UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

Cooperative Vehicular Networks for Intelligent Transportation Systems

by

Jieqiong Chen

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Sydney, Australia

2019

Certificate of Authorship/Originality

I certify that the work in this thesis has not been previously submitted for a degree nor has it been submitted as a part of the requirements for other degree except as fully acknowledged within the text.

I also certify that this thesis has been written by me. Any help that I have received in my research and in the preparation of the thesis itself has been fully acknowledged. In addition, I certify that all information sources and literature used are quoted in the thesis.

Production Note:
Signature removed prior to publication.

ABSTRACT

Cooperative Vehicular Networks for Intelligent Transportation Systems

by

Jieqiong Chen

Transportation systems are fundamental for the human society as they allow people and goods to move from one location to another. With an increasing volume of population and vehicles, current transportation systems are now facing a number of disruptive challenges such as congestion, crashes, air pollution and noise throughout the world. However, traditional solutions like expanding the present transportation systems by increasing the number of roads are recognized to be expensive, disruptive and involve protracted effort. Instead, intelligent transportation systems (ITS), with the goal of building a safer, more efficient and environmentally sustainable transportation system by incorporating state-of-the-art sensing, computing and communication technologies, is expected to be a better solution.

ITS are complex systems and they function in a broad range of areas through smartly sensing, analysing and disseminating different kinds of traffic information. Vehicular networks, which incorporate advanced communication technology with intelligent vehicles equipped with on-board units (OBUs) and intelligent roadside infrastructure, realise the function of large scale traffic information dissemination for ITS through vehicle to vehicle (V2V), vehicle to infrastructure (V2I) and infrastructure to infrastructure (I2I) communications. Therefore, as one of the most enabling tools to support ITS, vehicular networks play a crucial role in improving road safety, relieving traffic congestion, enhancing driving experience and reducing pollution.

Considering the critical impact information exchange poses on the transportation systems, vehicular network applications require particularly fast, reliable and secure message dissemination in the network. However, depending only on V2V or V2I communications may fail to meet these requirements. On one hand, the frequently changing topology of vehicular networks caused by the highly dynamic nature of vehicles and the lossy vehicular wireless channels resulting from fading, path loss and the fast movement of vehicles, would result in unreliable and intermittent V2V communications. On the other hand, V2I communications may have limited availability, especially in rural areas and in the initial deployment phase of vehicular networks due to the high cost of implementation and maintenance of infrastructure. These make research on employing cooperative communications within vehicular networks both interesting and important.

In this thesis, we focus on the design of cooperative vehicular networks for ITS to satisfy the requirement of disseminating data quickly, reliably and securely, in the conditions of sparse roadside infrastructure, high mobility, and intermittent connectivity. Firstly, we propose a cooperative communication strategy that explores the combined use of V2I communications, V2V communications, mobility of vehicles, and cooperation among vehicles and infrastructure, to facilitate data dissemination in vehicular networks. The network performance, measured by the achievable throughput when there exists only one vehicle with a download request in the network, and the achievable capacity when there exist multiple vehicles with download requests in the network respectively, are analysed. The results show that the proposed cooperative communication strategy significantly boosts the throughput (or capacity) of vehicular networks. Secondly, to protect secure message dissemination, we investigate topological approaches to keep the message dissemination in vehicular networks robust against insider attackers who may tamper with the message content. As a novel approach, we take the network topology into consideration when designing algorithms to check the integrity and consistency of messages. Overall, our work provides guidance on the optimum design of cooperative vehicular networks for ITS to achieve fast, reliable and secure message dissemination.

Dissertation directed by Professor Guoqiang Mao
School of Electrical and Data Engineering

Acknowledgements

The completion of this dissertation has been possible with the inspiration and encouragement from many people, to whom I am greatly indebted.

First of all, I overwhelmingly pay my immeasurable appreciation and deepest gratitude to my supervisor, Professor Guoqiang Mao, for his persistent guidance, valuable recommendations, generous advice, never-ending patience and ongoing support. I feel extremely fortunate to be mentored by him during my Ph.D. candidature. Without his guidance and persistent help, this work would not have been possible.

I am also grateful to other members in our research team, for their making a friendly working environment, and for their support, great assistance and valuable advice to my research.

I sincerely acknowledge the deepest love and ongoing support from my beloved family during my studies. I am really grateful to my parents, my parents-in-law, my sister, and my brother-in-law, for their unconditional love and continuing encouragement.

Finally, but mostly, I wish to express my deepest gratitude to my beloved husband Zhihui Jiang, who was always standing by me in my hard times during this work and giving me faith. Thanks for all the joy and beautiful memories he brought to our life, and his precious love. I consider myself extremely fortunate to have him in the journal of my life.

Jieqiong Chen

Sydney, Australia

January 2019

List of Publications

The following is a list of publications in refereed journals and conference proceedings produced during my Ph.D. candidature. In some cases, the journal papers contain material overlapping with the conference publications.

Journal Papers

J-1. **J. Chen**, G. Mao, C. Li, D. Zhang, "A Topological Approach to Secure Message Dissemination in Vehicular Networks," Accepted by *IEEE Transactions on Intelligent Transportation Systems*, Dec. 2018.

J-2. **J. Chen**, G. Mao, C. Li, W. Liang, D. Zhang, "Capacity of Cooperative Vehicular Networks with Infrastructure Support: Multi-user Case," *IEEE Transactions on Vehicular Technologies*, vol. 67, no. 2, pp. 1546 - 1560, Feb. 2018.

J-3. **J. Chen**, G. Mao, C. Li, A. Zafar, A. Y. Zomaya, "Throughput of infrastructure-based cooperative vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 11, pp. 2964-2979, Nov. 2017.

J-4. **J. Chen**, G. Mao, "On the Security of Warning Message Dissemination in Vehicular Ad Hoc Networks," *Journal of Communications and Information Networks*, vol. 2, no. 2, pp 4658, Jun. 2017

Conference Papers

C-1. **J. Chen**, G. Mao, "Secure Message Dissemination in Vehicular Networks: A Topological Approach," *IEEE Global Communications Conference (Globecom)*, 2018

C-2. **J. Chen**, G. Mao, C. Li, "Capacity of Infrastructure-based Cooperative Vehicular Networks," *IEEE Global Communications Conference (Globecom)*, 2017

C-3. **J. Chen**, A. Zafar, G. Mao, C. Li, "On the achievable throughput of cooperative vehicular networks," *IEEE International Conference on Communications (ICC)*, 2016

Contents

List of Figures

Chapter 1

Introduction

Road transportation systems are fundamental for the human society as they allow the movement of people and goods from one location to another using a variety of vehicles across different infrastructure systems. Along with the desired outputs of passenger trips and freight shipments, road transportation systems also bring adverse outcomes such as congestion, crashes, air pollution and noise. With an increasing volume of vehicles over the past decades, the concern on road traffic efficiency and safety have been growing. Consequently, extensive research has been conducted in investigating and developing new techniques to address these urgent transport issues. Among all techniques, Intelligent Transportation Systems (ITS), with the goal of alleviating road traffic problems using state-of-the-art sensing, computing and communication technologies to build a safer, more efficient and environmentally sustainable transport network, stand out as a popular and important one.

1.1 Research Background

1.1.1 Intelligent Transportation Systems

With an increasing transportation vehicles during past decades, traffic congestion and accidents have become significant transport issues all around the world. In particular, it is claimed in the 2015 Urban Mobility Report, published by the Texas A&M Transportation Institute [1], that traffic congestion caused urban Americans to travel an extra 6.9 billion hours and purchase an extra 3.1 billion gallons of fuel, for a congestion cost of $160 billion in 2014. Meanwhile, the 2017 report from

World Health Organization (WHO) pointed out that road traffic injuries have been the main cause of death among people aged 15 - 29 years. More than 1.2 million people die annually in highway-related crashes [2, 3]. These figures are likely to increase continually without significant changes made to the traditional transportation systems. However, traditional solutions like expanding the present transportation systems by increasing the number of roads are recognized to be expensive, disruptive and involve protracted effort [4]. To this end, ITS, which incorporate the advanced sensing, computing and communication technologies to a smart use of the transport network, is expected to be a better solution to the urgent road traffic problems.

The term ITS was coined in 1994 by the United States Department of Transportation, while its development can be traced back to the 1980s with programs that used sensing, information and communication technologies to achieve greater sophistication of traffic and transport, such as the program PROMETHEUS (PROgraM for European Traffic with Highest Efficiency and Unprecedented Safety) in Europe, the program IVHS (Intelligent Vehicle Highway Systems) in the United States, and the program RACS (Road Automobile Communication System) in Japan [5, 6]. ITS are complex systems and they function in a broad range of areas through sensing and perception, cognition, and actuation sub-systems to improve road safety, relieve traffic congestion, enhance driving experience and reduce pollution. Their functional areas mainly include (1) Advanced Traffic Management System (ATMS), which is the fundamental part of ITS that operates traffic congestion detection, traffic prediction and traffic control through collecting, analysing and disseminating real-time traffic data; (2) Advanced Traveller Information System (ATIS), which provides real-time traffic information like road conditions and optimal routings to travellers, so as to reduce congestion and optimize traffic flow; (3) Advanced Vehicle Control system, which incorporates sensors, computers and control systems to assist and alert drivers on collisions, so as to increase road safety; and (4) Advanced Public

Transportation System, which uses technologies from ATMS and ATIS to greatly enhance the accessibility of information to users of public transportation [7, 8].

1.1.2 Vehicular Networks

Vehicular networks, which incorporate advanced communication technology with intelligent vehicles equipped with on-board units (OBUs) and intelligent roadside infrastructure, serve as one of the most significant tools for the realization of ITS. With the assistance of wireless communication technology, vehicular networks allow different kinds of information exchanged among vehicles and roadside infrastructure through vehicle to vehicle (V2V), vehicle to infrastructure (V2I) and infrastructure to infrastructure (I2I) communications. Specially, vehicular network is a type of a conventional mobile ad-hoc network, but with nodes moving on predefined roads, and their trails are not too complicated. See. Fig. 1.1 for an illustration.

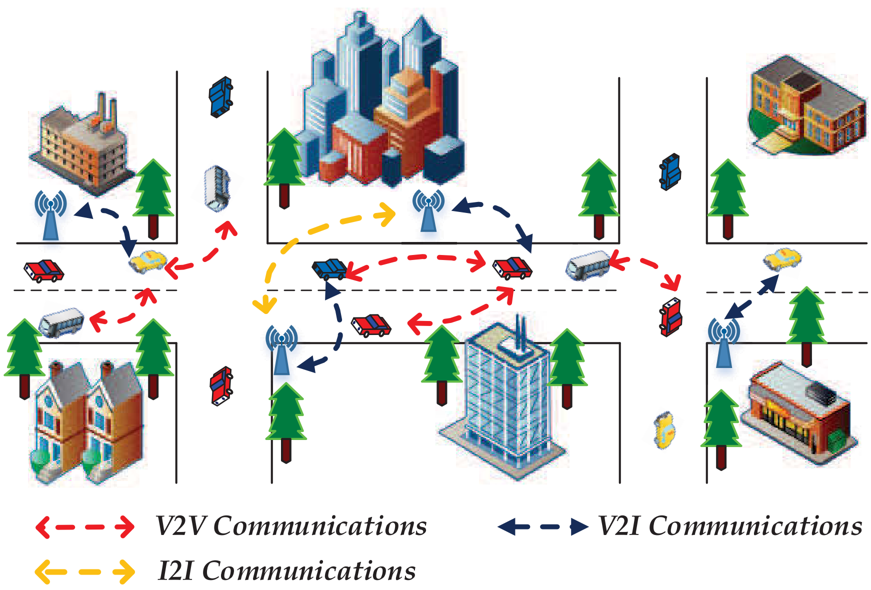


Figure 1.1 : An illustration of vehicular networks.

Vehicular networks provide great support to ITS by realising the function of large scale traffic information dissemination, and play a critical role in improving road safety, enhancing traffic efficiency, and increasing comfort and convenience to drivers and passengers [9, 10, 11, 12]. Specifically, by rapidly disseminating warning messages on real-time information about traffic accidents or road obstacles to all vehicles in vicinity through V2V communications, road safety can be greatly improved (i.e. traffic safety services). Moreover, a timely and effective exchange of sensed traffic information among vehicles keeps drivers informed about the traffic density at all locations throughout the city. This consequently assists in providing guidance to drivers on the optimal driving route to avoid traffic congestion. In the meantime, the roadside infrastructure can collect real-time traffic information from vehicles and other infrastructure, and this collected traffic data can be accessed by traffic planners for performing traffic analysis to make more accurate and reliable traffic predictions and traffic planning (i.e. traffic efficiency services). Furthermore, vehicular networks enable vehicles and roadside infrastructure to offer value-added services like digital maps with real-time traffic status and in-car entertainment services, which can greatly enhance the convenience and comfort of drivers and passengers (i.e. infotainment services).

The great benefits vehicular networks can provide to ITS have attracted academia, governments, and car manufacturers globally to initiate numerous activities in this field in the recent past. Different standardization bodies have devoted efforts into specifying vehicular communications, after dedicated spectrum at 5.9 GHz was allocated for ITS both in the US in 1999 and in Europe in 2008, respectively. Different families of standards include: 1) the Dedicated Short-Range Communications (DSRC) completed in 2010 in US [13]; 2) the Release 1 of the ETSI/CEN Cooperative-ITS (C-ITS) completed in 2013 in Europe [14]; and 3) 3GPP Cellular-V2X (C-V2X) completed in 2017 [15]. Both DSRC and C-ITS use the IEEE 802.11p [16] standard for the physical and data link layers, which is a short-range technology for V2V communications. The C-V2X standard embeds into vehicular environments the latest technologies in cellular communications, and it includes two radio interfaces. The cellular interface supports V2I communications, while the direct interface named sidelink/PC5 supports V2V communications based on direct LTE side-link [15]. In addition, many programs have been established widely for the future deployment of large-scale vehicular networks, e.g., the Connected Vehicle Pilot Deployment Program began in 2015 by the United States Department of Transportation to test cutting-edge connected vehicle technologies, and the Compass4D project in Europe, etc. [17]. Moreover, a large number of car manufacturers are also supplying vehicles with on-board computing and wireless communication devices. For instance, about 100,000 DSRC-equipped vehicles from Toyota and Lexus are on the road in Japan currently [18].

In spite of the benefits vehicular networks bring to ITS, the effective design and deployment of vehicular networks are still challenging. First, vehicular networks are only effective when every vehicle has equipped an OBU and the road side infrastructure is also deployed concurrently. However, due to the cost and budget constraints, it is expected to be a long journey for all car manufacturers to install OBUs on each vehicle, and for the government to upgrade the vehicular network infrastructure. Second, there are still some urgent technical issues remaining to be solved. Due to the high-mobility nature of vehicular network, randomness in wireless channel dynamics and link interferences, vehicular communications may become intermittent and unreliable. Consequently, it may lead vehicular network applications struggling to provide information as rapidly and reliably as required. Meanwhile, the vulnerability of wireless communications makes message dissemination in vehicular networks susceptible to malicious attacks, and these attacks could potentially result in catastrophic consequences like traffic congestion, traffic crashes, and even loss of lives. Extensive research needs to be conducted to protect secure message dissemination in vehicular networks.

Based on the aforementioned technical challenges remaining to be solved before vehicular networks can be fully deployed, it is pressing to investigate cooperative vehicular networks for ITS to satisfy the requirement of disseminating data quickly, reliably and securely in sparse roadside infrastructure, high mobility, and intermittent connectivity conditions, so as to pave ways for the realisation of ITS.

1.2 Research Motivation

1.2.1 Cooperative Communications

The requirement of fast, reliable and secure message dissemination in vehicular networks has attracted significant attention from both academia and industry to employ cooperative communications within vehicular networks by enabling vehicles and infrastructure to cooperate with each other [19]. V2I and V2V communications, as the two fundamental communication techniques, play critical roles in data dissemination in vehicular networks. V2V communications, such as Vehicular Ad-hoc NETworks (VANETs), are easy to deploy and suitable for real-time communications between nearby vehicles. However, it is well known that V2V communications may become unreliable when the number of hops in the communication becomes large [20]. They may also not be supported and incur a long communication delay when the vehicular density is low [21, 22]. Thus, infrastructure support in the form of Road-Side Units (RSUs), Wi-Fi Access Points (APs) or LTE base stations that provides V2I communications is both essential and beneficial. However, V2I communications have their own limitations. On one hand, LTE-based V2I communication is not applicable to a vehicular communication environment at this stage due to the high handover and collision issues caused by the high mobility of vehicles [23]. On the other hand, DSRC-based V2I communication may have limited availability, especially in rural areas and in the initial deployment phase of vehicular networks due to the high cost of implementation and maintainance of infrastructure [24]. Furthermore, an individual vehicle's sojourn time within the coverage of an infrastructure point may be short due to the high speed of vehicles and the limited coverage of vehicular communication infrastructure. The aforementioned factors may result in frequent interruptions in data transmissions, especially when downloading files of large size from the Internet, e.g., in-car entertainment services. Therefore, V2I and V2V communications have to co-exist and complement each other, and cooperative communication strategy should be designed to enable neighbouring vehicles and infrastructure to communicate cooperatively with each other to promote fast and efficient data dissemination in vehicular networks.

1.2.2 Vehicular Network Capacity

Capacity is one of the most important performance metrics in vehicular networks, because it characterizes the feasible data dissemination rate of the network. Investigating capacity of vehicular networks with infrastructure support assuming the cooperative communication strategy is both an interesting and a challenging problem, as the capacity is determined by the inter-play of multiple factors including V2I communications, V2V communications, density and mobility of vehicles, and cooperation among vehicles and infrastructure. Since the seminal work of Gupta and Kumar [25], extensive research on capacity has been conducted [26, 27, 28, 29]. However, the existing work either assumes that the number of vehicles or vehicular density is sufficiently large and utilizes asymptotic analysis to study the capacity scaling law, which is only applicable when the number of vehicles or vehicular density is sufficiently large; or considering the communication capacity (channel throughput) in the network. This motivates us to focus on an accurate analysis, instead of the scaling law, of the capacity of vehicular networks with a moderate number of vehicles or a finite vehicular density where the asymptotic analysis may not apply.

1.2.3 Vehicular Network Security

Even though vehicular network applications bring numerous benefits to ITS, accompanying these benefits is the urgent security issues that need to be paid special attention [30, 31]. Specifically, considering the vulnerability of wireless communications, message dissemination in vehicular networks is susceptible to malicious attacks, e.g., malicious vehicles who may spread false messages [32], tamper or drop the received messages [33, 34] to disrupt delivery of authentic messages. These attacks could potentially result in catastrophic consequences like traffic congestion, traffic crashes, even loss of lives, and therefore are significant security threats to transportation systems that must be thoroughly investigated before vehicular networks can be deployed.

Vehicular network security design should guarantee non-repudiation, authentication, information integrity, and in some specific application scenarios, confidentiality, to protect the network against attackers [31, 35]. Conventional security mechanisms largely based on cryptography and key management [36, 37] are dedicated to guarantee authentication to the network against outsider attackers, however are not feasible in protecting the integrity of disseminated messages when there exist insider attackers who possess valid certificates that can pass the authentication process conducted by the certification authorities [38, 39].

To keep the network message dissemination robust to insider attackers, the trustworthiness of each vehicle and the integrity of their transmitted messages are of great importance. Different from traditional security settings, in vehicular networks, information collection and dissemination are conducted by distributed vehicles. Quite often, information may be generated by or received from a vehicle that has never been encountered before. Moreover, the associated network topology of vehicular networks is constantly changing, considering that both V2V and V2I connections may emerge opportunistically. These unique characteristics may render the entity-based trust establishment approach, conducted at each vehicle by monitoring their instantaneous neighbours' behaviour, futile in vehicular networks because it is challenging to maintain a stable reputation value for an unknown and fast-moving vehicle. Furthermore, safety-related vehicular network applications usually require vehicles make a quick response to the messages [40]. In such cases, determining the truth-value of the disseminated messages is of more importance than detecting the malicious vehicles. Therefore, decision algorithms based on data consistency and integrity checks emerge, e.g., [41, 42, 43, 44, 45]. However, when a vehicle receives conflicting messages from different paths, it is not straightforward to tell which message is true if focusing only on the data while ignoring the underlying path information which describes where these messages come from. Actually, messages delivered by different paths can be correlated when the underlying paths share some common nodes. For instance, multiple false messages may result from the same malicious vehicle shared by multiple paths. Therefore, taking the underlying topological information into consideration is essential when designing decision algorithms for vehicles to conduct data consistency checks.

Based on the aforementioned essentiality and challenges in guaranteeing secure message dissemination in vehicular networks, investigating topological approaches to secure message dissemination in vehicular networks becomes a significant topic that is uniquely worthy of investigation.

1.3 Research Objectives and Contributions

From the aforementioned research background and motivation, this thesis focuses on the following research problems: (1) design cooperative vehicular networks for ITS to achieve fast and reliable data dissemination; and (2) design cooperative vehicular networks for ITS to achieve secure message dissemination. In the following of this section, the detailed research problems and the corresponding contributions will be elaborated.

For the first research problem, we consider a typical delay-tolerant application scenario where there is only one vehicle with a download request from the Internet, e.g., download a large-size video file. We aim to (1) design a cooperative communication strategy that utilizes V2I communications, V2V communications, and cooperation among infrastructure and vehicles, as well as mobility of vehicles to facilitate data transmission in the considered network; and (2) theoretically investigate the long-term data rate the target vehicle can achieve, i.e., the achievable throughput, assuming the proposed cooperative communication strategy. The main contributions regarding this research work are summarized as follows:

i. Propose a novel cooperative communication strategy, which incorporates V2I communications, V2V communications, mobility of vehicles, and cooperation among vehicles and infrastructure to facilitate data dissemination in vehicular networks;

ii. Develop an analytical framework to model and investigate the data dissemination process assuming the aforementioned cooperative communication strategy, and obtain the closed-form expression of the throughput achieved by the target vehicle in the network. The achieved result reveals the relationship between the throughput and its major performance-impacting parameters such as distance between two neighbouring infrastructure points, radio ranges of infrastructure and vehicles, transmission rates of V2I communications and V2V communications, and density of vehicles. To the best of our knowledge, this is the first research that accurately analyses the achievable throughput of vehicular networks with a finite vehicular density, as previous works on throughput of vehicular networks focused on studying the scaling law of the capacity when the vehicular density is sufficiently large.

iii. Demonstrate the fact that our proposed cooperative strategy significantly improves the throughput of vehicular networks, by comparing the throughput achieved assuming our proposed cooperative communication strategy with that achieved assuming non-cooperative strategy, or assuming existing cooperative communication strategy in literature.

iv. Shed light on the optimum deployment of vehicular network infrastructure in terms of the interval distance, and the optimum design of cooperative vehicular networks for ITS to achieve fast and reliable data dissemination.

The second research problem considered in this thesis is an extension of the first one. In this problem, we consider a delay-tolerant application scenario with a subset of vehicles, instead of only one, having download requests from the Internet. Each vehicle with a download request downloads a distinct large-size file from the Internet, and other vehicles without download requests assist the delivery of the files to the target vehicles. By adopting the cooperative communication strategy proposed in solving the first research problem, we aim to: (1) develop an analytical framework to model the data dissemination process; and (2) investigate the closed form expression of the network capacity. Different from the single vehicle scenario described as the first research problem, when there are multiple vehicles with download requests in the network, the possible contention and collision among vehicles in vehicular communications become both an important and challenging issue to study. The main contributions regarding this research work are summarized as follows:

i. Develop an analytical framework to model and investigate the data dissemination process assuming the cooperative communication strategy proposed in the first research problem, and obtain the closed-form expression of the capacity achieved by the target vehicles, which reveals the relationship between the capacity and its major performance-impacting parameters;

ii. Demonstrate the fact that our proposed cooperative strategy significantly improves the capacity of vehicular networks, by comparing the capacity of vehicular networks achieved assuming our proposed cooperative communication strategy with that achieved assuming non-cooperative strategy, or assuming existing cooperative communication strategy in literature.

iii. Shed light on the optimum deployment of vehicular network infrastructure in terms of their interval distance, and the optimum design of cooperative vehicular networks for ITS to achieve fast and reliable data dissemination.

The third research problem considered in this thesis is to protect secure message dissemination in vehicular networks. We consider the case when vehicular networks contain insider malicious vehicles that may tamper with the content of messages to disrupt their successful delivery, and we aim to investigate topology-based decision algorithms to keep vehicles from being misguided by false messages. To the best of our knowledge, this is the first work that takes the underlying topology information into consideration when checking the consistency of messages to protect secure message dissemination. The novelty and major contributions of this research work are summarized as follows:

i. Propose two topology-based message decision algorithms - the optimum decision algorithm and a heuristic decision algorithm - to cope with the issue of message inconsistency caused by insider malicious vehicles in the network, so as to reduce their impact on the message security performance.

ii. The proposed optimum decision algorithm is able to maximize the chance of making a correct decision on the message content, assuming the prior knowledge of the percentage of malicious vehicles in the network. The proposed heuristic decision algorithm is developed, and removes the need to know the aforementioned percentage of malicious vehicles, which could be difficult to estimate.

iii. Both of the proposed algorithms outperform existing decision algorithms which do not consider or only partially consider the topological information. Moreover, the proposed heuristic decision algorithm, which is rather easy to be implemented in practice, is sufficient to achieve a high security performance.

iv. Shed light on the optimum design of vehicular networks for ITS to achieve secure message dissemination.

1.4 Thesis Organization

This thesis is organised as follows: Chapter 2 presents a survey of related works, including cooperative vehicular networks, capacity of vehicular networks and security of vehicular networks. Chapter 3 proposes a novel cooperative communication strategy, and investigates the achievable throughput of vehicular networks when assuming there is only one vehicle in the network with a download request. Chapter 4 deals with the scenario that there are multiple vehicles in the network with download requests, and investigates the capacity of vehicular networks assuming the proposed cooperative communication strategy. Chapter 5 investigates topological approaches to protect secure message dissemination in vehicular networks. Two decision algorithms, the optimum decision algorithm and a heuristic decision algorithm, are proposed. Chapter 6 presents a brief summary of the thesis contents, and gives recommendation for future works.

Chapter 2

Literature Review

This chapter is dedicated to reviewing related works to this thesis, including works on cooperative vehicular networks, capacity of vehicular networks and security of vehicular networks.

2.1 Cooperative Vehicular Networks

Extensive work in the literature investigated the performance of vehicular networks, measured by the information propagation speed [21], transmission delay [46, 47, 48], downloaded data volume [49], packet reception rate [50], communication link quality [51], etc. Among the major techniques to enhance these performance measures, cooperative vehicular networks, which utilize both cooperations among vehicles and cooperations among infrastructure points, stand out as a popular and important technique. In the following, we review works on cooperation among vehicles, cooperation among infrastructure points, and cooperation among both vehicles and infrastructure points.

2.1.1 Cooperation among Vehicles

The following work investigated cooperative communications among vehicles in vehicular networks. In [49], Zhou *et al.* introduced a cooperative communication strategy using a cluster of vehicles on the highway to cooperatively download the same file from the infrastructure to enhance the probability of successful download. In [46], Zhu *et al.* studied using multiple nearby vehicles to collaboratively download data from an RSU and analysed the average download time using network coding techniques. In [50], Das *et al.* introduced a coalitional graph game to model cooperative message sharing among vehicles in vehicular networks and proposed a coalition formation algorithm to improve the packet reception rate and reduce transmission delay. In [52], Yan *et al.* developed a theoretical model to analyse the achievable channel data rate of VANETs for cooperative mobile content dissemination, also assisted by network coding techniques. They focused on the transmission throughput, i.e., the channel data rate in MAC layer. Li *et al.* [53] proposed a push-based popular content distribution scheme for vehicular networks, where large files are broadcast proactively from a few infrastructure points to vehicles inside an area of interest and further disseminated cooperatively among vehicles using V2V communications. In [54], Wang *et al.* introduced a coalitional graph game to model the cooperations among vehicles and proposed a coalition formation algorithm to implement the cooperation between vehicles for popular content distribution. In [55], Liu *et al.* utilized the location information of each vehicle and cooperation among vehicles to maximize the number of vehicles that successfully retrieve their requested data in vehicular networks.

2.1.2 Cooperation among Infrastructure Points

Infrastructure support plays a significant role in vehicular networks. Recent studies show that by adding a number of infrastructure points in a wireless (vehicular) network, the network performance will be improved. Specifically, in [26], Dai *et al.* showed that the capacity will be significantly improved and the delay will be reduced by utilizing infrastructure support compared to that without utilizing infrastructure. In [22], Reis *et al.* showed that significant benefits of RSUs in terms of connectivity and message dissemination will be achieved when the deployed RSUs are interconnected. In [56], Liu *et al.* applied V2X communication environments with infrastructure support to improve bandwidth efficiency and enhance data service performance.

Besides utilizing isolated infrastructure points, cooperation among infrastructure points is able to further improve network performance. Cooperation among infrastructure points can be achieved by caching different files or different parts of a file in different infrastructure points to help moving vehicles to download from the Internet. In [57], to fully utilize the wireless bandwidth provided by APs, Zhang and Yeo proposed a cooperative content distribution system for vehicles by using cooperative APs to distribute contents to moving vehicles. More specifically, by prefetching different data into the selected APs, vehicles can obtain the complete data from those selected APs when travelling through their coverage areas. In [58], Li *et al.* proposed a heuristic content distribution algorithm that caches data in different infrastructure points, to maximize the downloaded data size. In [59] and [60], the authors utilized cooperative infrastructure points in vehicular networks to maximize the success probability of download, utilizing a greedy algorithm and integer linear programming optimization respectively.

2.1.3 Cooperation among Both Vehicles and Infrastructure Points

Studies considering both vehicular cooperation and infrastructure cooperation are comparatively scarce. By exploring cooperation among vehicles and interconnected infrastructure points, Mershad et al. [61] designed an optimum routing algorithm to reduce end-to-end delay for delivering a packet from a source to its destination. In [62], Si *et al.* designed an optimum distributed data hopping mechanism to enable delay-tolerant data routing over a vehicular network. In [63], Wang *et al.* proposed a scheme that utilizes moving vehicles to serve as relays to assist data dissemination to a target vehicle, and the relay selection is conducted by the cooperative infrastructure points. They focus on reducing the transmission outage of the target vehicle.

2.2 Capacity of Vehicular Networks

Capacity is one of the most important performance metrics in vehicular networks, as it characterizes the feasible data dissemination rate of the network. Since the seminal work of Gupta and Kumar [25], extensive research on capacity has been conducted, e.g., [26, 29, 64, 65, 66, 67, 68]. Particularly, Gupta and Kumar [25] showed that the maximum throughput of static wireless networks is $\Theta(\frac{1}{\sqrt{n}})$ with n being the number of nodes in the network. In [66], Franceschetti et al. considered essentially the same random network as that in [25] except that nodes are now allowed to use two different transmission ranges. The link capacity is determined by the associated SINR through the Shannon–Hartley theorem. It was shown in [66] that the transport capacity and the per-node throughput can also reach $\Theta(n)$ and $\Theta(\frac{1}{\sqrt{n}})$ respectively even when nodes are randomly deployed. In [64], Grossglauser and Tse showed that by leveraging on the nodes' mobility, a per-node throughput of $\Theta(1)$ can be achieved at the expense of unbounded delay. Their work sparked tremendous interest in studying the capacity-delay tradeoffs in mobile networks assuming various mobility models and the obtained results often vary greatly with the different mobility models being considered, e.g., [69, 70, 71] and references therein. In [29], Mao *et al.* presented a simple relationship to estimate the capacity of both static and mobile networks, and developed a generic methodology for analyzing the network capacity that is applicable to a variety of different scenarios.

