = N_e(0) + \int_0^t \lambda(\hat{t})d\hat{t} - \int_0^t \mu(\hat{t})d\hat{t}, \quad t \in [0, \tau). \tag{5.3}$$

The departure rate function $\mu(t)$ is given by Lemma 9.

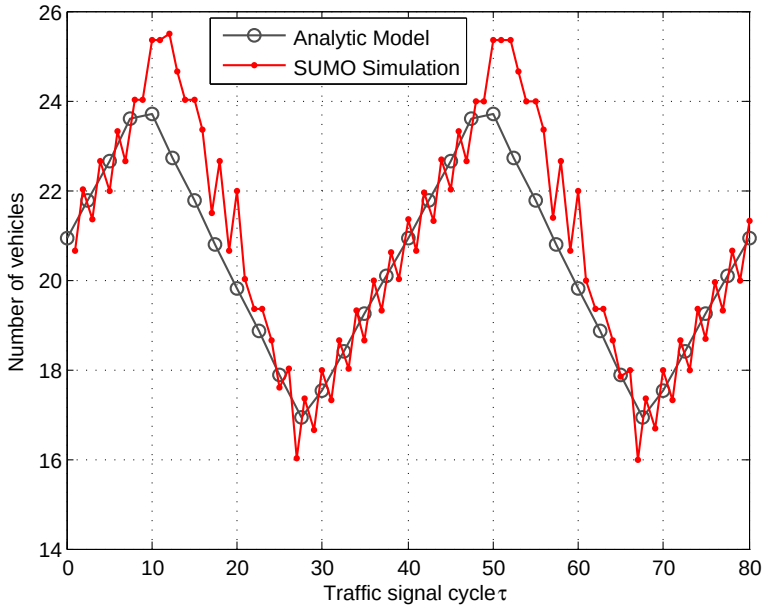


(a) $N_e(t)$



(b) $N_n(t)$

Figure 5.3: Comparison of vehicle dynamics in the `AreaOI` between our analytic results and simulations based on SUMO. $\rho_{max} = 0.12$ vehicle/m, $v = 14$ m/s, $\lambda = 0.17$ vehicle/s, $\tau_r = \tau_g = 40$ s [4], $\Delta = 1.3$ s and $R = 100$ m.



(c) $N(t)$

Figure 5.3: Comparison of vehicle dynamics in the `AreaOI` between our analytic results and simulations based on SUMO. $\rho_{max} = 0.12$ vehicle/m, $v = 14$ m/s, $\lambda = 0.17$ vehicle/s, $\tau_r = \tau_g = 40$ s [4], $\Delta = 1.3$ s and $R = 100$ m.

Lemma 9 *Vehicles (on the eastbound lane) depart from the* `AreaOI` *at the rate* $\mu(t)$, $t \in [0, \tau)$, *where*

$$\mu(t) = \begin{cases} 0, & t \in [0, \frac{R}{v} + \Delta); \\ \rho^* v, & t \in [\frac{R}{v} + \Delta, \frac{R}{v} + \Delta + t^*); \\ \lambda, & t \in [\frac{R}{v} + \Delta + t^*, \tau_g + \frac{R}{v}); \\ 0, & t \in [\tau_g + \frac{R}{v}, \tau), \end{cases}$$

where $\rho^* = \frac{\rho_{max}}{1+\rho_{max}\Delta v}$ *and* $t^* = \frac{\lambda(\tau_r+\Delta)}{\rho^* v - \lambda}$.

Proof. Note that the time interval $[0, \tau_g)$ is the `EGP` and $[\tau_g, \tau)$ is the `ERP`. At $t = \Delta$, the head-of-line vehicle queued in the previous `ERP` starts to pass through the intersection at the speed v. Since the head-of-line vehicle has to move a distance of R to depart from the `AreaOI`, during $[0, \frac{R}{v} + \Delta)$, there is no vehicle departures. At $t = \frac{R}{v} + \Delta$, the vehicle queue with the density ρ^* starts to depart from the `AreaOI`, where ρ^* can be easily determined by the following equality: $\frac{1}{\rho_{max}} + \Delta v = \frac{1}{\rho^*}$. Next, we determine the duration of this departure, denoted by t^*. Since at $t = \frac{R}{v} + \Delta + t^*$, $\rho^* v t^*$ vehicles have departed from the `AreaOI` and all the vehicles in the `AreaOI` are located in $[-R, R]$ with density ρ, we have the following equation with respect to t^* under the unsaturation assumption: $\lambda(\frac{R}{v} + \tau_r + \frac{R}{v} + \Delta + t^*) = \rho^* v t^* + \rho 2R$. From (5.2), we have $t^* = \frac{\lambda(\tau_r+\Delta)}{\rho^* v - \lambda}$. Therefore, during $[\frac{R}{v} + \Delta, \frac{R}{v} + \Delta + t^*)$, the departure rate is $\rho^* v$. Following the vehicle queue, vehicles depart at the arrival rate λ until $t = \tau_g + \frac{R}{v}$. Note that at $t = \tau_g$, the traffic signal turns from green to red. Again, there is no vehicle departure during $[\tau_g + \frac{R}{v}, \tau)$. □

Thus, from (5.3) and Lemma 9, we have

$$N_e(t) = \begin{cases} \lambda(\tau_r + R/v + t), & t \in [0, \frac{R}{v} + \Delta); \\ \lambda(\tau_r + R/v + t) - \rho^* v(t - R/v - \Delta), & \\ \qquad\qquad t \in [\frac{R}{v} + \Delta, \frac{R}{v} + \Delta + t^*); & \\ 2\lambda R/v, & t \in [\frac{R}{v} + \Delta + t^*, \tau_g + \frac{R}{v}); \\ \lambda(R/v + t - \tau_g), & t \in [\tau_g + \frac{R}{v}, \tau). \end{cases} \tag{5.4}$$

Under the same control of traffic signals, the vehicle flow on the westbound lane has the same behavior as that on the eastbound lane. Thus, $N_w(t) = N_e(t)$, where $N_w(t)$ is the number of vehicles at time t on the westbound lane in the `AreaOI`. Similarly, $N_n(t)$ and $N_s(t)$ are denoted for the northbound and southbound lanes, respectively. $N_s(t) = N_n(t)$

and $N_n(t)$ can be derived in the same way as $N_e(t)$. As $N(t)$ is a summation of $N_e(t)$, $N_w(t)$, $N_n(t)$, and $N_s(t)$, we can immediately have the following result.

Lemma 10 *Under the stop-and-go flow model and the unsaturation assumption, the number of vehicles in the* `AreaOI` *at time t, $t \in [0, \tau)$, is given by*

$$N(t) = \begin{cases} 2\lambda(\tau_r + 3R/v + t), & t \in [0, \frac{R}{v}); \\ 2\lambda(\tau_r + 2R/v + 2t), & t \in [\frac{R}{v}, \frac{R}{v} + \Delta); \\ 2\lambda(\tau_r + 2R/v + 2t) - 2\rho^* v(t - R/v - \Delta), & \\ \qquad\qquad t \in [\frac{R}{v} + \Delta, \frac{R}{v} + \Delta + t^*); & \\ 2\lambda(3R/v + t), & t \in [\frac{R}{v} + \Delta + t^*, \tau_g + \frac{R}{v}); \\ 2\lambda(2R/v + 2t - \tau_g), & t \in [\tau_g + \frac{R}{v}, \tau_g + \frac{R}{v} + \Delta); \\ 2\lambda(2R/v + 2t - \tau_g) - 2\rho^* v(t - \tau_g - R/v - \Delta), & \\ \qquad\qquad t \in [\tau_g + \frac{R}{v} + \Delta, & \\ \qquad\qquad\qquad \tau_g + \frac{R}{v} + \Delta + t^*); & \\ 2\lambda(3R/v + t - \tau_g), & t \in [\tau_g + \frac{R}{v} + \Delta + t^*, \tau). \end{cases}$$

To evaluate how our proposed stop-and-go flow model can reflect the major behavior of vehicle dynamics in the `AreaOI`, we perform simulations in an open-source traffic software SUMO[6] in which the car following model developed by Stefan Krauß [186] is used. We compare the obtained analytic results with simulation results, as shown in Fig. 5.3.

5.3.3 Sojourn Time

The sojourn time $S(\tau_a)$ is the time duration the tagged vehicle stays in an `AreaOI`. Note that the sojourn time is equal to the connection time for intersections with an AP deployed. $S(\tau_a)$ depends on the arrival time τ_a, which is uniformly distributed over the interval $[0, \tau)$. The analytic expression of $S(\tau_a)$ is given in the following lemma.

Lemma 11 *Under the stop-and-go flow model and the unsaturation assumption, the so-*

[6]SUMO is an open source, microscopic and continuous road traffic simulator designed to handle large road networks.

Table 5.2: NS-3&SUMO simulation parameters

SUMO Road Traffic				**NS-3**			
Parameter	***Value***	***Parameter***	***Value***	***Parameter***	***Value***	***Parameter***	***Value***
Acceleration	1.5 m/s^2	Deceleration	4.5 m/s^2	Wi-Fi standard	IEEE 802.11g	Rate Adaptation	AarfWifiManager
Minimum vehicle gap	3 m	Maximum velocity	14 m/s	Physical layer model	YansWifiPhy	Channel model	YansWifiChannel
Vehicle's netto-length	5 m	Repetition period	6 s	Maximum transmission level	40 mW ($\approx$ 130 m)	Data rate set	[1,2,5.5,6,9,11,12,18, 24,36,48,54] mbps
Car-following model	SUMOKrauß	Driver imperfection	0.1	Propagation Loss	Log-Distance model	Path-loss exponent	4
Green (red) period	10∼50 s	Amber period	1 s	Application model	OnOffApplication	Off-time period	0

journ time of the tagged vehicle given the arrival time τ_a is as follows.

$$S(\tau_a) = \begin{cases} \frac{\lambda(\tau_r + \frac{R}{v} + \tau_a)}{\rho^* v} + \frac{R}{v} + \Delta - \tau_a, \\ \qquad\qquad \tau_a \in [0, t^* + \Delta - \frac{R}{v}); \\ \frac{2R}{v}, \qquad\qquad \tau_a \in [t^* + \Delta - \frac{R}{v}, \tau_g - \frac{R}{v}); \\ \frac{\lambda(\tau_a - \tau_g + \frac{R}{v})}{\rho^* v} + \frac{R}{v} + \Delta + \tau - \tau_a, \\ \qquad\qquad \tau_a \in [\tau_g - \frac{R}{v}, \tau). \end{cases}$$

Proof. Note that the tagged vehicle arrives in the location $-R$ at τ_a. It can be seen that the tagged vehicle will be the head of line in the vehicle queue if $\tau_a = \tau_g - \frac{R}{v}$. When $\tau_a \in [\tau'_a, \tau_g - \frac{R}{v})$, the tagged vehicle passes through the intersection without stop ($S(\tau_a) = 2R/v$), where τ'_a satisfies the following equality: $\lambda(\tau'_a + \tau_r + \frac{R}{v}) = \rho^* v t^*$. Hence, $\tau'_a = t^* + \Delta - \frac{R}{v}$. If $\tau_a \in [0, \tau'_a)$, the tagged vehicle will stop and join the vehicle queue because the vehicle queue formed during the previous `ERP` has not been dissolved yet. The sojourn time $S(\tau_a)$ is thereby $\frac{\lambda(\tau_r + \frac{R}{v} + \tau_a)}{\rho^* v} + \frac{R}{v} + \Delta - \tau_a$. When $\tau_a \in [\tau_g - \frac{R}{v}, \tau)$, the sojourn time of the tagged vehicle can be expressed as $\underline{\tau - \tau_a} + \underline{\tau''_a}$. The first part denotes the time elapsed before the traffic signal turning from red to green, and the second part denotes the time the tagged vehicle continues the movement until it departs from the `AreaOI`. τ''_a can be obtained by $\tau''_a = \frac{\lambda(\tau_a - \tau_g + \frac{R}{v})}{\rho^* v} + \frac{R}{v} + \Delta$. □

Table 5.3: Stop-and-Go flow model parameters for analytic results

Parameter	***Value***	***Parameter***	***Value***
λ	0.17 vehicle/s	v	14 m/s
ρ_{max}	0.12 vehicle/m	Δ	1.3 s
τ_r	10∼50 s	τ_g	10∼50 s
$\varphi_{max}C$	9 mbps	R	130 m

5.3.4 Throughput Capacity Per Drive-thru

Based on Lemmas 10 and 11, we next derive the throughput capacity per drive-thru for the tagged vehicle.