Focusing on the capacity of vehicular networks, Lu *et al.* [71] presented a comprehensive overview of capacity-delay trade-offs under a variety of mobility models and scaling laws for hybrid wireless networks. In addition, they introduced recent progress in throughput capacity of emerging vehicular networks. In [27], Wang *et al.* studied urban vehicular networks with uniformly distributed RSUs and analyzed the asymptotic uplink throughput scaling law when the total number of vehicles is sufficiently large. In [28], Huang *et al.* introduced an Euclidean planar graph and used a practical geometric structure to study the asymptotic capacity of urban Vehicular Ad-Hoc NETworks (VANETs). In [72], Ni *et al.* proposed an interference-based capacity analysis for the abstracted 1-dimensional VANETs scenario assuming the classic Car-Following model to represent the dynamic change of the inter-vehicle distance. In [73], Yang. *et al.* studied both broadcast and unicast communication throughput of VANETs under various conditions in vehicular traffic and wireless communications, including traffic density, market penetration rates of equipped vehicles, percentages of senders, transmission ranges, and interference ratios. They defined and derived formulas for computing broadcast and unicast throughput of VANETs along either discrete or continuous traffic streams.

2.3 Security of Vehicular Networks

To protect secure message dissemination in vehicular networks against insider malicious vehicles, the trustworthiness of each vehicle and the integrity of each transmitted message are two major factors which need to be considered. Accordingly, three misbehaviour detection schemes are commonly adopted to help prevent the disseminated messages from being tampered with: entity-centric misbehaviour detection scheme, data-centric misbehaviour detection scheme, and a combined use of both. In the following, we will review works on these three schemes separately.

Entity-centric misbehaviour detection schemes are commonly conducted at each vehicle by monitoring their instantaneous neighbours' behavior to assess their trustworthiness level, so as to filter out malicious vehicles. In [74], Gazdar *et al.* proposed a dynamic and distributed trust model based on the use of a Markov chain to evaluate the evolution of each vehicle's trust value. In [75], Ahmed *et al.* proposed a trust framework to identify malicious nodes in the network by evaluating the trust value of each vehicle, where the trust includes node trust and recommendation trust.

In [76], motivated by the job market signalling model, Haddadou *et al.* proposed a distributed trust model for VANETs that is able to gradually detect all malicious nodes as well as boosting the cooperation of selfish nodes. In [77], Sedjelmaci *et al.* proposed a lightweight intrusion detection framework with the help of a clustering algorithm to overcome the challenges of intermittent and ad hoc monitoring and assessment processes caused by the high mobility and rapid topology change in vehicular networks.

Data-centric misbehaviour detection schemes focus on the consistency check of the disseminated data to filter out false data. In [41], Dietzel *et al.* indicated that redundant data forwarding paths are the most promising technique for effective data consistency check in a multi-hop information dissemination environment, and proposed three graph-theoretic metrics to measure the redundancy of dissemination protocols. In [42], Raya *et al.* proposed a framework for vehicular networks to establish data-centric trust, and evaluated the effectiveness of four data fusion rules. In [43], Huang *et al.* firstly demonstrated that information cascading and oversampling adversely affect the performance of trust management schemes in VANETs, and then proposed a novel voting scheme that takes the distance between the transmitter and receiver into account when assigning weight to the trust level of the received data. In [44], Zaidi *et al.* proposed a rogue node detection system for VANETs utilizing statistical techniques to determine whether the received data is authentic. In [45], Radak *et al.* applied a so-called cautious operator to deal with data received from different sources to detect dangerous events on the road. Their adopted cautious operator is an extension of the Demper-Shafer theory that is known to be superior in handling data coming from dependent sources.

A combined use of entity-centric and data-centric misbehaviour detection scheme makes use of both the trust level of vehicles and the consistency of received data to detect misbehaving vehicles and filter out incorrect messages. Works adopting the combined scheme are limited. In [78], Dhurandher *et al.* proposed a security algorithm using both node reputation and data plausibility checks to protect the network against attacks. The node reputation value is obtained by both direct monitoring and indirect recommendation from neighbours; and the data consistency check is conducted by comparing the received data with the sensed data by the vehicle's own sensors. In [79], Li *et al.* proposed an attack-resistant trust management scheme to evaluate the trustworthiness of both data and vehicles in VANETs. They adopted the Dempster-Shafer theory to combine the data received from different sources, and then used this combined result to update the trust value of vehicles.

In summary, all the above works on protecting vehicular networks from insider attackers either focused on node trust model establishment and management to detect misbehaving nodes in the network, or focused on methods to assess data from different sources to check their consistency, but did not take the underlying network topological information into consideration. Our work on network security distinguishes from theirs in that we focus on the received data itself, and utilize the underlying network topology information to design the decision algorithms for vehicles to check data consistency so as to maximally protect the authenticity of the disseminated messages.

Chapter 3

Throughput of Cooperative Vehicular Networks with Infrastructure Support: Single-user Case

In this chapter, we consider a scenario where there is a vehicle of interest (VoI) wanting to download a large-size file, e.g., a video, from the Internet, and all the other vehicles (termed *helpers*) assist its download using a cooperative communication strategy. The strategy explores the combined use of V2I communication, V2V communication, the mobility of vehicles and cooperation among vehicles and infrastructure to facilitate data transmission. The scenario being considered corresponds to the category of delay-tolerant applications. A detailed analysis for the achievable throughput was presented assuming the proposed cooperative communication strategy and the closed-form expression of achievable throughput (or its upper and lower bound) was obtained in three different regimes we classified in our analysis based on the relationship between the data rates of V2I communications, V2V communications, and the speeds of vehicles. Numerical and simulation results show that the proposed cooperative strategy can significantly improve the achievable throughput of vehicular networks even when traffic density is comparatively low.

The rest of this chapter is organized as follows: Section 3.1 introduces the system model, the proposed cooperative communication strategy and the problem formation. Theoretical analysis of the data dissemination process and the achievable throughput are provided in Section 3.2. Section 3.3 validates the analytical results using simulations and discuss the impact of major performance-impacting parameters. Section 3.4 summarizes this chapter.

3.1 System Model and Problem Formation

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

3.1.1 Network Model

We consider a highway with bi-directional traffic flows. The highway is modeled by an infinite line with roadside infrastructure, e.g., RSUs, Wi-Fi APs or LTE base stations, regularly deployed with equal distance d. The width of a lane is typically small compared with the transmission range of vehicles. Therefore, we ignore the road width and model multiple lanes in the same direction as one lane [80, 81, 82]. We further assume that all infrastructure points are connected to the Internet through wired or wireless backbone with much larger capacity than the vehicular network.

We adopt a widely used traffic model in highways [22, 82, 83], such that in each direction (eastbound and westbound), the distribution of vehicles follows a homogeneous Poisson process with densities ρ_1 and ρ_2 respectively. It follows that the inter-vehicle distances in each direction are exponentially distributed. This exponential inter-vehicle spacing distribution has been supported by some empirical study that shows it can accurately characterize the real traffic distribution when the traffic density is low or medium [82]. Besides, vehicles in each direction travel at the same constant speed of v_1 and v_2 respectively [22, 82, 84]. In real networks, individual vehicular speed may deviate from the mean speed, for example, described by a Gaussian speed model [21, 85]. However, such deviations, which result in vehicle overtaking, have only a minor impact on the throughput being studied as shown later in our simulation. The system model is illustrated in Fig. 3.1.

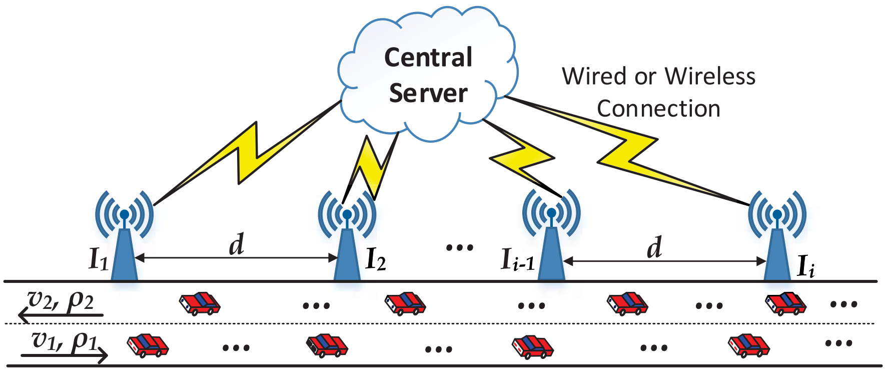


Figure 3.1 : An illustration of the system model for a bi-directional highway with infrastructure regularly deployed with equal distance d: single-user case.

3.1.2 Wireless Communication Model

Both V2I and V2V communications are considered. All infrastructure points have the same radio range, denoted by r_I; and all vehicles have the same radio range, denoted by r_0 where $r_I > r_0$, which reflects the fact that infrastructure typically has stronger communication capabilities. A pair of vehicles (or vehicle and infrastructure) can directly connect with each other if and only if (iff) their Euclidean distance is not larger than the radio range r_0 (or r_I) [47, 86]. This adopted unit disk model has been extensively used in the field [22, 81, 86, 87]. It grossly captures the fact that wireless communications only occur between nearby devices, namely, the closer two devices become, the easier (with higher probability) it is for them to establish a connection. This simplification allows us to omit physical layer details and focus on the topological impact of vehicular networks on the throughput, which is the main performance determining factor. We will show later in the simulation that the unit disk model assumption has little impact on the throughput.

We consider that each vehicle has a single antenna so that they can not transmit and receive at the same time [26, 54]. The antenna is omnidirectional so that the signal transmitted by a vehicle is able to reach all the potential receivers within its coverage. Besides, we adopt unicast transmission such that each infrastructure (or vehicle) can only transmit information to one vehicle at a time. Both unicast and broadcast are important in vehicular networks [88]. For some safety-related applications, e.g., disseminating a message about an accident on the road, or when some vehicles request common content, it is better to use broadcast (or multicast) to inform as many vehicles as possible. In addition to broadcast, unicast is also important and commonly adopted in vehicular networks to transmit data from a single source to a single destination [46, 89]. The scenario being considered in this chapter corresponds to a unicast scenario. Furthermore, it has been shown in [25] that whether the infrastructure transmits to one vehicle at a time, or divides its bandwidth among multiple users and transmits to multiple users at the same time, does not affect throughput calculation. Therefore, the unicast scheme we considered is immaterial to our analysis. In addition, as the major focus of this work is to investigate the impact of the topological aspects of the vehicular network on the achievable throughput, similar as [47, 82, 90], we ignore the packet loss issue. We refer readers to [91, 92] for relevant work on packet loss.

For the interference model, we assume that V2I and V2V communications are allocated different channels so that there is no mutual interference between them. In addition, we adopt the widely used Protocol Interference Model [93] where a transmitter cannot transmit if there are other transmitters transmitting within its interference range. In our work, the inter-infrastructure distance is large so that infrastructure can transmit simultaneously without causing any mutual interference. For the V2V communications, noting that the VoI is the only receiver of V2V communications, therefore, under the Protocol Interference Model, there will at most one transmitter transmitting its data to the VoI at a randomly chosen time instant, which implies that there will be no interference caused by other simultaneous transmitters. Our work assuming the Protocol Interference Model can be readily extended to another widely used Physical Interference Model (also known as the Signal-to-Interference-plus-Noise Ratio (SINR) Model) because it has been established in [93] that any spatio-temporal scheduling satisfying the Protocol Interference Model can also meet the requirement of the corresponding Physical Interference Model when some parameters, like the interference range, transmission power and the SINR threshold are appropriately selected. Furthermore, the MAC protocol associated with the Protocol Interference Model is the Carrier-Sense Multiple Access (CSMA) scheme. As we consider a single VoI only in this work, collisions, which are major concerns of the MAC protocol, have little impact on the achievable throughput of the VoI.

We assume V2I and V2V communicate at a constant data rate w_I and w_V respectively [25, 29, 64]. For time-varying channels, the values of w_I and w_V can be replaced by the respective time-averaged data rate of V2I and V2V communications and our analysis still applies. Indeed, analysis later in this chapter will show that depending on the relationship between w_I, w_V and the speeds of vehicles in both directions, the system can be classified into three regimes: one regime where the throughput is mainly limited by the data rate of V2I communications, i.e., w_I; one regime where the throughput is mainly restrained by the data rate of V2V communications, w_V; and another regime where the throughput is determined both by the data rate of V2I communications, w_I, and the data rate of V2V communications, w_V. In the meantime, only one-hop communications are considered. This can be explained by the fact that in the specific scenario being considered, there is only one vehicle with a download request (VoI), all other vehicles (helpers) assist the VoI to receive more data. Any new data in the vehicular network must come from the infrastructure. Therefore, allowing multi-hop V2V communications between the VoI and helpers, i.e., allowing V2V communications between helpers, only helps to balance the distribution of information among helpers but do not increase the net amount of information available in the network. Moreover, even though allowing more than one hop communication between the VoI and infrastructure is beneficial to the VoI's data downloading because it allows the VoI having longer connection time with the infrastructure, the improvement is expected to be marginal, especially when the traffic density is small, which has been verified by our simulation result as shown later.

3.1.3 Cooperative Communication Strategy

Now we introduce the cooperative communication strategy considered in this chapter. In particular, we consider a scenario where a VoI wants to download a large file, e.g., a video, from a remote server, and analyze the throughput that can be achieved by the VoI via a combined use of V2I communications, V2V communications, vehicular mobility and cooperations among vehicles and infrastructure.

The scenario being considered corresponds to a vehicular network where only a small number of vehicles have requests for large-file downloads. Another scenario that has been widely considered in the literature, commonly known as *the saturated traffic scenario*, considers that all vehicles have requests for download. The saturated traffic scenario is often used in analysing the capacity of the network [29, 64]. We point out that for the particular problem being considered, i.e., downloading a large file from a remote server, the saturated traffic scenario constitutes a trivial case and offer the following intuitive explanation for that. Note that when downloading files from a remote server, the new information (e.g., parts of the files) must come from the infrastructure points. V2V communications only help to balance the distribution of information among vehicles and do not increase the net amount of information available in the system. Therefore, when all vehicles have download requests, it can be easily established that the optimum strategy that maximizes the capacity is for each vehicle to download its requested file directly from the infrastructure. When only a single vehicle or a small number of vehicles have download requests, the situation becomes more intriguing. In this situation, other vehicles may help to retrieve information from the infrastructure when these vehicles enter into the coverage of their respective infrastructure points and then deliver the information to the VoI(s) outside the coverage of the infrastructure. In this way, the net amount of new information available in the system is boosted, therefore increasing the throughput (capacity) of the VoI(s).

As mentioned in the beginning of this subsection, the VoI wants to download a large file from the remote server. This requested large file may be first split into multiple pieces and transmitted to different infrastructure points so that each infrastructure point has a different piece of data, which enables cooperation among infrastructure. Data delivered to an infrastructure point may be further split and transmitted to either the VoI or helpers when they move into its coverage, so that helpers have different pieces of data from each other and from the VoI. Data received by the helpers will be transmitted to the VoI when they encounter the VoI, which exploits the mobility of vehicles and V2V communications to achieve vehicular cooperation. Since vehicles in the same direction move at the same constant speed, the inter-vehicle distances in the same lane remain the same at any time instant. This follows that vehicles in the same direction of the VoI can only offer limited help to the VoI because only vehicles within the coverage of the VoI can offer help. Therefore, in this work, we only consider vehicles in the opposite direction of the VoI that will receive data from infrastructure and can transmit the received data to the VoI as helpers.

To present the cooperative communication strategy, it suffices to consider two consecutive infrastructure points along the travel direction of the VoI collaborating to deliver data. We denote the nearest infrastructure point along the travel direction of the VoI by I_1 and the second nearest one by I_2. When the VoI is in the coverage of I_1, it receives data directly from I_1. In the meantime, the helpers may also receive different pieces of data from I_2 when they move through the coverage of I_2. When the VoI moves outside the coverage of I_1, it may continue to receive data from helpers using V2V communications. Of course, when the VoI moves along its direction, the two infrastructure points participating in the cooperative communication are also updated. In this way, V2I communications between the VoI and infrastructure, between helpers and their respective infrastructure points, V2V communications between the VoI and helpers, as well as vehicular mobility are coherently combined to maximize the throughput of the VoI. In our considered network scenario, V2I communications by both the VoI and helpers are essential to retrieve data from the infrastructure. V2V communications only help to assist the VoI to retrieve more data from the Internet and deliver the data retrieved by helpers to the VoI. V2V communications can not increase the net amount of data in the network. Furthermore, we consider that some practical issues like out of sequence data delivery can be handled by techniques such as network coding (e.g., [94]) so that we can focus on the main theme of the work without the need for considering their impacts.

3.1.4 Problem Formation

Now we give a formal definition of the throughput studied in this chapter. Consider an arbitrarily chosen time interval $[0, t]$ and denote the amount of data received by the VoI as $D(t)$, which includes data received from both infrastructure and helpers. In this work, we are interested in finding the long-term achievable throughput of the VoI, using our cooperative communication strategy, where the long-term throughput, denoted by η, is formally defined as follows:

$$\eta = \lim_{t \to \infty} \frac{D(t)}{t}. \tag{3.1}$$

Without loss of generality, we assume that the VoI travels at speed v_1, and the helpers travel at speed v_2 and have vehicular density ρ_2. We define the time interval starting from the time instant when the VoI enters into the coverage of one infrastructure point to the time instant when the VoI enters into the coverage of the next infrastructure point as *one cycle*. By using cycles as the basic blocks, the entire data receiving process of the VoI can be modelled by a renewal reward process [95]. Each cycle in the renewal reward process consists of one V2I communication process, followed by a V2V communication process, and the reward is the amount of data received by the VoI during each cycle. This follows that the throughput can be calculated as follows:

$$\eta = \lim_{t\to\infty} \frac{D(t)}{t} = \frac{E[D_I] + E[D_V]}{E[T]}, \tag{3.2}$$

where $E[D_I]$ and $E[D_V]$ are respectively the expected amount of data received by the VoI during the V2I communication process and the V2V communication process in one cycle, and $E[T]$ is the expected time of one cycle, which can be calculated as $E[T] = \frac{d}{v_1}$. Since when the VoI is covered by an infrastructure point, it will only use V2I communication, $E[D_I]$ can be readily obtained as follows:

$$E[D_I] = \frac{2r_I w_I}{v_1}. \tag{3.3}$$

Using (3.2) and (3.3), the problem of calculating the achievable throughput by the VoI can be transformed into the problem of calculating the expected amount of data that can be received by the VoI from V2V communications in one cycle. Without loss of generality, we call the two infrastructure points I_1 and I_2 respectively as defined earlier. Because of the unicast transmission model we adopt, during V2V communications between the VoI and helpers, the amount of data two adjacent helpers can deliver to the VoI become correlated when their inter-vehicle distance is smaller than $2r_0$. Meanwhile, the amount of delivered data is further limited by the amount of data each helper receives from I_2, which can also be correlated because during helpers' V2I communications, the amount of data received by adjacent helpers become correlated when their inter-vehicle distance is smaller than $2r_I$. This complicated correlation structure is quite intricate for statistical analysis. In this work, we handle the challenge by formulating the V2V data delivering process in one cycle as a constrained optimization problem, with the goal of obtaining the maximum amount of data received by the VoI from helpers and finding the corresponding scheduling scheme, which includes a V2I transmission scheme for helpers and a V2V transmission scheme, to reach this maximum value. In the following, we will show the formation of the constrained optimization problem.

Denote by n the number of helpers encountered by the VoI during a cycle and n is a Poissonly distributed random integer. Denote by V_1 the first helper encountered by the VoI when the VoI moves outside the coverage of I_1, by V_2 the second helper, and so on. Denote the distance between two consecutive helpers V_i and V_{i+1} by $l_i, i = 1, ...n-1$. See Fig. 3.2 for an illustration. Furthermore, denote by D_i the amount of data received by helper V_i from I_2, and denote by Y_i the amount of data delivered by V_i to the VoI. We first consider the situation that n is a fixed integer and $l_i, i = 1, ...n-1$ are known values, i.e., corresponding to a specific instance of these random values, and then extend to consider the more general situation that n and $l_i, i = 1, ...n-1$ are random values.

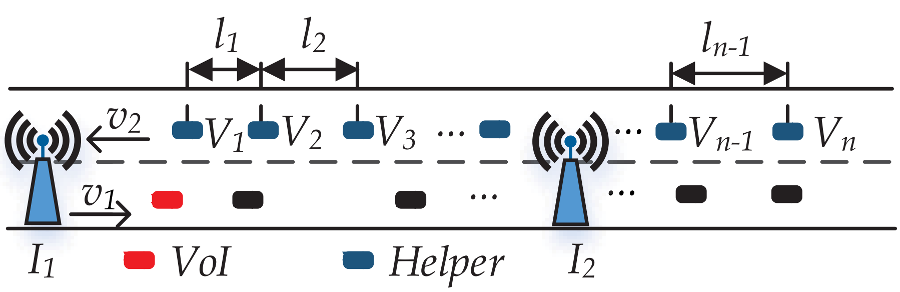


Figure 3.2 : An illustration of helpers encountered by the VoI during one V2V communication cycle and their interval distance.

Without considering the boundary case, caused by helpers located near the borders of the coverage area of infrastructure points, the problem of finding the maximum amount of data received by the VoI from V2V communications in one cycle, given n and $l_i, i = 1, ...n-1$, can be formulated as the following optimization problem. We will show later in the simulation that the boundary case has negligible impact on the achievable throughput as we are focusing on the long-term throughput. The optimization is taken over the set of all possible scheduling schemes. Without causing any confusion, we drop the notation for the set of all possible scheduling schemes to have a simpler expression.

$$\max \sum_{i=1}^{n} Y_i \tag{3.4}$$

$$s.t. 0 \leq D_i \leq \frac{2r_I}{v_2} w_I,\ i = 1, 2, ...n \tag{3.5}$$

$$\sum_{i=k_1}^{k_2} D_i \leq \frac{\sum_{i=k_1}^{k_2-1} \min\{l_i, 2r_I\} + 2r_I}{v_2} w_I,\ 1 \leq k_1 \leq k_2 \leq n \tag{3.6}$$

$$0 \leq Y_i \leq \frac{2r_0}{v_1 + v_2} w_V, \quad i = 1, 2, ...n \tag{3.7}$$

$$Y_i \leq D_i, \quad i = 1, 2, ...n \tag{3.8}$$

$$\sum_{i=k_1}^{k_2} Y_i \leq \frac{\sum_{i=k_1}^{k_2-1} \min\{l_i, 2r_0\} + 2r_0}{v_1 + v_2} w_V,\ 1 \leq k_1 \leq k_2 \leq n \tag{3.9}$$

In the above optimization problem, $\sum_{i=1}^{n} Y_i$ is the total amount of data received by the VoI from helpers during one cycle. Constraint (3.5) gives the maximum and the minimum amount of data received by each helper from infrastructure point I_2. Constraint (3.6) gives an upper bound on the amount of data that any $k_2 - k_1 + 1$ consecutive helpers can receive from I_2, where the term $\frac{\sum_{i=k_1}^{k_2-1} \min\{l_i, 2r_I\} + 2r_I}{v_2}$ gives the total amount of time these $k_2 - k_1 + 1$ consecutive helpers can receive data from I_2. Particularly, due to the randomness of vehicle distributions, it may happen that there exists a void region of larger than $2r_I$, which has no vehicle (helper). When the void region occurs, helpers may not be able to receive data continuously from

I_2. Therefore, all two constraints (3.5) and (3.6) must be considered to completely describe the V2I communication between I_2 and helpers. Constraint (3.6) also captures the correlation that may occur during the data receiving process of adjacent helpers, which has been explained earlier. Similarly, constraint (3.7) gives the maximum and minimum amount of data that can be received by the VoI from each helper. Constraint (3.8) implies that the amount of data each helper can deliver to the VoI cannot exceed the data it receives from I_2. Constraint (3.9) gives the upper bound of the amount of data the VoI can receive from any $k_2 - k_1 + 1$ consecutive helpers.

3.2 Analysis of V2V Communication Process and Achievable Throughput

The data received by the VoI, whether through V2I communications directly from the infrastructure or through V2V communications from helpers, ultimately comes from the infrastructure. Intuitively, as we increase the data rate of V2I communications, w_I, from a very small value while keeping other parameters constant, initially the throughput will be limited by the data rate of V2I communications. We call this regime the *Infrastructure-Limited Regime*. As we further increase the value of w_I, we will reach a *Transitional Regime* where both the data rate of V2I communications and the data rate of V2V communications play major roles in determining the throughput of the VoI. If we increase the value of w_I further to a very large value, V2I communications will no longer be a bottleneck in determining the throughput of the VoI. Instead, the data rate of V2V communication, w_V, becomes the determining factor of the VoI throughput. We call this regime the *V2V-Limited Regime*.

It is evident from the optimization problem (3.4) that the amount of data received from V2V communications by the VoI given fixed n and $l_i, i = 1, ...n$, $\sum_{i=1}^{n} Y_i$, satisfies:

$$\sum_{i=1}^{n} Y_i \leq \min\{D_{Vu1}, D_{Vu2}\} \tag{3.10}$$

where $D_{Vu1} = \frac{\sum_{i=1}^{n-1}\min\{l_i, 2r_I\}+2r_I}{v_2} w_I$ comes from a combination of constraints (3.6) and (3.8), representing the maximum amount of data all helpers can receive from infrastructure; and $D_{Vu2} = \frac{\sum_{i=1}^{n-1}\min\{l_i, 2r_0\}+2r_0}{v_1+v_2} w_V$ comes from constraint (3.9), representing the maximum amount of data all helpers can deliver to the VoI through V2V communications without considering the limitation of the amount of data they receive. When $0 < w_I \leq \frac{r_0 w_V v_2}{r_I(v_1+v_2)}$, we have $D_{Vu1} \leq D_{Vu2}$, which implies that the amount of data the VoI can receive from helpers is limited by the the amount of data helpers can receive through their V2I communications, thus limited by w_I. Thus, we define the *Infrastructure-Limited Regime*: $0 < w_I \leq \frac{r_0 w_V v_2}{r_I(v_1+v_2)}$. Similarly, when $w_I \geq \frac{w_V v_2}{v_1+v_2}$, we have $D_{Vu1} > D_{Vu2}$, which implies that the amount of data the VoI can receive from helpers is limited by the amount of data the helpers can deliver through V2V communications, thus limited by w_V. Therefore, we define the *V2V-Limited Regime*: $w_I \geq \frac{w_V v_2}{v_1+v_2}$. The rest of the region forms the *Transitional Regime*: $\frac{r_0 w_V v_2}{r_I(v_1+v_2)} < w_I < \frac{w_V v_2}{v_1+v_2}$.

In the following subsections, we analyse the achievable throughput by the VoI under each regime separately.

3.2.1 Infrastructure-Limited Regime

In this subsection, we first analyze the maximum amount of data that can be received from helpers by the VoI in one cycle by solving the optimization problem (3.4), and find the corresponding scheduling scheme to achieve this maximum solution given fixed n and $l_i, i = 1, ...n-1$. Then, we extend to consider that n and $l_i, i = 1, ...n-1$ are random values, corresponding the Poisson distribution of vehicles, and analyse the achievable throughput under the obtained optimal scheduling scheme.

The following theorem summarizes the major result of V2V communication process analysis.

Theorem 3.1. *In the Infrastructure-Limited regime, given fixed* n *and* $l_i, i = 1, 2, ...n-1$*, the maximum amount of data the VoI can receive from all* n *helpers in one cycle is given by*

$$\left(\sum_{i=1}^{n} Y_i\right)_1^* = \frac{\sum_{i=1}^{n-1}\min\{l_i, 2r_I\} + 2r_I}{v_2} w_I \tag{3.11}$$

where $(\sum_{i=1}^{n} Y_i)_1^*$ *is the respective* $\sum_{i=1}^{n} Y_i$ *associated with its optimum value and we use the subscript* 1 *to mark the Infrastructure-Limited regime and superscript* $*$ *to mark the optimum value.*

Furthermore, there exists a V2I transmission scheme for helpers and a V2V transmission scheme to reach the above maximum amount of received data for the VoI, satisfying:

$$\begin{cases} Y_{1i}^* = D_{1i}^* = \frac{\min\{l_i, 2r_I\}}{v_2} w_I, \quad i = 1, 2, ...n-1 \\ Y_{1n}^* = D_{1n}^* = \frac{2r_I}{v_2} w_I \end{cases} \tag{3.12}$$

where D_{1i}^* *and* $Y_{1i}^*, i = 1, ..., n$ *are the respective* D_i *and* $Y_i, i = 1, ..., n$ *associated with the optimum solution.*

Proof. In the Infrastructure-Limited regime, with conditions $w_I \leq \frac{r_0 w_V v_2}{r_I(v_1+v_2)}$ and $r_I > r_0$, we have $\frac{2r_I}{v_2} w_I \leq \frac{2r_0}{v_1+v_2} w_V$ and $\frac{\min\{l_i, 2r_I\}}{v_2} w_I \leq \frac{\min\{l_i, 2r_0\}}{v_1+v_2} w_V, i = 1, ...n-1$. It follows that:

$$\frac{\sum_{i=1}^{n-1}\min\{l_i, 2r_I\} + 2r_I}{v_2} w_I \leq \frac{\sum_{i=1}^{n-1}\min\{l_i, 2r_0\} + 2r_0}{v_1 + v_2} w_V \tag{3.13}$$

The above equation implies that in the optimization problem (3.4), constraint (3.7) is redundant and constraint (3.9) can be replaced with a tighter constraint after merging the two constraints (3.6) and (3.8). The new constraints for the optimization problem (3.4) under Infrastructure-Limited regime are shown as follows:

$$0 \le Y_i \le D_i \le \frac{2r_I}{v_2} w_I, \quad i = 1, 2, ...n \tag{3.14}$$

$$\sum_{i=k_1}^{k_2} Y_i \le \sum_{i=k_1}^{k_2} D_i \le \frac{\sum_{i=k_1}^{k_2-1} \min\{l_i, 2r_I\} + 2r_I}{v_2} w_I, \ 1 \le k_1 \le k_2 \le n \tag{3.15}$$

Constraint (3.15) indicates that an upper bound of $\sum_{i=1}^{n} Y_i$ can be expressed by $\frac{\sum_{i=1}^{n-1} \min\{l_i, 2r_I\}+2r_I}{v_2} w_I$. In the following, we will show that this upper bound is exactly the optimum solution of $\sum_{i=1}^{n} Y_i$ in the Infrastructure-Limited regime and can be reached under some scheduling scheme.

Noting that the upper bound of $\sum_{i=1}^{n} Y_i$, $\frac{\sum_{i=1}^{n-1} \min\{l_i, 2r_I\}+2r_I}{v_2} w_I$, is the sum of n separate components, with each component smaller than or equal to $\frac{2r_I}{v_2} w_I$. Therefore, when each Y_i is equal to D_i, and is further equal to the corresponding component forming $\frac{\sum_{i=1}^{n-1} \min\{l_i, 2r_I\}+2r_I}{v_2} w_I$, i.e., when each Y_i and each $D_i, i = 1, ...n$ are given by (3.12), the value of $\sum_{i=1}^{n} Y_i$ will reach its upper bound $\frac{\sum_{i=1}^{n-1} \min\{l_i, 2r_I\}+2r_I}{v_2} w_I$ while satisfying the constraints in the optimization problem (3.4). This leads to the expression of (3.11).