Definition 3 *Throughput capacity per drive-thru: the maximum number of bits received by the tagged vehicle from the Wi-Fi AP during one typical drive-thru of Wi-Fi coverage.*

Wi-Fi transmission based on the IEEE 802.11 protocol adopts an adaptive modulation scheme with different transmission bit rates, depending on the communication distance from the AP. However, to reduce the complexity of our model computation, we consider a non-adaptive scheme with constant transmission bit rate in a fixed AP coverage range. The non-adaptive transmission rate is also analytically considered in [187]. In addition, the contention-based MAC protocol, i.e., IEEE 802.11 DCF (distributed coordination function), is adopted to schedule parallel transmissions. To characterize the protocol overhead (including the overhead of physical layer), we introduce an empirical efficiency factor $\varphi \in (0, 1)$, which can be obtained from real-world measurements or through theocratical analysis. We consider that the tagged vehicle share the Wi-Fi resource equally with other vehicles in the AP's coverage range. Hence, the throughput capacity per drive-thru achieved by the tagged vehicle is given by

$$F(\tau_a) = \int_{\tau_a}^{\tau_a + S(\tau_a)} \frac{\varphi_{max} C}{N(t \bmod \tau)} dt, \tag{5.5}$$

where φ_{max} is the maximum efficiency factor for a given transmission rate and C is the AP's transmission rate. $\varphi_{max}C$ thereby indicates the maximum bit rate that can be utilized for data transmission. For example, according to [188], $\varphi_{max} = 5/11$ for IEEE 802.11b and $C = 11$ Mbps, as the theoretical maximum throughput is shown to be 5Mbps for a 11Mbps transmission rate. It is worthy noting that the throughput capacity per drive-thru depends on the arrival time of the tagged vehicle. That is to say with different arrival time to the `AreaOI` (with an AP deployed), the tagged vehicle achieves different throughput capacity

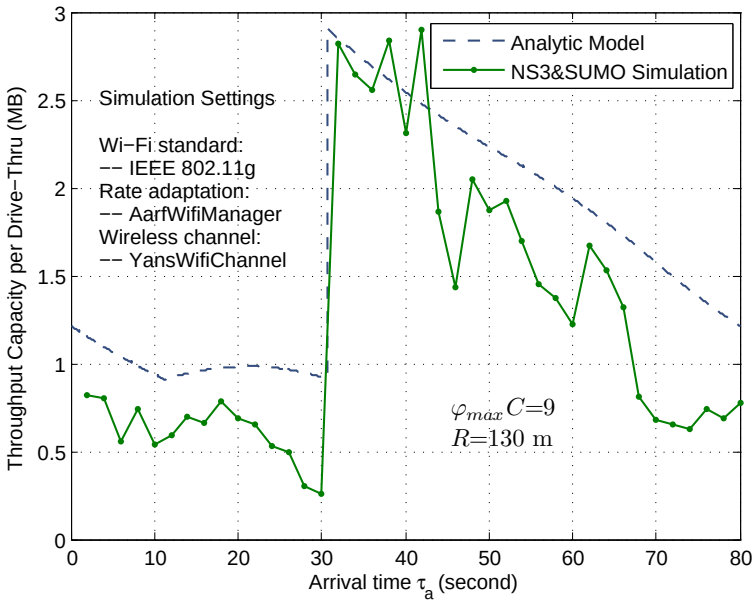


Figure 5.4: $F(\tau_a)$ vs τ_a $[\tau_g = \tau_r = 40$ s].

per drive-thru, reflecting the impact of traffic signals. Particularly, there is a significant throughput gain for vehicles stopping at intersections due to red signals, as shown in the following simulation and analytic results, which is an important finding of this study.

To validate our modeling and analysis, we conduct simulations in the network simulator NS-3 and the road traffic simulator SUMO. We first use SUMO to generate the mobility trace file of vehicle traffic in one `IArea`. And then, the trace file is used as an input for network simulations in NS-3. Simulation parameters and model parameters for analytic results are respectively given in Table II and Table III. Fig. 5.4 presents the analytic and NS-3&SUMO simulation results on throughput capacity per drive-thru with respect to the arrival time. We adopt the IEEE 802.11g standard and adaptive data rates up to 54 Mbps in the simulation. As the results shown in Fig. 5.4, vehicles arriving at the `AreaOI` with an arrival time around 31 seconds ($\tau_a \approx 31$) in a TSC, which thereby stop at the intersection due to the red signal so as to prolong the connection time with the AP, can achieve a much higher throughput capacity per drive-thru (approximately three times as high as the lowest one ($\tau_a \approx 11$) analytically), indicating a significant throughput gain. From Fig. 5.4, it can be seen that even with the none-adaptive data rate and simplified MAC operation, our theoretical and simulation measurements still match well in terms of general trends, which demonstrates the validity of our proposed modeling approach on throughput analysis of a

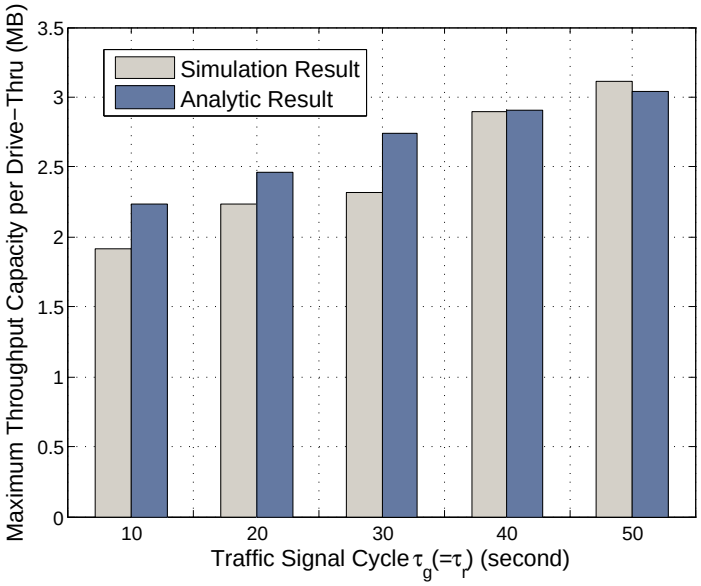


Figure 5.5: $\max F(\tau_a)$, $\tau_a \in [0, \tau)$ vs τ_g.

vehicle driving through a Wi-Fi coverage considering the impact of traffic signals.

Fig. 5.5 presents the maximum throughput capacity per drive-thru, i.e.,

$$F_{max} = \max_{\tau_a \in [0,\tau)} F(\tau_a),$$

under different configurations of TSC. It is intuitive that with a longer `EGP` or `ERP`, vehicles achieve a larger F_{max} due to the prolonged connection time with the AP. However, the gain is not very significant since the increase of `EGP` or `ERP` also incurs a larger number of vehicles waiting at the intersection due to the red signal so as to increase the number of vehicles contending for Wi-Fi resources. As shown in Fig. 5.5, the impact of increasing the connection time after all dominates the impact of increasing the number of vehicles in the AP's coverage. Fig. 5.6 presents the throughput gain, which is the maximum value over the mean value of throughput capacity per drive-thru with respect to τ_a (mathematically defined as $\max_{\tau_a \in [0,\tau)} F(\tau_a)/[\frac{1}{\tau}\int_0^\tau F(\tau_a)d\tau_a]$), under different configurations of TSC. It can be seen that with a longer `EGP` or `ERP`, we have a higher throughput gain. For example, for $\tau_g = \tau_r = 40$ s, the maximum throughput capacity per drive-thru (achieved when $\tau_a \approx 31$ s) is 1.9 times the mean value by simulation and 1.8 times by analysis, demonstrating a significant impact of traffic signals.

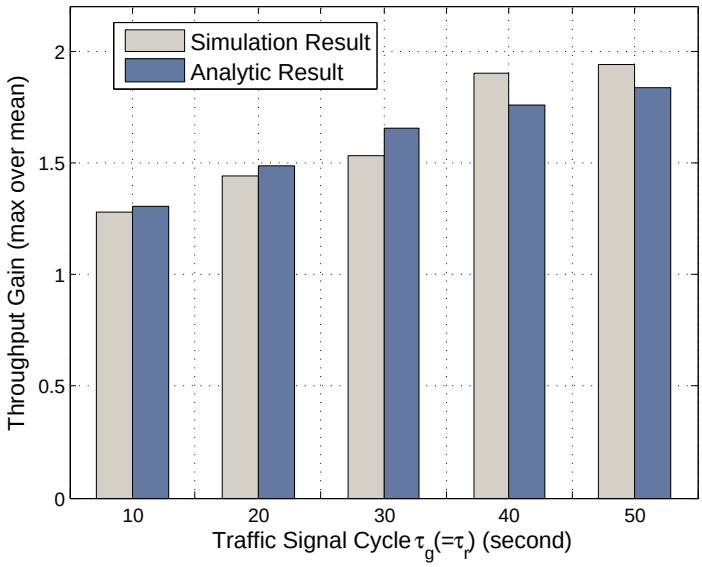


Figure 5.6: Throughput gain (max over mean) vs τ_g.

5.3.5 Time-Average Throughput Capacity

The tagged vehicle moves along the road and experiences alternate disconnected period and connected period (being inside the `AreaOI` with an AP deployed). We are interested in the average bit rate of the tagged vehicle received from the APs over a long time, e.g., the entire travel time. Mathematically, we present the definition of time-average throughput capacity.

Definition 4 *Time-average throughput capacity: the maximum average bit rate received by the tagged vehicle from the drive-thru Wi-Fi networks in a long term, which is given mathematically by*

$$F_W = \lim_{t\to\infty} \frac{\widetilde{F}(t)}{t}, \tag{5.6}$$

where $\widetilde{F}(t)$ is the total number of bits received by time t.

We next derive the time-average throughput capacity. The `IArea`s on the route of the tagged vehicle are indexed by $1, 2, \ldots n$, as shown in Fig. 5.1. Let T_n denote the time from

the departure of the $(n-1)$-th `AreaOI` to the departure of the n-th `AreaOI`. Thus, we have

$$T_n = (L - 2R)/v + S(\tau_a^n), \tag{5.7}$$

where τ_a^n is the arrival time to the n-th `AreaOI`. It can be seen that T_1, T_2, ..., T_n are independent and identically distributed (i.i.d) random variables with a common distribution T under the independence assumption. Further, we denote by $\widetilde{F}_n$ the throughput capacity achieved in the n-th `AreaOI`. $\widetilde{F}_n = F(\tau_a^n)$ with probability p_{ap}; and $\widetilde{F}_n = 0$ with probability $1 - p_{\mathsf{ap}}$. The number of `AreaOI`s passed through by the tagged vehicle by time t is denoted by $\{I(t), t \geq 0\}$. Then, we have

$$\sum_{n=1}^{I(t)} \widetilde{F}_n \leq \widetilde{F}(t) < \sum_{n=1}^{I(t)+1} \widetilde{F}_n. \tag{5.8}$$

Since $\widetilde{F}_n$ can be considered as the reward earned during the time period of T_n, we model $\{\widetilde{F}(t); t > 0\}$ as a renewal reward process with inter-renewal time $\{T_n; n \geq 1\}$. The inter-renewal times have a finite expectation $E[T] < \infty$. $\widetilde{F}_1$, $\widetilde{F}_2$, ..., $\widetilde{F}_n$ are i.i.d random variables with a common distribution $\widetilde{F}$. The following lemma holds for renewal reward processes.

Lemma 12 *(Theorem 5.4.1 in [189]) Consider a renewal reward process $\{R(t); t > 0\}$ with expected inter-renewal time $E[X] = \overline{X} < \infty$. If each R_n is a random variable with $E[R_n] < \infty$, then with probability 1,*

$$\lim_{t\to\infty} \frac{R(t)}{t} = \frac{E[R_n]}{\overline{X}}. \tag{5.9}$$

Proposition 1 *Under the independence assumption, with probability 1, the time-average throughput capacity of the tagged vehicle is given by*

$$F_W = \frac{p_{\mathsf{ap}} \int_0^{\tau} F(\tau_a) d\tau_a}{\tau(L - 2R)/v + \int_0^{\tau} S(\tau_a) d\tau_a}. \tag{5.10}$$

Proof. According to Lemma 12, $F_W = E[\widetilde{F}]/E[T]$, as obviously we have $E[\widetilde{F}] < \infty$ and $E[T] < \infty$. Specifically, from Lemma 11 and (5.7), $E[T] = \frac{L-2R}{v} + E[S(\tau_a)] =$

$\frac{L-2R}{v} + \frac{1}{\tau}\int_0^\tau S(\tau_a)d\tau_a$. From (5.11), $E[\widetilde{F}] = p_{\mathsf{ap}}\frac{1}{\tau}\int_0^\tau F(\tau_a)d\tau_a$. Note that τ_a is uniformly distributed over $[0, \tau)$. We omit the tedious calculations of these two integrals. □

It can be seen that the time-average throughput capacity of Wi-Fi network is determined by the urban environment (L), Wi-Fi deployment (p_{ap}), Wi-Fi coverage R, and connection time depending on traffic signal operation (τ_g, τ_r) and vehicle traffic (λ, v). The analytic and simulation results of F_W in terms of deployment scale are shown in Fig. 5.7. It can be seen that the time-average throughput capacity increases with a larger Wi-Fi deployment. Note that the analytic result is quite optimistic. This is because we use an empirical efficiency factor to simplify the MAC and physical layer operation of Wi-Fi. Our analytic result can be considered as an upper bound of Wi-Fi throughput performance.