It remains to demonstrate that there exists a scheduling scheme to reach this optimum solution specified in (3.11). To this end, we show that (3.12) readily leads to the design of an optimal transmission scheme. Specifically, a scheduling scheme which schedules both V2I and V2V transmissions on a first-in-first-out (FIFO) basis can achieve the optimum solution (3.11). We acknowledge that the optimum scheduling algorithm that achieves the optimum solution (3.11) may not be unique. When other performance metrics are considered, e.g., delay, the earliest deadline first scheme may have better delay performance while achieving the same throughput. Particularly, in the FIFO scheduling scheme, each helper starts its V2I communication once it enters the coverage of infrastructure point I_2 *and* there is no other helper preceding it communicating with the infrastructure point I_2, and stops when the helper leaves the coverage of the infrastructure point I_2, which lead to that each helper will receive an amount of data shown as each $D^*_{1i}, i = 1, 2, ...n$ in Eq. (3.12). Similarly, for V2V communications, the VoI receives data from each helper one by one when there exists at least one helper within its coverage on a FIFO basis. Once a helper starts to deliver its data to the VoI, it will stop until it has transmitted all its data to the VoI *or* when it leaves the coverage of the VoI, which leads to the case that the data the VoI receives from each helper is shown as each $Y^*_{1i}, , i = 1, 2, ...n$ in Eq. (3.12). Noting that Eq. (3.12) leads to the optimum solution (3.11), it can be readily established that the aforementioned scheduling scheme achieves the maximum amount of received data for the VoI specified in (3.11). □

Remark 3.1. *Note that* (3.11) *and the corresponding scheduling scheme that satisfies* (3.12) *are valid for any value of* n *and the corresponding* $l_i, i = 1, ...n - 1$.

On the basis of Theorem 3.1, we now analyse the achievable throughput by the VoI considering that both n and the corresponding $l_i, i = 1, ...n - 1$ are random values. A brute force approach of computing the achievable throughput will first consider that n is a Poisson random variable, then conditioned on each value of n (noting that conditional on a specific instance of n, helpers become uniformly distributed and hence l_is, $i = 1, ...n - 1$, become correlated), evaluate the joint distribution of the random variables min $\{l_i, 2r_I\}, i = 1, ...n-1$, and finally transform the conditional value into an unconditional one using the total probability theorem and the Poisson distribution of n. This will result in a very complicated analysis. In the following, we use simpler techniques by resorting to the concept of *clusters*, defined later in the text, to analyse the achievable throughput.

We designate the time instant when the VoI leaves the coverage of I_1 as $t = 0$ and define its moving direction as the positive (right) direction of the coordinate system. Furthermore, we define the point to the right of I_1 and at a distance $r_I - r_0$ to I_1 as the origin of the coordinate system. It follows from the above that the time instant when the VoI enters into I_2's coverage will be $t_1 = \frac{d-2r_I}{v_1}$. Noting that the relative speed of the VoI to the helpers travelling in the opposite direction is $v_1 + v_2$, therefore the relative distance travelled by the VoI, relative to the helpers in the opposite direction which all travel at the same constant speed of v_2 during $[0, t_1]$, is given by $\frac{(d-2r_I)}{v_1}(v_1 + v_2)$. The random number of helpers encountered by the VoI, who may deliver data to the VoI, during $[0, t_1]$, is determined by the parameter s:

$$s = \frac{(d - 2r_I)(v_1 + v_2)}{v_1} + r_0, \tag{3.16}$$

where the r_0 term is due to the consideration that when the VoI exits the coverage of I_1 and is located at coordinate r_0 (and at time instant $t = 0$), the *helper(s)* to the left of the VoI and within a distance r_0 to the VoI may possibly deliver data to the VoI too. Thus, all helpers in the opposite direction that the VoI may encounter during its V2V communication process in one cycle are within road segment $[0, s]$.

As explained in the beginning of this subsection, we use the concept of *clusters* to simplify our analysis. A *cluster* is defined as a maximal set of helpers located within road segment $[0, s]$ *and* the distance between any two adjacent helpers is smaller than or equal to $2r_I$. Forming clusters in this way allows us to remove the complexity associated with the computation of the joint distribution of $\min\{l_i, 2r_I\}, i = 1, ...n-1$ because within each cluster, we have $\min\{l_i, 2r_I\} = l_i, i = 1, 2, ...$ For each cluster, we only need to focus on the length of the cluster rather than the individual inter-vehicle distances. There may be multiple clusters within road segment $[0, s]$ and a cluster may contain a single vehicle only. Denote the coordinate of the first helper that can transmit data to the VoI since $t = 0$ by l_0. Due to the memoryless property of the exponential distribution of inter-vehicle distances, l_0 has the same exponential distribution as other l_is, $i = 1, ...n - 1$ and the starting position of a vehicle in a cluster does not affect the distribution of the length of the cluster. Denote by K_1 the random non-negative integer representing the number of clusters the VoI will encounter in one cycle. Furthermore, denote by $L_j^{(1)}, j = 1, ..., K_1$, the length of each cluster, which are identically and independently distributed (i.i.d), and by $g_j^{(1)}, j = 1, ...K_1$ the length of each gap between two adjacent clusters, which are also i.i.d. See Fig. 3.3 for an illustration.

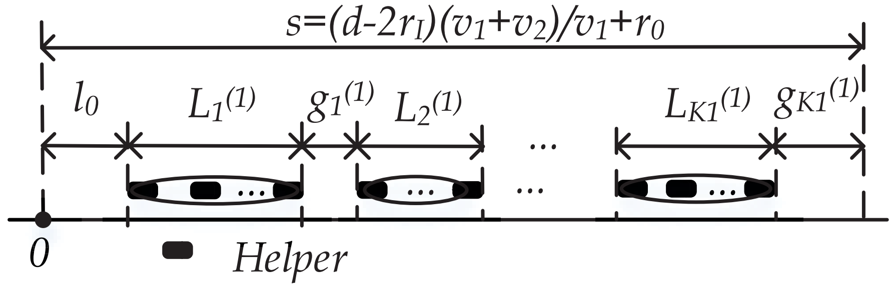


Figure 3.3 : An illustration of clusters formed by the helpers.

Noting that (3.11) is also valid for any subset of helpers within road segment $[0, s]$ adopting the scheduling scheme described in the proof of Theorem 3.1, therefore the amount of data each cluster of helpers delivers to the VoI, denoted by $R_j^{(1)}, j = 1, ...K_1$, can be obtained as follows (recall that the analysis is conducted for the Infrastructure-Limited regime):

$$R_j^{(1)} = \frac{L_j^{(1)} + 2r_I}{v_2} w_I, j = 1, ...K_1 \tag{3.17}$$

It follows that the amount of data received by the VoI from helpers in one cycle, denoted by D_{V1}, can be readily calculated by summing the amount of data received by the VoI from each cluster of helpers:

$$D_{V1} = \sum_{j=1}^{K_1} R_j^{(1)} = \sum_{j=1}^{K_1} \frac{L_j^{(1)} + 2r_I}{v_2} w_I. \tag{3.18}$$

Noting that in (3.18), both the number of clusters in road segment $[0, s]$, K_1, and the length of each cluster, $L_j^{(1)}$, are random variables, and are not independent. If we *approximately* consider they are independent with each other, then from (3.18), the expected amount of data received by the VoI from V2V communications in one cycle, $E[D_{V1}]$, can be calculated as follows:

$$E[D_{V1}] = E[K_1] \cdot \frac{E[L_j^{(1)}] + 2r_I}{v_2} w_I, \tag{3.19}$$

where $E[K_1]$ is the expected number of clusters in the road segment $[0, s]$. The accuracy of this approximation is verified by simulation, see Fig. 3.4 below. It shows that the approximation will marginally increase the result of $E[D_{V1}]$, which in turn, will lead to a marginally increase in the achievable throughput in this regime.

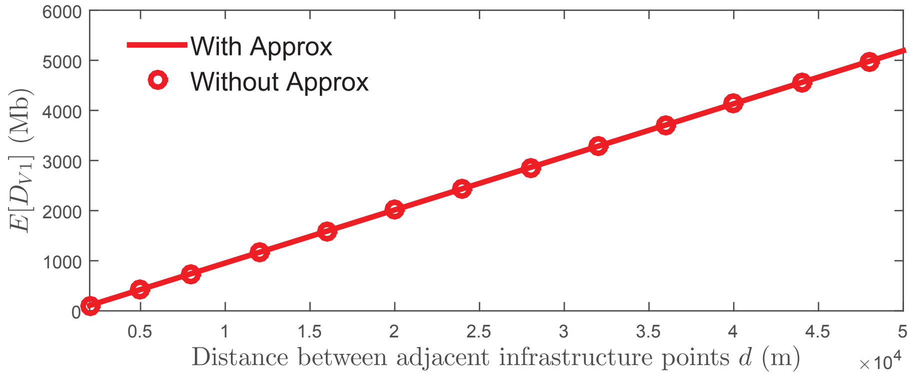


Figure 3.4 : A comparison between the result of $E[D_{V1}]$ with and without the approximation.

As $L_j^{(1)}, j = 1, ...K_1$ and $g_j^{(1)}, j = 1, ...K_1$ are both i.i.d and each $L_j^{(1)}$ and $g_j^{(1)}$ are also mutually independent, then according to the Generalized Wald's equality [95, Theorem 4.5.2], when $s \gg E[L_j^{(1)}] + E[g_j^{(1)}]$, $E[K_1]$ can be approximately calculated as follows:

$$E[K_1] = \frac{s - E[l_0]}{E[L_j^{(1)}] + E[g_j^{(1)}]}. \tag{3.20}$$

By putting (3.20) into (3.19), we have

$$E[D_{V1}] = \frac{s - E[l_0]}{E[L_j^{(1)}] + E[g_j^{(1)}]} \times \frac{E[L_j^{(1)}] + 2r_I}{v_2} w_I, \tag{3.21}$$

where the values of $E[L_j^{(1)}]$ and $E[g_j^{(1)}]$ have been given by [82]:

$$E[L_j^{(1)}] = \left(e^{2\rho_2 r_I} - 1\right) \left(\frac{1}{\rho_2} - \frac{2r_I e^{-2\rho_2 r_I}}{1 - e^{-2\rho_2 r_I}}\right), \tag{3.22}$$

and

$$E[g_j^{(1)}] = 2r_I + \frac{1}{\rho_2}. \tag{3.23}$$

As mentioned before, due to the memoryless property of exponential distribution, l_0 has the same distribution as l_i [96], i.e., we have:

$$E[l_0] = \frac{1}{\rho_2}. \tag{3.24}$$

Combining (3.21)-(3.24), we can obtain:

$$E[D_{V1}] = \frac{\left[\frac{(d-2r_I)(v_1+v_2)}{v_1} + r_0 - \frac{1}{\rho_2}\right](1 - e^{-2\rho_2 r_I})\, w_I}{v_2}. \tag{3.25}$$

By plugging equations (3.3) and (3.25) into (3.2), we have the achievable throughput in the Infrastructure-Limited regime, denoted by η_1, as follows:

$$\eta_1 = \frac{2r_I w_I + c_1}{d}, \tag{3.26}$$

where

$$c_1 = \frac{\left[(d-2r_I)(v_1+v_2) + r_0 v_1 - \frac{v_1}{\rho_2}\right](1 - e^{-2\rho_2 r_I})\, w_I}{v_2}.$$

3.2.2 V2V-Limited Regime

Now we analyse the achievable throughput in the V2V-Limited regime. Similar to subsection 3.2.1, in this subsection, we first analyse the maximum amount of data that can be received from helpers by the VoI in one cycle by solving the optimization problem (3.4), and then find the corresponding scheduling scheme to achieve this maximum solution given fixed n and $l_i, i = 1, \ldots n-1$. Finally, we extend to consider that n and $l_i, i = 1, \ldots n-1$ are random values, corresponding to Poisson distribution of vehicles, and analyse the achievable throughput under the proposed scheduling scheme.

The following theorem summarizes the main result of V2V communication process analysis.

Theorem 3.2. *In the V2V-Limited regime, given fixed* n *and* $l_i, i = 1, ...n-1$*, the maximum amount of data the VoI can receive from all* n *helpers in one cycle is given by*

$$\left(\sum_{i=1}^{n} Y_i\right)_2^* = \frac{\sum_{i=1}^{n-1} \min\{l_i, 2r_0\} + 2r_0}{v_1 + v_2} w_V, \tag{3.27}$$

where $(\sum_{i=1}^{n} Y_i)_2^*$ *is the respective* $\sum_{i=1}^{n} Y_i$ *associated with its optimum value and we use the subscript* 2 *to mark the V2V-Limited regime.*

Furthermore, there exists a V2I transmission scheme for helpers and a V2V transmission scheme to reach the above maximum amount of received data for the VoI, satisfying:

$$\begin{cases} D_{2i}^* = \frac{\min\{l_i, 2r_I\}}{v_2} w_I, i = 1, 2, ...n-1 \\ D_{2n}^* = \frac{2r_I}{v_2} w_I \\ Y_{2i}^* = \frac{\min\{l_i, 2r_0\}}{v_1+v_2} w_V, i = 1, 2...n-1 \\ Y_{2n}^* = \frac{2r_0}{v_1+v_2} w_V \end{cases} \tag{3.28}$$

where D_{2i}^* *and* $Y_{2i}^*, i = 1, ..., n$ *are the respective* D_i *and* $Y_i, i = 1, ..., n$ *associated with the optimum solution.*

Proof. In the V2V-Limited regime, with conditions $w_I \geq \frac{w_V v_2}{v_1+v_2}$ and $r_I > r_0$, we have $\frac{2r_0}{v_1+v_2} w_V < \frac{2r_I}{v_2} w_I$ and $\frac{\min\{l_i, 2r_0\}}{v_1+v_2} w_V \leq \frac{\min\{l_i, 2r_I\}}{v_2} w_I, i = 1, ...n-1$. It follows that:

$$\frac{\sum_{i=1}^{n-1} \min\{l_i, 2r_0\} + 2r_0}{v_1 + v_2} w_V < \frac{\sum_{i=1}^{n-1} \min\{l_i, 2r_I\} + 2r_I}{v_2} w_I, \tag{3.29}$$

Then from constraints (3.6), (3.8), (3.9) and inequality (3.29), we can conclude that the value of $\sum_{i=1}^{n} Y_i$ in optimization problem (3.4) is upper bound by $\frac{\sum_{i=1}^{n-1} \min\{l_i, 2r_0\} + 2r_0}{v_1+v_2} w_V$. In the following, we will show that this upper bound is exactly the optimum solution of $\sum_{i=1}^{n} Y_i$ in the V2V-Limited regime and can be reached under some scheduling scheme.

Noting that $\frac{\sum_{i=1}^{n-1} \min\{l_i, 2r_0\} + 2r_0}{v_1+v_2} w_V$ is the sum of n separate components, with each component is not larger than $\frac{2r_0}{v_1+v_2} w_V$. Therefore, when each Y_i is equal to the corresponding part of the n component forming the upper bound of $\sum_{i=1}^{n} Y_i$, $\frac{\sum_{i=1}^{n-1} \min\{l_i, 2r_0\} + 2r_0}{v_1+v_2} w_V$, and each D_i is equal to the corresponding part of the n component forming the upper bound of $\sum_{i=1}^{n} D_i$ shown in (3.6), i.e., when each Y_i and $D_i, i = 1, ...n$ are given by (3.28), the value of $\sum_{i=1}^{n} Y_i$ will reach its upper bound $\frac{\sum_{i=1}^{n-1} \min\{l_i, 2r_0\} + 2r_0}{v_1+v_2} w_V$ while satisfying all constraints in optimization problem (3.4). This leads to expression of (3.27).

Now we show that there exists a scheduling scheme to reach the maximum solution specified in (3.27). To this end, we show that (3.28) readily leads to the design of the scheduling scheme. Particularly, the scheduling scheme schedules both helpers' V2I communication and V2V communications on a FIFO basis. Specifically, the V2I transmission scheme for helpers is the same as that for the Infrastructure-Limited regime, leading to the case that each helper will receive an amount of data shown as each $D_{2i}^*, i = 1, 2, ...n$ in Eq. (3.28). For V2V communications, the VoI starts to receive data from a helper once it enters this helper's coverage *and* has retrieved all data from the previous helper or has left the previous helper's coverage, and stops when the VoI leaves the coverage of the current helper or has retrieved all data of the current helper, leading to the case that the data the VoI receives from each helper is shown as each $Y_{2i}^*, i = 1, 2, ...n$ in Eq. (3.28). Noting that Eq. (3.28) leads to the optimum solution (3.27), it can be readily established that the aforementioned scheduling scheme achieves the maximum amount of the received data for the VoI specified in (3.27). This completes the proof. $\square$

Similarly as Theorem 3.1, Theorem 3.2 is also valid for any value of n and the corresponding $l_i, i = 1, ...n-1$.

On the basis of Theorem 3.2, we now analyze the achievable throughput by the VoI considering that both n and the corresponding $l_i, i = 1, ...n-1$ are random values. In the V2V-Limited regime, we define a *cluster* to be a maximal set of helpers located within road segment $[0, s]$ *and* the distance between any two adjacent helpers is smaller than or equal to $2r_0$. The reason that we define the clusters differently from that in the Infrastructure-Limited regime is that in this regime, it is the correlation in the V2V communication process (and the associated difficulty in determining the joint distribution of $\min\{l_i, 2r_0\}, i = 1, ...n-1$) that plays a dominating effect on determining the achievable throughput. By defining clusters in the above way, within each cluster, we have $\min\{l_i, 2r_0\} = l_i, i = 1, 2, ...$

In the V2V-Limited regime, the amount of data each cluster of helpers delivers to the VoI, denoted by $R_j^{(2)}$, can be calculated as follows:

$$R_j^{(2)} = \frac{L_j^{(2)} + 2r_0}{v_1 + v_2} w_V, j = 1, ...K_2 \tag{3.30}$$

where $L_j^{(2)}, j = 1, ...K_2$ is the random variable representing the length of the j-th cluster, and K_2 is the random integer representing the number of clusters the VoI will encounter in one cycle.

Utilizing the same approximation method as that used to calculate $E[D_{V1}]$ in the Infrastructure-Limited regime, i.e., approximately consider that K_2 and $L_j^{(2)}$ are independent in this regime, the expected amount of data received by the VoI from helpers in one cycle in the V2V-Limited regime can be obtained as follows:

$$E[D_{V2}] = \frac{s - E[l_0]}{E[L_j^{(2)}] + E[g_j^{(2)}]} \times \frac{E[L_j^{(2)}] + 2r_0}{v_1 + v_2} w_V, \tag{3.31}$$

where $E[L_i^{(2)}]$ and $E[g_i^{(2)}]$ are given by:

$$E[L_j^{(2)}] = \left(e^{2\rho_2 r_0} - 1\right)\left(\frac{1}{\rho_2} - \frac{2r_0 e^{-2\rho_2 r_0}}{1 - e^{-2\rho_2 r_0}}\right), \tag{3.32}$$

and

$$E[g_j^{(2)}] = 2r_0 + \frac{1}{\rho_2}. \tag{3.33}$$

Combing (3.24) and (3.31)-(3.33), we have:

$$E[D_{V2}] = \frac{\left[\frac{(d-2r_I)(v_1+v_2)}{v_1} + r_0 - \frac{1}{\rho_2}\right]\left(1 - e^{-2\rho_2 r_0}\right) w_V}{v_1 + v_2}. \tag{3.34}$$

By plugging (3.3), (3.34) into (3.2), the achievable throughput in the V2V-Limited regime, denoted by η_2, can be obtained as follows:

$$\eta_2 = \frac{2r_I w_I + c_2}{d}, \tag{3.35}$$

where

$$c_2 = \frac{\left[(d - 2r_I)(v_1 + v_2) + r_0 v_1 - \frac{v_1}{\rho_2}\right]\left(1 - e^{-2\rho_2 r_0}\right) w_V}{v_1 + v_2}.$$

3.2.3 Transitional Regime

Now we analyse the achievable throughput in the transitional regime where the analysis is more intricate than that for the Infrastructure-Limited and the V2V-Limited regime. Particularly, in the transitional regime, both V2V communications and helpers' V2I communications contribute to determining the achievable throughput of the VoI. Therefore both the correlation in the amount of data received by adjacent helpers from infrastructure and in the amount of data received by the VoI from adjacent helpers, as explained in Section 3.1.4, need to be considered. This makes finding the optimum solution for the optimization problem (3.4) more challenging. Therefore, in this subsection, instead of analysing the exact achievable throughput, we analyse its upper and lower bound. In the following, we will analyse the upper and the lower bound of the achievable throughput separately.

Upper bound of the achievable throughput

As shown in (3.10), an upper bound of $\sum_{i=1}^{n} Y_i$ is given by:

$$\sum_{i=1}^{n} Y_i \leq \min\{D_{Vu1}, D_{Vu2}\} \tag{3.36}$$

That is, the upper bound of data amount received by the VoI from helpers is determined by the smaller value of the amount of data received by the helpers from their V2I communications, D_{Vu1}, and the amount of data helpers can deliver to the VoI in V2V communications (without considering the limitation of the amount of data they receive), D_{Vu2}. It is shown in Theorem 3.1 and Theorem 3.2 that D_{Vu1} and D_{Vu2} are exactly the maximum amount of data the VoI can receive from helpers in the Infrastructure-Limited regime and V2V-Limited regime respectively. Therefore, according to the throughput calculation analysis given in subsection 3.2.1 and 3.2.2, when $\sum_{i=1}^{n} Y_i$ is upper bounded by D_{Vu1} (or D_{Vu2}), the corresponding achievable throughput of the VoI will be upper bounded by the achievable throughput in the Infrastructure-Limited regime, η_1, (or the achievable throughput in the V2V-Limited regime, η_2). It follows that an upper bound of the achievable throughput by the VoI in the transitional regime, denoted by η_{3u}, is given by:

$$\eta_{3u} = \min\{\eta_1, \eta_2\} = \begin{cases} \eta_1, & \eta_1 \leq \eta_2 \\ \eta_2, & \eta_1 > \eta_2 \end{cases} \tag{3.37}$$

Putting (3.26) and (3.35) into (3.37) and simplifying, we have:

$$\eta_{3u} = \begin{cases} \frac{2r_I w_I + c_1}{d}, & \frac{r_0 w_V v_2}{r_I(v_1+v_2)} < w_I \leq \frac{1-e^{-2\rho_2 r_0}}{1-e^{-2\rho_2 r_I}} \times \frac{w_V v_2}{v_1+v_2} \\ \frac{2r_I w_I + c_2}{d}, & \frac{1-e^{-2\rho_2 r_0}}{1-e^{-2\rho_2 r_I}} \times \frac{w_V v_2}{v_1+v_2} < w_I < \frac{w_V v_2}{v_1+v_2} \end{cases} \tag{3.38}$$

with c_1 and c_2 as given in the earlier analysis.

Remark 3.2. *Equation* (3.38) *shows that* $\frac{1-e^{-2\rho_2 r_0}}{1-e^{-2\rho_2 r_I}} \times \frac{w_V v_2}{v_1+v_2}$ *is a transition point for the value of* w_I *to determine the upper bound of achievable throughput in the transitional regime, whose value depends on the helpers' density* ρ_2*. Specifically, when* $\rho_2 \to 0$, $\frac{1-e^{-2\rho_2 r_0}}{1-e^{-2\rho_2 r_I}} \times \frac{w_V v_2}{v_1+v_2} \to \frac{r_0 w_V v_2}{r_I(v_1+v_2)}$*; when* ρ_2 *increases, the gap between* $\frac{1-e^{-2\rho_2 r_0}}{1-e^{-2\rho_2 r_I}} \times \frac{w_V v_2}{v_1+v_2}$ *and* $\frac{r_0 w_V v_2}{r_I(v_1+v_2)}$ *becomes larger and the gap between* $\frac{1-e^{-2\rho_2 r_0}}{1-e^{-2\rho_2 r_I}} \times \frac{w_V v_2}{v_1+v_2}$ *and* $\frac{w_V v_2}{v_1+v_2}$ *becomes smaller; and when* $\rho_2 \to \infty$, $\frac{1-e^{-2\rho_2 r_0}}{1-e^{-2\rho_2 r_I}} \times \frac{w_V v_2}{v_1+v_2} \to \frac{w_V v_2}{v_1+v_2}$.

Lower bound of the achievable throughput

In this part, we first analyse the lower bound of the maximum amount of data that can be received from helpers by the VoI and the corresponding scheduling scheme to achieve this lower bound given fixed n and $l_i, i = 1, ...n-1$. Then we extend to consider that n and $l_i, i = 1, ...n-1$ are random values, corresponding to a Poisson distribution of vehicles, and analyse the lower bound of the achievable throughput.

Theorem 3.3. *In the transitional regime, given fixed n and $l_i, i = 1, ...n-1$, a lower bound of the maximum amount of data the VoI can receive from n helpers in one cycle is given by*

$$\left(\sum_{i=1}^{n} Y_i\right)_3^* \geq \sum_{i=1}^{n-1} \min\left\{\frac{l_i}{v_2}w_I, \frac{2r_0}{v_1+v_2}w_V\right\} + \frac{2r_0}{v_1+v_2}w_V \tag{3.39}$$

where $(\sum_{i=1}^{n} Y_i)_3^*$ is the respective $\sum_{i=1}^{n} Y_i$ associated with its optimum value and we use the subscript 3 to mark the transitional regime.

Furthermore, there exists a V2I transmission scheme for helpers and a V2V transmission scheme to achieve the above lower bound of the maximum amount of data for the VoI, satisfying:

$$\begin{cases} D_{3i}^* = \frac{\min\{l_i, 2r_I\}}{v_2}w_I, & i = 1, 2, ...n-1 \\ D_{3n}^* = \frac{2r_I}{v_2}w_I & \\ Y_{3i}^* = \min\left\{\frac{l_i w_I}{v_2}, \frac{2r_0 w_V}{v_1+v_2}\right\}, & i = 1, 2...n-1 \\ Y_{3n}^* = \frac{2r_0}{v_1+v_2}w_V & \end{cases} \tag{3.40}$$

where D_{3i}^* and $Y_{3i}^*, i = 1, ..., n$ are the respective D_i and $Y_i, i = 1, ..., n$ associated with the optimum solution.

Proof. We find the lower bound of the maximum amount of data received by the VoI from n helpers in one cycle by constructing a specific V2I transmission scheme and analyse the corresponding value of $\sum_{i=1}^{n} Y_i$ achieved under this scheme. As this value of $\sum_{i=1}^{n} Y_i$ is obtained under a specific V2I transmission scheme, it may not be the maximum value for the original optimization problem (3.4) because of a lack of consideration of all possible V2I transmission schemes for helpers, but will form a lower bound of the maximum value of $\sum_{i=1}^{n} Y_i$ for the the original optimization problem (3.4). In the following, we will first construct a specific V2I transmission scheme, and then analyse the optimum amount of data received by the VoI from helpers under this specific V2I transmission scheme, as well as finding a corresponding V2V transmission scheme to reach the lower bound specified in the theorem.

It has been described in the proof of Theorem 3.1 and Theorem 3.2 that the V2I transmission scheme for helpers to reach the corresponding optimum throughput under the Infrastructure-Limited and the V2V-Limited regime are the same, and this V2I transmission scheme satisfies the following equations:

$$\begin{cases} D_i = \frac{\min\{l_i, 2r_I\}}{v_2} w_I, \quad i = 1, 2, ...n-1 \\ D_n = \frac{2r_I}{v_2} w_I. \end{cases} \tag{3.41}$$

We adopt this same V2I transmission scheme for helpers in the transitional regime as well. It follows that the amount of data received by each helper from infrastructure, D_i, $i = 1, ..., n$, is given by (3.41). With condition $\frac{r_0 w_V v_2}{r_I(v_1+v_2)} < w_I < \frac{w_V v_2}{v_1+v_2}$ for the transitional regime, constraints (3.7) and (3.8) in the optimization problem (3.4) can be replaced with a tighter constraint after putting in (3.41), which are shown as follows:

$$\begin{aligned} Y_i &\leq \min\left\{D_i, \frac{2r_0}{v_1+v_2} w_V\right\} \\ &= \min\left\{\frac{l_i}{v_2} w_I, \frac{2r_I}{v_2} w_I, \frac{2r_0}{v_1+v_2} w_V\right\} \\ &= \min\left\{\frac{l_i}{v_2} w_I, \frac{2r_0}{v_1+v_2} w_V\right\}, i = 1, 2, ...n-1 \end{aligned} \tag{3.42}$$

and

$$0 \leq Y_n \leq \frac{2r_0}{v_1+v_2} w_V \tag{3.43}$$

Formula (3.42) and (3.43) indicate that the upper bound of $\sum_{i=1}^{n} Y_i$ can be expressed by $\sum_{i=1}^{n-1} \min\left\{\frac{l_i}{v_2} w_I, \frac{2r_0}{v_1+v_2} w_V\right\} + \frac{2r_0}{v_1+v_2} w_V$. In the following, we will show that this upper bound is exactly the optimum solution of $\sum_{i=1}^{n} Y_i$ under the adopted V2I transmission scheme and we can find a corresponding V2V transmission scheme to achieve this upper bound.

With condition $w_I < \frac{w_V v_2}{v_1+v_2}$, we have:

$$\min\left\{\frac{l_i}{v_2}w_I, \frac{2r_0}{v_1+v_2}w_V\right\} \leq \min\left\{\frac{l_i}{v_1+v_2}w_V, \frac{2r_0}{v_1+v_2}w_V\right\} \tag{3.44}$$

This follows that:

$$\begin{aligned}\sum_{i=1}^{n} Y_i &\leq \sum_{i=1}^{n-1} \min\left\{\frac{l_i}{v_2}w_I, \frac{2r_0}{v_1+v_2}w_V\right\} + \frac{2r_0}{v_1+v_2}w_V \\ &\leq \frac{\sum_{i=1}^{n-1} \min\{l_i, 2r_0\} + 2r_0}{v_1+v_2}w_V,\end{aligned} \tag{3.45}$$

which shows that when $\sum_{i=1}^{n} Y_i$ is not larger than $\sum_{i=1}^{n-1} \min\left\{\frac{l_i}{v_2}w_I, \frac{2r_0}{v_1+v_2}w_V\right\} + \frac{2r_0}{v_1+v_2}w_V$, the constraint (3.9) in optimization problem (3.4) will also be satisfied. Thus, when each $Y_i, i = 1, ...n$, is equal to its upper bound shown in (3.42) and (3.43), and when each $D_i, i = 1, ..n$ is given by (3.41), i.e., when each Y_i and $D_i, i = 1, ...n$ are given by (3.40), the value of $\sum_{i=1}^{n} Y_i$ will reach its upper bound $\sum_{i=1}^{n-1} \min\left\{\frac{l_i}{v_2}w_I, \frac{2r_0}{v_1+v_2}w_V\right\} + \frac{2r_0}{v_1+v_2}w_V$ while satisfying all constraints in the optimization problem (3.4). This leads to (3.39).

Now we show that there exists a V2V transmission scheme to achieve the lower bound specified in (3.39). To this end, it can be readily shown from (3.40) that the same V2V transmission scheme described in the proof of Theorem 3.1 satisfies (3.40), therefore can realize the maximum amount of received data for the VoI specified in (3.39) under the specific V2I transmission scheme. This completes the proof. □

On the basis of Theorem 3.3, we now analyze the lower bound of the achievable throughput by the VoI considering that both n and the corresponding $l_i, i = 1, ...n-1$ are random values. Similarly to the analysis in subsections 3.2.1 and 3.2.2, calculating the lower bound of the achievable throughput directly according to (3.39) is challenging due to the complexity associated with analysing the joint distribution of $\min\left\{\frac{l_i}{v_2}w_I, \frac{2r_0}{v_1+v_2}w_V\right\}, i = 1, ...n-1$. In this regime, we define a cluster to be a maximal set of helpers located within road segment $[0, s]$ and the distance between any two adjacent helpers is smaller than or equal to $\frac{2r_0 w_V v_2}{w_I(v_1+v_2)}$. It follows that within each cluster, $\min\left\{\frac{l_i}{v_2}w_I, \frac{2r_0}{v_1+v_2}w_V\right\} = \frac{l_i}{v_2}w_I, i = 1, 2, ...$, therefore removing the above challenge.

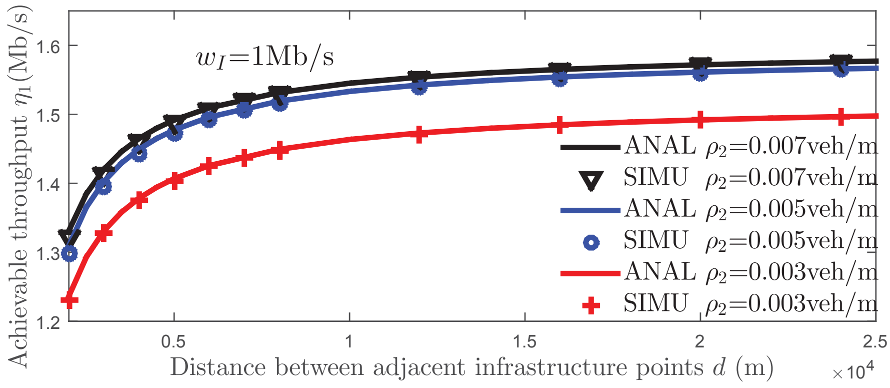
Utilizing the same approximation method as that used to calculate $E[D_{V1}]$ and $E[D_{V2}]$ in the Infrastructure-Limited regime and the V2V-Limited regime respectively, the lower bound of the achievable throughput in transitional regime, denoted by η_{3l}, is obtained as follows:

$$\eta_{3l} = \frac{2r_I w_I + c_3}{d}, \tag{3.46}$$

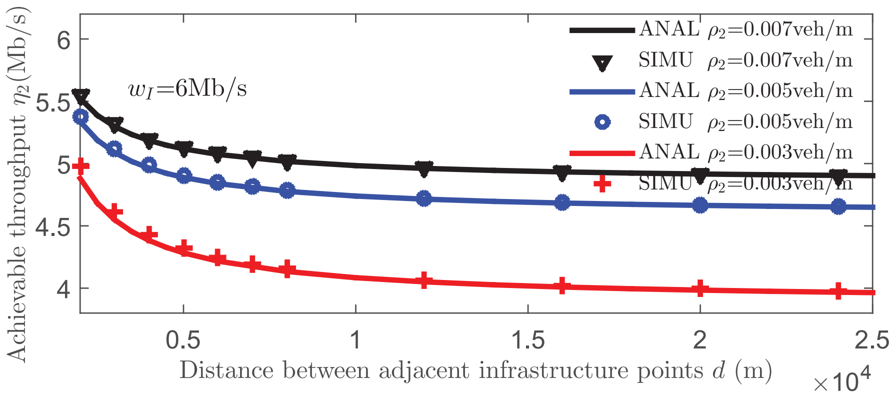
where $c_3 = \frac{\left[(d-2r_I)(v_1+v_2)+r_0v_1-\frac{v_1}{\rho_2}\right]\left(1-e^{-\frac{2\rho_2 r_0 w_V v_2}{w_I(v_1+v_2)}}\right)w_I}{v_2}$.

3.3 Simulation and Discussion

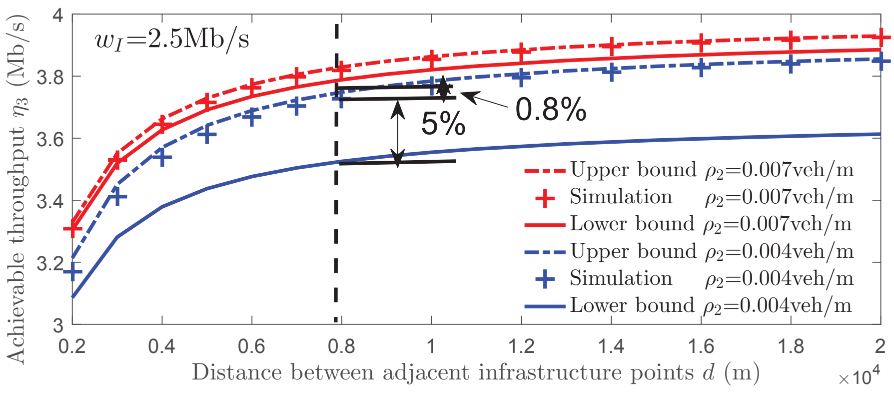
In this section we use Monte Carlo simulations to verify the accuracy of the analysis and establish the applicability of the theoretical analysis for more general scenarios beyond the ideal assumptions (e.g., constant speed, unit disk model, and constant channel condition, etc.) used in the analysis. Specifically, 20 infrastructure points are regularly deployed along the highway and the distance between adjacent infrastructure points, d, is varied from 2km to 50km. The helpers' density ρ_2 varies from 0 to 0.1veh/m and the speed of the VoI and helpers are v_1=15m/s and v_2=25m/s respectively. The radio range of infrastructure and vehicles are 500m and 250m (typical radio ranges using DSRC [49]) respectively. The transmission rate of V2V communications is w_V=5Mb/s and the transmission rate of V2I communications w_I varies from 0 to 10Mb/s to allow us to cover all three regimes. Each simulation is repeated 2000 times and the average value is shown in the plot.



(a) Infrastructure-Limited Regime



(b) V2V-Limited Regime



(c) Transitional Regime

Figure 3.5 : A comparison between our analytical results and the simulation results under each regime, with different helpers' density ρ_2.

Fig. 3.5 shows a comparison between analytical results and simulation results under each regime. Specifically, Fig. 3.5(a) and Fig. 3.5(b) compare the achievable throughput obtained from our analysis and the simulation result, where the simulation is conducted assuming the scheduling scheme described in the proof of Theorem 3.1 and Theorem 3.2 respectively. Fig. 3.5(c) compares the upper and lower bound of the achievable throughput obtained from analysis and the optimum throughput in the simulation. It is shown that in the Infrastructure-Limited and V2V-Limited regime, the analytical results match very well with simulations especially when the distance of two neighbouring infrastructure points, d, is large. This confirms that the approximations used in the earlier analysis to obtain the analytical results have negligible impact on the accuracy of the analytical results. In the transitional regime, there is a small gap between the simulated optimum throughput and its upper and lower bound we obtained, e.g, when ρ_2=0.004veh/m and d=8km in this case, the difference between the optimum throughput from the simulation and its upper (or lower) bound is only around 0.8% (or 5%), and the gap decreases with the increase of helpers' density. This shows that even though the upper bound in the transitional regime is not achievable, it is quite close to the optimum throughput.

Interestingly, Fig. 3.5(a) and 3.5(c) show that in the Infrastructure-Limited regime and the transitional regime, the achievable throughput increases when d increases while Fig. 3.5(b) shows that in the V2V-Limited regime, the achievable throughput decreases when d increases. This can be explained that while keeping other parameters constant, an increase in d on one hand will improve the amount of data received by the VoI from V2V communications, which improves the achievable throughput; on the other hand, it will increase the amount of time spent in one cycle by the VoI, which reduces the achievable throughput. When w_I is small, the amount of data received by the VoI from V2V communications is small due to the limitation of the amount of data received by helpers from infrastructure. Therefore, an increase in d will lead to an larger rate of increase in the total amount of received data by the VoI than the rate of increase in the amount of time spent in one cycle, which results in the overall increase of the achievable throughput, shown as Fig. 3.5(a) and 3.5(c). However, when w_I is large, the amount of data the VoI can receive from V2V communications is comparatively large, an increase in d has marginal impact on the data amount received by the VoI. Therefore, an increase in d will lead to an smaller rate of increase in the total amount of received data by the VoI than the rate of increase in the amount of time spent in one cycle, which results in the overall decrease of the achievable throughput, shown as Fig. 3.5(b).

Fig. 3.5 also gives insight into the optimum choice of distance between infrastructure points. It is obvious from these figures that when d increases beyond a certain threshold, e.g., d=10km in our case, an increase in d has limited impact on the achievable throughput. This can be explained by the fact that when d is small ($d <$ 10km in the simulation), the amount of data received by the VoI from V2V communications is relatively small compared with that received from V2I communications, especially when traffic density is low (here average ρ_2=0.005veh/m). It follows that the VoI's achievable throughput is mainly dominated by its V2I communications. However, with the increase of d, the increase of data received from V2V communications makes V2I communication's dominating impact subdued, which in turn leads to the subtle variation of the throughput.

Fig. 3.6 compares the achievable throughput (the lower bound for the transitional-regime is used) using our cooperative communication strategy (labelled as With coop) with its non-cooperative counterpart (labelled as Without coop). The non-cooperative counterpart is conducted by setting the helpers' density $\rho_2 = 0$ because when there are no helpers in the vehicular network, there will be no cooperative communications. It is shown that even when helpers' density is low, e.g., ρ_2=0.002veh/m, the throughput achieved by utilizing our cooperative communica-

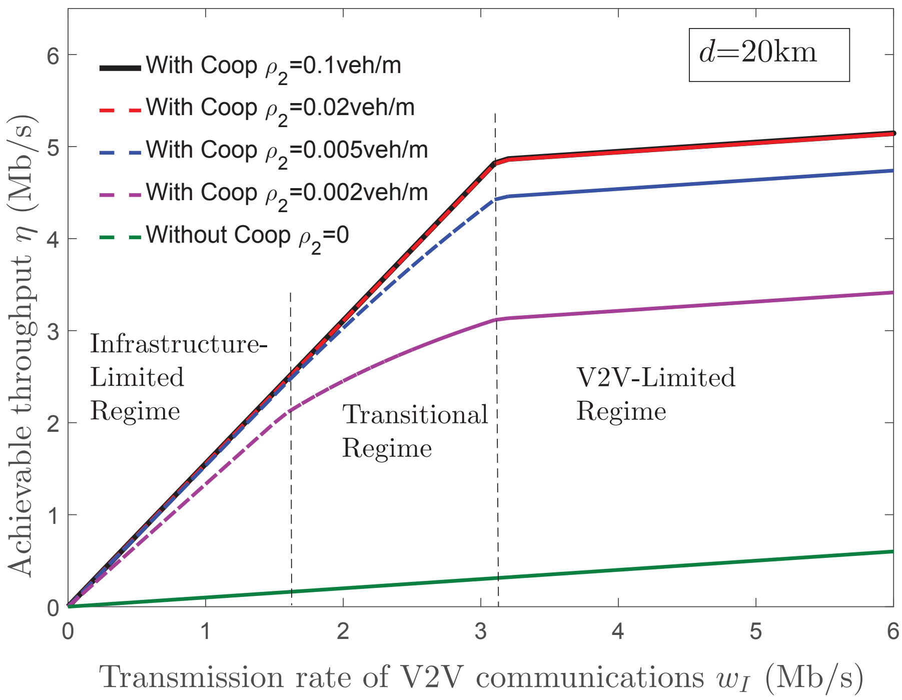


Figure 3.6 : A comparison between the throughput achieved from vehicular networks with and without cooperative communication.

tion strategy is around 15 times larger when w_I=3Mb/s and around 10 times larger when w_I=6Mb/s than that achieved without cooperative communications. This gives an important conclusion that our cooperative communication strategy can significantly improve the throughput even when vehicular density is low.

Fig. 3.6 also reveals the relationship between the achievable throughput and helpers' density ρ_2. Importantly, we can see that a higher density is beneficial to the throughput because a higher ρ_2 will enhance the connectivity of vehicular networks, which leads to higher chance of V2V cooperative communications. However, when ρ_2 increases beyond a certain threshold, e.g., ρ_2=0.005veh/m in this case, a further increase in ρ_2 has only marginal impact on the achievable throughput. This is due to the fact that when ρ_2 is large enough for the VoI to find at least one helper in its coverage at any time point, increasing the density (which will lead to more helpers within the VoI's coverage at one time) is no longer helpful to improve the throughput because the VoI can only receive data from one vehicle at one time and the total amount of time the VoI can receive data from V2V communication will be the same.

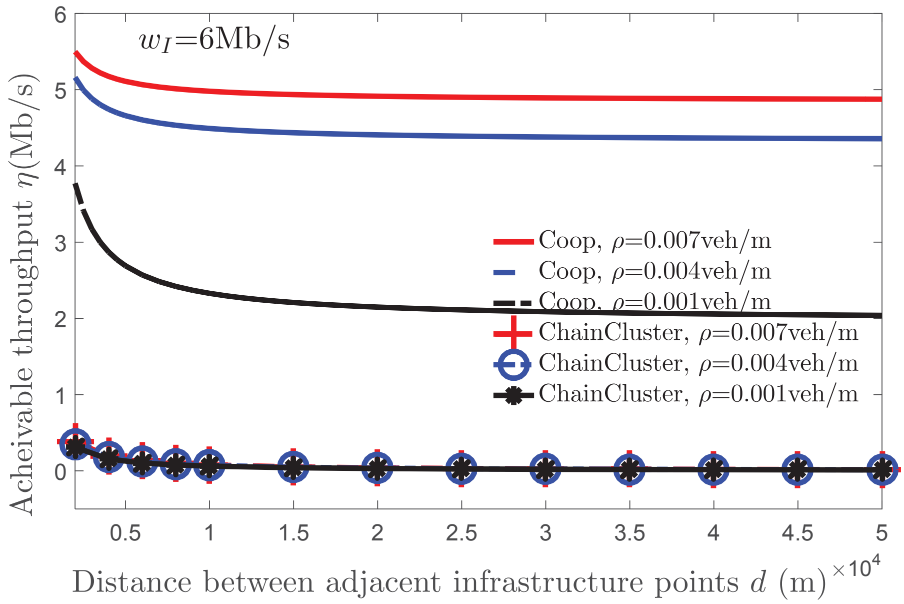


Figure 3.7 : A comparison between the throughput achieved from our proposed strategy and that from the strategy proposed in [49].

Fig. 3.7 compares the achievable throughput assuming our proposed cooperative communication strategy (labelled as Coop) with that assuming the cooperative strategy proposed in [49] (labelled as ChainCluster) in the V2V-Limited regime. Specifically, the strategy proposed in [49] utilized vehicles moving in the same direction as the target vehicle (VoI) to form a cluster to help the VoI's download. A vehicle can be chosen into the cluster if and only if it can connect to the VoI via a multi-hop path. It can be seen that the throughput achieved by the VoI assuming our cooperative communication strategy is much larger than that achieved assuming the strategy proposed in [49]. This is due to the fact that in [49], the authors only used the cooperation among vehicles moving in the same direction and within the same cluster of the VoI, while in our strategy, both cooperation among infrastructure and cooperation of all vehicles travelling in the opposite direction of the VoI are fully utilized to help the VoI's download, which significantly improves the achievable throughput of the VoI.

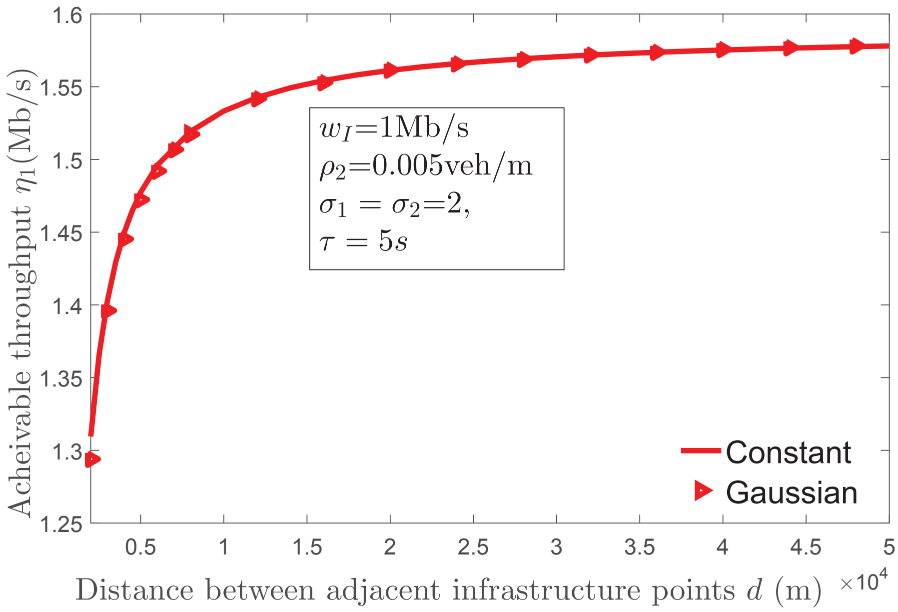


Figure 3.8 : A comparison between throughput achieved from the constant speed model and the time-varying speed model which follows Gaussian distribution.

Fig. 3.8 shows a comparison of the achievable throughput from the constant speed model (labelled as Constant Speed) and the time-varying speed model (labelled as Gaussian Speed) under the Infrastructure-Limited regime. The time-varying speed of vehicles in each lane follows Gaussian distributions, defined as: $v_1^{'} \sim N(v_1, \sigma_1^2)$ and $v_2^{'} \sim N(v_2, \sigma_2^2)$, where v_1 and v_2 are the constant speed used in our analysis, and σ_1^2 and σ_2^2 are the mean variance of the mean speed v_1 and v_2 respectively. To model the slight deviations from the mean speed, we set σ_1=σ_2=2 and the speed-change time interval τ=5s. The figure shows that when individual vehicular speed deviates slightly from the mean speed, it has marginal impact on the achievable throughput.

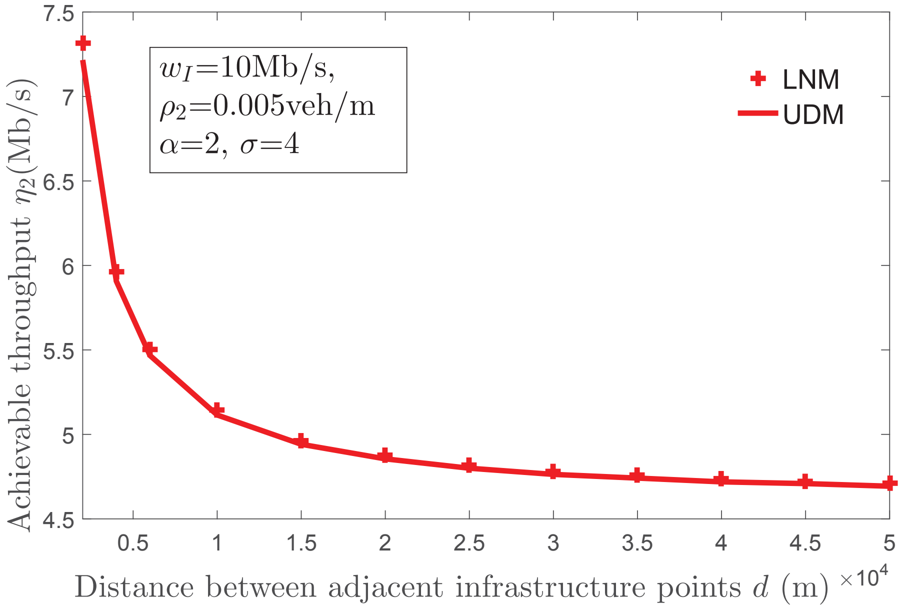


Figure 3.9 : A comparison between throughput achieved from the unit disk model and the log-normal connection model.

Fig. 3.9 gives a comparison of throughput achieved assuming the unit disk model (labelled as UDM) and that assuming the log-normal connection model (labelled as LNM) in the V2V-Limited regime, and shows that the unit disk model assumption has little impact on the throughput. The parameters of log-normal connection model are set as path loss exponent α=2 and standard deviation σ=4 [97]. It is shown that the system assuming the log-normal connection model has a slightly higher achievable throughput than that assuming the unit disk model, which coincides with the results in [97] that the log-normal connection model is beneficial to information delivery in vehicular networks. The reason behind this phenomenon is that the log-normal connection model introduces a Gaussian variation of the transmission range around the mean value, which implies a higher chance for the VoI to be connected to helpers located further away.

Fig. 3.10 compares the throughput achieved by allowing only one-hop communication and allowing both k-hop ($k = 2, 3, 5$) V2I communications between the VoI

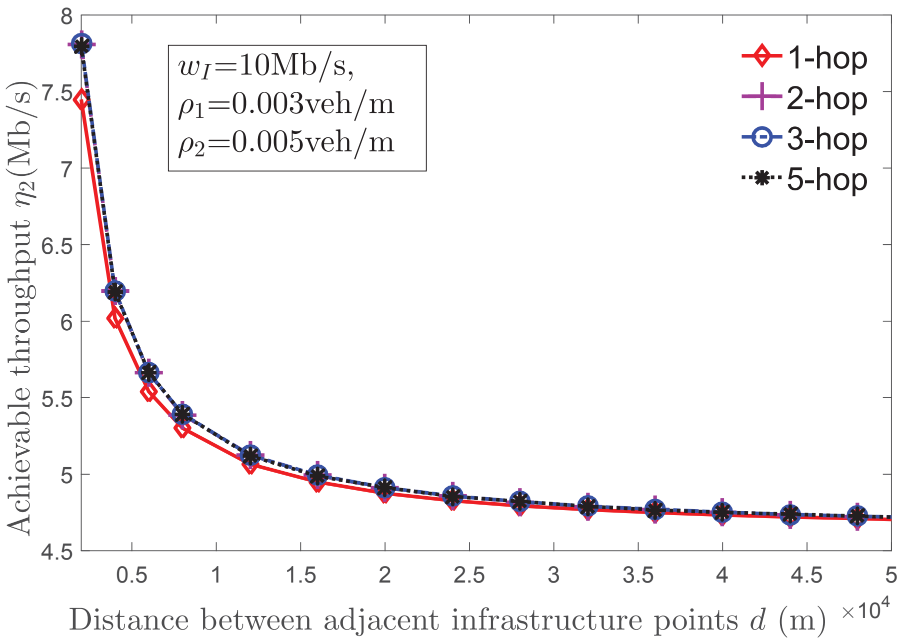


Figure 3.10 : A comparison between throughput achieved when allowing one-hop communication and multi-hop communications.

and infrastructure, and k-hop V2V communication between the VoI and helpers. It is shown that allowing multi-hop communications beyond one hop has little impact on the throughput. Particularly, as pointed out in the end of Section 3.1.2, in our considered scenario, allowing multi-hop V2V communication only helps to balance the distribution of information among helpers but do not increase the net amount of information available in the network. The marginal increase in the achievable throughput comes from multi-hop V2I communications between the VoI and infrastructure, because it allows the VoI having longer connection time with the infrastructure.

Fig. 3.11 compares throughput achieved from the constant channel model with that from the time-varying channel model, and shows that our analysis under the constant channel model is applicable to a more realistic time-varying channel model which considers both fading and path loss. Specifically, for the time-varying chan-

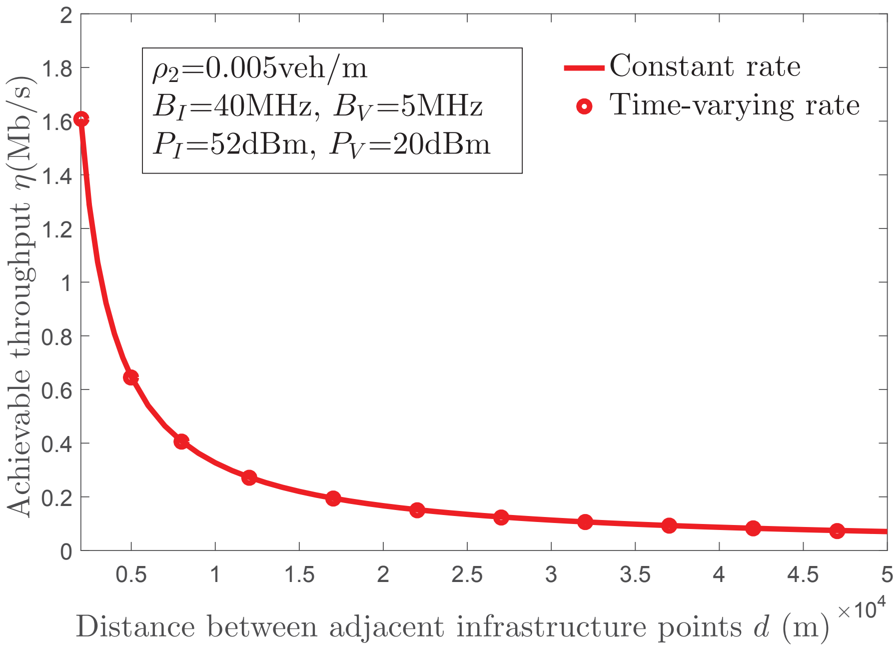


Figure 3.11 : A comparison between throughput achieved from constant channel model and time-varying channel model which considering Rayleigh fading and path loss.

nel model, we adopt the model used in [54] that considers Rayleigh fading and path loss, from which the transmission rate is given by $w'_I = B_I \log_2 \left(1 + P_I|\beta d_i^{-2}|^2\right)$ and $w'_V = B_V \log_2 \left(1 + P_V|\beta d_{ij}^{-2}|^2\right)$, with the bandwidth and transmit power of each infrastructure and vehicle being B_I=40MHz, P_I=52dBm and B_V=5MHz, P_V=20dBm [98] respectively. Parameter β is the Gaussian random variable with mean 0 and variance 1 and d_i, d_{ij} are the distances between a vehicle and its associated infrastructure point, between vehicle and vehicle when conducting V2I and V2V communications respectively. The above settings of B_I, P_I, B_V and P_V imply that the network is in the V2V-Limited regime. By dividing the total coverage length of the transmitter (infrastructure or vehicle) into K (here we set K=1000) small segments, the average channel throughput w_I and w_V in the time-varying channel model can be obtained by averaging the transmission rates of all segments. This obtained average throughput w_I and w_V are then used in our constant channel model. It is obvious from Fig. 3.11 that the achievable throughput from the above two channel models match each other. This phenomenon can be explained by equation (3.35) which shows that the achievable throughput in V2V-Limited regime is a linear function of w_I and w_V. Then it follows that $E[\eta(w_I, w_V)] = \eta(E[w_I], E[w_V])$, which implies that for time-varying channels, the time-varying values of $w_I^{'}$ and $w_V^{'}$ can be replaced by the respective time-averaged throughput of V2I and V2V communications and our analysis still applies.

3.4 Summary

This chapter proposed a cooperative communication strategy for vehicular networks with a finite vehicular density by utilizing V2I communications, V2V communications, mobility of vehicles, and cooperations among vehicles and infrastructure to facilitate data dissemination. A detailed analysis was presented when there exists one vehicle with a download request in the network, and the closed-form expression of the achievable throughput (or its upper and lower bound) by the target vehicle was obtained in three different regimes we classified in our analysis based on the relationship between the data rates of V2I communications, V2V communications, and the speeds of vehicles. Numerical and simulation results show that the proposed cooperative strategy can significantly improve the achievable throughput of vehicular networks even when traffic density is low. Simulation results show that our analysis can be extended to more realistic models such as the time-varying speed model, the log-normal shadowing model and the time-varying channel model considering fading and path loss. Our results shed insight on the optimum design of cooperative vehicular networks for ITS to achieve fast and reliable data dissemination.

Chapter 4

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

In this chapter, we extend the work introduced in Chapter 3 and consider a typical delay-tolerant application scenario with a subset of vehicles, termed Vehicles of Interest (VoIs), having download requests, e.g., videos, from the Internet. Each VoI downloads a distinct large-size file from the Internet and other vehicles without download requests, termed *helpers*, assist the delivery of the files to the VoIs. Different from the single-VoI scenario, when there are multiple vehicles with download requests, the possible contention and collision among vehicles in vehicular communications become both important and challenging issues to study. We analyse the capacity of vehicular networks with a finite traffic density adopting the cooperative communication strategy proposed in Chapter 3. Our result shows that the proposed cooperative strategy can improve the capacity of vehicular networks, and the improvement is more pronounced when the proportion of vehicles with download requests is low.

The rest of this chapter is organized as follows: Section 4.1 introduces the system model, the proposed cooperative communication strategy and the problem formation. Theoretical analysis is provided in Section 4.2. In Section 4.3, we validate the analytical result using simulations and conduct numerical analysis to discuss our result and its insight. Section 4.4 summarizes this chapter.

4.1 System Model and Problem Formation

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

4.1.1 Network Model

We continue to follow on the network model adopted in Chapter 3 that in each direction (eastbound and westbound) of the considered highway, the distribution of vehicles follows a homogeneous Poisson process with densities ρ_1 and ρ_2 respectively, and vehicles in each direction travel at the same constant speed of v_1 and v_2 respectively. Therefore, as a ready consequence of the superposition property of Poisson processes [99], all vehicles on the highway are also Poissonly distributed with density $\rho = \rho_1 + \rho_2$. We assume that the proportion of VoIs travelling towards each direction is p $(0 < p < 1)$. Therefore, VoIs and helpers respectively have traffic density $p\rho$ and $(1 - p)\rho$.

Even though the analysis is based on a straight linear highway, our obtained result can be readily applied to each separate road in 2-Dimensional/3-Dimensional (2D/3D) scenarios, e.g., grid 2D network [48] and 3D multi-level network [90], and the network capacity can be readily achieved by summing the capacity achieved from each separate road. Moreover, we will show later in the analysis that the capacity achieved in this work does not depend on the speed distribution of vehicles, and our analysis also applies to other time-varying speed model, e.g., Gaussian speed model [21, 85], as long as the resulting spatial distribution of vehicles is stationary. The system model is illustrated in Fig. 4.1.

4.1.2 Wireless Communication Model

We continue to follow on the wireless communication model we assumed in Chapter 3, i.e., we assume (1) the unit disk wireless connection model with infrastructure

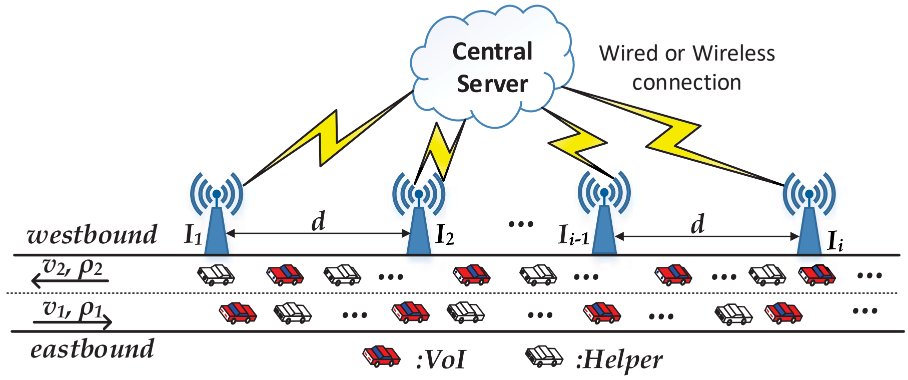


Figure 4.1 : An illustration of the system model for a bi-directional highway with infrastructure regularly deployed with equal distance d: multi-user case.

and vehicles having radio range r_I and r_0 respectively; (2) half-duplex antenna for vehicles and infrastructure; (3) one-hop and unicast transmission scheme; (4) constant data rate w_I and w_V for V2I and V2V communication respectively; and (5) CSMA MAC protocol with sensing range R_c for V2V communications.

We consider a V2I transmitting scheme where the infrastructure will transmit its data to VoIs first, i.e., helpers can receive data from infrastructure only when there is no VoI within the coverage of infrastructure. This V2I scheduling scheme makes the VoIs achieve the maximum data rate from infrastructure. For V2V communications, helpers function as transmitters and VoIs as receivers. A transmitter can choose a receiver from either direction within its transmission range.