5.4 Benchmark: Cellular Macrocell Service

The performance of cellular macrocell service is considered as a benchmark in the study of cost-effectiveness of Wi-Fi network. Thus, we use the networking scenario with cellular macrocell BSs providing full service coverage for performance comparison. A macrocell enables cellular services relying on a high-power cellular BS [190]. For apple-to-apple comparison, we assume that the considered cellular network only serves vehicle users. We focus on the analysis of average downlink capacity achieved by the tagged vehicle.

Definition 5 *Average downlink capacity: the maximum average downlink data rate received from cellular BSs in a long term, which is given mathematically by*

$$F_C = \lim_{t\to\infty}\frac{1}{t}\int_0^t U(t)dt, \tag{5.11}$$

where $U(t)$ is the instantaneous maximum data rate at time t.

5.4.1 Spectrum Efficiency

The maximum data rate $U(t)$ can be determined by $U(t) = b(t)\Upsilon(t)$, where $b(t)$ is acquired transmission bandwidth and $\Upsilon(t)$ is the maximum spectrum efficiency (bits/s/Hz), at time t. The maximum spectral efficiency is theoretically governed by the Shannon capacity. However, the Shannon capacity is not achievable in reality due to limited coding block

length, non-avoidable system overhead, etc. [191]. Following [192], we adopt a modified Shannon capacity formula,

$$\Upsilon = \mathtt{BW}_e \cdot \sigma \cdot \log_2(1 + \frac{\mathtt{SINR}}{\mathtt{SINR}_e}), \tag{5.12}$$

where $\mathtt{BW}_e$ is the system bandwidth efficiency, $\mathtt{SINR}_e$ is the efficiency of signal-to-interference-plus-noise ratio (SINR), and σ is a correction factor due to the dependency between $\mathtt{SINR}_e$ and $\mathtt{SINR}$. Formula (5.12) can be used to approximate the maximum spectrum efficiency of real-world cellular systems with different settings. For example, for an LTE cellular system with single antenna transmissions and Round Robin scheduling, $\mathtt{BW}_e\sigma = 0.56$ and $\mathtt{SINR}_e = 2.0$ [192].

5.4.2 Distribution of SINR

The SINR on the wireless link between cellular BS and the tagged vehicle is an important basis to determine the downlink capacity. The interference experienced by the vehicle comes from the transmission of other-cell BSs. In urban areas with densely deployed BSs, other-cell interference is a major impediment to high spectrum efficiency [193]. To model the deployment of macrocell BSs in the considered area Ω, we consider a homogeneous Poisson point process (PPP) of density ξ, in which each point represents a location of BS. Modeling BS location as a PPP is widely adopted in the literature, e.g., [194] and [195], which is able to characterize the variety of macrocell size due to differences in transmission power, tower height, etc.. We consider the same vehicle density ρ as in the drive-thru Wi-Fi scenario. The impact of traffic signals on vehicle density is not considered here since cellular macrocell is much larger than Wi-Fi coverage and thereby not sensitive to such variations of vehicle density. Considering Rayleigh fading on other-cell interference, the complementary cumulative distribution function (CCDF) of SINR is given in [194], i.e.,

$$\mathbf{Pr}(\mathtt{SINR} > Z) = \frac{1}{1 + Z^{\frac{2}{\beta}} \int_{Z^{-\frac{2}{\beta}}}^{\infty} \frac{1}{1+u^{\beta/2}} du}, \tag{5.13}$$

where $\beta > 2$ is called the path-loss exponent. Typically, we have $\beta = 4$ for urban environments [174]. Thus, (5.13) can be further simplified as follows.

$$\mathbf{Pr}(\mathtt{SINR} > Z) = \frac{1}{1 + \sqrt{Z}(\pi/2 - \arctan(1/\sqrt{Z}))} \tag{5.14}$$

The distribution of downlink SINR given in (5.14) is for the configuration of single transmit and single receive antenna. In addition, the thermal noise is ignored, as the urban cellular network is often interference-limited.

5.4.3 Average Downlink Capacity

To derive the average downlink capacity of the tagged vehicle, we consider a simple bandwidth sharing model: every vehicle can obtain a constant bandwidth b_0 from BSs, i.e., $b(t) = b_0$. Given that each BS provides a bandwidth of B, the total bandwidth resource in Ω is thus $\xi\Omega B$. As the total number of vehicles in Ω is given by $\Omega \cdot 4L\rho/L^2$ (approximately Ω contains Ω/L^2 `IArea`s and each `IArea` contains $4L\rho$ vehicles), we have $b_0 = \frac{\xi BL}{4\rho}$. Considering the spectrum efficiency and distribution of SINR, the following result of average downlink capacity is obtained.

Proposition 2 *The average downlink capacity of the tagged vehicle is given by*

$$F_C = \mathcal{U}_1 \int_{r>0} \frac{1}{1+\sqrt{\mathcal{U}_2}(\pi/2 - \arctan(1/\sqrt{\mathcal{U}_2}))} dr, \tag{5.15}$$

where $\mathcal{U}_1 = \frac{\xi BL\mathtt{BW}_e\sigma}{4\rho}$ *and* $\mathcal{U}_2 = \mathtt{SINR}_e(2^r - 1)$.

Proof.

$$\begin{aligned} F_C &= \lim_{t\to\infty} \frac{1}{t} \int_0^t b(t)\Upsilon(t)dt = b_0 E[\Upsilon] \\ &= \frac{\xi BL}{4\rho} E\left[\mathtt{BW}_e\sigma \log_2\left(1 + \frac{\mathtt{SINR}}{\mathtt{SINR}_e}\right)\right] \\ &= \frac{\xi BL\mathtt{BW}_e\sigma}{4\rho} \int_{r>0} \mathbf{Pr}\left(\log_2\left(1 + \frac{\mathtt{SINR}}{\mathtt{SINR}_e}\right) > r\right) dr \\ &= \frac{\xi BL\mathtt{BW}_e\sigma}{4\rho} \int_{r>0} \mathbf{Pr}\left(\mathtt{SINR} > \mathtt{SINR}_e(2^r - 1)\right) dr. \end{aligned}$$

The third equality holds due to $E[X] = \int_{x>0} \mathbf{Pr}(X > x)dx$ for a nonnegative random variable X. From (5.14), the proposition holds. □

In the analysis of average downlink capacity, the inter-cell interference management techniques are not considered, such as frequency reuse. However, these advanced techniques are indeed beneficial for improving the spectrum efficiency and available data rates.

Especially, the performance of LTE system is more limited by other-cell interference compared to 3G cellular systems [196], implying that interference management is also necessary for LTE systems. Hence, our result given in Proposition 2 is conservative and can be considered as a lower bound of average downlink capacity. In providing the benchmark cellular service for comparison, small cell service on top of the existing macrocell service is not considered. We give a discussion on heterogenous networks in Section 5.6. Under the same comparison level with Fig. 5.7, the analytic and simulation results of $\digamma_C$ in terms of deployment scale are shown in Fig. 5.8. It can be seen that the average downlink capacity increases with a larger celluar deployment in terms of ξ. We can also notice the conservativeness of our analytic result that we have discussed.

5.5 Cost-Effectiveness Analysis

We examine the cost-effectiveness of Wi-Fi deployment in this section. For service delay of `Wi-Fi` scenario, mathematically, $D_W = E\big[\{\min t_0 : \widetilde{F}(t_0) > G\}\big]$. Since it is difficult to obtain the distribution of $\widetilde{F}(t)$ and given that $\mathbf{Pr}(\lim_{t\to\infty} \frac{\widetilde{F}(t)}{t} = \digamma_W) = 1$, D_W is approximated to be $\frac{G}{\digamma_W}$ for a large G, e.g., one hundred MBytes. Similarly, $D_C \approx \frac{G}{\digamma_C}$. Hence, the `NSD` $\alpha \approx \digamma_C/\digamma_W$. The cost-effectiveness η depends on the TCO (including CAPEX and OPEX) of infrastructure node and the number of infrastructure node deployed in each scenario. The CAPEX includes the cost of equipment, planning, installation, commissioning, etc., and the OPEX includes the cost of site rental, power, maintenance, etc. [197]. Based on the cost model provided in [197], the ratio of the TCO of a Wi-Fi AP to the TCO of a macro 3-Sector LTE BS is around 12%, i.e., $e_W/e_C \approx 0.12$. To evaluate the cost-effectiveness of a `Wi-Fi` scenario, we fix the cellular deployment (Ω, N_C) and thereby the $\digamma_C$ is determined. Given that the average number of APs deployed in a `Wi-Fi` scenario is $p_{\mathsf{ap}}\Omega/L^2$, the explicit relation between η and α is given by

$$\eta = 1 - \frac{e_W}{e_C}\frac{\Omega\digamma_C\big(\tau(L-2R)/v + \int_0^\tau S(\tau_a)d\tau_a\big)}{\alpha N_C L^2 \int_0^\tau \digamma(\tau_a)d\tau_a}. \tag{5.16}$$

This is the main result of this chapter, which presents the tradeoff between η (TCO savings) and α (service degradation). It can be seen that η depends on both controllable (e.g., Wi-Fi deployment (p_{ap}, R, α)) and uncontrollable (e.g., the urban environment (L, τ), vehicle traffic statistics (λ, v)) parameters. The analytic and simulation results on η are shown in Fig. 5.9. The gap between theory and simulation is due to the conservative result on $\digamma_C$ given in Proposition 2 (see Fig. 5.8). Through these results, the MNO is

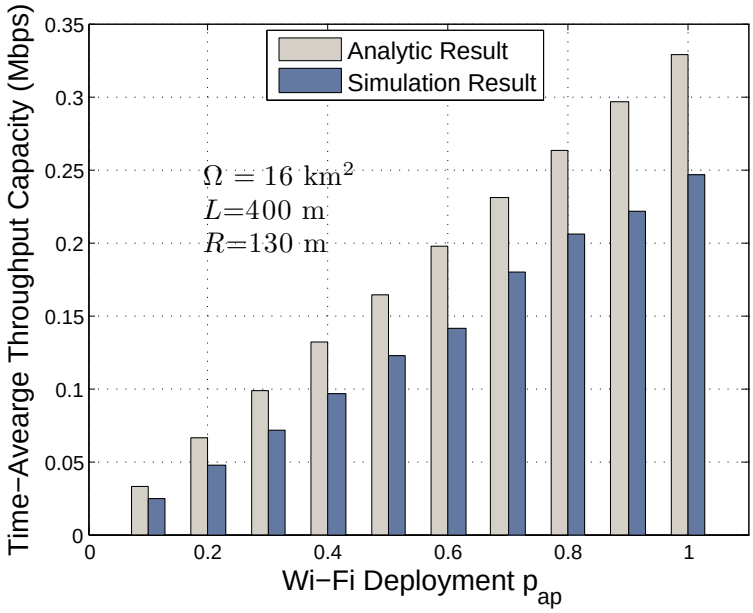


Figure 5.7: F_W vs p_{ap}.

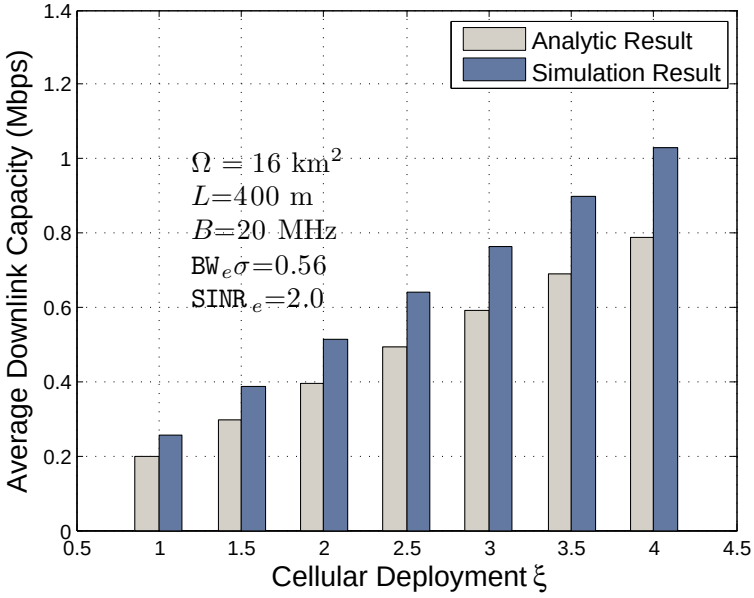


Figure 5.8: F_C vs ξ.

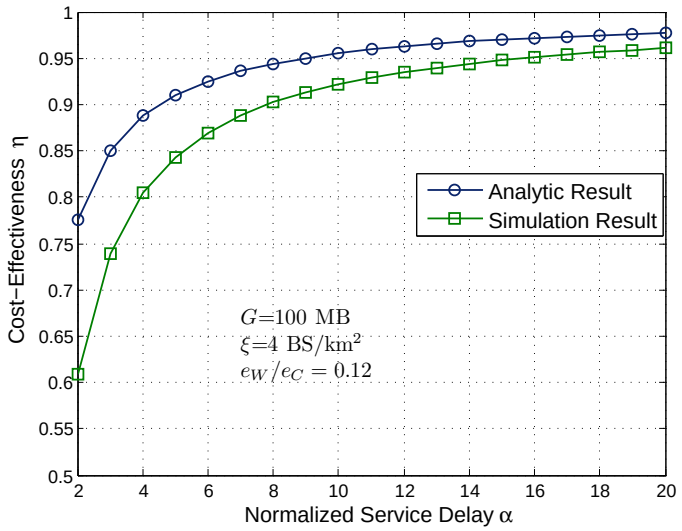


Figure 5.9: Tradeoff between cost-effectiveness and normalized service delay.

able to deploy Wi-Fi network according to the required level of service, and to have the knowledge of the corresponding TCO savings (compared to cellular solution) immediately. For example, theoretically from Fig. 5.9, almost 90% of the TCO can be saved if the average service delay of Wi-Fi deployment is 3X larger than that of cellular deployment, demonstrating the great potential of Wi-Fi solution for vehicle users in terms of cost-effectiveness. Fig. 5.10 analytically presents the tradeoff between η and α under different configurations of TSC. We can clearly see the impact of traffic signals on the tradeoff between the cost-effectiveness of Wi-Fi deployment and the service degradation. However, the impact is not significant especially for a large `NSD`. While for a single drive-thru, there exists a significant throughput gain due to the impact of traffic signals.

5.6 Discussion

Small cells and heterogenous network: Both small cells and Wi-Fi are cost-effective solutions to massive increase of mobile data demand. However, the feasibility of outdoor small cells for mobile user at vehicular speed is not clear yet. In our study, we do not consider deploying cellular BSs and Wi-Fi APs in one scenario for the purpose of apple-to-apple comparison and explicitly showing the great potential of Wi-Fi. We believe that a heterogenous network with cellular macrocell service for coverage and small cells and Wi-Fi for capacity would be desired.

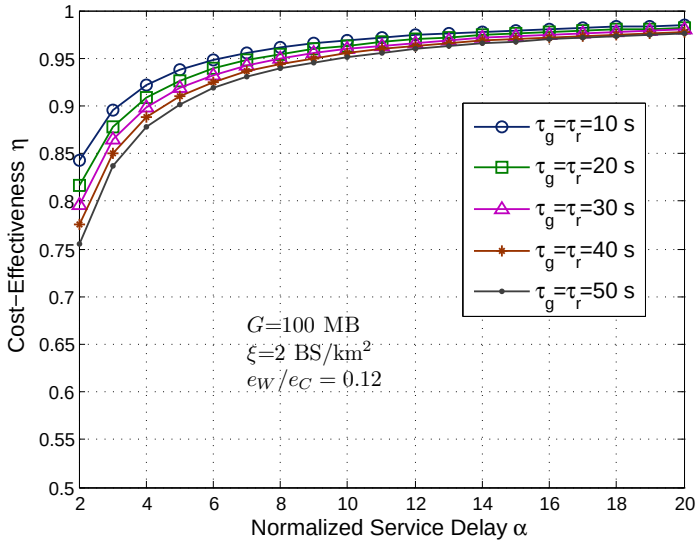


Figure 5.10: Tradeoff between η and α under different configurations of TSC.

V2V communications: In this chapter, although we do not consider store-carry-and-forward communications between vehicles, which may incur additional cost/complexity for opportunistic data exchange, V2V communications can indeed benefit Internet traffic delivery. For example, the vehicles in the coverage of Wi-Fi hotspots can help to relay the traffic so as to virtually extend the Wi-Fi service region.

5.7 Summary

In this chapter, we have investigated the cost-effectiveness of Wi-Fi solution for vehicular Internet access considering the tradeoff with the user's satisfaction. In `Wi-Fi` scenario, we have particularly studied the fundamental impact of traffic signals at intersection on Wi-Fi access. By examining the tradeoff between cost-effectiveness and normalized service delay, we have demonstrated that Wi-Fi has great potential to serve vehicle users with a much lower TCO. Our results provide a quick and efficient way of determining the Wi-Fi deployment strategy and the corresponding TCO savings. Future work includes large-scale simulations with real-world data set of vehicle mobility in urban scenarios, and cost-effectiveness analysis taking V2V communications into consideration.

Chapter 6

Conclusions and Future Work

In this chapter, we summarize the main concepts and results presented in this thesis and highlight future research directions.

6.1 Conclusions

In this thesis, we have investigated the scaling laws of vehicular networks. Based on the analysis and discussion provided, we present the following remarks.

- The study of capacity scaling laws plays an vital role in understanding the fundamental properties of VANETs. Instead of showing the exact performance for a specific network, theoretical capacity bounds present the performance limit of networks with optimal operations, and thereby guide the network design and deployment. For large-scale VANETs, capacity scaling results can also be applied to predict the network performance, at least in the order sense.

- Vehicle mobility imposes many challenges in vehicular networking, but also brings opportunities, one of which is to facilitate the store-carry-and-forward data delivery to support delay-tolerant applications, such as traffic conditions update, advertisements and social media dissemination, without any assistance of infrastructure. It has been shown that the wireless network is not scalable without mobility [58]; while the network will be truely scalable with i.i.d (extreme) mobility [88]. So in theory, we would like to know if the vehicular network is scalable for serving delay-tolerant

data delivery considering vehicle mobility, which is definitely time-correlated, map-restricted, and related to the people's social life. We argue that the capacity scaling highly depends on the social features of vehicle mobility. From the capacity and delay analysis, we also know that according to different mobility patterns, it is beneficial to design suitable packet forwarding schemes.

- Wireless access technologies cannot be circumvented when it comes to VANETs, since providing Internet services is the major solution to meet the ever-increasing mobile data demand of vehicle users. The capacity-cost study provides some basic ideas for the choice of access infrastructures. The main conclusion is that the deployment of BSs or WMBs is cost-effective to offer a low-speed downlink rate to vehicles; nonetheless, when providing a high-speed Internet access, the deployment of RAPs outperforms the other two alternatives in terms of deployment costs. Therefore, we believe that non-cellular access infrastructure will play an increasingly important role in offering a cost-effective data pipe for vehicles, especially for supporting high-bandwidth applications. Moreover, we show a significant throughput gain for vehicles stopping at intersections due to red signals by considering the impact of traffic signals in the Wi-Fi deployment.

6.2 Future Research Directions

In this thesis, we focus on the scaling law study of VANETs with and without access infrastructure, respectively, mainly based on the theoretical analysis. Our future work includes extensive simulation validations base on trace data of real-world scenarios and further digging up the implication on network design and operation. Despite existing studies on capacity analysis of VANETs, many issues remain unclear. For example, when jointly considering more complex street patten and inhomogeneous vehicle densities, it might be difficult to determine the throughput capacity and network delay. Moreover, due to the emergent and public nature of safety applications, broadcasting leads an important role in disseminating safety messages to vehicles in proximity. The study of broadcast capacity is another research interest. We close this chapter with additional three thoughts on future research directions in the field of scaling laws.

- The design, analysis and deployment of wireless networks necessitate a general understanding of capacity scaling laws. Existing works often adopt different methodologies and sets of assumptions and models in developing capacity scaling laws, which may yield custom-designed solutions without universal properties that can be applied to

other types of wireless networks. To better understand the impact of various settings and techniques on capacity scaling laws, it would be useful to provide a unified framework. Two research works have been performed toward this end: the study of capacity scaling laws under a generalized physical model [198] and the establishment of a simple set of criteria that can be used to determine the capacity for various physical layer technologies under the protocol model [199].

- The Shannon capacity was achieved by considering arbitrarily delay and vanishingly small error probability. In [55], Andrews *et al.* referred to a throughput-delay-reliability (TDR) triplet, since these quantities are interrelated. Thus, the throughput capacity of wireless networks would likely be constrained by these two fundamental quantities—delay and reliability jointly. Actually, the link reliability has been considered in studies of transmission capacity [200–202] which is the spatial intensity of attempted transmissions under a target outage of wireless links. The tradeoff between throughput capacity, delay, and reliability should be investigated. However, this is much more challenging.

- Investigations on throughput capacity and network delay of emerging wireless networks are also promising. Particular characteristics of networks being studied often make the problem very challenging, such as road geometry and vehicle density in vehicular networks. In addition to the aforementioned cognitive radio networks and vehicular networks, femtocell networks [203] and smart grid have also gained much interest recently, both of which have complex network architecture and heterogenous communication devices, making the study of scaling laws a demanding task.

References

[1] A. Zemlianov and G. De Veciana, "Capacity of ad hoc wireless networks with infrastructure support," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 3, pp. 657–667, 2005.

[2] B. Liu, P. Thiran, and D. Towsley, "Capacity of a wireless ad hoc network with infrastructure," in *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2007, pp. 239–246.

[3] D. Shila, Y. Cheng, and T. Anjali, "Throughput and delay analysis of hybrid wireless networks with multi-hop uplinks," in *Proceedings of IEEE INFOCOM*, Shanghai, China, April 2011.

[4] J. A. Bonneson, S. R. Sunkari, and M. P. Pratt, "Traffic Signal Operations Handbook," Texas Transportation Institute, Texas A & M University System, Tech. Rep., 2009.

[5] Transparency Market Research, *Connected Car Market -Global Industry Analysis, Size, Share, Growth, Trends and Forecast, 2013-2019*, 2013.

[6] D. Schrank, B. Eisele, and T. Lomax, "TTI's 2012 Urban Mobility Report," *Texas A&M Transportation Institute and the Texas A&M University System*, 2012.

[7] J. Anda, J. LeBrun, D. Ghosal, C.-N. Chuah, and M. Zhang, "VGrid: vehicular adhoc networking and computing grid for intelligent traffic control," in *Proc. IEEE VTC Spring*, Stockholm, Sweden, June 2005.

[8] C. Olaverri-Monreal, P. Gomes, R. Fernandes, F. Vieira, and M. Ferreira, "The see-through system: A VANET-enabled assistant for overtaking maneuvers," in *Proc. IEEE Intelligent Vehicles Symposium (IV)*, San Diego, CA, USA, June 2010.

[9] A. Balasubramanian, R. Mahajan, A. Venkataramani, B. N. Levine, and J. Zahorjan, "Interactive wifi connectivity for moving vehicles," in *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 4. ACM, 2008, pp. 427–438.

[10] J. Lin, S. Chen, Y. Shih, and S. Chen, "A study on remote on-line diagnostic system for vehicles by integrating the technology of OBD, GPS, and 3G," *World Academy of Science, Engineering and Technology*, vol. 56, p. 56, 2009.

[11] M. Ramadan, M. Al-Khedher, and S. Al-Kheder, "Intelligent anti-theft and tracking system for automobiles," *International Journal of Machine Learning and Computing*, vol. 2, no. 1, February 2012.

[12] Telefnica, *Connected Car Industry Report*, 2013.

[13] "eCall." [Online]. Available: http://ec.europa.eu/digital-agenda/en/ecall-time-saved-lives-saved

[14] [Online]. Available: http://www.nhtsa.gov/About+NHTSA/Press+Releases/2014/USDOT+to+Move+Forward+with+Vehicle-to-Vehicle+Communication+Technology+for+Light+Vehicles

[15] G. Karagiannis, O. Altintas, E. Ekici, G. Heijenk, B. Jarupan, K. Lin, and T. Weil, "Vehicular networking: A survey and tutorial on requirements, architectures, challenges, standards and solutions," *IEEE Communications Surveys & Tutorials,*, no. 99, pp. 1–33, 2011.

[16] J. Kenney, "Dedicated short-range communications (DSRC) standards in the united states," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1162–1182, 2011.

[17] H. Hartenstein, K. Laberteaux, and I. Ebrary, *VANET: vehicular applications and inter-networking technologies*. Wiley Online Library, 2010.

[18] H. Hartenstein and K. Laberteaux, "A tutorial survey on vehicular ad hoc networks," *IEEE Communications Magazine*, vol. 46, no. 6, pp. 164–171, 2008.

[19] F. Bai and B. Krishnamachari, "Exploiting the wisdom of the crowd: localized, distributed information-centric VANETs," *IEEE Communications Magazine*, vol. 48, no. 5, pp. 138–146, 2010.

[20] T. H. Luan, X. Shen, and F. Bai, "Integrity-oriented content transmission in highway vehicular ad hoc networks," in *Proc. IEEE INFOCOM*, Turin, Italy, April 2013.

[21] M. Boban, T. T. Vinhoza, M. Ferreira, J. Barros, and O. K. Tonguz, "Impact of vehicles as obstacles in vehicular ad hoc networks," *IEEE J. Selected Areas in Communications*, vol. 29, no. 1, pp. 15–28, 2011.

[22] L. Cheng, B. Henty, D. Stancil, F. Bai, and P. Mudalige, "Mobile vehicle-to-vehicle narrow-band channel measurement and characterization of the 5.9 ghz dedicated short range communication (dsrc) frequency band," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 8, pp. 1501–1516, 2007.