4.1.3 Cooperative Communication Strategy

The cooperative communication strategy that explores the combined use of V2I communications between the VoIs and infrastructure, between helpers and infrastructure, V2V communications between the VoIs and helpers, cooperations among infrastructure and among vehicles, as well as vehicular mobility to maximize the capacity of the VoIs has been described in detail in Chapter 3. When there are multiple vehicles with download requests in the network, we assume there is a central server that has full knowledge of the network topology and data transmission process to guarantee that the data the helpers receive from infrastructure is the data required by the VoIs they will encounter. This assumption, which may cause large wireless communication overhead, is required to establish the maximum data rate, i.e., capacity, that can be achieved by the VoIs because it assumes a perfect scheduling of data items for both V2I and V2V communications. Some practical issues like out of sequence data delivery and missing packets can be handled by techniques such as network coding (e.g., [94]) so that we can focus on the main theme of the work without the need for considering their impacts. Therefore, when the VoIs are in the coverage of infrastructure, they receive data directly from the infrastructure. In the meantime, the helpers may also receive different pieces of data from the infrastructure when they obtain access to the infrastructure. When the VoIs move outside the coverage of infrastructure, they may continue to receive data from helpers, exploiting the mobility of vehicles and V2V communications. We do not consider the case that VoIs share their received content with others during V2V communication as it does not improve the capacity considering the fact that different VoIs request different files in our considered scenario, and the fact that a larger number of relay nodes, which leads to a larger number of hops between the source and destination, is detrimental to the achievable capacity [29]. Even though the proposed cooperative communication scheme is simple, all the major topological parameters have been taken into consideration.

4.1.4 Problem Formation

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

$$\eta_c = \max_{\chi \in \Phi} \eta^{\chi} = \max_{\chi \in \Phi} \lim_{t \to \infty} \frac{D^{\chi}(t)}{t}. \tag{4.1}$$

4.2 Analysis of the Capacity

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

We define the area covered by one infrastructure point (termed *V2I Area*) and the adjacent area between two consecutive infrastructure points but not covered by the infrastructure point (termed *V2V Area*) as a *cycle*, which has length d. See Fig. 4.2 for an illustration. It follows from the renewal theory [95] that the long-term achievable capacity by the VoIs from each cycle, denoted by η_{cycle}, is identical and the total capacity achieved in a given highway segment with length $L \gg d$ can be readily calculated by (ignoring the trivial fact that $\frac{L}{d}$ may not be an integer):

$$\eta_c = \frac{L}{d} \eta_{\text{cycle}}. \tag{4.2}$$

From (4.2), to calculate the total capacity achievable by the VoIs from a highway segment with length L, it suffices to calculate the capacity achieved by the VoIs from one cycle, which includes capacity achieved both from V2I communications and V2V communications, given as follows:

$$\eta_{\text{cycle}} = \lim_{t \to \infty} \frac{D_{V2I}(t)}{t} + \lim_{t \to \infty} \frac{D_{V2V}(t)}{t}, \tag{4.3}$$

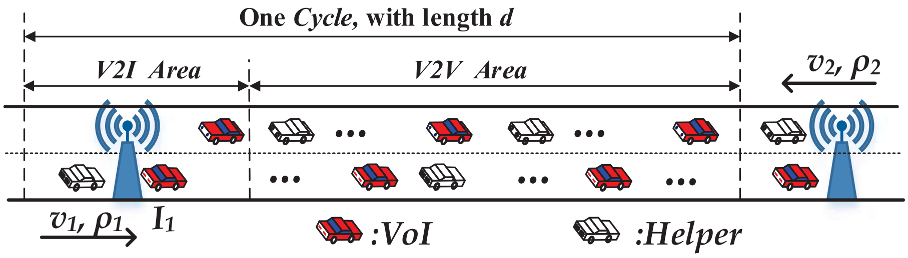


Figure 4.2 : An illustration of one cycle, which includes V2I Area and V2V Area.

where $D_{V2I}(t)$ and $D_{V2V}(t)$ are respectively the maximum expected amount of data received by the VoIs from infrastructure in the V2I Area and from helpers in the V2V Area during time period t. In the following, we will focus on studying one cycle entirely contained within the highway segment of length L, termed the *cycle of interest*. We will first calculate the two terms on the right hand side of (4.3) separately, and then combine both terms to obtain the final expression of the achievable capacity.

4.2.1 Capacity Achieved from V2I Communications

Without loss of generality, we call the infrastructure point located in our cycle of interest I_1. We assume that time is divided into time slots of equal length $\triangle t$, and $\triangle t$ is sufficiently small that we can approximately regard vehicles as stationary during each time slot. Denote by $q_1(i), i = 1, 2, ...$ a discrete random variable representing the fraction of time that VoIs' V2I communication with I_1 happens during the ith time slot $[(i-1)\triangle t, i\triangle t)$. Recall that to make the VoIs achieve the maximum data rate from the infrastructure, we adopt a V2I communication scheme that infrastructure delivers its data directly to the VoIs as long as there are VoIs within its coverage. Therefore, $q_1(i)$ is equal to 1 when there exist at least one VoIs within the coverage of I_1 during the ith time slot; otherwise it is equal to 0. This follows that the maximum expected amount of data the VoIs can obtain from one cycle through V2I communications during time period $[0, t]$ can be calculated by (ignoring the trivial fact that $\frac{t}{\triangle t}$ may not be an integer):

$$D_{V2I}(t) = \lim_{\triangle t \to 0} E\left(w_I \sum_{i=1}^{t/\triangle t} q_1(i) \triangle t \right) \tag{4.4}$$

According to the ergodicity and stationarity properties of homogeneous Poisson point processes [100], the time average of $q_1(i)$ is equal to the probability that there is at least one VoI within the coverage of I_1 at a randomly chosen time slot, denoted by $\bar{q}_1$. Considering the Poisson distribution of vehicles, it can be readily shown that $\bar{q}_1 = 1 - e^{-p\rho 2r_I}$. Therefore, we have

$$\lim_{t\to\infty} \frac{\lim_{\triangle t \to 0} E\left(\sum_{i=1}^{t/\triangle t} q_1(i) \triangle t \right)}{t} = \bar{q}_1 = 1 - e^{-p\rho 2r_I}. \tag{4.5}$$

Combing (4.4)-(4.5), we have the long-term capacity achieved by the VoIs from one cycle through their V2I communications:

$$\lim_{t\to\infty} \frac{D_{V2I}(t)}{t} = w_I \left(1 - e^{-p\rho 2r_I} \right). \tag{4.6}$$

4.2.2 Capacity Achieved from V2V Communications

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

Theorem 4.1. *The capacity the VoIs can achieve through V2V communications from one cycle is given by:*

$$\lim_{t\to\infty} \frac{D_{V2V}(t)}{t} = \min\left\{ \lim_{t\to\infty} \frac{D_{I\_H}(t)}{t}, \lim_{t\to\infty} \frac{D_V(t)}{t} \right\}, \tag{4.7}$$

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

Since the bottleneck is either in the V2V communications between VoIs and helpers, represented by $\lim_{t\to\infty}\frac{D_V(t)}{t}$, or in the V2I communications between helpers and infrastructure, represented by $\lim_{t\to\infty}\frac{D_{I\_H}(t)}{t}$, the proof of Theorem 4.1 follows readily. More specifically, imagine the V2V communication process between helpers and VoIs as a single-queue queuing system. The rate the helpers receive data from infrastructure, $\lim_{t\to\infty}\frac{D_{I\_H}(t)}{t}$, is equivalent to the incoming rate of the queue. The rate helpers deliver data to the VoIs, $\lim_{t\to\infty}\frac{D_V(t)}{t}$, is equivalent to the processing speed of the queue. The outgoing rate of the queue, $\lim_{t\to\infty}\frac{D_{V2V}(t)}{t}$, is equal to either the incoming rate or the processing speed.

From Theorem 4.1, to obtain the capacity achieved by the VoIs through V2V communications with the helpers from one cycle, it remains to calculate the long-term data rate achieved by the helpers from one cycle through their V2I communications, $\lim_{t\to\infty}\frac{D_{I\_H}(t)}{t}$, and the maximum long-term data rate the VoIs can achieve through V2V communications from one cycle without considering the limitation of the amount of data the helpers received, $\lim_{t\to\infty}\frac{D_V(t)}{t}$. In the following, we will calculate these two terms separately.

Calculation of $\lim_{t\to\infty}\frac{D_{I\_H}(t)}{t}$

Denote by $q_2(i)$ a discrete random variable, which is equal to 1 when helper's V2I communication happens during the i-th time slot $[(i-1)\triangle t, i\triangle t)$, $i=1,2,...$, otherwise it is equal to 0. Similar to the analysis in section 4.2.1, we have

$$\lim_{t\to\infty}\frac{D_{I\_H}(t)}{t}=\lim_{t\to\infty}\frac{w_I\lim_{\triangle t\to 0}E\left(\sum_{i=1}^{t/\triangle t}q_2(i)\triangle t\right)}{t}=w_I\bar{q}_2, \tag{4.8}$$

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

$$\bar{q}_2=e^{-p\rho 2r_I}\left(1-e^{-(1-p)\rho 2r_I}\right)=e^{-p\rho 2r_I}-e^{-\rho 2r_I}. \tag{4.9}$$

Combing (4.8) and (4.9), we have:

$$\lim_{t\to\infty}\frac{D_{I\_H}(t)}{t}=w_I\bar{q}_2=w_I\left(e^{-p\rho 2r_I}-e^{-\rho 2r_I}\right). \tag{4.10}$$

Calculation of $\lim_{t\to\infty}\frac{D_V(t)}{t}$

In this part, we analyse the maximum data rate achieved by the VoIs through V2V communications from one cycle area without considering the amount of data each helper has.

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

Denote by $N_p^{\chi}(i)$ the number of simultaneous helper-VoI pairs in the V2V Area during the i-th time slot $[(i-1)\triangle t, i\triangle t)$, where the superscript $\chi\in\Phi$ denotes the scheduling algorithm that selects the simultaneously active helper-VoI pairs and Φ

denotes the set of all scheduling algorithms. Using the analysis in [29], it can be shown that:

$$\begin{aligned}
\lim_{t\to\infty} \frac{D_V(t)}{t} &= \max_{\chi\in\Phi} \lim_{t\to\infty} \frac{D_V^\chi(t)}{t} \\
&= \max_{\chi\in\Phi} \lim_{t\to\infty} \frac{w_V \lim_{\triangle t\to 0} E\left(\sum_{i=1}^{t/\triangle t} N_p^\chi(i)\triangle t\right)}{t} \\
&= \max_{\chi\in\Phi} w_V E\left[N_p^\chi\right], \qquad (4.11)
\end{aligned}$$

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

From (4.11), the maximum value of $\lim_{t\to\infty} \frac{D_V^\chi(t)}{t}$ is achieved when using an optimum V2V scheduling algorithm that schedules as many active helper-VoI pairs as possible. Therefore, in the following analysis, we shall establish the optimum V2V scheduling algorithm and the maximum $E\left[N_p^\chi\right]$ that can be achieved by the algorithm. Specifically, we will first find an optimum V2V scheduling scheme, denoted by χ_{opt}, that leads to the maximum number of simultaneously active helper-VoI pairs, and then calculate $E\left[N_p^{\chi_{opt}}\right]$ under this optimum algorithm. Without loss of generality, we designate the left boundary point of the V2V Area, i.e., the point to the right of infrastructure point I_1 and at a distance r_I to I_1, as the origin of the coordinate system, and the east (right) direction as positive ($+x$) direction. The following theorem summarizes the optimum scheduling scheme.

Theorem 4.2. *An optimum scheduling scheme χ_{opt}, which leads to the maximum number of simultaneously active helper-VoI pairs in V2V Area is as follows: select active helper-VoI pairs in order from left to the right. First, choose the first helper to the right of the origin that has at least one VoI within its coverage as the first transmitter, and the left-most VoI within the coverage of that helper as its receiver. The next transmitter is the nearest helper to the current transmitter, and satisfies the following conditions: 1) the distance between this helper and the current transmitter* *is no smaller than R_c; 2) it can find at least one VoIs within its coverage, which is different from the receiver of the current transmitter. If there are multiple VoIs, always choose the leftmost VoI. Repeat the above process until the rightmost border of the V2V Area is reached.*

Proof. See Appendix A. □

Remark 4.1. *Note that the optimum scheduling algorithm that achieves the maximum number of active helper-VoI pairs may not be unique.*

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

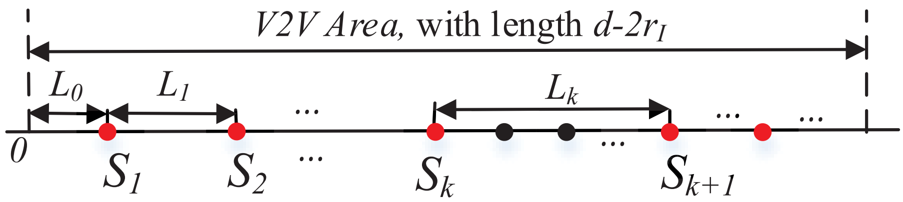


Figure 4.3 : An illustration of the distribution of distances between two consecutive simultaneous transmitters.

Denote by $S_k \in [0, d-2r_I], k = 1, 2, ...$ the position of the k-th transmitter (helper in the active helper-VoI pair) under the optimum scheduling scheme χ_{opt}. Denote by $L_k, k = 1, 2, ...$ the distance between the k-th and the $(k+1)$-th transmitter, and L_0 the distance between the first transmitter and the origin. See Fig. 4.3 for an illustration. It is straightforward that $L_k, k = 1, 2, ...$ are identically and independently distributed and are also independent of L_0. Note that the distribution of L_0 is not the same as that of $L_k, k = 1, 2, ...$ This is due to the fact that when selecting the first transmitter assuming χ_{opt}, we directly choose the first (leftmost) helper within $[0, d-2r_I]$ that has at least one VoI within its coverage. In contrast, when $k > 1$, we select the k-th transmitter from helpers located within $[S_{k-1} + R_c, d - 2r_I]$, which imposes an additional condition that $L_k > R_c$. It can be shown that the expected number of helper-VoI pairs, $E\left[N_P^{\chi_{opt}}\right]$, is exactly the expected number of renewals of a renewal counting process in the V2V Area, with a delay length of $E[L_0]$ (as L_0 has a different distribution from $L_k, k = 1, 2, ...$) and each renewal has an average length $E[L_k], k = 1, 2, ...$ Using the renewal theory [95], $E\left[N_p^{\chi_{opt}}\right]$ can be calculated as follows:

$$E\left[N_p^{\chi_{opt}}\right] = \sum_{n=1}^{\infty} Pr(\sum_{i=0}^{n-1} L_i \leq d - 2r_I), \tag{4.12}$$

An alternative way of obtaining (4.12) is by noting that the inner term of (4.12), i.e., $Pr(\sum_{i=0}^{n-1} L_i \leq d - 2r_I)$, gives the cumulative distribution function (cdf) of the maximum number of simultaneously active helper-VoI pairs assuming the optimum scheduling χ_{opt}. It then follows that the summation of the cdf gives the expected value of $N_p^{\chi_{opt}}$.

Equation (4.12) shows that to calculate $E\left[N_p^{\chi_{opt}}\right]$, we first need to calculate the distribution of each $L_k, k = 0, 1, 2, ...$ The following two theorems characterize the probability density function (pdf) of L_k, denoted by $f_{L_k}(x), k = 1, 2...$ and the pdf of L_0, denoted by $f_{L_0}(x)$ respectively.

Theorem 4.3. *Consider a bi-directional vehicular network with vehicular densities ρ_1 and ρ_2, with p percentage of vehicles being the VoIs and the remaining $1 - p$ percentage of vehicles being the helpers. Furthermore, each vehicles' radio range is r_0 and sensing range is R_c. The distance between two consecutive transmitters (helpers in two consecutive active helper-VoI pairs), $L_k, k \geq 1$, under the optimum scheduling scheme χ_{opt}, has the pdf as follows:*

$$f_{L_k}(x) = \begin{cases} \sum_{m=1}^{\infty} f(x - R_c; m, (1-p)\rho) Pr(m_k = m), & x \geq R_c \\ 0 & x < R_c \end{cases}, \tag{4.13}$$

where $f(x;k,\alpha) = \frac{\alpha^k x^{k-1} e^{-\alpha x}}{(k-1)!}$ *is the pdf of Erlang distribution with shape parameter* k *and rate parameter* α*, and* m_k *is the (random) number of helpers within* $[S_k + R_c, \min\{S_{k+1}, d - 2r_I\})$*. The probability mass function of* m_k *is given by:*

$$Pr(m_k = 1) \approx 1 - e^{-p\rho 2r_0}, \tag{4.14}$$

and for $m \geq 2$*,*

$$Pr(m_k = m) \approx e^{-p\rho 2r_0}\left(p - pe^{-\rho 2r_0}\right)\left(1 - p + pe^{-\rho 2r_0}\right)^{m-2}. \tag{4.15}$$

Moreover, when $R_c \geq 2r_0$*, the approximation in* (4.14) *and* (4.15) *becomes accurate and can be replaced by equality.*

Proof. See Appendix B. □

Theorem 4.4. *Under the same setting as that described in Theorem 4.3, the distance between the first transmitter and the origin under the optimum scheduling scheme* χ_{opt}*,* L_0*, has the pdf as follows:*

$$f_{L_0}(x) = \sum_{m=1}^{\infty} f\left(x; m, (1-p)\rho\right) Pr(m_0 = m), \tag{4.16}$$

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

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

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

Proof. The cdf of L_0 can be derived using the same method as that used in the proof of Theorem 4.3. Particularly, we have $L_0 = \sum_{i=0}^{m_0-1} l_{0,i}$, and in this case, define $h_{0,i} = \min\{l_{0,i}, 2r_0\}, i = 1, ...m_0 - 1$, $h_{0,0} = \min\{l_{0,0}, r_0\} + r_0$, and define $H_{0,m} = \sum_{i=1}^{m-2} h_{0,i} + h_{0,0}$. See proof of Theorem 4.3 for definitions of these parameters. The proof follows. □

From (4.13) and (4.16), we can see that the pdf of L_0 and $L_k, k = 1, 2, ...$ are both in the forms of rather complicated expressions. Accordingly, the computation of the distribution of $\sum_{i=0}^{n-1} L_i, n = 1, 2, ...$, which is required for computing $E\left[N_p^{\chi_{opt}}\right]$ and relies on the joint the distribution of $L_k, n = 1, 2, ...$, can become even more intricate. In our case, assuming that d is much larger compared with the distance between two consecutive simultaneous transmitters, i.e., we have $d \gg E[L_k]$. It is then reasonably accurate to calculate the value of $E\left[N_p^{\chi_{opt}}\right]$ *approximately* using the Elementary Delayed Renewal Theorem [95, Theorem 5.8.4], shown as follows:

$$E\left[N_p^{\chi_{opt}}\right] \approx \frac{d - 2r_I}{E[L_k]}, \tag{4.17}$$

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

According to the pdf of L_k provided in (4.13), the expectation of $L_k, k = 1, 2, ...$ can be readily calculated as follows:

$$\begin{aligned}
E[L_k] &= E\left[R_c + \sum_{i=0}^{m_k-1} l_{k,i}\right] \\
&= R_c + \sum_{m=1}^{\infty} E\left[\sum_{i=0}^{m-1} l_{k,i}\right] Pr(m_k = m) \\
&= R_c + \sum_{m=1}^{\infty} \frac{m}{(1-p)\rho} Pr(m_k = m) \\
&= R_c + \frac{p - pe^{-\rho 2r_0} + e^{-p\rho 2r_0}}{(1-p)p\rho\left(1 - e^{-\rho 2r_0}\right)},
\end{aligned} \tag{4.18}$$

where the first step results by using (B.1) and the second step is obtained by using the total probability theorem; the third step is obtained due to fact that $\sum_{i=0}^{m-1} l_{k,i}$ is the sum of m i.i.d. exponential random variable with mean $\frac{1}{(1-p)\rho}$, which is independent of m_k; and the last step results by plugging in $Pr(m_k = m)$ shown as (B.12).

Putting (4.18) into (4.17) and simplifying it, we have

$$E\left[N_p^{X_{opt}}\right] = \frac{(1-p)p\rho(1-e^{-\rho 2r_0})(d-2r_I)}{(1-p)p\rho(1-e^{-\rho 2r_0})R_c + p - pe^{-\rho 2r_0} + e^{-p\rho 2r_0}}. \tag{4.19}$$

Combining (4.11) and (4.19), we have:

$$\lim_{t\to\infty}\frac{D_V(t)}{t} = \frac{w_V(1-p)p\rho(1-e^{-\rho 2r_0})(d-2r_I)}{(1-p)p\rho(1-e^{-\rho 2r_0})R_c + p - pe^{-\rho 2r_0} + e^{-p\rho 2r_0}}. \tag{4.20}$$

4.2.3 Achievable Capacity

In this subsection, we first give the final result of the capacity achieved by the VoIs by combing the results from Section 4.2.1 and 4.2.2; then we analyse the capacity achieved by eastbound and westbound VoIs separately to demonstrate the relationship between capacity achieved by eastbound and westbound VoIs and their vehicular density.

Total Achievable Capacity

Combining the capacity achieved by the VoIs from V2I communications and V2V communications from one cycle with length d shown in Section 4.2.1 and 4.2.2, the total capacity achieved by the VoIs from a highway segment with length L can be readily obtained as follows:

$$\begin{aligned}\eta_c =& \frac{L}{d}\lim_{t\to\infty}\frac{D_{V2I}(t)+D_{V2V}(t)}{t}\\ =& \frac{L}{d}\min\left\{w_I\left(1-e^{-\rho 2r_I}\right), w_I\left(1-e^{-p\rho 2r_I}\right) + \frac{w_V c_2(d-2r_I)}{c_2R_c + p - pe^{-\rho 2r_0} + e^{-p\rho 2r_0}}\right\},\end{aligned} \tag{4.21}$$

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

Remark 4.2. *It is interesting to note from* (4.21) *that the achievable capacity does not depend on the speed of vehicles, which appears to be counter-intuitive at first sight. This can be explained from the data dissemination process. As* (4.6) *and* (4.11) *show, both the capacity achieved by the VoIs from infrastructure and from helpers only depend on the spatial distribution of vehicles. In our system, the vehicles' arrival follows a Poisson process and the vehicles move at a constant speed. Therefore, the spatial distribution of the vehicles are both stationary and ergodic [100] (ignoring the finite length of the road segment L). It follows that the capacity that can be achieved by the VoIs is independent of vehicular speed. This observation implies that when vehicles arrive following a Poisson process, our analysis assuming the constant speed model is also applicable to other time-varying speed models, e.g., the Gaussian speed model [21], as long as the resulting spatial distribution of vehicles is time-invariant, i.e., stationary.*

Remark 4.3. *In the extreme case when all vehicles have download requests, i.e., when $p = 1$, the achievable capacity from one cycle is $\eta_{cycle} = \eta_{max} = w_I \left(1 - e^{-\rho 2 r_I}\right)$, which is exactly the achievable capacity when all vehicles directly receive data from the infrastructure without cooperative V2V communications. It can be further established that when p is greater than a certain threshold, cooperative V2V communications between the helpers and the VoIs are of little use in boosting the capacity. This can be explained by that in the particular scenario considered in the chapter, all new data comes from outside the vehicular network. V2V communications between the helpers and the VoIs only help to extend the communication range of the VoIs when there is no VoI in the infrastructure's coverage and balance data among the VoIs, but cannot increase the net amount of data available in the vehicular network. Therefore, when the density of the VoIs is high, the probability that there is no VoI in the infrastructure's coverage is negligible and thus it is more beneficial for the VoIs to retrieve data directly from the infrastructure. In this situation, V2V cooperative communications offer little benefit in boosting the vehicular network capacity. Unsurprisingly, following the argument outlined earlier, when the vehicular density is sufficiently large, even for a small value of p, the benefit of V2V cooperative com-*

munications vanishes very quickly. This can be also validated using (4.21)*.*

Capacity achieved by eastbound and westbound VoIs

In this subsection, we analyse the capacity achieved from one cycle by all eastbound and westbound VoIs separately, denoted by η_e and η_w respectively. The following theorem summarizes the results.

Theorem 4.5. *The capacities achieved from one cycle by the eastbound and westbound VoIs, are proportional to the traffic density of eastbound vehicles and westbound vehicles respectively, which are given by:*

$$\eta_e = \eta_{cycle} \cdot \frac{\rho_1}{\rho_1 + \rho_2}, \tag{4.22}$$

and

$$\eta_w = \eta_{cycle} \cdot \frac{\rho_2}{\rho_1 + \rho_2}. \tag{4.23}$$

Proof. See Appendix C. □

Remark 4.4. *It is interesting to note that the capacities achieved by VoIs travelling in each direction are strictly proportional to the vehicular densities of that direction respectively. This can be explained by the fact that when the vehicular density in a direction increases, there is a higher chance for the VoIs travel in that direction to communicate with the infrastructure and to receive more data indirectly from the helpers via V2V communication. As a VoI travels in a direction is statistically indistinguishable from another VoI travels in the same direction, this result suggests that the achievable throughput by a (any) VoI, no matter which direction the VoI is travelling in, will be statistically the same.*

4.3 Simulation and Discussion

In this section we conduct Monte-Carlo simulations to establish the accuracy of the theoretical analysis and discuss its insights. Specifically, we set the length of a highway segment L=100km. Eastbound and westbound vehicles move at constant speeds of v_1=20m/s and v_2=25m/s respectively. The radio ranges of infrastructure points and vehicles are 400m and 200m (typical radio ranges using DSRC [49]) respectively. The data rates of V2I and V2V communications are w_I=20Mb/s and w_V=2Mb/s. Each simulation is repeated 2000 times and the average value is shown in the plot.

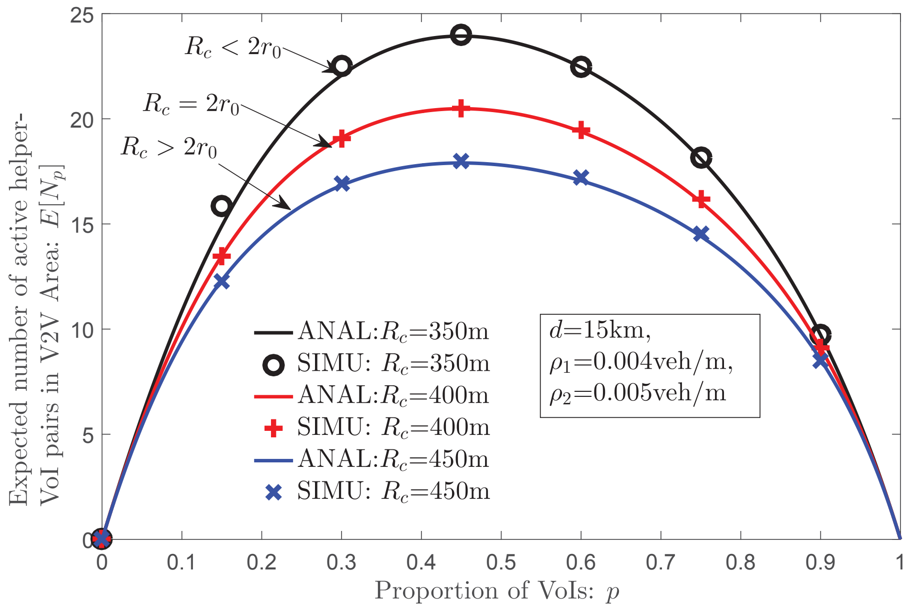
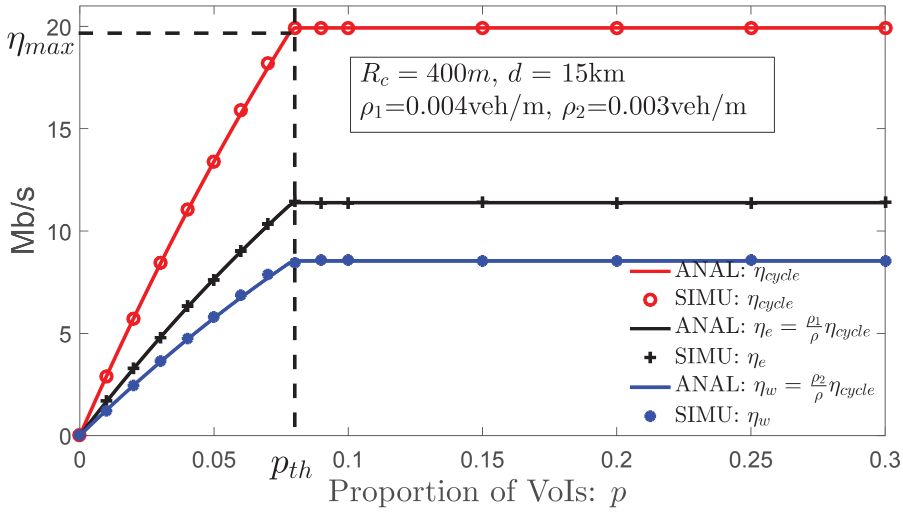


Figure 4.4 : A comparison of the expected number of simultaneously active helper-VoI pairs in one V2V Area with respect to the proportion of VoIs p between simulation and analysis, for different sensing ranges R_c.

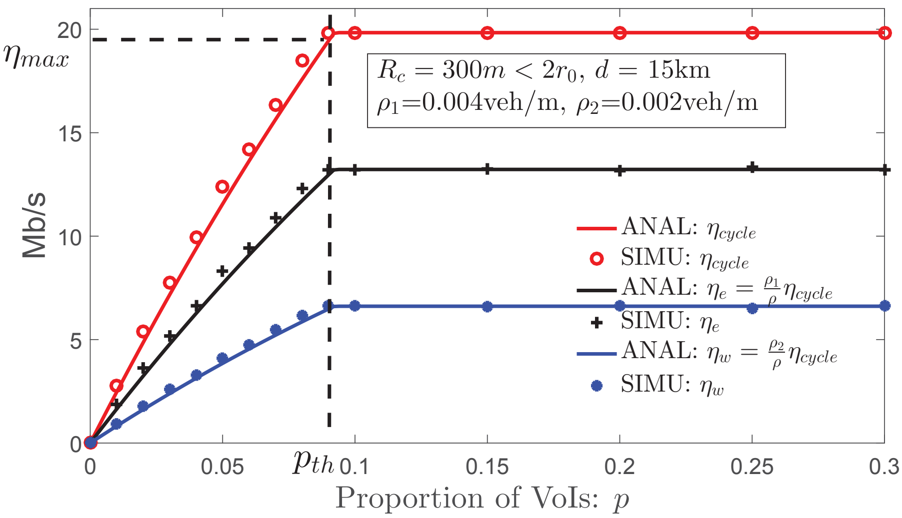
Fig. 4.4 shows a comparison between the expected number of simultaneous active helper-VoI pairs in a V2V Area from analysis and simulation, under three different sensing ranges R_c. It is shown that the analytical result match perfectly with the simulation result when $R_c \geq 2r_0$. When $R_c < 2r_0$, there is a marginal gap between the simulation and analytical result, and the gap reduces with an increase of p. This is due to the fact that when $R_c < 2r_0$, other things being equal, with an increase of p, the density of helpers becomes smaller. Therefore, it is less likely to occur the scenario discussed in the proof of Theorem 4.3 that the VoI of the k-th ($k \geq 1$) helper-VoI pair is located within the coverage of the helper $V_{k,1}$, helper $V_{k,2}, \cdots$. Consequently, the approximations used in (4.14) and (4.15) become more accurate. Furthermore, as expected, a higher value of sensing range R_c results in a lower expected number of simultaneous helper-VoI pairs.