[23] C. Mecklenbräuker, A. Molisch, J. Karedal, F. Tufvesson, A. Paier, L. Bernadó, T. Zemen, O. Klemp, and N. Czink, "Vehicular channel characterization and its implications for wireless system design and performance," *Proceedings of the IEEE*, no. 99, pp. 1–24, 2011.

[24] J. Yin, T. ElBatt, G. Yeung, B. Ryu, S. Habermas, H. Krishnan, and T. Talty, "Performance evaluation of safety applications over dsrc vehicular ad hoc networks," in *Proc. ACM VANET*, Philadelphia, PA, USA, October 2004.

[25] V. Taliwal, D. Jiang, H. Mangold, C. Chen, and R. Sengupta, "Empirical determination of channel characteristics for DSRC vehicle-to-vehicle communication," in *Proc. ACM VANET*, Philadelphia, PA, USA, October 2004.

[26] F. Bai, D. D. Stancil, and H. Krishnan, "Toward understanding characteristics of dedicated short range communications (DSRC) from a perspective of vehicular network engineers," in *Proc. ACM MobiCom*, Chicago, IL, USA, September 2010.

[27] X. Wu, S. Subramanian, R. Guha, R. White, J. Li, K. Lu, A. Bucceri, and T. Zhang, "Vehicular Communications Using DSRC: Challenges, Enhancements, and Evolution," *IEEE J. Selected Areas of Communications*, vol. 31, no. 9, pp. 399–408, 2013.

[28] T. Jiang, H.-H. Chen, H.-C. Wu, and Y. Yi, "Channel modeling and inter-carrier interference analysis for V2V communication systems in frequency-dispersive channels," *Mobile Networks and Applications*, vol. 15, no. 1, pp. 4–12, 2010.

[29] F. Yu and S. Biswas, "Self-configuring TDMA protocols for enhancing vehicle safety with DSRC based vehicle-to-vehicle communications," *IEEE J. Selected Areas of Communications*, vol. 25, no. 8, pp. 1526–1537, 2007.

[30] N. Lu, Y. Ji, F. Liu, and X. Wang, "A dedicated multi-channel MAC protocol design for VANET with adaptive broadcasting," in *Proceedings of IEEE WCNC*, Sydney, Australia, April 2010.

[31] H. Omar, W. Zhuang, and L. Li, "VeMAC: A TDMA-based MAC protocol for reliable broadcast in VANETs," *IEEE Trans. on Mobile Computing*, vol. 12, no. 9, pp. 1724–1736, September 2012.

[32] S. Bharati and W. Zhuang, "CAH-MAC: Cooperative ADHOC MAC for Vehicular Networks," *IEEE J. Selected Areas of Communications*, vol. 31, no. 9, pp. 470–479, September 2013.

[33] V. Bychkovsky, B. Hull, A. Miu, H. Balakrishnan, and S. Madden, "A measurement study of vehicular internet access using in situ wi-fi networks," in *Proceedings of ACM MobiCom*, 2006.

[34] A. Jameel, A. Fuchs, and M. Stuempfle, "Internet multimedia on wheels: connecting cars to cyberspace," in *Proc. IEEE ITSC*, Boston, MA, USA, November 1997.

[35] A. Jameel, M. Stuempfle, D. Jiang, and A. Fuchs, "Web on wheels: Toward internet-enabled cars," *IEEE Computer*, vol. 31, no. 1, pp. 69–76, 1998.

[36] R. Lind, R. Schumacher, R. Reger, R. Olney, H. Yen, M. Laur, and R. Freeman, "The Network Vehicle-a glimpse into the future of mobile multi-media," *IEEE Aerospace and Electronic Systems Magazine*, vol. 14, no. 9, pp. 27–32, 1999.

[37] "Car Connectivity Consortium, MirrorLink." [Online]. Available: http://www.mirrorlink.com/

[38] *BMW ConnectedDrive*. [Online]. Available: http://www.bmw.com/com/en/insights/technology/connecteddrive/2013/index.html

[39] J. Ott and D. Kutscher, "Drive-thru Internet: IEEE 802.11b for Automobile," in *Proc. IEEE INFOCOM*, Hong Kong, China, March 2004.

[40] ——, "The 'Drive-thru' Architecture: WLAN-based Intemet Access on the Road," in *Proc. IEEE VTC Spring*, May 2004.

[41] R. Gass, J. Scott, and C. Diot, "Measurements of in-motion 802.11 networking," in *Proc. IEEE Workshop on Mobile Computing Systems and Applications*, Semiahmoo Resort, WA, USA, April 2005.

[42] D. Hadaller, S. Keshav, T. Brecht, and S. Agarwal, "Vehicular opportunistic communication under the microscope," in *Proc. ACM MobiSys*, San Juan, Puerto Rico, June 2007.

[43] J. Eriksson, H. Balakrishnan, and S. Madden, "Cabernet: vehicular content delivery using WiFi," in *Proceedings of ACM MobiCom*, 2008.

[44] T. H. Luan, X. Ling, and X. Shen, "Mac in motion: impact of mobility on the mac of drive-thru internet," *IEEE Trans. on Mobile Computing,*, vol. 11, no. 2, pp. 305–319, 2012.

[45] J. Zhao, T. Arnold, Y. Zhang, and G. Cao, "Extending drive-thru data access by vehicle-to-vehicle relay," in *Proc. ACM VANET*, San Francisco, CA, USA, September 2008.

[46] O. Trullols-Cruces, M. Fiore, and J. Barcelo-Ordinas, "Cooperative download in vehicular environments," *IEEE Trans. on Mobile Computing*, vol. 11, no. 4, pp. 663–678, 2012.

[47] F. Malandrino, C. Casetti, C. Chiasserini, and M. Fiore, "Optimal content downloading in vehicular networks," *IEEE Trans. on Mobile Computing*, vol. 12, no. 7, pp. 1377–1391, 2012.

[48] Z. Zheng, P. Sinha, and S. Kumar, "Sparse wifi deployment for vehicular internet access with bounded interconnection gap," *IEEE/ACM Trans. on Networking*, vol. 20, no. 3, pp. 956–969, 2012.

[49] T. H. Luan, L. X. Cai, J. Chen, X. Shen, and F. Bai, "Vtube: Towards the media rich city life with autonomous vehicular content distribution," in *Proc. IEEE SECON*, Salt Lake City, UT, USA, June 2011.

[50] M. Wada, T. Yendo, T. Fujii, and M. Tanimoto, "Road-to-vehicle communication using led traffic light," in *Proc. IEEE Intelligent Vehicles Symposium*, June 2005.

[51] S. Iwasaki, C. Premachandra, T. Endo, T. Fujii, M. Tanimoto, and Y. Kimura, "Visible light road-to-vehicle communication using high-speed camera," in *Proc. IEEE Intelligent Vehicles Symposium*, Eindhoven, the Netherlands, June 2008.

[52] C. B. Liu, B. Sadeghi, and E. W. Knightly, "Enabling vehicular visible light communication (V2LC) networks," in *Proc. ACM VANET*, Las Vegas, USA, September 2011.

[53] S. Rajagopal, R. D. Roberts, and S.-K. Lim, "IEEE 802.15. 7 visible light communication: Modulation schemes and dimming support," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 72–82, 2012.

[54] A. El Gamal and Y. Kim, *Network information theory*. Cambridge University Press, 2011.

[55] J. Andrews, S. Shakkottai, R. Heath, N. Jindal, M. Haenggi, R. Berry, D. Guo, M. Neely, S. Weber, S. Jafar *et al.*, "Rethinking information theory for mobile ad hoc networks," *IEEE Communications Magazine*, vol. 46, no. 12, pp. 94–101, 2008.

[56] C. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.

[57] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, 1979.

[58] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, 2000.

[59] A. Goldsmith, M. Effros, R. Koetter, M. Médard, A. Ozdaglar, and L. Zheng, "Beyond shannon: the quest for fundamental performance limits of wireless ad hoc networks," *IEEE Communications Magazine*, vol. 49, no. 5, pp. 195–205, 2011.

[60] P. Li, M. Pan, and Y. Fang, "The capacity of three-dimensional wireless ad hoc networks," in *Proceedings of IEEE INFOCOM*, Shanghai, China, April 2011.

[61] H. Pishro-Nik, A. Ganz, and D. Ni, "The capacity of vehicular ad hoc networks," in *Proceedings of Allerton Conference*, 2007.

[62] M. Nekoui, A. Eslami, and H. Pishro-Nik, "Scaling laws for distance limited communications in vehicular ad hoc networks," in *Proceedings of IEEE ICC*, 2008, pp. 2253–2257.

[63] G. Zhang, Y. Xu, X. Wang, X. Tian, J. Liu, X. Gan, H. Yu, and L. Qian, "Multicast capacity for hybrid VANETs with directional antenna and delay constraint," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 818–833, 2012.

[64] M. Wang, H. Shan, L. Cai, N. Lu, X. Shen, and F. Bai, "Throughput capacity of VANETs by exploiting mobility diversity," in *Proceedings of IEEE ICC*, Ottawa, Canada, June 2012.

[65] N. Lu, T. Luan, M. Wang, X. Shen, and F. Bai, "Bounds of asymptotic performance limits of social-proximity vehicular networks," *IEEE/ACM Transactions on Networking*, to appear.

[66] N. Lu, N. Zhang, N. Cheng, X. Shen, J. W. Mark, and F. Bai, "Vehicles meet infrastructure: Towards capacity-cost tradeoffs for vehicular access networks," *IEEE Transactions on Intelligent Transportation Systems*, to appear.

[67] T. Cover, J. Thomas, J. Wiley *et al.*, *Elements of information theory*. Wiley Online Library, 1991, vol. 306.

[68] M. Franceschetti, O. Dousse, D. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 1009–1018, 2007.

[69] S. Yi, Y. Pei, and S. Kalyanaraman, "On the capacity improvement of ad hoc wireless networks using directional antennas," in *Proceedings of ACM MobiHoc*, 2003.

[70] P. Li, C. Zhang, and Y. Fang, "The capacity of wireless ad hoc networks using directional antennas," *IEEE Transactions on Mobile Computing*, vol. 10, no. 10, pp. 1374–1387, 2011.

[71] C. Peraki and S. Servetto, "On the maximum stable throughput problem in random networks with directional antennas," in *Proceedings of ACM MobiHoc*, 2003.

[72] H. Sadjadpour, Z. Wang, and J. Garcia-Luna-Aceves, "The capacity of wireless ad hoc networks with multi-packet reception," *IEEE Transactions on Communications*, vol. 58, no. 2, pp. 600–610, 2010.

[73] J. Garcia-Luna-Aceves, H. Sadjadpour, and Z. Wang, "Challenges: towards truly scalable ad hoc networks," *Proceedings of ACM MobiCom*, 2007.

[74] Z. Wang, H. Sadjadpour, and J. Garcia-Luna-Aceves, "The capacity and energy efficiency of wireless ad hoc networks with multi-packet reception," in *Proceedings of MobiHoc*. ACM, 2008, pp. 179–188.

[75] ——, "Fundamental limits of information dissemination in wireless ad hoc networks–part II: Multi-packet reception," *IEEE Transactions on Wireless Communications*, vol. 10, no. 3, pp. 803–813, 2011.

[76] S. Aeron and V. Saligrama, "Wireless ad hoc networks: Strategies and scaling laws for the fixed SNR regime," *IEEE Transactions on Information Theory*, vol. 53, no. 6, pp. 2044–2059, 2007.

[77] A. Ozgur, O. Lévêque, and D. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3549–3572, 2007.

[78] J. Ghaderi, L. Xie, and X. Shen, "Hierarchical cooperation in ad hoc networks: Optimal clustering and achievable throughput," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3425–3436, 2009.

[79] U. Niesen, P. Gupta, and D. Shah, "On capacity scaling in arbitrary wireless networks," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 3959–3982, 2009.

[80] M. Franceschetti, M. Migliore, and P. Minero, "The capacity of wireless networks: information-theoretic and physical limits," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3413–3424, 2009.

[81] S. Lee and S. Chung, "Capacity scaling of wireless ad hoc networks: Shannon meets maxwell," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1702–1715, 2012.

[82] K. Lu, S. Fu, Y. Qian, and H. Chen, "On capacity of random wireless networks with physical-layer network coding," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 5, pp. 763–772, 2009.

[83] J. Liu, D. Goeckel, and D. Towsley, "The throughput order of ad hoc networks employing network coding and broadcasting," in *IEEE Proceedings of MILCOM*, Washington, USA, October 2006.

[84] ——, "Bounds on the gain of network coding and broadcasting in wireless networks," in *IEEE Proceedings of INFOCOM*, Anchorage, USA, May 2007.

[85] A. Keshavarz-Haddadt and R. Riedi, "Bounds on the benefit of network coding: Throughput and energy saving in wireless networks," in *IEEE Proceedings of INFOCOM*, Phoenix, USA, March 2008.