Fig. 4.5 shows a comparison of the capacities achieved from one cycle by all VoIs, by all eastbound VoIs, and by all westbound VoIs. Both Fig. 4.5(a) and Fig. 4.5(b) show that the analytical results match very well with simulations. Furthermore, it can be seen that the capacity achieved by the VoIs travelling towards each direction (η_e and η_w respectively) is exactly proportional to the traffic density in that direction, as predicted by Theorem 4.5.

Fig. 4.5 also reveals the relationship between the capacity and the proportion of VoIs p, and shows that the capacity increases to its maximum value when the proportion of VoIs is larger than a certain threshold. Beyond that threshold, a further increase in p has little impact on the capacity. Specifically, as shown in Fig. 4.5(a), when p is small, the capacity increases sharply with an increase of p; however, when p increases beyond a certain threshold, e.g., p_{th}=0.08 in this case, a further increase in p has no impact on the capacity. This can be explained by that when $p < p_{th}$, the number of VoIs is insufficient to retrieve all the data received by the helpers from their V2I communications. That is, vehicular networks offer more data (capacity) than that can be retrieved by the VoIs and the capacity is limited by the V2V communications between the VoIs and the helpers. Therefore, an increase in p would significantly increase the number of simultaneous active helper-VoI pairs and consequently boost the capacity. However, when the proportion of VoIs reaches a certain threshold, VoIs can retrieve almost all the data received by the helpers from their V2I communications. In this case, the capacity achieved by the VoIs approaches its maximum $\eta_{max} = w_I(1 - e^{-\rho 2 r_I})$, which is equal to the average data



(a) Case of $R_c \geq 2r_0$

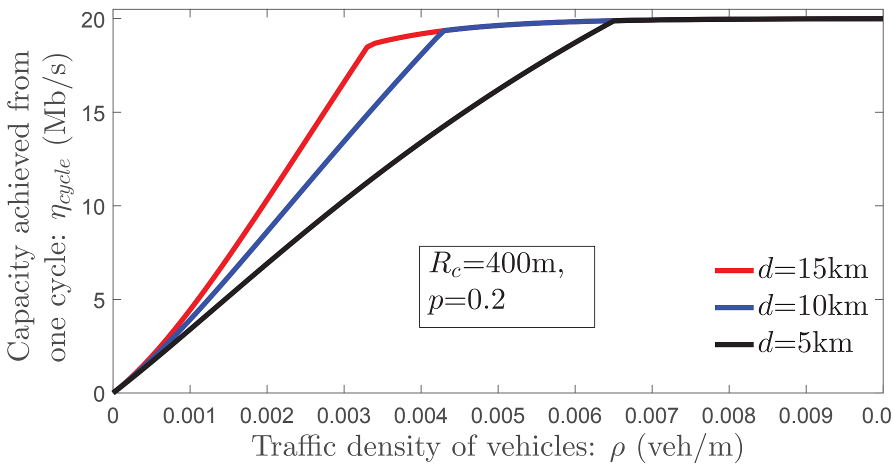


(b) Case of $R_c < 2r_0$

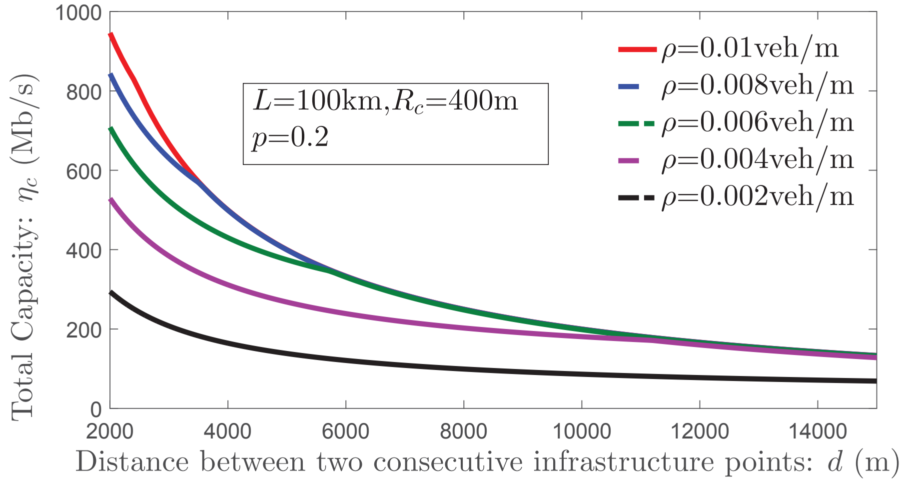
Figure 4.5 : A comparison of the capacities achieved from one cycle by all VoIs, by all VoIs in the eastbound direction, and by all VoIs in the westbound direction as a function of the proportion of VoIs p.

rate the infrastructure point delivers its data to all vehicles, including both VoIs and helpers.

Fig. 4.6 demonstrates the relationship between the capacity and the inter-



(a) Achievable capacity from one *cycle* under different d.



(b) Total achievable capacity under different traffic density.

Figure 4.6 : Relationship between capacity, distance between adjacent infrastructure points and vehicular density.

infrastructure distance, and gives insight into the optimum vehicular network infrastructure deployment in terms of their inter-distances under different vehicular density. It is shown in Fig. 4.6(a) that when the vehicular density ρ is small, a larger d will lead to a larger capacity achieved from one cycle because a large d will increase the capacity achieved by the VoIs from V2V communications with helpers. However with an increase of ρ, the capacity achieved by the VoIs from one cycle under different values of d differ marginally and converge to the same maximum value. This can be explained by considering that when ρ is large, most of the VoIs can receive data directly from the infrastructure, and the contribution from V2V communications with helpers becomes less significant. Even though an increase in d would help to boost the capacity achieved from one cycle when the vehicular density is small, Fig. 4.6(b) shows that the total achievable capacity decreases with an increase of d. This is due to the fact that an increase in d on one hand brings marginal improvement on the capacity achieved from one cycle, on the other hand, it reduces the number of cycles, which consequently leads to a reduction in the total capacity. Furthermore, it can be seen that to achieve the same capacity, when the vehicular density is larger, the inter-infrastructure distance needs to be higher. Therefore, when determining the optimum deployment of vehicular network infrastructure, it is important to take the vehicular density into account, e.g., in areas where the vehicular density is usually large, by utilizing a cooperative communication strategy, the number of infrastructure points can be reduced.

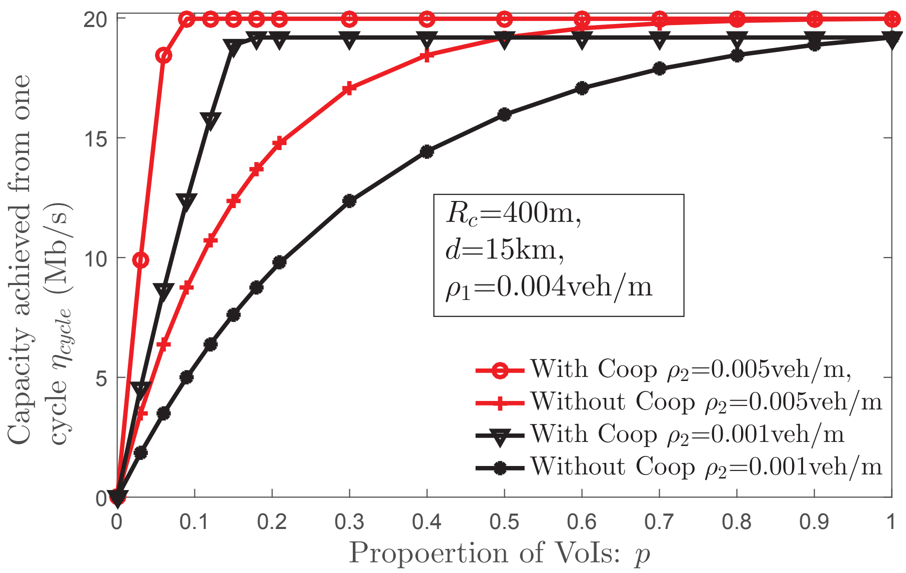


Figure 4.7 : A comparison between the capacity achieved from one cycle, with and without cooperative communications.

Fig. 4.7 compares the capacity from one cycle using our cooperative communication strategy (labelled as With Coop) with its non-cooperative counterpart (labelled as Without Coop), and shows that our cooperative communication strategy can improve the capacity, even when there is only a small proportion of vehicles with download requests, i.e., a small p. The result for the non-cooperative counterpart is obtained by letting the VoIs only receive data from infrastructure. It is shown that with an increase in p, the proposed cooperative communication strategy becomes less effective in improving the capacity. This is due to the fact that a larger p leads to a smaller number of helpers, which results in a reduction in the amount of data the helpers can help to retrieve from the infrastructure. Thus, the contribution to the capacity from the proposed cooperative communication strategy becomes less significant. Furthermore, we can see that under the same network setting, without using the cooperative communication strategy, only when all vehicles have download requests, i.e., $p = 1$, the maximum capacity $\eta_{max} = w_I(1 - e^{-\rho 2r_I})$ can be achieved. In contrast, with the cooperative communication strategy, this maximum capacity η_{max} can be achieved even when a small proportion of vehicles have download requests. This validates the effectiveness of cooperative communications to boost network performance.

Fig. 4.8 compares the achievable capacity assuming our proposed cooperative communication strategy (labelled as Coop) with that assuming the strategy proposed in [49] (labelled as ChainCluster). It can be seen that our scheme achieves better performance in terms of the achievable capacity than that proposed in [49]. This is due to the fact that in [49], the authors only used the cooperation among vehicles moving in the same direction and within the same cluster of the VoIs, while in our strategy, both cooperation among infrastructure and cooperation of vehicles travelling in each direction are utilized to help the VoIs' download, which improves the capacity achieved by the VoIs.

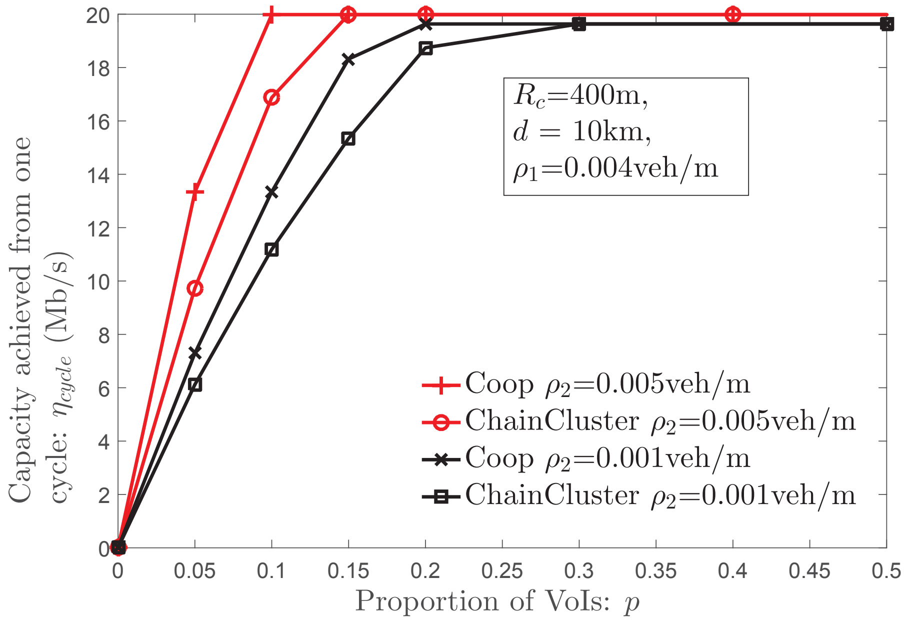


Figure 4.8 : A comparison between the capacity achieved from one cycle assuming our proposed strategy and that assuming the strategy proposed in [49].

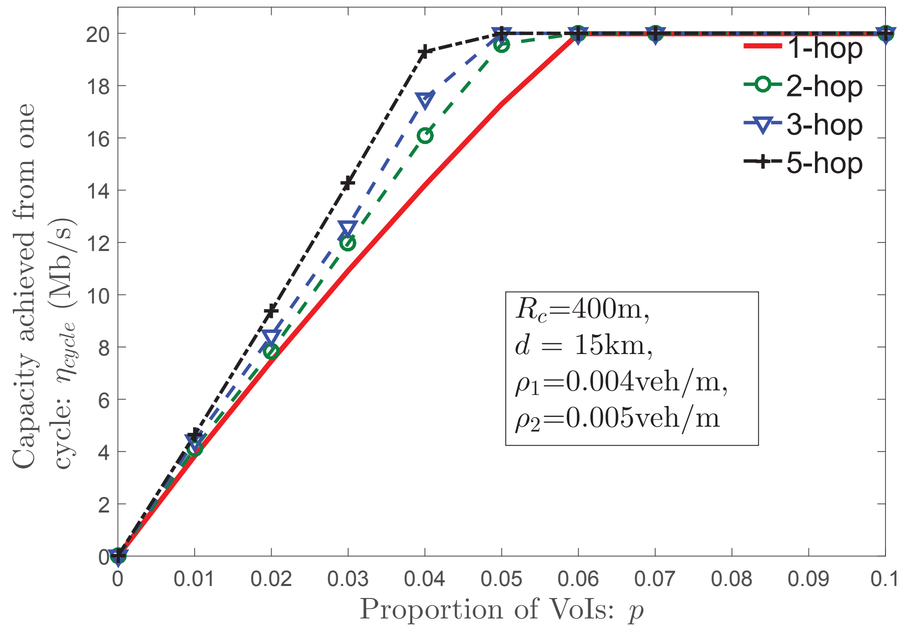


Figure 4.9 : A comparison between capacity achieved from one cycle when allowing one-hop communication and multi-hop communications.

Fig. 4.9 compares the capacity achieved by allowing only one-hop communication and allowing both k-hop ($k = 2, 3, 5$) V2I communications between the VoIs and infrastructure and k-hop V2V communication between the VoIs and helpers. It is shown that allowing multi-hop communications beyond one hop has little impact on the capacity. This can be explained by the fact that in the specific scenario being considered, there are only a subset of vehicles with download requests (VoIs), and all other vehicles (helpers) assist the VoIs to receive more data. Any new data in the vehicular network must come from the infrastructure. Therefore, allowing multi-hop V2V communications only helps to balance the distribution of information among vehicles but does not increase the net amount of information available in the network. The marginal increase in the achievable capacity comes from multi-hop V2I communications between the VoIs and infrastructure, because it allows the VoIs to have a longer connection time (via some intermediate vehicles) with the infrastructure. This increase only occurs when the proportion of VoIs, p, is smaller than a threshold, e.g., $p_{th} = 0.06$ in the considered scenario.

Fig. 4.10 compares capacity achieved from the constant channel model with that from the time-varying channel model used in [54], and shows that our analysis under the constant channel model is applicable to a more realistic time-varying channel model. The parameters of the time-varying channel model have been given in Section 3.3. It is obvious from Fig. 4.10 that the achievable capacity from the above two channel models match each other. This phenomenon can be explained by equation (4.21) which shows that the achievable capacity is a linear function of w_I and w_V. Then, it follows that $E[\eta_c(w_I, w_V)] = \eta_c(E[w_I], E[w_V])$, which implies that for time-varying channels, the time-varying values of $w_I^{'}$ and $w_V^{'}$ can be replaced by the respective time-averaged throughput of V2I and V2V communications and our analysis still applies.

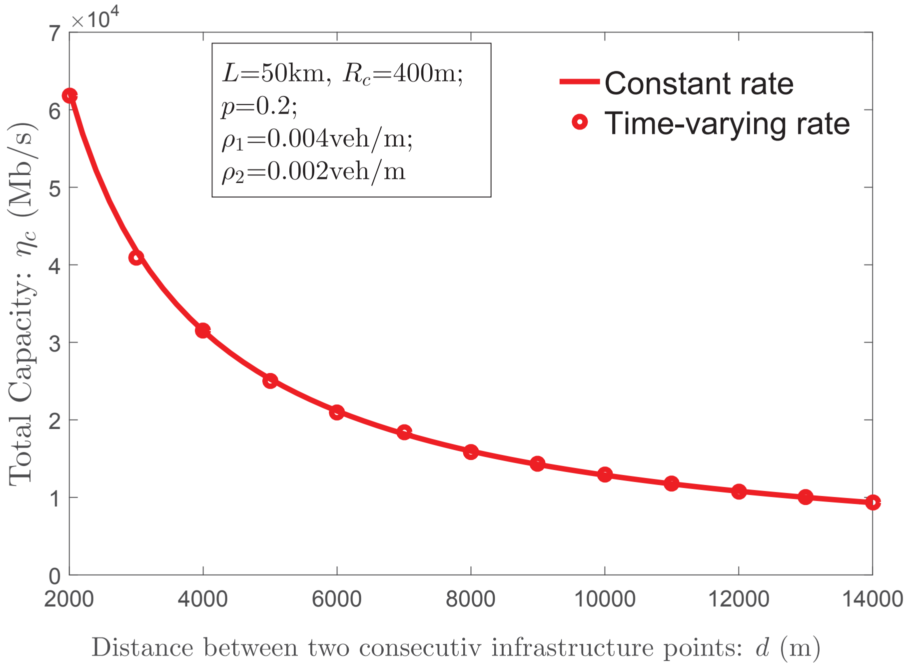


Figure 4.10 : A comparison between capacity achieved from the constant channel model and the time-varying channel model which considers Rayleigh fading and path loss.

4.4 Summary

In this chapter, we extended the work introduced in Chapter 3 and considered a scenario with multiple vehicles having download requests in the network. We analysed the capacity of vehicular networks with a finite traffic density adopting the cooperative communication strategy proposed in Chapter 3, and obtained its closed-form expression. Our result showed that the proposed cooperative strategy can improve the capacity of vehicular networks, and the improvement is more pronounced when the proportion of vehicles with download requests is low. Moreover, our result sheds insight into the optimum design of cooperative vehicular networks for ITS to achieve fast and reliable data dissemination.

Chapter 5

A Topological Approach to Secure Message Dissemination in Vehicular Networks

In this chapter, we consider vehicular networks containing insider malicious vehicles that may tamper with the content of messages to disrupt their successful delivery. We are interested in investigating topology-based decision algorithms to keep vehicles from being misguided by false messages. By incorporating the underlying network topology information, we propose an optimal decision algorithm that is able to maximize the chance of making a correct decision on the message content, assuming the prior knowledge of the percentage of malicious vehicles in the network. Furthermore, a novel heuristic decision algorithm is proposed that can make decisions without the aforementioned knowledge of the percentage of malicious vehicles. Simulations are conducted to compare the security performance achieved by our proposed decision algorithms with that achieved by existing ones that do not consider or only partially consider the topological information, to verify the effectiveness of the algorithms. Our results show that by incorporating the network topology information, the security performance can be much improved. This work shed light on the optimum design of cooperative vehicular networks for ITS to achieve secure message dissemination.

The rest of this chapter is organized as follows: Section 5.1 introduces the system model and the problem formation. The optimum decision algorithm and the heuristic decision algorithm are presented in Section 5.2 and Section 5.3 respectively. In Section 5.4, we conduct simulations to validate the effectiveness of our proposed decision algorithms and discuss their insight. Section 5.5 summarizes this chapter.

5.1 System Model and Problem Formation

In this section, we first introduce the system model, including the network model, message dissemination model, and the attack model. Then, we give a rigorous description of the research problem addressed in this chapter.

5.1.1 Network and Message Dissemination Model

We consider a vehicular network where each vehicle has a unique ID number that is registered in certification authority to represent its identity, and vehicles cannot forge their own or other vehicles' ID numbers. That is, we assume all vehicles are legitimate vehicles that have passed the authentication process conducted by the certification authority [42, 44].

Specifically, consider that there is a vehicle in the network (termed as the source vehicle) intending to deliver a message about the road condition to inform other vehicles further away. The road condition information can be abnormal situations, e.g., congestion, hazardous road conditions such as traffic accident, slippery road, etc., or normal situation, e.g., uncongested traffic. We assume that the content of message takes value from $\{0, 1\}$, and 1 represents abnormal road condition and 0 represents normal road condition. When the source vehicle sends the source message m_0 on road condition, it will also send the location information applied to the road condition along with message m_0, to help other vehicles in the network make a better route choice. It is worth noting that the road situation can also be described as a multi-variable vector and these variables can be correlated [42], e.g., one such variable can be traffic congestion state and another can be accident state. We denote the content of message on road condition transmitted by the source vehicle, which represents the actual road condition, by m_0, $m_0 \in \{0, 1\}$. Other vehicles do not know the true value of m_0 a priori.

The message is forwarded from the source vehicle in a broadcast and multi-hop [86, 101] manner to other vehicles with the help of relay vehicles. Relay vehicles can be any vehicle along the message propagation path. Each time when a relay vehicle receives a message from its neighbor, it will forward the message. That is, when a relay vehicle gets the same message several times (from different neighbors), it will forward as many times as it receives the message. Multi-path forwarding makes it challenging for the attackers to influence all message forwarding paths [41], therefore helps to improve the message security. When a vehicle transmits the message on road condition to other vehicles, it transmits its identity information, i.e., ID number, along with the message. This is commonly adopted in vehicular network applications and can be achieved by some standard signature approach [10, 39]. Using this, any vehicle in the network is able to obtain an integrity-protected path list that records the relay vehicles of the corresponding received message on road condition.

5.1.2 Attack Model

Vehicles in the network can be classified into two categories: *normal vehicles,* which behave normally and will forward the received message without any alteration, and *malicious vehicles,* which may tamper the received message. It is assumed that vehicles cannot forge their own or other vehicles' ID numbers, and the path list transmitted along with the message is protected by signature approach. Therefore, malicious vehicles can only tamper the content part of the message, but cannot tamper the path list record. Malicious vehicles are uniformly and randomly distributed in the system with proportion p. It follows that the probability of a vehicle being a malicious vehicle is p, independent of the event that another distinct vehicle is a malicious vehicle.

We assume that the source vehicle is normal and only relay vehicles may be malicious. We acknowledge there is possibility that the source vehicle can be malicious, and it is also an important scenario when investigating the vehicular network message security issue. In this work, our main focus is to design a topological approach to address the message inconsistency issue resulted from the message dissemination process. Therefore, we assume the source vehicle is normal, and only relay vehicles can be malicious. We will leave the work that removes this assumption as our future work. Besides, we assume the normal vehicles do not know which vehicles are normal or malicious. On the other hand, we consider the most unfavourable situation for secure message dissemination in vehicular networks that malicious vehicles not only know which vehicles are malicious, but also are capable of communicating with each other via back channels of infinite bandwidth [102]. That is, we assume malicious vehicles collaborate with each other and they also know what the correct message m_0 transmitted by the source vehicle is. Therefore, if a malicious vehicle receives the correct message, it will tamper it to the incorrect one, i.e., different from message m_0; and if it receives the incorrect message, it will directly forward it to others. This implies that as long as a message is relayed by at least one malicious vehicle, the message would be incorrect. Fig. 5.1 gives a simple example of message dissemination process when there are insider attackers in the network.

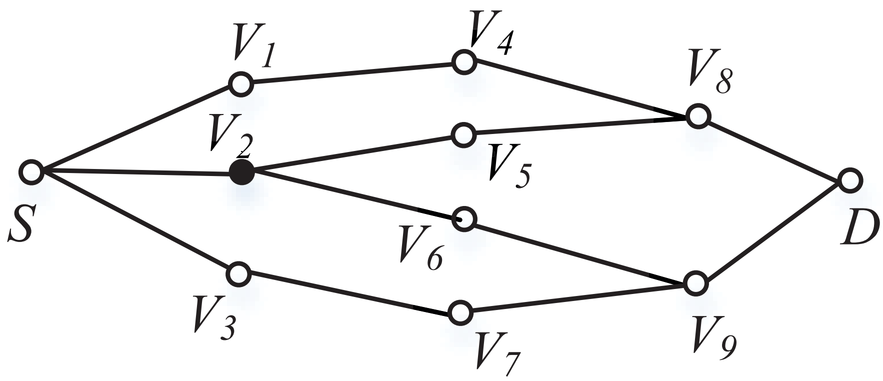


Figure 5.1 : An illustration of a vehicular network when there exists a malicious vehicle V_2 who would tamper with the content of message.

5.1.3 Problem Formation

Now we give a detailed description of the research problem considered in this chapter.

We consider that there is a vehicle, which is several hops away from the source vehicle, trying to make a decision on the message content when it receives several copies of message, and we call it the destination vehicle. Note that the destination vehicle can be any vehicle along the message dissemination path. From the time instant the destination vehicle receives the first message, it waits time period T to receive more messages before making a final decision. T characterizes the response time requirement on the decision, and a larger T potentially allows the vehicle to receive more messages. We assume the destination vehicle receives k copies of the message during its waiting time period T and the number of relay vehicles that participate in relaying the k copies of message from the source vehicle to the destination vehicle is n. The value of k will impact the integrity of the decision, and we will discuss its impact later in the simulation. In the following analysis, we regard k and n are constants, namely, we consider a network containing n relay vehicles and k paths between the source vehicle and the destination vehicle.

Denote the k messages on road condition received by the destination vehicle by M_i, $i = 1, 2, ...k$, $M_i \in \{0, 1\}$. As each message on road condition is transmitted together with a specific delivery path from the source vehicle to the destination vehicle, we number the corresponding paths by $L_1, L_2, ...L_k$. In addition, we number the relay vehicles by $V_1, V_2, ...V_n$. A vehicle V_i may belong to one or more paths.

Note that due to the existence of malicious vehicles who may tamper with the content of the message, the k copies of the message received by the destination vehicle can be in conflict instead of being consistent with each other. Furthermore, with the potential existence of some shared relay vehicles in different paths, the k

messages received from k different paths may not be independent. These correlations are all contained in the information of message dissemination paths. Therefore, we construct a topology matrix to represent the underlying network topological correlation. Specifically, based on the path information derived from the received messages, the destination vehicle can readily construct a $k \times n$ topology matrix B, where each row represents a path, each column a node (vehicle), and the (i, j)-th entry B_{ij} being an indicator whether vehicle V_j belongs to path L_i:

$$B_{ij} = \begin{cases} 1, & \text{if vehicle } V_j \text{ belongs to path } L_i \\ 0, & \text{else} \end{cases} \tag{5.1}$$

In this chapter, we are interested in investigating decision algorithms for the destination vehicle to make a correct decision on the content of the disseminated message against attacks from malicious vehicles by utilizing the underlying network topology information. Denote by d, $d \in \{0, 1\}$ the final decision on the content of message made by the destination vehicle. If the decision is the same as the source message, i.e., if $d = m_0$, we say the destination vehicle makes a correct decision, otherwise we say it makes an incorrect decision. We use the probability of correct decision, denoted by P_{succ}, as the performance metric to measure the secure message dissemination performance, and P_{succ} can be formally defined as follows:

$$P_{succ} = \Pr(d = 1, m_0 = 1) + \Pr(d = 0, m_0 = 0) \tag{5.2}$$

In the following two sections, we will propose two decision algorithms to improve the message security performance in vehicular networks by utilizing the underlying network topology information. First, we will propose an optimum decision algorithm that maximizes the probability of correct decision P_{succ} based on Bayes decision theory, assuming the percentage of malicious vehicles in the network is known. A detailed implementation of the algorithm will be provided to illustrate how a destination vehicle makes the decision on the message content according to this prior knowledge and the network topology information. Then, we will introduce a heuristic decision algorithm based on the Maximum Likelihood Estimation. This heuristic decision algorithm will enable a vehicle to make a decision when receiving conflicting messages purely based on network topology information, without the need for knowing the percentage of malicious vehicles, which can be difficult to estimate in some circumstances. Therefore, the heuristic algorithm is easier to implement in practice.

5.2 Optimum Decision Algorithm

In this section, we propose a decision algorithm which aims to optimize the secure message dissemination performance in terms of maximizing the probability of a correct decision P_{succ}, that is,

$$\max \; P_{succ}, \tag{5.3}$$

where P_{succ} is given by (5.2).

In the following, we will first present the optimum decision algorithm followed by a detailed proof to prove its optimality, and then we will introduce its detailed implementation and discuss its limitation in practical realization.

5.2.1 Optimum Decision Algorithm

The following theorem summarizes the optimum decision algorithm we are investigating to maximize P_{succ}.

Theorem 5.1. *Consider a destination vehicle receives k copies of messages $M_1 = m_1, M_2 = m_2, ...M_k = m_k$. Given the prior knowledge of the probabilities that the occurrence of abnormal event of interest, e.g., traffic congestion, are $P_1 = Pr(m_0 = 1)$, and $P_0 = 1 - P_1 = Pr(m_0 = 0)$, which can be estimated from empirical knowledge*

[103], the optimum decision algorithm that leads to (5.3) *can be shown as follows:*

$$d = \begin{cases} 1, & \frac{Pr(M_1=m_1,...M_k=m_k|m_0=1)}{Pr(M_1=m_1,...M_k=m_k|m_0=0)} > \frac{P_0}{P_1} \\ 0, & \frac{Pr(M_1=m_1,...M_k=m_k|m_0=1)}{Pr(M_1=m_1,...M_k=m_k|m_0=0)} < \frac{P_0}{P_1} \end{cases}, \tag{5.4}$$

and when $\frac{Pr(M_1=m_1,...M_k=m_k|m_0=1)}{Pr(M_1=m_1,...M_k=m_k|m_0=0)} = \frac{P_0}{P_1}$, *d is randomly chosen from 0 and 1 with equal probability.*

Proof. As introduced in [104, 105], the objective of a binary Bayes decision problem is to minimize the expectation of the decision cost, denoted by $U(d, m_0)$. Let U_{ij}, $i = 0, 1$, $j = 0, 1$, represents the cost of declaring the final result $d = i$ when actually the source message $m_0 = j$, and U_{ij} can be negative to represent the benefits of making a correct decision. As a ready consequence of the total probability theorem, the expectation of the decision cost $U(d, m_0)$ can be expressed as follows:

$$U(d, m_0) = \sum_{i=0}^{1} \sum_{j=0}^{1} U_{ij} \Pr\left(d = i,\ m_0 = j\right). \tag{5.5}$$

When assuming $U_{01} > U_{11}$ and $U_{10} > U_{00}$, which is reasonable considering the cost of making an incorrect decision is usually larger than that making a correct decision, the optimum decision algorithm that minimizes the expectation of the decision cost made by the destination vehicle given its k copies of received message $M_1 = m_1, M_2 = m_2, ...M_k = m_k$, is given by [105]:

$$d = \begin{cases} 1 & \frac{\Pr(M_1=m_1,...M_k=m_k|m_0=1)}{\Pr(M_1=m_1,...M_k=m_k|m_0=0)} > \frac{P_0(U_{10}-U_{00})}{P_1(U_{01}-U_{11})} \\ 0, & \frac{\Pr(M_1=m_1,...M_k=m_k|m_0=1)}{\Pr(M_1=m_1,...M_k=m_k|m_0=0)} < \frac{P_0(U_{10}-U_{00})}{P_1(U_{01}-U_{11})} \end{cases}, \tag{5.6}$$

where $\Pr\left(M_1 = m_1, ...M_k = m_k|m_0 = 1\right)$ and $\Pr\left(M_1 = m_1, ...M_k = m_k|m_0 = 0\right)$ are the two conditional probabilities of the occurrence of event $M_1 = m_1, M_2 = m_2, ...M_k = m_k$, which characterize the correlations between received messages. Meanwhile, when a tie occurs, namely, when $\frac{\Pr(M_1=m_1,...M_k=m_k|m_0=1)}{\Pr(M_1=m_1,...M_k=m_k|m_0=0)} = \frac{P_0(U_{10}-U_{00})}{P_1(U_{01}-U_{11})}$, d is randomly chosen from 0 and 1 with equal probability.

From (5.5), when assuming the cost of making a correct decision is 0 and making an incorrect decision is 1, namely, by assuming $U_{00} = U_{11} = 0$ and $U_{01} = U_{10} = 1$, we have:

$$U(d, m_0) = \Pr(d = 0, m_0 = 1) + \Pr(d = 1, m_0 = 0) = 1 - P_{succ}. \tag{5.7}$$

It follows that a minimization of the expectation of the decision cost, is equivalent to a maximization of the probability of correct decision, namely, we have

$$\min\ U(d, m) \iff \max\ P_{succ} \tag{5.8}$$

Therefore, the optimum decision algorithm we are investigating for the optimization problem (5.3) is exactly the decision algorithm that provides a solution to the classical Bayes decision problem in a special case, shown as (5.4), which finalizes the proof. □

Remark 5.1. *It can be seen from* (5.4) *that, given the probabilities of the occurrence of an abnormal event of interest, P_0 and P_1, the decision on $d = 1$ or $d = 0$ depends on the ratio $\frac{Pr(M_1=m_1,...M_k=m_k|m_0=1)}{Pr(M_1=m_1,...M_k=m_k|m_0=0)}$. That is, given a set of received messages $M_1 = m_1, M_2 = m_2, ...M_k = m_k$, the destination vehicle needs to calculate the probability that the event $M_1 = m_1, ...M_k = m_k$ occurs if the true message m_0 is 1, denoted as $Pr(M_1 = m_1, ...M_k = m_k|m_0 = 1)$ and the probability that the event occurs if the true message m_0 is 0, denoted as $Pr(M_1 = m_1, ...M_k = m_k|m_0 = 0)$. A decision on d is then made by comparing the value of $\frac{Pr(M_1=m_1,...M_k=m_k|m_0=1)}{Pr(M_1=m_1,...M_k=m_k|m_0=0)}$ and $\frac{P_0}{P_1}$. Therefore, calculating of the two probabilities is the critical part of implementing the algorithm in practice.*

In summary, the optimum decision algorithm for the destination vehicle to maximally make a correct decision on the message content works as detailed in Algorithm 5.1, where the details of calculating the two terms $\Pr(M_1 = m_1, ...M_k = m_k|m_0 = 1)$ and $\Pr(M_1 = m_1, ...M_k = m_k|m_0 = 0)$ will be given in the following subsection.

Algorithm 5.1 Optimum Decision Algorithm

Input: $M_1, M_2, ...M_k$, P_0, P_1, p

Output: d

1: Construct topology matrix B based on the paths information derived from the received k copies of message;

2: Calculate the two probabilities $\Pr(M_1 = m_1, ...M_k = m_k | m_0 = 1)$ and $\Pr(M_1 = m_1, ...M_k = m_k | m_0 = 0)$ according to (5.10) and (5.13) respectively, given the network topology information and a prior knowledge on the proportion of malicious vehicles in the network;

3: **if** $\frac{\Pr(M_1=m_1,...M_k=m_k|m_0=1)}{\Pr(M_1=m_1,...M_k=m_k|m_0=0)} > \frac{P_0}{P_1}$ **then**

4: $d = 1$;

5: **else if** $\frac{\Pr(M_1=m_1,...M_k=m_k|m_0=1)}{\Pr(M_1=m_1,...M_k=m_k|m_0=0)} < \frac{P_0}{P_1}$ **then**

6: $d = 0$;

7: **else**

8: d is randomly chosen from 0 and 1 with equal probability;

9: **end if**

5.2.2 Algorithm Implementation

In this section, we will introduce the detailed implementation of the proposed optimum decision algorithm. As discussed in Remark 5.1, the foremost step is to calculate $\Pr(M_1 = m_1, ...M_k = m_k | m_0 = 1)$ and $\Pr(M_1 = m_1, ...M_k = m_k | m_0 = 0)$ as they are prerequisite to obtaining the final decision d.

The main idea behind the calculation of $\Pr(M_1 = m_1, ...M_k = m_k | m_0 = 1)$ and $\Pr(M_1 = m_1, ...M_k = m_k | m_0 = 0)$ is as follows. We classify vehicles into three different types based on the paths they belong to. We call a vehicle a *Type 0* (or *Type 1*) vehicle if it only belongs to paths that deliver messages with content 0 (or 1) to the destination vehicle, and a vehicle is a *Type 2* vehicle (if any) if it belongs to at least one path that delivers a message with content 0 and one path that delivers a message with content 1 to the destination vehicle. This follows that, the paths which deliver messages with content 1 to the destination vehicle contain Type 1 and Type 2 vehicles, and the paths which deliver messages with content 0 to the destination vehicle contain Type 0 and Type 2 vehicles. Therefore, by separating the paths according to the delivered message contents, it can be readily concluded that given $m_0 = 1$, all the Type 1 and Type 2 vehicles are normal vehicles, meanwhile malicious vehicles only exist among Type 0 vehicles. Then, by listing and analysing all the different combinations of malicious vehicles among the Type 0 vehicles, we can obtain the result of our target conditional probability $\Pr\left(M_1 = m_1, ...M_k = m_k | m_0 = 1\right)$. The idea of calculating $\Pr\left(M_1 = m_1, ...M_k = m_k | m_0 = 0\right)$ is with the same.

In the following, we will first demonstrate the method of constructing topology matrix B based on the above idea, and then calculate the two probabilities $\Pr\left(M_1 = m_1, ...M_k = m_k | m_0 = 1\right)$ and $\Pr\left(M_1 = m_1, ...M_k = m_k | m_0 = 0\right)$. Without loss of generality, we assume that among the k copies of messages $M_1 = m_1, ...M_k = m_k$ received by the destination vehicle, there are exactly k_1, messages with content 1 and the other $k - k_1$ messages with content 0. Note that $k_1 = 0$ and $k_1 = k$ are both trivial cases implying no conflict on the received messages so that the decision is straightforward, therefore we only consider the case when $0 < k_1 < k$.

Constructing the topology matrix B

Specifically, recall the definition of the topology matrix given in (5.1), that each row corresponds to a path and each column corresponds to a vehicle. Based on the idea discussed above to calculate the probabilities $\Pr\left(M_1 = m_1, ...M_k = m_k | m_0 = 1\right)$ and $\Pr\left(M_1 = m_1, ...M_k = m_k | m_0 = 0\right)$, we construct the network topology matrix

B in the following form:

$$B = \begin{bmatrix} B_1 & B_{s_1} & \mathbf{0} \\ \mathbf{0} & B_{s_0} & B_0 \end{bmatrix}, \tag{5.9}$$

where B_1, B_0, B_{s_1} and B_{s_0}, if existing, are non-zero matrices, and $\begin{bmatrix} B_1 & B_{s_1} & \mathbf{0} \end{bmatrix}$ is a $k_1 \times n$ sub-matrix corresponding to the paths that deliver messages with content 1 to the destination vehicle, and $\begin{bmatrix} \mathbf{0} & B_{s_0} & B_0 \end{bmatrix}$ is a $(k - k_1) \times n$ sub-matrix corresponding to the paths that deliver messages with content 0 to the destination vehicle. Meanwhile, the columns of B_1 and B_0 correspond to vehicles that only belong to paths that deliver messages with content 1 and that deliver messages with content 0 to the destination vehicle respectively, i.e., Type 1 vehicles and Type 0 vehicles respectively. The columns of sub-matrix $\begin{bmatrix} B_{s_1} \\ B_{s_0} \end{bmatrix}$ correspond to all the Type 2 vehicles. Assume that the number of Type 1 and Type 0 vehicles are n_1 and n_0 respectively, $0 \leq n_1 + n_0 \leq n$, and the number of Type 2 vehicles is $n_2 = n - n_1 - n_0$. It follows that matrix B_1 and B_0 are with size $k_1 \times n_1$ and $(k - k_1) \times n_0$ respectively, and matrix $\begin{bmatrix} B_{s_1} \\ B_{s_0} \end{bmatrix}$ is with size $k \times (n - n_1 - n_0)$.

It is worth noting that the above arrangement of columns and rows of matrix B represent a numbering sequence of vehicles and paths, so that it does not change the underlying topology in terms of paths information. Besides, the sub-matrix B_1 can be non-existent if $n_1 = 0$, i.e., when the paths that deliver messages with content 0 to the destination vehicle contains all the n vehicles in the network. Under this circumstance, $B = \begin{bmatrix} B_{s_1} & \mathbf{0} \\ B_{s_0} & B_0 \end{bmatrix}$. Similarly, the sub-matrix B_0 (or $\begin{bmatrix} B_{s_1} \\ B_{s_0} \end{bmatrix}$) can also be non-existent when $n_0 = 0$ (or $n_2 = 0$).

Calculation of $\boldsymbol{Pr}\,(M_1 = m_1, ...M_k = m_k | m_0 = 1)$ ***and***
$\boldsymbol{Pr}\,(M_1 = m_1, ...M_k = m_k | m_0 = 0)$

In this part, we show the method of calculating the two conditional probabilities $\Pr(M_1 = m_1, ...M_k = m_k | m_0 = 1)$ and $\Pr(M_1 = m_1, ...M_k = m_k | m_0 = 0)$ based on the constructed topology matrix B. The following two theorems summarize the results.

Theorem 5.2. *Consider a destination vehicle receives k copies of message $M_1 = m_1, M_2 = m_2, ...M_k = m_k$, and among which k_1 messages are with content 1 and the other $k - k_1$ messages are with content 0, $0 < k_1 < k$. Conditioned on the source message $m_0 = 1$, the conditional probability of the occurrence of event $M_1 = m_1, ...M_k = m_k$ can be calculated as follows:*

$$
\begin{aligned}
&Pr\,(M_1 = m_1, ...M_k = m_k | m_0 = 1) \\
&= \begin{cases} (1-p)^{n-n_0} \cdot \left[\sum_{i=1}^{n_0} a_i \cdot p^i \,(1-p)^{n_0 - i}\right], & n_0 > 0 \\ 0, & n_0 = 0 \end{cases},
\end{aligned}
\tag{5.10}
$$

where n_0 is the number of vehicles that only belong to paths that deliver messages with content 0 to the destination vehicle, i.e., the number of Type 0 vehicles in the network, and $a_i, i = 1, 2, ...n_0$ is the number of combinations that exactly i malicious Type 0 vehicles leading to the occurrence of event $M_1 = m_1, ...M_k = m_k$.

Proof. When $n_0 = 0$, there are no Type 0 vehicles in the network, which implies that the paths that deliver messages with content 1 to the destination vehicle contains all the n vehicles in the network, and the topology matrix $B = \begin{bmatrix} B_1 & B_{s_1} \\ \mathbf{0} & B_{s_0} \end{bmatrix}$. Under this circumstance, conditioned on the source message $m_0 = 1$, when the event that k_1 messages are with content 1 occurs, all the n vehicles in the network should be normal vehicles. This leads to the event that the other $k - k_1$ messages are with content 0 occurs with probability 0. Therefore, we have $\Pr\left(M_1 = m_1, ...M_k = m_k | m_0 = 1\right) = 0$ when $n_0 = 0$.

When $n_0 > 0$, from the topology matrix B, we can conclude that if the matrix $\begin{bmatrix} B_{s_1} \\ B_{s_0} \end{bmatrix}$ exists, then the corresponding Type 2 vehicles should be all normal vehicles. This can be readily concluded by the fact that there is out of possibility for two paths sharing the same malicious vehicle would deliver different contents. Therefore, malicious vehicles exist either among Type 1 vehicles or among Type 0 vehicles.

Given the source message $m_0 = 1$, all the Type 1 vehicles should be normal vehicles to result in the occurrence of the event that the k_1 paths delivering messages with correct content 1. Malicious vehicles can only exist among Type 0 vehicles. Besides, the malicious Type 0 vehicles should be able to compromise all the $k - k_1$ paths (corresponding to the sub-matrix $\begin{bmatrix} \mathbf{0} & B_{s_0} & B_0 \end{bmatrix}$) to resulting in the occurrence of the event that all the $k - k_1$ paths delivering messages with incorrect content 0. Therefore, any combination of malicious vehicles should satisfy the following condition: by implementing element-wise *union* to their corresponding columns in sub-matrix B_0, i.e., implementing element-wise Boolean operation *OR* to them, the result should be a column with each entry equal to 1.

Note that the number of malicious type 0 vehicles can be any integer among $[1, n_0]$. We denote by event e_i that randomly choosing i columns from sub-matrix B_0 and then conducting element-wise union operation to them, there results in a column with each entry equal to 1. Denote by $a_i, i = 1, 2, ...n_0$ the total number of combinations that event e_i occurs. Therefore, we have

$$a_i = \sum_{j=1}^{z_i} I\left(\text{event } e_i \text{ occurs}\right), \tag{5.11}$$

where $z_i = \begin{pmatrix} n_0 \\ i \end{pmatrix}$, and $I(x)$ is an indicator function that $I(x) = 1$, when x is true;

otherwise $I(x) = 0$.

It then follows from the combination theory [96] that :

$$\Pr\left(M_1 = m_1, ...M_k = m_k | m_0 = 1\right) = (1-p)^{n-n_0} \cdot \left[\sum_{i=1}^{n_0} a_i \cdot p^i (1-p)^{n_0 - i}\right], \quad (5.12)$$

where the first part corresponds to the probability that the k_1 paths deliver messages with correct content 1, so that all the $n - n_0$ vehicles contained in these k_1 paths are therefore normal vehicles; and the second part is the probability that the $k - k_1$ paths deliver messages with incorrect content 0, which is a summation of all the probabilities of different malicious vehicle combinations. □

Theorem 5.3. *Consider a destination vehicle receives k copies of message $M_1 = m_1, M_2 = m_2, ...M_k = m_k$, and among which k_1 messages are with content 1 and the other $k - k_1$ messages are with content 0, $0 < k_1 < k$. Conditioned on the source message $m_0 = 0$, the conditional probability of the occurrence of event $M_1 = m_1, ...M_k = m_k$ can be calculated as follows:*

$$\begin{aligned} &Pr\left(M_1 = m_1, ...M_k = m_k | m_0 = 0\right) \\ &= \begin{cases} (1-p)^{n-n_1} \cdot \left[\sum_{i=1}^{n_1} b_i \cdot p^i (1-p)^{n_1 - i}\right], & n_1 > 0 \\ 0, & n_1 = 0 \end{cases}, \end{aligned} \quad (5.13)$$

where n_1 is the number of vehicles that only belong to paths that deliver messages with content 1 to the destination vehicle, i.e., the number of Type 1 vehicles in the network, and $b_i, i = 1, 2, ...n_1$ is the number of combinations that exactly i malicious Type 1 vehicles lead to the occurrence of event $M_1 = m_1, ...M_k = m_k$.

Denote by event $e_i^{'}$ that randomly choosing i columns from sub-matrix B_1 and then conducting element-wise union operation to them, there results a column with each entry equal to 1. Denote by $b_i, i = 1, 2, ...n_1$ the total number of combinations that event $e_i^{'}$ occurs. Then we have

$$b_i = \sum_{j=1}^{z_i^{'}} I \left(\text{event } e_i^{'} \text{ occurs}\right), \tag{5.14}$$

where $z_i^{'} = \begin{pmatrix} n_1 \\ i \end{pmatrix}$. Therefore, this theorem can be readily proved following the same method as that used in the proof of Theorem 5.2, and hence is ignored.

5.2.3 A Discussion of the Optimum Algorithm

From the analysis in Section 5.2.2, we can see that the value of n_0, n_1, and a_i, $i = 1, 2, ...n_0$ in (5.10), b_i, $i = 1, 2, ...n_1$ in (5.13) can be obtained from the network topology matrix. That is, when the k received messages $M_1 = m_1, ...M_k = m_k$, and the network topology is given, the value of n_0, n_1, a_i, $i = 1, 2, ...n_0$, and b_i, $i = 1, 2, ...n_1$ are all determined. However, the exact values of $\Pr(M_1 = m_1, ...M_k = m_k | m_0 = 1)$ and $\Pr(M_1 = m_1, ...M_k = m_k | m_0 = 0)$, shown as also (5.10) and (5.13), also depend on the proportion of malicious vehicles p in the network, which usually, is not easy to be obtained or estimated as a prior knowledge. In the following, we use a simple example to show the dependency on p of the proposed optimum decision algorithm.

Consider a network that contains overall 7 independent paths from the source vehicle to the destination vehicle. The first three paths, containing 1, 8 and 15 vehicles respectively deliver messages with content 1 to the destination vehicle, and the other four paths, containing 6 vehicles each, deliver messages with content 0 to the destination vehicle. See Fig. 5.2 for an illustration.

According to (5.10) and (5.13), we have:

$$\begin{aligned} &\Pr(M_1 = M_2 = M_3 = 1, M_4 = ... = M_7 = 0 | m_0 = 1) \\ =&(1-p)^{1+8+15} \cdot \left[1 - (1-p)^6\right]^4 \\ =&(1-p)^{24} \cdot \left[1 - (1-p)^6\right]^4, \end{aligned} \tag{5.15}$$

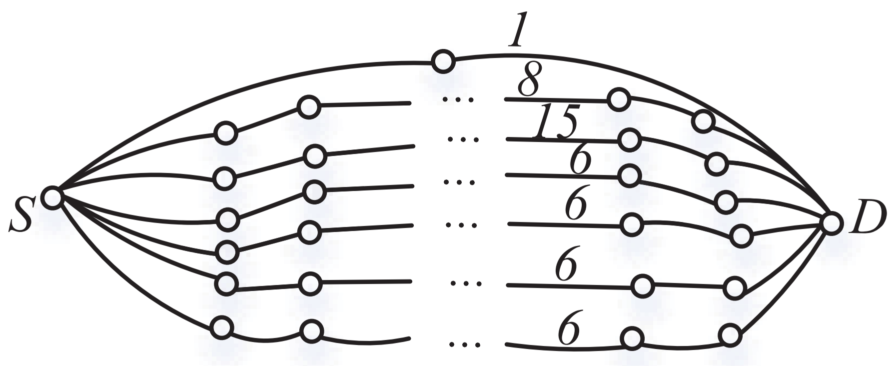


Figure 5.2 : An illustration of a vehicular network that contains 7 independent paths from the source vehicle to the destination vehicle, containing 1, 8, 15, 6, 6, 6, and 6 vehicles respectively.

and

$$\begin{aligned}
&\Pr\left(M_1 = M_2 = M_3 = 1, M_4 = ... = M_7 = 0 | m_0 = 0\right)\\
=&(1-p)^{6\times 4} \cdot \left[1-(1-p)\right]\left[1-(1-p)^8\right]\left[1-(1-p)^{15}\right]\\
=&(1-p)^{24} p \left[1-(1-p)^8\right]\left[1-(1-p)^{15}\right]. \qquad (5.16)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{\Pr\left(M_1 = M_2 = M_3 = 1, M_4 = ... = M_7 = 0 | m_0 = 1\right)}{\Pr\left(M_1 = M_2 = M_3 = 1, M_4 = ... = M_7 = 0 | m_0 = 0\right)}\\
=&\frac{1-(1-p)^6}{p\left[1-(1-p)^8\right]\left[1-(1-p)^{15}\right]} \qquad (5.17)
\end{aligned}$$

Let

$$f_1(p) = 1-(1-p)^6 \qquad (5.18)$$

and

$$f_2(p) = p\left[1-(1-p)^8\right]\left[1-(1-p)^{15}\right], \qquad (5.19)$$

and plot them with different value of p, see Fig. 5.3 for an illustration. We can see that the value of $\frac{\Pr(M_1=m_1,...M_k=m_k|m_0=1)}{\Pr(M_1=m_1,...M_k=m_k|m_0=0)}$ depends on the percentage of malicious vehicles in the network. Specifically, it is shown in Fig. 5.3 that when p is smaller than a threshold, e.g., $p_{th} = 0.092$ in this case, the value of $\frac{\Pr(M_1=m_1,...M_k=m_k|m_0=1)}{\Pr(M_1=m_1,...M_k=m_k|m_0=0)} = \frac{f_1(p)}{f_2(p)}$ is smaller than 1, while when p is larger than the threshold, the value of $\frac{\Pr(M_1=m_1,...M_k=m_k|m_0=1)}{\Pr(M_1=m_1,...M_k=m_k|m_0=0)} = \frac{f_1(p)}{f_2(p)}$ is larger than 1, and will further increase with an increase of p. Therefore, given the network topology, the decision based on (5.4) relies on the value of p. This illustrates that the value of p is indispensable in adopting the optimum decision algorithm to achieve an accurate decision result.

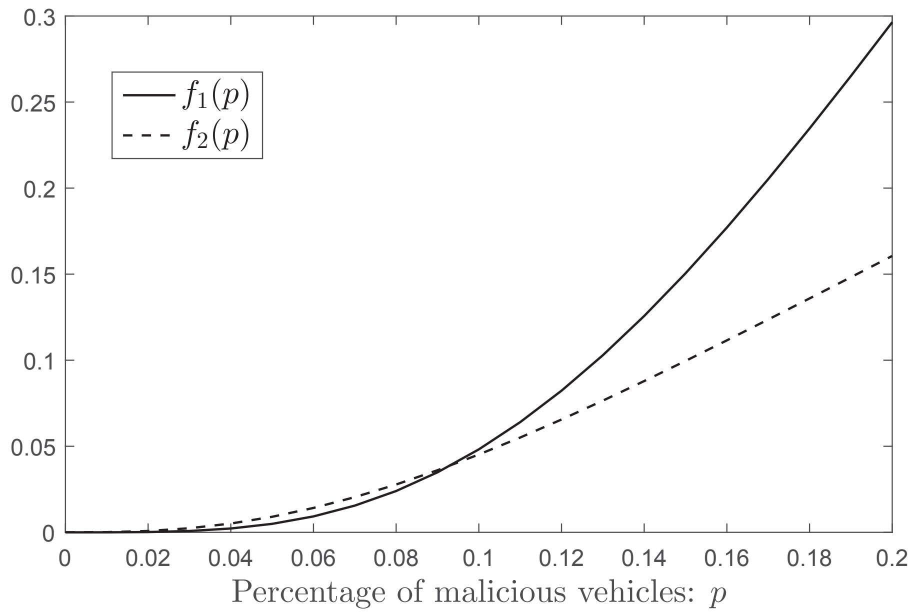


Figure 5.3 : An illustration to show that the percentage of malicious vehicles is indispensable in implementing the optimum decision algorithm to achieve an accurate decision result.

5.3 Heuristic Decision Algorithm

As discussed in the Section 5.2.3, the implementation of the optimum decision algorithm proposed in the last section relies on prior knowledge of the percentage of malicious vehicles p in the network, which is usually not easy to be obtained or estimated. In this section, to avoid the dependence on p, we propose a heuristic decision algorithm for the destination vehicle to make a decision when receiving conflicting messages purely based on network topology information.

The heuristic decision algorithm is derived from the principle of Maximum Likelihood Estimation [106], which can be described as follows:

$$d = \begin{cases} 1, & \frac{\Pr(M_1=m_1,...M_k=m_k|m_0=1)}{\Pr(M_1=m_1,...M_k=m_k|m_0=0)} > 1 \\ 0, & \frac{\Pr(M_1=m_1,...M_k=m_k|m_0=1)}{\Pr(M_1=m_1,...M_k=m_k|m_0=0)} < 1 \end{cases}, \tag{5.20}$$

where $M_1 = m_1, ...M_k = m_k$ are the k messages received by the destination vehicle, m_0 is the source message and d is the decision made by the destination vehicle. When $\frac{\Pr(M_1=m_1,...M_k=m_k|m_0=1)}{\Pr(M_1=m_1,...M_k=m_k|m_0=0)} = 1$, d is randomly chosen from 0 and 1 with equal probability.

Based on the received messages $M_1 = m_1$, $M_2 = m_2$, ..., $M_k = m_k$ and the path information obtained from messages, the method of constructing the topology matrix B is identical to that introduced in Section 5.2.2, i.e., $B = \begin{bmatrix} B_1 & B_{s_1} & \mathbf{0} \\ \mathbf{0} & B_{s_0} & B_0 \end{bmatrix}$. Therefore, by combining (5.10), (5.13) and (5.20), it can be readily shown that $d = \begin{cases} 0, & n_0 = 0 \\ 1, & n_1 = 0 \end{cases}$, and when $n_0 > 0$ and $n_1 > 0$,

$$\begin{aligned} \frac{\Pr(M_1 = m_1, ...M_k = m_k|m_0 = 1)}{\Pr(M_1 = m_1, ...M_k = m_k|m_0 = 0)} =& \frac{(1-p)^{n-n_0} \cdot \left[\sum_{i=1}^{n_0} a_i \cdot p^i (1-p)^{n_0-i}\right]}{(1-p)^{n-n_1} \cdot \left[\sum_{i=1}^{n_1} b_i \cdot p^i (1-p)^{n_1-i}\right]} \\ =& \frac{\sum_{i=1}^{n_0} a_i \cdot \left(\frac{p}{1-p}\right)^i}{\sum_{i=1}^{n_1} b_i \cdot \left(\frac{p}{1-p}\right)^i}. \end{aligned} \tag{5.21}$$

Recall that both sub-matrices $\begin{bmatrix} B_1 & B_{s_1} & \mathbf{0} \end{bmatrix}$ and $\begin{bmatrix} \mathbf{0} & B_{s_0} & B_0 \end{bmatrix}$ correspond to a sub-network of the considered network and the common nodes shared by the two sub-networks (if any) cannot be malicious vehicles. Therefore, when considering the potential malicious vehicle combinations, we avoid these common nodes and only focus on the sub-matrices B_1 and B_0. Specifically, we regard the network corresponding to sub-matrices B_1 and B_0 as networks where each row represents a complete path and each column represents a vehicle, denoted by T_1 and T_0 respectively. In the following, with a twist of the vertex-cut [107] terminology from graph theory which defines a vertex set whose removal would disconnect the graph, we define *malicious cut set*, *size* of a malicious cut set, and *minimal malicious cut set* of a network in this work, and demonstrate that the parameter a_i, $1 \leq i \leq n_0$ and b_i, $1 \leq i \leq n_1$ in (5.21), which was defined in (5.11) and (5.14), are exactly the number of malicious cut sets with size i of the network T_0 and T_1 respectively.

Definition 5.1. *A malicious cut set of a network is a combination of vehicles, where if all vehicles in the set are malicious vehicles all paths of the network can be compromised. The size of a malicious cut set is the number of vehicles contained in the set. A minimal malicious cut set is a malicious cut set with the smallest size.*

It is worth noting that the network may have multiple malicious cut sets and multiple minimal malicious cut sets. Consider the network shown in Fig. 5.4 for example. Vehicle sets $\{V_1, V_2, V_3\}$, $\{V_4, V_5, V_6, V_7\}$, and $\{V_8, V_9\}$ (to name a few) are all malicious cut sets of the network, and a minimal malicious cut set is the malicious cut set $\{V_8, V_9\}$ with size 2. Therefore, to compromise all paths of this network, the minimum number of malicious vehicles needed is 2.

Based on Definition 5.1, if a vehicle set is a malicious cut set, then each path of the network contains at least one vehicle belonging to this set. Recall that a_i (or b_i) represents the number of combinations that randomly choosing i columns from sub-matrix B_0 (B_1) and then conducting element-wise union to them, there results a column with each entry equal to 1. That is, a_i (or b_i) represents the number of combinations that by choosing i vehicles from Network T_0 (or T_1) to form a vehicle set, each path of network T_0 (or T_1) contains at least one vehicle belonging to this set. Therefore, a_i, $1 \leq i \leq n_0$ and b_i, $1 \leq i \leq n_1$ are exactly the number of malicious

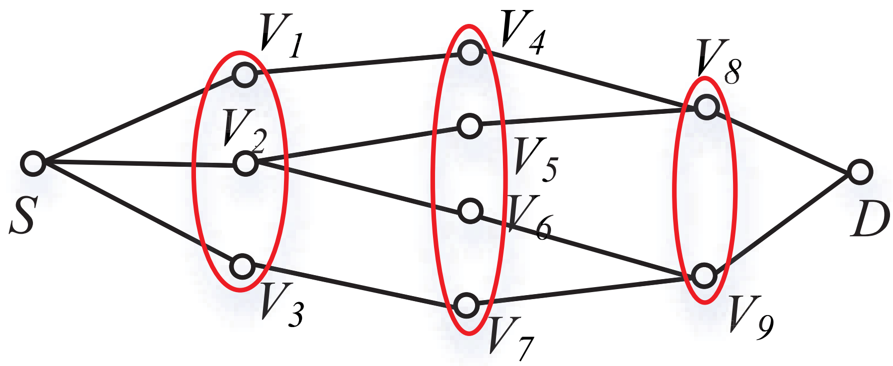


Figure 5.4 : An illustration to show the malicious cut sets and minimum malicious cut sets of a network.

cut sets with size i of the network T_0 and T_1 respectively.

According to the properties of malicious cut sets, it can be readily obtained that $a_i = 0$ if $a_{i+1} = 0$, and $a_{i+1} > 0$, if $a_i > 0$. Similarly, we have $b_i = 0$ if $b_{i+1} = 0$, and $b_{i+1} > 0$, if $b_i > 0$.

Define

$$r_0 = \min\{i: \ a_i > 0\}, \quad 1 \leq r_0 \leq n_0 \tag{5.22}$$

and

$$r_1 = \min\{i: \ b_i > 0\}, \quad 1 \leq r_1 \leq n_1, \tag{5.23}$$

the smallest integer that satisfies $a_i > 0$ and $b_i > 0$ respectively. Therefore, r_0 is the size of the minimal malicious cut set of network T_0, and a_{r_0} is the number of minimal malicious cut sets of network T_0. Similarly, r_1 is the size of the minimal malicious cut set of network T_1, and b_{r_1} is the number of minimal malicious cut sets of network T_1. This follows that

$$\frac{\Pr\left(M_1 = m_1, ...M_k = m_k | m_0 = 1\right)}{\Pr\left(M_1 = m_1, ...M_k = m_k | m_0 = 0\right)} = \frac{\sum_{i=r_0}^{n_0} a_i \cdot \left(\frac{p}{1-p}\right)^i}{\sum_{i=r_1}^{n_1} b_i \cdot \left(\frac{p}{1-p}\right)^i} \approx \frac{a_{r_0}\left(\frac{p}{1-p}\right)^{r_0}}{b_{r_1}\left(\frac{p}{1-p}\right)^{r_1}}, \tag{5.24}$$

where the first step is obtained from the fact that $a_1 = a_2 = ...a_{r_0-1} = 0$, $a_{r_0} > 0$, and $b_1 = b_2 = ...b_{r_1-1} = 0$, $b_{r_1} > 0$, and the second step is obtained by only keeping the first item of both the numerator and denominator. Considering the fact that when p is small, the probability that there are $i+1$ malicious vehicles in the network is much smaller than the probability that there are i malicious vehicles in the network, therefore, this approximation is quite accurate.