[86] J. Liu, D. Goeckel, and D. Towsley, "Bounds on the throughput gain of network coding in unicast and multicast wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 5, pp. 582–592, 2009.

[87] H. Zhang and J. Hou, "Capacity of wireless ad-hoc networks under ultra wide band with power constraint," in *IEEE Proceedings of INFOCOM*, Miami, USA, March 2005.

[88] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 477–486, 2002.

[89] R. Negi and A. Rajeswaran, "Capacity of power constrained ad-hoc networks," in *IEEE Proceedings of INFOCOM*, Hong Kong, China, March 2004.

[90] X. Tang and Y. Hua, "Capacity of ultra-wideband power-constrained ad hoc networks," *IEEE Transactions on Information Theory*, vol. 54, no. 2, pp. 916–920, 2008.

[91] S. Diggavi, M. Grossglauser, and D. Tse, "Even one-dimensional mobility increases the capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 3947–3954, 2005.

[92] S. Jafar, "Too much mobility limits the capacity of wireless ad hoc networks," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 3954–3965, 2005.

[93] P. Kyasanur and N. Vaidya, "Capacity of multi-channel wireless networks: impact of number of channels and interfaces," in *Proceedings of ACM MobiCom*, 2005.

[94] ——, "Capacity of multichannel wireless networks under the protocol model," *IEEE/ACM Transactions on Networking*, vol. 17, no. 2, pp. 515–527, 2009.

[95] M. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proceedings of ACM MobiCom*, 2005.

[96] S. Toumpis and A. Goldsmith, "Large wireless networks under fading, mobility, and delay constraints," in *Proceeding of IEEE INFOCOM*, Hongkong, China, March 2004.

[97] M. Ebrahimi, M. Maddah-Ali, and A. Khandani, "Throughput scaling laws for wireless networks with fading channels," *IEEE Transactions on Information Theory*, vol. 53, no. 11, pp. 4250–4254, 2007.

[98] R. Gowaikar, B. Hochwald, and B. Hassibi, "Communication over a wireless network with random connections," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 2857–2871, 2006.

[99] S. Cui, A. Haimovich, O. Somekh, H. Poor, and S. Shamai, "Throughput scaling of wireless networks with random connections," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3793–3806, 2010.

[100] R. Gowaikar and B. Hassibi, "Achievable throughput in two-scale wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 7, pp. 1169–1179, 2009.

[101] R. Jaber and J. Andrews, "A lower bound on the capacity of wireless erasure networks," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6502–6513, 2011.

[102] C. Hu, X. Wang, Z. Yang, J. Zhang, Y. Xu, and X. Gao, "A geometry study on the capacity of wireless networks via percolation," *IEEE Transactions on Communications*, vol. 58, no. 10, pp. 2916–2925, 2010.

[103] P. Gupta, P. Kumar *et al.*, "Internets in the sky: The capacity of three dimensional wireless networks," *Communications in Information and Systems*, vol. 1, no. 1, pp. 33–49, 2001.

[104] P. Li, M. Pan, and Y. Fang, "Capacity bounds of three-dimensional wireless ad hoc networks," *IEEE/ACM Transactions on Networking*, vol. 20, no. 4, pp. 1304–1315, 2012.

[105] Z. Wang, H. Sadjadpour, and J. Garcia-Luna-Aceves, "A unifying perspective on the capacity of wireless ad hoc networks," in *Proceedings of IEEE INFOCOM*, Phoenix, USA, April 2008.

[106] A. Keshavarz-Haddad, V. Ribeiro, and R. Riedi, "Broadcast capacity in multihop wireless networks," in *Proceedings of ACM MobiCom*, 2006.

[107] R. Zheng, "Asymptotic bounds of information dissemination in power-constrained wireless networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 1, pp. 251–259, 2008.

[108] X. Li, J. Zhao, Y. Wu, S. Tang, X. Xu, and X. Mao, "Broadcast capacity for wireless ad hoc networks," in *Proceedings of IEEE MASS*, Atlanta, USA, September 2008.

[109] S. Li, Y. Liu, and X. Li, "Capacity of large scale wireless networks under gaussian channel model," in *Proceedings of ACM MobiCom*, 2008.

[110] X. Li, "Multicast capacity of wireless ad hoc networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 3, pp. 950–961, 2009.

[111] C. Wang, X. Li, C. Jiang, S. Tang, Y. Liu, and J. Zhao, "Scaling laws on multicast capacity of large scale wireless networks," in *Proceedings of IEEE INFOCOM*, Rio de Janeiro, Brazil, April 2009.

[112] S. Shakkottai, X. Liu, and R. Srikant, "The multicast capacity of large multihop wireless networks," *IEEE/ACM Transactions on Networking*, vol. 18, no. 6, pp. 1691–1700, 2010.

[113] U. Niesen, P. Gupta, and D. Shah, "The balanced unicast and multicast capacity regions of large wireless networks," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2249–2271, 2010.

[114] X. Wang, W. Huang, S. Wang, J. Zhang, and C. Hu, "Delay and capacity tradeoff analysis for motioncast," *IEEE/ACM Transactions on Networking*, vol. 19, no. 5, pp. 1354–1367, 2011.

[115] X. Wang, L. Fu, and C. Hu, "Multicast performance with hierarchical cooperation," *IEEE/ACM Transactions on Networking*, vol. 20, no. 3, pp. 917–930, 2012.

[116] D. Nie, "A survey on multicast capacity of wireless ad hoc networks," 2009. [Online]. Available: http://iwct.sjtu.edu.cn/personal/xwang8/research/nieding/survey.pdf

[117] G. Sharma, R. Mazumdar, and N. Shroff, "Delay and capacity trade-offs in mobile ad hoc networks: A global perspective," *IEEE/ACM Transactions on Networking*, vol. 15, no. 5, pp. 981–992, 2007.

[118] D. Ciullo, V. Martina, M. Garetto, and E. Leonardi, "Impact of correlated mobility on delay-throughput performance in mobile ad-hoc networks," in *Proceedings of IEEE INFOCOM*, San Diego, USA, March 2010.

[119] S. Ross, *Introduction to probability models*. Academic press, 2009.

[120] K. Lee, Y. Kim, S. Chong, I. Rhee, and Y. Yi, "Delay-capacity tradeoffs for mobile networks with Lévy walks and Lévy flights," in *Proceedings of IEEE INFOCOM*, San Diego, USA, March 2010.

[121] I. Rhee, M. Shin, S. Hong, K. Lee, S. Kim, and S. Chong, "On the Levy-walk nature of human mobility," *IEEE/ACM Transactions on Networking*, vol. 19, no. 3, pp. 630–643, 2011.

[122] M. Neely and E. Modiano, "Capacity and delay tradeoffs for ad hoc mobile networks," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1917–1937, 2005.

[123] L. Ying, S. Yang, and R. Srikant, "Optimal delay–throughput tradeoffs in mobile ad hoc networks," *IEEE Transactions on Information Theory*, vol. 54, no. 9, pp. 4119–4143, 2008.

[124] X. Lin and N. Shroff, "The fundamental capacity-delay tradeoff in large mobile ad hoc networks," in *Proceedings of 3rd Annual Mediterranean Ad Hoc Networking Workshop*, Bodrum, Turkey, June 2004.

[125] A. El Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Optimal throughput-delay scaling in wireless networks-part I: The fluid model," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2568–2592, 2006.

[126] G. Sharma and R. Mazumdar, "Scaling laws for capacity and delay in wireless ad hoc networks with random mobility," in *Proceedings of IEEE ICC*, Paris, France, June 2004.

[127] X. Lin, G. Sharma, R. Mazumdar, and N. Shroff, "Degenerate delay-capacity tradeoffs in ad-hoc networks with brownian mobility," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2777–2784, 2006.

[128] P. Li, Y. Fang, J. Li, and X. Huang, "Smooth trade-offs between throughput and delay in mobile ad hoc networks," *Mobile Computing, IEEE Transactions on*, vol. 11, no. 3, pp. 427–438, 2012.

[129] P. Li, Y. Fang, and J. Li, "Throughput, delay, and mobility in wireless ad hoc networks," in *Proceedings of IEEE INFOCOM*, San Diego, USA, March 2010.

[130] M. Garetto, P. Giaccone, and E. Leonardi, "On the capacity of ad hoc wireless networks under general node mobility," in *Proceedings of IEEE INFOCOM*, Anchorage, USA, May 2007.

[131] ——, "Capacity scaling of sparse mobile ad hoc networks," in *Proceedings of IEEE INFOCOM*, Phoenix, USA, April 2008.

[132] ——, "Capacity scaling in ad hoc networks with heterogeneous mobile nodes: the super-critical regime," *IEEE/ACM Transactions on Networking*, vol. 17, no. 5, pp. 1522–1535, 2009.

[133] ——, "Capacity scaling in ad hoc networks with heterogeneous mobile nodes: the subcritical regime," *IEEE/ACM Transactions on Networking*, vol. 17, no. 6, pp. 1888–1901, 2009.

[134] M. Garetto and E. Leonardi, "Restricted mobility improves delay-throughput tradeoffs in mobile ad hoc networks," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 5016–5029, 2010.

[135] A. Ozgur and O. Lévêque, "Throughput-delay tradeoff for hierarchical cooperation in ad hoc wireless networks," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1369–1377, 2010.

[136] C. Comaniciu and H. Poor, "On the capacity of mobile ad hoc networks with delay constraints," *IEEE Transactions on Wireless Communications*, vol. 5, no. 8, pp. 2061–2071, 2006.

[137] B. Liu, Z. Liu, and D. Towsley, "On the capacity of hybrid wireless networks," in *Proceedings of IEEE INFOCOM*, San Francisco, USA, March 2003.

[138] U. Kozat and L. Tassiulas, "Throughput capacity of random ad hoc networks with infrastructure support," in *Proceedings of ACM MobiCom*, 2003, pp. 55–65.

[139] A. Agarwal and P. Kumar, "Capacity bounds for ad hoc and hybrid wireless networks," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 3, pp. 71–81, 2004.

[140] S. Toumpis, "Capacity bounds for three classes of wireless networks: asymmetric, cluster, and hybrid," in *Proceedings of ACM MobiHoc*, 2004, pp. 133–144.

[141] P. Li and Y. Fang, "Impacts of topology and traffic pattern on capacity of hybrid wireless networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 12, pp. 1585–1595, 2009.

[142] P. Li, C. Zhang, and Y. Fang, "Capacity and delay of hybrid wireless broadband access networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 2, pp. 117–125, 2009.

[143] G. Zhang, Y. Xu, X. Wang, and M. Guizani, "Capacity of hybrid wireless networks with directional antenna and delay constraint," *IEEE Transactions on Communications*, vol. 58, no. 7, pp. 2097–2106, 2010.

[144] W. Shin, S. Jeon, N. Devroye, M. Vu, S. Chung, Y. Lee, and V. Tarokh, "Improved capacity scaling in wireless networks with infrastructure," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5088–5102, 2011.

[145] X. Wang and Q. Liang, "On the throughput capacity and performance analysis of hybrid wireless networks over fading channels," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2930–2940, 2013.

[146] W. Huang, X. Wang, and Q. Zhang, "Capacity scaling in mobile wireless ad hoc network with infrastructure support," in *Proceedings of IEEE ICDCS*, Genoa, Italy, June 2010.

[147] P. Li and Y. Fang, "The capacity of heterogeneous wireless networks," in *Proceedings of IEEE INFOCOM*, San Diego, USA, March 2010.

[148] P. Zhou, X. Wang, and R. Rao, "Asymptotic capacity of infrastructure wireless mesh networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 8, pp. 1011–1024, 2008.

[149] O. Dousse, M. Franceschetti, and P. Thiran, "On the throughput scaling of wireless relay networks," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2756–2761, 2006.

[150] L. Law, K. Pelechrinis, S. Krishnamurthy, and M. Faloutsos, "Downlink capacity of hybrid cellular ad hoc networks," *IEEE/ACM Transactions on Networking*, vol. 18, no. 1, pp. 243–256, 2010.

[151] P. Li, X. Huang, and Y. Fang, "Capacity scaling of multihop cellular networks," in *Proceedings of IEEE INFOCOM*, Shanghai, China, April 2011.

[152] N. Zhang, N. Lu, N. Cheng, J. W. Mark, and X. S. Shen, "Cooperative spectrum access towards secure information transfer for crns," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 11, pp. 2453–2464, 2013.

[153] M. Vu and V. Tarokh, "Scaling laws of single-hop cognitive networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 8, pp. 4089–4097, 2009.

[154] S. Jeon, N. Devroye, M. Vu, S. Chung, and V. Tarokh, "Cognitive networks achieve throughput scaling of a homogeneous network," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5103–5115, 2011.

[155] C. Yin, L. Gao, and S. Cui, "Scaling laws for overlaid wireless networks: a cognitive radio network versus a primary network," *IEEE/ACM Transactions on Networking*, vol. 18, no. 4, pp. 1317–1329, 2010.

[156] W. Huang and X. Wang, "Capacity scaling of general cognitive networks," *to appear in IEEE/ACM Transactions on Networking*.

[157] Y. Li, X. Wang, X. Tian, and X. Liu, "Scaling laws for cognitive radio network with heterogeneous mobile secondary users," in *Proceedings of IEEE INFOCOM*, Orlando, USA, March 2012.

[158] N. Sarafijanovic-Djukic, M. Pidrkowski, and M. Grossglauser, "Island hopping: Efficient mobility-assisted forwarding in partitioned networks," in *Proceedings of IEEE SECON*, Reston, USA, September 2006.

[159] S. Kostof and R. Tobias, *The city shaped*. Thames and Hudson London, 1991.

[160] A. Siksna, "The effects of block size and form in north american and australian city centres," *Urban Morphology*, vol. 1, pp. 19–33, 1997.

[161] M. Neely, E. Modiano, and C. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 89–103, 2005.

[162] R. Urgaonkar and M. Neely, "Network capacity region and minimum energy function for a delay-tolerant mobile ad hoc network," *IEEE/ACM Transactions on Networking*, vol. 19, no. 4, pp. 1137–1150, 2011.

[163] D. Slaughter, *Difference Equations to Differential Equations*. University Press of Florida, 2009.

[164] R. Motwani and P. Raghavan, *Randomized algorithms*. Chapman & Hall/CRC, 2010.

[165] V. Vapnik and A. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory of Probability and its Applications*, vol. 16, p. 264, 1971.

[166] V. Vapnik, *Statistical learning theory*. Wiley-Interscience, 1998.

[167] "OnStar LTE." [Online]. Available: https://www.onstar.com/us/en/4glte/

[168] QUALCOMM, *Rising to Meet the 1000x Mobile Data Challenge*. Whitepaper, QUALCOMM Incorporated, San Diego, CA, USA, 2013.

[169] A. Mahajan, N. Potnis, K. Gopalan, and A. Wang, "Modeling VANET deployment in urban settings," in *Proceedings of ACM MSWiM*, 2007, pp. 151–158.

[170] N. Banerjee, M. Corner, D. Towsley, and B. Levine, "Relays, base stations, and meshes: enhancing mobile networks with infrastructure," in *Proceedings of ACM MobiCom*, 2008.

[171] F. Malandrino, C. Casetti, C.-F. Chiasserini, and M. Fiore, "Content downloading in vehicular networks: What really matters," in *Proceedings of IEEE INFOCOM*, Shanghai, China, April 2011.

[172] C. Stefanović, D. Vukobratović, F. Chiti, L. Niccolai, V. Crnojević, and R. Fantacci, "Urban infrastructure-to-vehicle traffic data dissemination using uep rateless codes," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 1, pp. 94–102, 2011.

[173] I. W.-H. Ho, K. K. Leung, and J. W. Polak, "Stochastic model and connectivity dynamics for vanets in signalized road systems," *IEEE/ACM Transactions on Networking (TON)*, vol. 19, no. 1, pp. 195–208, 2011.

[174] H. Xia, "A simplified analytical model for predicting path loss in urban and suburban environments," *IEEE Transactions on Vehicular Technology*, vol. 46, no. 4, pp. 1040–1046, 1997.

[175] J. Lee, R. Mazumdar, and N. Shroff, "Joint resource allocation and base-station assignment for the downlink in CDMA networks," *IEEE/ACM Transactions on Networking*, vol. 14, no. 1, pp. 1–14, 2006.

[176] F. Baccelli and B. Blaszczyszyn, *Stochastic Geometry and Wireless Networks Volume I: THEORY*. NOW: Foundations and Trends in Networking, 2010.

[177] ——, *Stochastic Geometry and Wireless Networks Volume II: APPLICATIONS*. NOW: Foundations and Trends in Networking, 2010.

[178] M. Haenggi and R. Ganti, *Interference in large wireless networks*. Now Publishers Inc, 2009.

[179] K. Johansson, "Cost effective deployment strategies for heterogeneous wireless networks," Ph.D. dissertation, Kommunikationsteknik, Kungliga Tekniska högskolan, 2007.

[180] T. Mangel, M. Michl, O. Klemp, and H. Hartenstein, "Real-world measurements of non-line-of-sight reception quality for 5.9 GHz IEEE 802.11 p at intersections," in *Communication Technologies for Vehicles*. Springer, 2011, pp. 189–202.

[181] "Wickedly Fast Wi-Fi." [Online]. Available: http://www.wickedlyfastwifi.com/

[182] F. Viti, *The dynamics and the uncertainty of delays at signals*. Netherlands TRAIL Research School, 2006.

[183] J. S. van Leeuwaarden, "Delay analysis for the fixed-cycle traffic-light queue," *Transportation Science*, vol. 40, no. 2, pp. 189–199, 2006.

[184] J. D. Fricker and R. K. Whitford, *Fundamentals of transportation engineering*. Pearson Prentice Hall, 2004.

[185] M. Van Aerde and H. Rakha, "Multivariate calibration of single regime speed-flow-density relationships," in *Proc. the 6th Vehicle Navigation and Information Systems Conference*, 1995, pp. 334–341.

[186] S. Krauß, "Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics," Ph.D. dissertation, Universitat zu Koln., 1998.

[187] W. L. Tan, W. C. Lau, O. Yue, and T. H. Hui, "Analytical models and performance evaluation of drive-thru internet systems," *IEEE JSAC*, vol. 29, no. 1, pp. 207–222, 2011.

[188] J. Jun, P. Peddabachagari, and M. Sichitiu, "Theoretical maximum throughput of ieee 802.11 and its applications," in *IEEE International Symposium on Network Computing and Applications*, 2003.

[189] R. G. Gallager, *Stochastic Processes: Theory for Applications*. Cambridge University Press, 2013.

[190] J. H. Schiller, *Mobile communications*. Pearson Education, 2003.

[191] E. Dahlman, S. Parkvall, J. Skold, and P. Beming, *3G evolution: HSPA and LTE for mobile broadband*. Academic press, 2010.

[192] P. Mogensen, W. Na, I. Z. Kovács, F. Frederiksen, A. Pokhariyal, K. I. Pedersen, T. Kolding, K. Hugl, and M. Kuusela, "LTE capacity compared to the shannon bound," in *Proc. IEEE VTC (Spring)*, 2007.

[193] S. Mukherjee, "Distribution of downlink sinr in heterogeneous cellular networks," *IEEE JSAC*, vol. 30, no. 3, pp. 575–585, 2012.

[194] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. on Communications*, vol. 59, no. 11, pp. 3122–3134, 2011.

[195] X. Lin, R. Ganti, P. Fleming, and J. Andrews, "Towards understanding the fundamentals of mobility in cellular networks," *IEEE Trans. on Wireless Communications*, vol. 12, no. 4, pp. 1686–1698, 2013.

[196] D. Astély, E. Dahlman, A. Furuskar, Y. Jading, M. Lindstrom, and S. Parkvall, "LTE: the evolution of mobile broadband," *IEEE Communications Magazine*, vol. 47, no. 4, pp. 44–51, 2009.

[197] M. Paolini and S. Fili, *The economics of small cells and Wi-Fi offload*. Report, SenzaFili Consulting, Sammamish, WA, USA, 2012.

[198] C. Wang, C. Jiang, X. Li, S. Tang, and P. Yang, "General capacity scaling of wireless networks," in *Proceedings of IEEE INFOCOM*, Shanghai, China, April 2011.

[199] C. Jiang, Y. Shi, Y. Hou, W. Lou, S. Kompella, and S. Midkiff, "Toward simple criteria to establish capacity scaling laws for wireless networks," in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 774–782.

[200] S. Weber, X. Yang, J. Andrews, and G. De Veciana, "Transmission capacity of wireless ad hoc networks with outage constraints," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4091–4102, 2005.

[201] A. Hunter, J. Andrews, and S. Weber, "Transmission capacity of ad hoc networks with spatial diversity," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5058–5071, 2008.

[202] S. Weber, J. Andrews, and N. Jindal, "An overview of the transmission capacity of wireless networks," *IEEE Transactions on Communications*, vol. 58, no. 12, pp. 3593–3604, 2010.

[203] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Communications Magazine*, vol. 46, no. 9, pp. 59–67, 2008.

[204] A. Abdrabou and W. Zhuang, "Probabilistic delay control and road side unit placement for vehicular ad hoc networks with disrupted connectivity," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 1, pp. 129–139, 2011.

[205] G. Alfano, M. Garetto, and E. Leonardi, "Capacity scaling of wireless networks with inhomogeneous node density: Upper bounds," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 7, pp. 1147–1157, 2009.

[206] G. Alfano, M. Garetto, E. Leonardi, and V. Martina, "Capacity scaling of wireless networks with inhomogeneous node density: lower bounds," *IEEE/ACM Transactions on Networking*, vol. 18, no. 5, pp. 1624–1636, 2010.

[207] ——, "Capacity scaling of wireless networks with inhomogeneous node density: Lower bounds," *IEEE/ACM Transactions on Networking*, vol. 18, no. 5, pp. 1624–1636, 2010.

[208] N. Bansal and Z. Liu, "Capacity, delay and mobility in wireless ad-hoc networks," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 2. IEEE, 2003, pp. 1553–1563.

[209] D. Benyamina, A. Hafid, and M. Gendreau, "Wireless Mesh Networks Design-A Survey," *IEEE Communications Surveys & Tutorials*, no. 99, pp. 1–12, 2011.

[210] S. Céspedes, N. Lu, and X. Shen, "VIP-WAVE: On the feasibility of ip communications in 802.11 p vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, to appear.

[211] J. Camp, J. Robinson, C. Steger, and E. Knightly, "Measurement driven deployment of a two-tier urban mesh access network," in *Proceedings of ACM MobiSys*, 2006, pp. 96–109.

[212] B. Chen and M. Chan, "Mobtorrent: A framework for mobile internet access from vehicles," in *Proceedings of IEEE INFOCOM*, Rio de Janeiro, Brazil, April 2009.

[213] R. Chen, W. Jin, and A. Regan, "Broadcasting safety information in vehicular networks: issues and approaches," *IEEE Network*, vol. 24, no. 1, pp. 20–25, 2010.

[214] E. Duarte-Melo, A. Josan, M. Liu, D. Neuhoff, and S. Pradhan, "The effect of node density and propagation model on throughput scaling of wireless networks," in *Information Theory, 2006 IEEE International Symposium on*. IEEE, pp. 1693–1697.

[215] M. Garetto, A. Nordio, C. Chiasserini, and E. Leonardi, "Information-theoretic capacity of clustered random networks," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7578–7596, 2011.

[216] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. Thomas, J. Andrews, P. Xia, H. Jo *et al.*, "Heterogeneous cellular networks: From theory to practice," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 54–64, 2012.

[217] C. Hu, X. Wang, and F. Wu, "Motioncast: on the capacity and delay tradeoffs," in *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2009, pp. 289–298.

[218] D. Jiang, V. Taliwal, A. Meier, W. Holfelder, and R. Herrtwich, "Design of 5.9 ghz DSRC-based vehicular safety communication," *IEEE Wireless Communications*, vol. 13, no. 5, pp. 36–43, 2006.

[219] M. Li, Z. Yang, and W. Lou, "Codeon: Cooperative popular content distribution for vehicular networks using symbol level network coding," *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 1, pp. 223–235, 2011.

[220] X. Lin, "Simplification of network dynamics in large systems," Ph.D. dissertation, Purdue University, August 2005.

[221] X. Lin and N. Shroff, "The fundamental capacity-delay tradeoff in large mobile ad hoc networks," in *Proceedings of Third Annual Mediterranean Ad Hoc Networking Workshop*, 2004, p. 2004.

[222] J. Liu, X. Jiang, H. Nishiyama, and N. Kato, "Generalized two-hop relay for flexible delay control in manets," *IEEE/ACM Transactions on Networking*, vol. 20, no. 6, pp. 1950–1963, 2012.

[223] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "VeMail: A message handling system towards efficient transportation management," in *Proceedings of IEEE WCNC*, Shanghai, China, April 2013.

[224] N. Lu, T. Luan, M. Wang, X. Shen, and F. Bai, "Capacity and delay analysis for social-proximity urban vehicular networks," in *Proceedings of IEEE INFOCOM*, Orlando, USA, March 2012.

[225] R. Lu, X. Lin, X. Liang, and X. Shen, "A dynamic privacy-preserving key management scheme for location-based services in VANETs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 1, pp. 127–139, 2012.

[226] R. Lu, X. Lin, T. Luan, X. Liang, and X. Shen, "Pseudonym changing at social spots: An effective strategy for location privacy in VANETs," *IEEE Trans. on Vehicular Technology*, vol. 61, no. 1, pp. 86–96, 2012.

[227] T. Luan, L. Cai, J. Chen, X. Shen, and F. Bai, "VTube: Towards the Media Rich City Life with Autonomous Vehicular Content Distribution," in *Proceedings of IEEE SECON*, Salt Lake City, USA, June 2011.

[228] T. Luan, X. Ling, and X. Shen, "MAC in motion: impact of mobility on the MAC of drive-thru internet," *IEEE Transactions on Mobile Computing*, no. 99, pp. 305–319, 2011.

[229] J. Mammen and D. Shah, "Throughput and delay in random wireless networks with restricted mobility," *Information Theory, IEEE Transactions on*, vol. 53, no. 3, pp. 1108–1116, 2007.

[230] F. Martinez, C. Toh, J. Cano, C. Calafate, and P. Manzoni, "Emergency services in future intelligent transportation systems based on vehicular communication networks," *IEEE Intelligent Transportation Systems Magazine*, vol. 2, no. 2, pp. 6–20, 2010.

[231] B. Nazer and M. Gastpar, "Reliable physical layer network coding," *Proceedings of the IEEE*, no. 99, pp. 1–23, 2011.

[232] A. Ozgur, O. Lévêque, and D. Tse, "Linear capacity scaling in wireless networks: Beyond physical limits?" in *Information Theory and Applications Workshop (ITA), 2010*. IEEE, 2010, pp. 1–10.

[233] A. Paier, R. Tresch, A. Alonso, D. Smely, P. Meckel, Y. Zhou, and N. Czink, "Average downstream performance of measured IEEE 802.11 p infrastructure-to-vehicle links," in *Proceedings of IEEE ICC*, Cape Town, South Africa, May 2010.

[234] E. Perevalov and R. Blum, "Delay limited capacity of ad hoc networks: Asymptotically optimal transmission and relaying strategy," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 2. IEEE, 2003, pp. 1575–1582.

[235] A. Ramamoorthy, J. Shi, and R. Wesel, "On the capacity of network coding for random networks," *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2878–2885, 2005.

[236] C. Robinson, L. Caminiti, D. Caveney, and K. Laberteaux, "Efficient coordination and transmission of data for cooperative vehicular safety applications," in *Proceedings of the 3rd international workshop on Vehicular ad hoc networks*. ACM, 2006, pp. 10–19.

[237] F. Ros, P. Ruiz, and I. Stojmenovic, "Reliable and efficient broadcasting in vehicular ad hoc networks," in *Proceedings of IEEE VTC Spring*, Barcelona, Spain, April 2009.

[238] V. Vapnik, *Estimation of dependences based on empirical data*. Springer-Verlag New York Inc, 2006.

[239] W. Viriyasitavat, F. Bai, and O. Tonguz, "Dynamics of network connectivity in urban vehicular networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 3, pp. 515–533, 2011.

[240] X. W. W. Huang, "Throughput and delay scaling of general cognitive networks," in *IEEE INFOCOM 2011, to appear, Shanghai, China.*, 2011.

[241] X. Wang, W. Huang, S. Wang, J. Zhang, and C. Hu, "Delay and capacity tradeoff analysis for motioncast," *Networking, IEEE/ACM Transactions on*, no. 99, pp. 1–1.

[242] Z. Wang, H. Sadjadpour, J. Garcia-Luna-Aceves, and S. Karande, "Fundamental limits of information dissemination in wireless ad hoc networks–part I: single-packet reception," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 5749–5754, 2009.

[243] S. Weber, X. Yang, J. Andrews, and G. De Veciana, "Transmission capacity of wireless ad hoc networks with outage constraints," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4091–4102, 2005.

[244] L. Xie and P. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Transactions on Information Theory*, vol. 50, no. 5, pp. 748–767, 2004.

[245] C. Zhang, X. Zhu, and Y. Fang, "On the improvement of scaling laws for large-scale MANETs with network coding," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 5, pp. 662–672, 2009.

[246] N. Zhang, N. Lu, R. Lu, J. W. Mark, and X. Shen, "Energy-efficient and trust-aware cooperation in cognitive radio networks," in *Proceedings of IEEE ICC*, Ottawa, CA, June 2012.

[247] "KPMG's Global Automotive Executive Survey." [Online]. Available: http://www.kpmg.com/GE/en/IssuesAndInsights/ArticlesPublications/Documents/Global-automotive-executive-survey-2012.pdf

List of Publications

Book/Book Chapter

[B1]. **N. Lu**, and X. Shen, "Capacity Analysis of Vehicular Communication Networks," **SpringerBriefs in Electrical and Computer Engineering**, Springer Verlag, 2013. (ISBN: 978-1-4614-8396-0)

[B2]. N. Zhang, N. Cheng, **N. Lu**, H. Zhou, J.W. Mark, and X. Shen,"Opportunistic Spectrum Access in Multi-Channel Cognitive Radio Networks," **Multimedia Over Cognitive Radio Networks: Algorithms, Protocols, and Experiments**, CRC Press, to appear.

Journal Papers

[J1]. **N. Lu**, N. Cheng, N. Zhang, X. Shen, J.W. Mark, and Fan Bai, "Wi-Fi Hotspot at Signalized Intersection: Cost-Effectiveness for Vehicular Internet Access," **submitted to IEEE Transactions on Vehicular Technology**.

[J2]. **N. Lu**, N. Cheng, N. Zhang, X. Shen, and J.W. Mark, "Connected Vehicles: Solutions and Challenges," **IEEE Internet of Things Journal**, Vol. 1, No. 4, pp. 289-299, 2014.

[J3]. **N. Lu**, T.H. Luan, M. Wang, X. Shen, and F. Bai, "Bounds of Asymptotic Performance Limits of Social-Proximity Vehicular Networks," **IEEE/ACM Transactions on Networking**, Vol. 22, No. 3, pp. 812-825, 2014.

[J4]. **N. Lu** and X. Shen, "Scaling Laws for Throughput Capacity and Delay in Wireless Networks – A Survey,"

IEEE Communications Surveys and Tutorials, Vol. 16, No. 2, pp. 642-657, 2014.

[J5]. **N. Lu**, N. Zhang, N. Cheng, X. Shen, J.W. Mark, and F. Bai, "Vehicles Meet Infrastructure: Towards Capacity-Cost Tradeoffs for Vehicular Access Networks," **IEEE Transactions on Intelligent Transportation Systems**, Vol. 14, No. 3, pp. 1266-1277, 2013.

[J6]. C. Chen, J. Yan, **N. Lu**, Y. Wang, X. Yang, and X. Guan, "Ubiquitous Monitoring for Industrial Cyber-Physical Systems over Relay Assisted Wireless Sensor Networks," **IEEE Transactions on Emerging Topics in Computing**, minor revision.

[J7]. N. Zhang, N. Cheng, **N. Lu**, H. Zhou, X. Shen, and J.W. Mark, "Risk-aware Cooperative Spectrum Access for Multi-Channel Cognitive Radio Networks," **IEEE Journal on Selected Areas in Communications**, to appear.

[J8]. R. Du, C. Chen, B. Yang, **N. Lu**, X. Guan, and X. Shen, "Effective Urban Traffic Monitoring by Vehicular Sensor Networks", **IEEE Transactions on Vehicular Technology**, to appear.

[J9]. N. Cheng, **N. Lu**, N. Zhang, X. Shen, and J.W. Mark, "Vehicular WiFi Offloading: Challenges and Solutions," **Vehicular Communications (Elsevier)**, Vol. 1, No. 1, pp. 13-21, 2014.

[J10]. N. Cheng, N. Zhang, **N. Lu**, X. Shen, J.W. Mark, and F. Liu, "Opportunistic Spectrum Access for CR-VANETs: A Game Theoretic Approach," **IEEE Transactions on Vehicular Technology**, Vol. 63, Issue 1, pp. 237-251, 2014.

[J11]. N. Zhang, **N. Lu**, N. Cheng, J.W. Mark, and X. Shen, "Cooperative Spectrum Access Towards Secure Information Transfer for CRNs," **IEEE Journal on Selected Areas in Communications**, Vol. 31, No. 11, pp. 2453-2464, 2013.

[J12]. S. Cspedes, **N. Lu**, and X. Shen, "VIP-WAVE: On the Feasibility of IP Communications in 802.11p Vehicular Networks," **IEEE Transactions on Intelligent Transportation Systems**, vol. 14, no. 1, pp. 82-97, 2013.

[J13]. K. Liu, J. Guo, **N. Lu**, F. Liu, X. Wang, and P. Wang, "RAMC: A RSU-assisted multi-channel coordination MAC protocol for VANET,"
IEICE Transactions on Communications, vol. 94, no. 1, pp. 203-214, 2011.

Conference Papers

[C1]. N. Cheng, **N. Lu**, N. Zhang, X. Shen, and J.W. Mark, "Opportunistic WiFi Offloading in Vehicular Environment: A Queueing Analysis,"
GLOBECOM 2014 — IEEE Global Communications Conference, Austin, TX, USA, 2014.

[C2]. N. Zhang, N. Cheng, **N. Lu**, H. Zhou, J.W. Mark, and X. Shen, "Cooperative Cognitive Radio Networking for Opportunistic Channel Access,"
GLOBECOM 2013 — IEEE Global Communications Conference, Atlanta, USA, 2013.

[C3]. N. Cheng, **N. Lu**, N. Zhang, X. Shen, and J.W. Mark, "Vehicle-Assisted Data Delivery for Smart Grid: An Optimal Stopping Approach,"
ICC 2013 — IEEE International Conference on Communications, Budapest, Hungary, 2013.

[C4]. Z. Liu, Y. Mao, **N. Lu**, Y. Ji, and X. Shen, "Resource Allocation for WWAN Video Multicast with Cooperative Local Repair",
ICC 2013 — IEEE International Conference on Communications, Budapest, Hungary, 2013.

[C5]. **N. Lu**, N. Cheng, N. Zhang, X. Shen, and J.W. Mark, "VeMail: A Message Handling System Towards Efficient Transportation Management,"
WCNC 2013 — IEEE Wireless Communications and Networking Conference, Shanghai, China, 2013.

[C6]. N. Zhang, **N. Lu**, N. Cheng, J.W. Mark, and X. Shen, "Cooperative Networking Towards Secure Communications for CRNs,"
WCNC 2013 — IEEE Wireless Communications and Networking Conference, Shanghai, China, 2013.

[C7]. N. Zhang, **N. Lu**, N. Cheng, J.W. Mark, and X. Shen, "Towards Secure Communications in Cooperative Cognitive Radio Networks",

ICCC 2013 — IEEE/CIC International Conference on Communications in China, Xi'An, China, 2013.

[C8]. **N. Lu**, T.H. Luan, M. Wang, X. Shen, and F. Bai, "Capacity and Delay Analysis for Social-Proximity Urban Vehicular Networks,"
INFOCOM 2012 — IEEE International Conference on Computer Communications, Orlando, USA, 2012. (Acceptance ratio: 18%)

[C9]. N. Zhang, **N. Lu**, R. Lu, J.W. Mark, and X. Shen, "Energy-Efficient and Trust-aware Cooperation in Cognitive Radio Networks,"
ICC 2012 — IEEE International Conference on Communications, Ottawa, Canada, 2012.

[C10]. M. Wang, H. Shan, L.X. Cai, **N. Lu**, X. Shen, and F. Bai, "Throughput Capacity of VANETs by Exploiting Mobility Diversity",
ICC 2012 — IEEE International Conference on Communications, Ottawa, Canada, 2012.

[C11]. N. Cheng, **N. Lu**, P. Wang, X. Wang, and F. Liu, "A QoS-provision multi-channel MAC in RSU-assisted vehicular networks,"
VNC 2011 — IEEE Vehicular Networking Conference, Amsterdam, The Netherlands, 2011.

[C12]. **N. Lu**, Y. Ji, F. Liu, and X. Wang, "A Dedicated Multi-channel MAC Protocol Design for VANET with Adaptive Broadcasting,"
WCNC 2010 — IEEE Wireless Communications and Networking Conference, Sydney, Australia, 2010.

[C13]. N. Cheng, **N. Lu**, K. Liu, X. Wang, and F. Liu, "A Prioritized Resource Scheduling Scheme for Throughput-Sensitive Applications in VANET,"
WiCOM 2010 — International Conference on Wireless Communications, Networking and Mobile Computing, Chengdu, China, 2010.

[C14]. **N. Lu**, X. Wang, P. Wang, P. Lai, and F. Liu, "A Distributed Reliable Multi-channel MAC Protocol for Vehicular Ad Hoc Networks,"
IV 2009 — IEEE Intelligent Vehicles Symposium, Xi'an, China, 2009.

[C15]. P. Lai, X. Wang, **N. Lu**, and F. Liu, "A reliable broadcast routing scheme based on mobility prediction for VANET,"
IV 2009 — IEEE Intelligent Vehicles Symposium, Xi'an, China, 2009.

Vita

Ning Lu received the B.Sc. and M.Sc. degrees from Tongji University, Shanghai, China, in 2007 and 2010, respectively, both in electrical engineering. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His current research interests include scaling laws, real-time scheduling, and cross-layer design for vehicular networks. Mr. Lu served as a Technical Program Committee Member for IEEE PIMRC'12, WCSP'13, WCSP'14, ICNC'15, and ICC'15.