Note that when p is small, we have $\frac{p}{1-p} \ll 1$. Therefore, when $r_0 \neq r_1$, whether the value of $\frac{a_{r_0}\left(\frac{p}{1-p}\right)^{r_0}}{b_{r_1}\left(\frac{p}{1-p}\right)^{r_1}}$ shown as (5.24) is larger than 1 is dominantly determined by the value of $r_0 - r_1$. Specifically, when $r_0 < r_1$, we have $\left(\frac{p}{1-p}\right)^{r_0 - r_1} \gg 1$. In this case, the coefficient $\frac{a_{r_0}}{b_{r_1}}$ plays a marginal role and therefore $\frac{a_{r_0}\left(\frac{p}{1-p}\right)^{r_0}}{b_{r_1}\left(\frac{p}{1-p}\right)^{r_1}} > 1$; when $r_0 > r_1$, we have $\left(\frac{p}{1-p}\right)^{r_0 - r_1} \ll 1$, and therefore $\frac{a_{r_0}\left(\frac{p}{1-p}\right)^{r_0}}{b_{r_1}\left(\frac{p}{1-p}\right)^{r_1}} < 1$. On the contrary, when $r_0 = r_1$, whether the value of $\frac{a_{r_0}\left(\frac{p}{1-p}\right)^{r_0}}{b_{r_1}\left(\frac{p}{1-p}\right)^{r_1}}$ is larger than 1 would heavily depend on the value of the coefficient $\frac{a_{r_0}}{b_{r_1}}$. Consequently, we have

$$\frac{\Pr\left(M_1 = m_1, \ldots M_k = m_k | m_0 = 1\right)}{\Pr\left(M_1 = m_1, \ldots M_k = m_k | m_0 = 0\right)} \approx \frac{a_{r_0}\left(\frac{p}{1-p}\right)^{r_0}}{b_{r_1}\left(\frac{p}{1-p}\right)^{r_1}}$$

$$\begin{cases} > 1, & r_0 < r_1 \\ < 1, & r_0 > r_1 \\ = \frac{a_{r_0}}{b_{r_1}}, & r_0 = r_1 \end{cases}, \tag{5.25}$$

which shows that to compare the values of $\Pr\left(M_1 = m_1, \ldots M_k = m_k | m_0 = 1\right)$ and $\Pr\left(M_1 = m_1, \ldots M_k = m_k | m_0 = 0\right)$, we only need to compare the values of r_0 and r_1, namely, the size of minimal malicious cut set of networks T_0 and T_1 when $r_0 \neq r_1$, or the value of a_{r_0} and b_{r_1}, namely, the number of minimal malicious cut sets of networks T_0 and T_1 when they have the same size in terms of minimal malicious cut set.

From Menger's Theorem [107], the size of the minimal vertex-cut whose removal would disconnect two non-adjacent vertices, is equal to the maximum number of vertex-independent paths between these two non-adjacent vertices. Therefore, it can be concluded that the size of the minimal malicious cut set of a network is also equal to the maximum number of node-disjoint paths in the network between the source vehicle and the destination vehicle. Therefore, r_0 and r_1 are also the maximum number of node-disjoint paths which exist in networks T_0 and T_1 respectively. Note that calculating the maximum number of vertex-disjoint paths from source to destination is a special case of finding the maximum flow problem by setting every vertex capacity as 1 [107]. Therefore, the values of r_0 and r_1 can be readily obtained by existing maximum flow algorithms, e.g., introduced in [107, 108, 109]. When $r_0 = r_1$, a_{r_0} and b_{r_1} can be obtained by the exhaustive search algorithm according to their definitions given by (5.11) and (5.14).

In summary, by combining (5.20) and (5.25), the decision rule of our proposed heuristic algorithm can be shown as

$$d = \begin{cases} 1, & (r_0 < r_1) \text{ } or \text{ } (r_0 = r_1, \ a_{r_0} > b_{r_1}) \\ 0, & (r_0 > r_1) \text{ } or \text{ } (r_0 = r_1, \ a_{r_0} < b_{r_1}) \end{cases}, \tag{5.26}$$

and when $r_0 = r_1$, and $a_{r_0} = b_{r_1}$, d is randomly chosen from 0 and 1 with equal probability.

Remark 5.2. *It is worth noting that the topology of the network corresponding to a topology matrix B_1 may not be unique. For instance, a topology matrix* $B = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ *can correspond to both networks shown as Fig. 5.5. However, the malicious cut sets of the networks with different topology remain the same as there is a one-to-one correspondence between each malicious cut set and a combination of columns from the topology matrix that an element-wise union of them leading to a column with each entry equal to 1. That is, as long as networks have the same topology matrix B, they would have the same malicious cut sets. Therefore, the network T_1 (or T_0) corresponding to sub-matrix B_1 (or B_0) may not be unique, however it does not effect their malicious cut set analysis.*

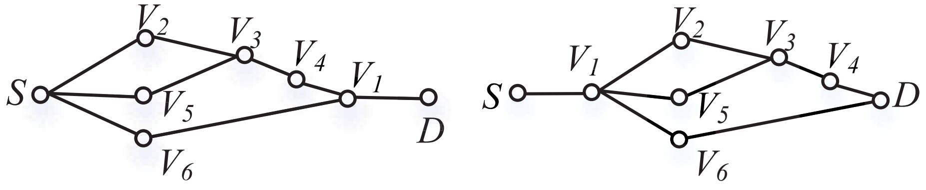


Figure 5.5 : An illustration of two networks that have the same topology matrix.

Remark 5.3. *The implication of the heuristic decision algorithm* (5.26) *can also be explained as follows. Given two networks that deliver conflicting message contents, by removing the common nodes shared by these two networks and regarding each path after the removal of the common nodes as a new complete path, there results in two new independent networks that deliver conflicting message contents. Therefore, a decision can be made by comparing the robustness of the two new networks. Note that a smaller size of the minimal malicious cut set of a network implies fewer minimal malicious vehicles are required to compromise that network, and consequently, a higher probability that incorrect messages will be delivered. Therefore, the decision will always be chosen as the message delivered by the network with a lower probability of being compromised.*

From (5.26), we can see that the decision result is now entirely determined by the network topology, and is independent of the proportion of malicious vehicles in the network. That is, the proposed heuristic decision algorithm is purely topology-based which is easy to be implemented in practice. In summary, the heuristic decision algorithm works as detailed in Algorithm 5.2.

Algorithm 5.2 Heuristic Decision Algorithm

Input: $M_1, M_2, ...M_k$

Output: d

Construct topology matrix B based on the path information derived from the received k copies of the message;
Based on the constructed topology matrix B, calculate r_0 and r_1 according to the maximum flow algorithm;
if $r_0 < r_1$ **then**
    $d = 1;$
else if $r_0 > r_1$ **then**
    $d = 0;$
else
    Calculate a_{r_0} and b_{r_1} based on the definition given by (5.11) and (5.14);
    if $a_{r_0} > b_{r_1}$ **then**
        $d = 1;$
    else if $a_{r_0} < b_{r_1}$ **then**
        $d = 0;$
    else
        d is randomly chosen from 0 and 1 with equal probability;
    end if
end if

5.4 Simulation and Discussion

In this section, we conduct simulations based on real traffic data to establish the validity of the decision algorithms proposed in Section 5.2 and Section 5.3, as well as to demonstrate the application of our proposed algorithms in a real life situation to indicate their usefulness. We utilize the real traffic data collected by inductive loop

Figure 5.6 : An illustration of the target road segment.

detectors in Taipei city, including 1-minute averaged vehicular speed passing each loop detector, and the volume, i.e., the number of vehicles, passing the corresponding loop detectors during this 1 minute, to model real life traffic. Specifically, we choose the traffic data at 17:00 pm (peak hour, the traffic is congested) on the road segment, named Jianguo North Road, covered by 6 loop detectors (labeled by 'Loop 1' to 'Loop 6'), to build the considered vehicular network for our simulation. See Fig. 5.6 for an illustration. The interval distance between two consecutive loop detectors are 700m, 550m, 300m, 280m, and 240m respectively. Vehicles are moving in the direction from Loop 6 to Loop 1. We assume that at the specific minute we focused on, vehicles located between two loop detectors move at the same constant speed, the same as the average speed passing the next loop during this minute. For instance, we approximately regard vehicles located between Loop 1 and Loop 2 at 17:00pm travel at the same constant speed as the average speed vehicles passing Loop 1 at 17:00 pm.

We assume the source vehicle is exactly located at the location of Loop 1 when it detects the congestion at its location, and it would like to disseminate the congestion information to other vehicles moving towards the congestion area. Therefore, messages are disseminated in the direction from Loop 1 to Loop 6, which is opposite to the traveling direction of vehicles. Vehicles communicate with their neighbors adopting the unit disk model [86, 47] with a transmission range $r_0 = 250$m [21]. Each relay vehicles in the network has a probability p to be a malicious vehicle. We focus on a destination vehicle located at a distance L from the source vehicle, and track the probability of correct decision made by the destination vehicle. From the time instant the destination vehicle receives the first message reporting road condition, it waits a fixed time period T to receive more number of messages before it starts to make a decision. The per-hop transmission delay is assumed to be $\beta = 4$ms [21].

At each simulation, a topology matrix B can be constructed based on the underlying network topology. Therefore, given the malicious vehicle distribution and the topology information, the content of the k messages $M_1, M_2, ...M_k$ received by the destination vehicle is determined. The destination vehicle then makes a decision given the received messages and the derived underlying topology information according to our proposed decision algorithms at each simulation. The decision result can be either correct or incorrect. The simulation is repeated 5000 times and the proportion of the correct decision, i.e., the probability of correct decision, is plotted.

In the following, we first compare our proposed two decision algorithms, and then we study the effects of topology information, and some performance-impacting parameters on the algorithms. The performance-impacting parameters including the proportion of malicious vehicle in the network, and the choice of waiting time by the destination vehicle before it starts to make the decision.

5.4.1 Comparison of the Two Proposed Algorithms

In this section, we compare the message security performance achieved by the two proposed decision algorithms to provide insight on the optimum decision algorithm design for secure message dissemination.

Fig. 5.7 compares the probability of a correct decision achieved by the proposed optimum decision algorithm (labeled as Optimum Algorithm) and by the proposed pure topology-based heuristic decision algorithm (labeled as Heuristic Algorithm) respectively. It is shown that when the percentage of malicious vehicles in the network is small, e.g., when $p < 0.2$ in this case, the message security performance achieved by the optimum decision algorithm is only slightly better than the performance achieved by the heuristic decision algorithm. This implies that the heuristic decision algorithm, purely based on network topology information and easily implemented in practice, is sufficient to achieve a high message security performance for vehicular networks.

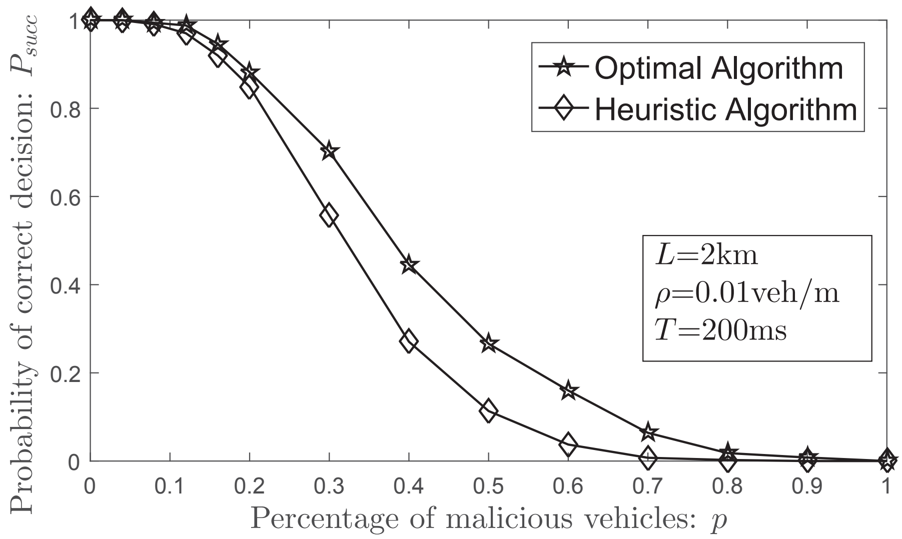


Figure 5.7 : A comparison of the probabilities of a correct decision achieved by the optimum decision algorithm proposed in Section 5.2, and by the heuristic decision algorithm proposed in Section 5.3.

5.4.2 Impact of Topology Information

To evaluate the effectiveness of our proposed algorithms that take the underlying topology information into consideration, we compare the security performance, in terms of the probability of a correct decision made by the destination vehicle, achieved by our proposed algorithms described by Algorithm 5.1 and 5.2 respectively, with that achieved by existing weighted voting algorithms like the weighted voting algorithm proposed in [105] (labeled with WV: MMSE) that considers partial correlation between messages, the weighted voting algorithm proposed in [43] (labeled with WV: $w \propto \alpha^{h-1}$) that does not consider the underlying topology information causing the correlation between messages, and the majority voting (a special case of weighted voting by assigning identical weights to each vote) that totally ignores the underlying topological correlation. Specifically, the weighted voting algorithm proposed in [105] sets weight to each message as $w_i = \sum_{j=1}^{k} C_{ij}^{(-1)} \left(\sum_{r,j=1}^{k} C_{rj}^{(-1)} \right)^{-1}$, where C is the error covariance matrix whose (i,j)th entry is defined by the error covariance between message M_i and message M_j, calculated by $C_{ij} = E\left[(M_i - m_0)(M_j - m_0)\right]$. C^{-1} is the inverse matrix of the error covariance matrix C, and $C_{ij}^{(-1)}$ is the (i,j)th entry of the matrix C^{-1}. The weighted voting algorithm proposed in [43] simply assigns weight to each message as $w_i = \frac{\alpha^{h_i-1}}{\sum_j \alpha^{h_j-1}}$, where $\alpha \in (0,1)$ is a weighting factor to reduce the oversampling impact caused by messages generated from the same source and h_i is the number of hops travelled by the ith message from the source to the destination.

It can be seen in Fig. 5.8 that both our proposed algorithms outperform the weighted voting algorithms proposed in [105], [43] and the majority voting algorithm, which demonstrates that our algorithms taking into account topology information and correlation between different copies of message are able to effectively improve the robustness of vehicle networks against attacks from malicious vehicles.

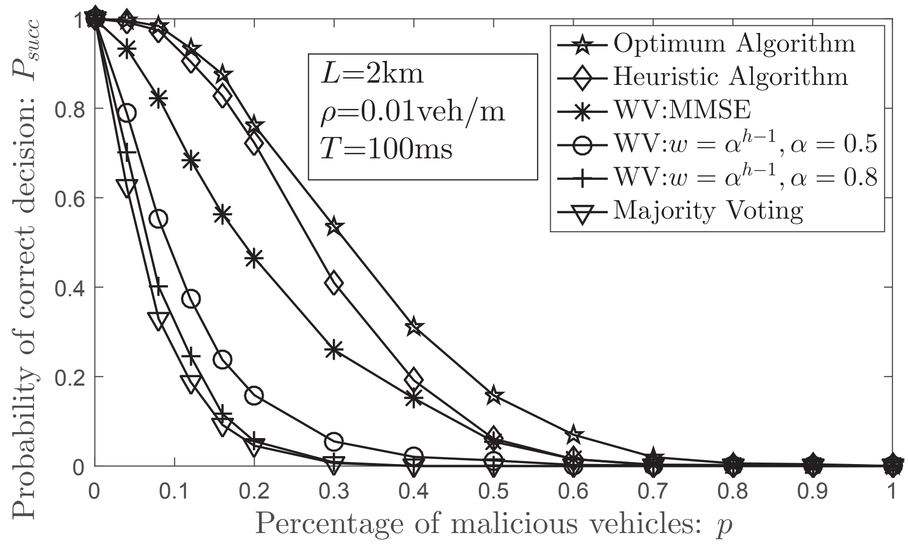


Figure 5.8 : A comparison of the probability of a correct decision achieved assuming our proposed algorithms and that achieved assuming other existing weighted voting algorithms.

5.4.3 Impact of the Percentage of Malicious Vehicles

Fig. 5.8 reveals the relationship between the probability of a correct decision P_{succ} and the percentage of malicious vehicles in the network, p. It can be seen that the probability of a correct decision made by the destination vehicle decreases to its minimum value $P_{succ} = 0$ when the proportion of malicious vehicles in the network is larger than a certain threshold. Beyond that threshold, a further increase in p has little impact on the security performance. Specifically, as shown in Fig. 5.8, when p is small, the security performance achieved assuming the optimum decision algorithm decreases with an increase of p; however, when p increases beyond a certain threshold, a further increase in p has no impact on the security performance. This can be explained by the fact that the more malicious vehicles exist in the network, the more tampered copies of message will be delivered, and therefore a lower chance for the destination vehicle to make a correct decision regardless of what algorithm it adopts. Furthermore, when the number of malicious vehicles in the network reaches a certain threshold, most of the message dissemination paths will be compromised. In this case, the destination vehicle will totally misguided by the incorrect messages and the message security performance approaches its minimum value $P_{succ} = 0$.

5.4.4 Impact of the Waiting Time Period

As mentioned in Section 5.1.3, the waiting time period T the destination vehicle waits before it starts to make a decision is an important parameter that should balance the trade-off between the response time requirement and the integrity of the decision. Therefore, in this section, we study the impact of the waiting time period T on the security performance assuming the two proposed algorithms, under different traffic densities.

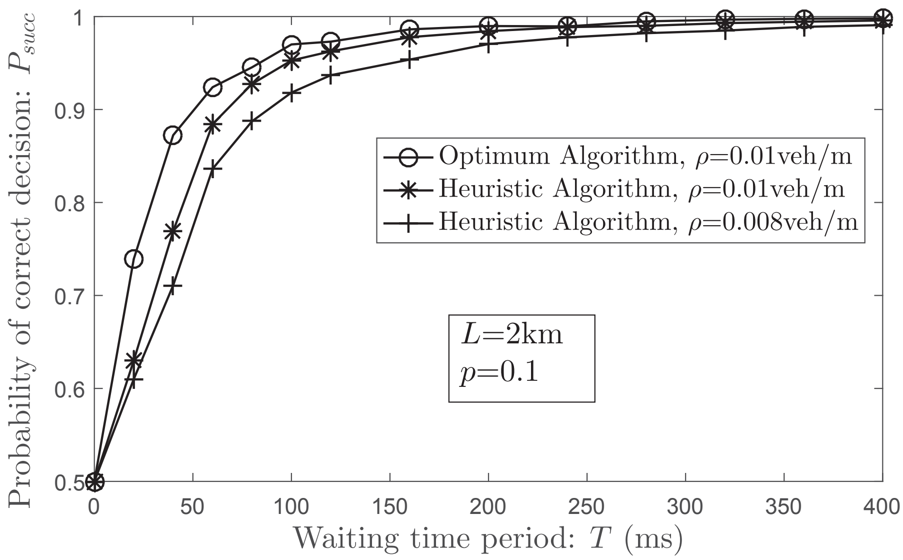


Figure 5.9 : An illustration of the relationship between the probability of a correct decision and the waiting time period the destination vehicle waits before it starts to make a final decision by adopting the proposed two algorithms respectively.

Fig. 5.9 demonstrates the relationship between the probability of a correct decision, P_{succ}, and the waiting time period T the destination vehicle waits before it starts to make a decision, assuming our two proposed algorithms respectively, and gives insight into the choice of waiting time by the destination vehicle. Importantly, we can see that for both algorithms, a greater waiting time is beneficial to the secure message dissemination because a longer waiting time potentially implies a larger number of received messages. This consequently, brings more information on the underlying network topology, and therefore leads to a more robust result from the data consistency check. However, when T increases beyond a certain threshold T_{th}, e.g., in the case of ρ=0.01veh/m, $T_{th} = 100ms$ when adopting the proposed optimum decision algorithm and $T_{th} = 150ms$ when adopting the proposed heuristic decision algorithm when, a further increase in T has marginal (less than 5%) impact on the probability of a correct decision. This is due to the fact that when T is larger than a threshold, the marginal return brought by waiting a longer time to the security performance is diminishing. Furthermore, it can be seen that to achieve the same message security performance, when the vehicular density is lower, the waiting time needs to be longer. Therefore, when determining the waiting time period, it is important to take the vehicular density into account, e.g., in areas where the vehicular density is large, the waiting time can be reduced. Thus, Fig. 5.9 exhibits a guide on the choice of waiting time period for destination vehicles.

5.5 Summary

By utilizing underlying network topology information, this chapter proposed two decision algorithms - the optimum decision algorithm and a heuristic decision algorithm - to address the issue of message inconsistency caused by insider malicious vehicles that would tamper the content of disseminated messages in the network. The proposed optimum decision algorithm is able to effectively help a destination vehicle maximally make a correct decision on the content of message, given the network topology information and the prior knowledge of the percentage of malicious vehicles in the network. The proposed heuristic decision algorithm further enables a vehicle to make a decision based on network topology information only and without the need for knowing the percentage of malicious vehicles which can be difficult to estimate in some circumstances, at a modest cost in performance. Therefore, the heuristic algorithm is easier to implement in practice. Simulations based on real traffic data were conducted to evaluate the effectiveness of two algorithms. It was demonstrated that the heuristic decision algorithm is able to achieve a security performance close to that achieved by the optimum decision algorithm, especially when the percentage of malicious vehicles is small. By comparing the two proposed algorithms with existing algorithms that do not consider the underlying topological information or only partially consider message correlation, it was shown that the proposed algorithms greatly outperform existing ones. Moreover, the impact of some key parameters on the performance of the proposed algorithms was discussed, including the percentage of malicious vehicles, and the waiting time the destination vehicle waits before making a final decision. A deeper insight revealed in our work is that messages coming from redundant paths are not equal and messages coming from diversified and independent paths carry more information than those from correlated paths. In this sense, we consider our work is just a first step towards the big direction of harnessing the network topology information to improve the vehicular network security. In the future, we would like to utilize more traveling information of vehicles in the network, like location, speed, direction, .etc, to design a more comprehensive topological approach, so as to further improve the vehicular network security.

Chapter 6

Conclusion

This thesis had presented works on cooperative vehicular networks for ITS, including investigating a novel cooperative communication strategy for vehicular networks to facilitate data dissemination, and novel topology-based decision algorithms for vehicular networks to protect secure message dissemination. In the following, the key results and findings of this thesis are summarised.

In Chapter 3, “Throughput of Cooperative Vehicular Networks with Infrastructure Support: Single-user Case”, we proposed a cooperative communication strategy for vehicular networks with a finite vehicular density by utilizing V2I communications, V2V communications, mobility of vehicles, and cooperations among vehicles and infrastructure to facilitate data dissemination. A detailed analysis for the achievable throughput was presented when considering a typical delay-tolerant application scenario that there is only one vehicle has a download request from the Internet, and all the other vehicles assist its download using the proposed cooperative communication strategy. The closed-form expression of achievable throughput (or its upper and lower bound) was obtained in three different regimes we classified in our analysis. The results showed that the proposed cooperative strategy can significantly improve the achievable throughput of vehicular networks even when traffic density is low.

Next, in Chapter 4, “Capacity of Cooperative Vehicular Networks with Infrastructure Support: Multi-user Case”, we extended the work introduced in Chapter 3 and considered a scenario with a subset of vehicles having download requests from the Internet. Each vehicle with a download request downloads a distinct large-size file from the Internet and other vehicles without download requests assist the delivery of the files to them. An analytical framework was developed and the capacity of vehicular networks with a finite traffic density adopting the cooperative communication strategy was analysed. The results have showed that the proposed cooperative strategy can improve the capacity of vehicular networks, and the improvement is more pronounced when the proportion of vehicles with download requests are low.

Finally, in Chapter 5, "A Topological Approach to Secure Message Dissemination in Vehicular Networks", we proposed two decision algorithms that utilize the underlying network topology information to address the issue of message inconsistency caused by malicious vehicles that would tamper with the content of disseminated messages, so as to protect secure message dissemination in vehicular networks and keep the network robust against insider attackers. The optimum decision algorithm proposed is able to maximize the chance of making a correct decision on the message content, assuming the prior knowledge of the percentage of malicious vehicles in the network. The heuristic decision algorithm proposed enables vehicles to make decisions without the aforementioned knowledge of the percentage of malicious vehicles, therefore is easier to implement in practice. We demonstrated that the heuristic decision algorithm is able to achieve a security performance close to that achieved by the optimum decision algorithm, especially when the percentage of malicious vehicles in the network is small. By comparing the two proposed algorithms with existing algorithms that do not consider the underlying topological information or only partially consider message correlation, we showed that our proposed algorithms greatly outperform existing ones.

In addition to the key results and findings summarised above, there are still some research problems to be investigated in the future. First, it would be interesting to investigate the optimum road side smart infrastructure deployment scheme, to better serve the needs of ITS. Our work has provided guidance on the optimum deployment of road side infrastructure in terms of their interval distance, that in areas where the vehicular density is usually large, by utilizing a cooperative communication strategy, the number of infrastructure points can be reduced. Indeed, deploying road side smart infrastructure to support ITS is a challenging problem that should take multiple factors into consideration, e.g., the penetration rate of smart vehicles in the city, the economic issues, etc. Therefore, it is an on-going and worthwhile topic which deserves further study. Second, more interesting results related to cooperative vehicular network design could be investigated when considering multiple metrics instead of optimizing only a single objective. This is motivated by the fact that in practice, a trade-off may exist between multiple metrics, such as throughput and delay, throughput and fairness, security and privacy, etc.. Therefore, optimally designing cooperative vehicular networks to balance conflicting metrics is both interesting and challenging. Furthermore, approaches to protect vehicular network security require further investigation. For instance, our proposed topological approaches can be extended to hybrid approaches that integrate traffic information from heterogeneous sources, e.g., from messages on traffic state received through vehicular communications, and from the local traffic state obtained through built-in sensors. In this case, to decide whether there is an abnormal traffic condition in the target location of the road, the challenge lies in analysing the correlation between the heterogeneous information, and designing a hybrid decision algorithm to appropriately combine them.

Appendices

Appendix A

Proof of Theorem 4.2 in Chapter 4.2.2

Recall that we set the point to the right of infrastructure point I_1 and at a distance r_I to I_1 (the left boundary point of the V2V Area) as the origin of the coordinate system, and the right direction as the positive ($+x$) direction. Denote by $X_k, k = 1, 2, ...$ the location of the i-th transmitter (helper of the active helper-VoI pair), numbered from left to the right, under the optimum scheduling scheme χ_{opt}. Denote by $Y_k, k = 1, 2, ...$ the location of the i-th transmitter under an arbitrary scheduling scheme $\chi^{'}$. It follows that $X_1 < X_2 < \cdots < X_k < X_{k+1} < \cdots$ and $Y_1 < Y_2 < \cdots < Y_k < Y_{k+1} < \cdots$. See Fig. A.1 for an illustration. In the following, we prove that χ_{opt} described in Theorem 4.2 is an optimum scheduling scheme that would lead to the maximum number of active helper-VoI pairs by recursion that $X_k \leq Y_k$ holds for any $k = 1, 2, ...$

For $k = 1$, noting that according to the scheduling scheme χ_{opt}, the first transmitter is the leftmost helper in the V2V Area that has at least one VoI within its coverage. Therefore, it follows readily that $X_1 \leq Y_1$.

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

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

(ii) Case $X_{n+1} > Y_n$: in this case, under the scheduling scheme χ_{opt}, the $(n+1)$-th transmitter is the *nearest helper* to the right of the n-th transmitter satisfying simultaneous transmission conditions: it is outside the sensing range of the n-th transmitter who are located at X_n *and* has at least one VoI within its transmis-

(a) Results under the proposed selection scheme χ_{opt}

(b) Results under another selection scheme $\chi^{'}$

Figure A.1 : An illustration of the distribution of simultaneous transmitters, where the triangular points represent the helpers that are chosen as simultaneous transmitters and the dots represent the helpers that are not chosen as transmitters.

sion range that is different from the VoI that the n-th transmitter transmits to. Therefore, there is no helper within road segment (X_n, X_{n+1}) that can transmit simultaneously. If $Y_{n+1} < X_{n+1}$, a contradiction must occur. Thus, $X_{n+1} \leq Y_{n+1}$.

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

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

Appendix B

Proof of Theorem 4.3 in Chapter 4.2.2

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

Denote by $V_{k,1}$ the first helper located in the road segment $[S_k + R_c, d - 2r_I]$, by $V_{k,2}$ the second helper, and so on. Denote by $l_{k,i}, i = 1, 2, ...$ the distance between helpers $V_{k,i}$ and $V_{k,i+1}$ and by $l_{k,0}$ the distance between helper $V_{k,1}$ and the point $S_k + R_c$. See Fig. B.1 for an illustration. As an easy consequence of the Poisson distribution of helpers, $l_{k,i}, i = 1, 2, ...$ follows identical and independent exponential distribution with a mean value $\frac{1}{(1-p)\rho}$, and further due to the memoryless property of exponential distribution [96], $l_{k,0}$ also has the same distribution as $l_{k,i}, i = 1, 2, ...$

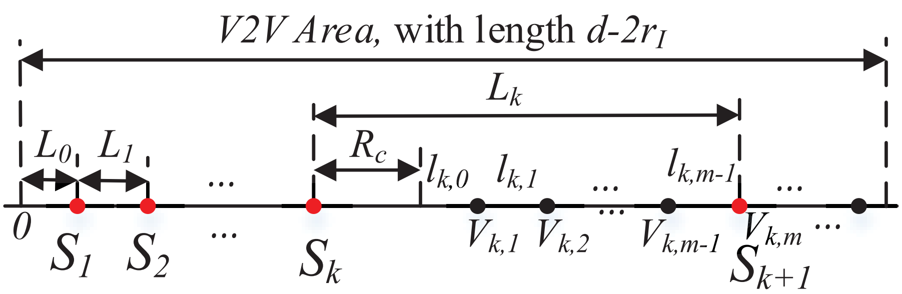


Figure B.1 : An illustration of the distribution of distances between two consecutive simultaneous transmitters.

Consider the k-th transmitter and suppose the $(k+1)$-th transmitter is exactly the m_k-th helper located within the road segment $[S_k + R_c, d - 2r_I]$, where m_k is a random integer. Note that the distribution of m_k is independent of $l_{k,i}, i = 1, 2, ...,$ but is determined by the distribution of VoIs within the coverage of each helper $V_{k,1}$, $V_{k,2}$, $\cdots$. Therefore,

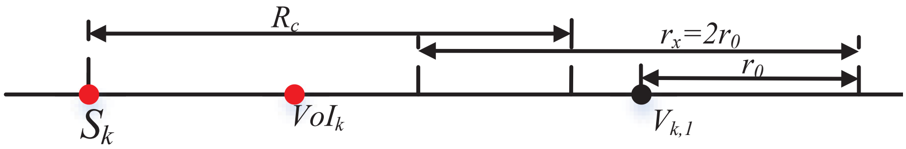
$$L_k = R_c + \sum_{i=0}^{m_k - 1} l_{k,i}. \tag{B.1}$$

From (B.1), to calculate the distribution of L_k, it remains to calculate the distribution of m_k, $Pr(m_k = m)$. We compute the distribution of m_k in the following paragraphs.

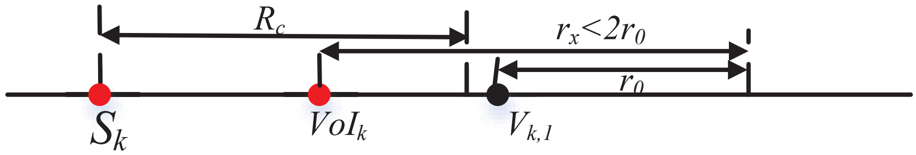
(i) when $m = 1$, it means that the first helper $V_{k,1}$ located within road segment $[S_k + R_c, d - 2r_I]$ is chosen as the $(k+1)$-th transmitter. This implies that there should be at least one VoIs different from the VoI of the kth active helper-VoI pair within the coverage of helper $V_{k,1}$.

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

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



(a) $R_c \geq 2r_0$



(b) $R_c < 2r_0$

Figure B.2 : An illustration of the case that the helper $V_{k,1}$ is chosen as the $(k+1)$-th transmitter.

on our result. This approximation allows us to write:

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

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

$$\begin{aligned}
&Pr(m_k = m) \\
\approx &Pr\,(\text{no VoI within the coverage of } V_{k,i}, i = 1, ...m-1 \\
&\cap\ \exists\text{VoI within the coverage of } V_{k,m}) \\
= &Pr\left(\text{no VoI in } \left[\sum_{i=1}^{m-2} \min\left\{l_{k,i}, 2r_0\right\} + 2r_0\right] \cap\ \exists\text{VoI in } \min\left\{l_{k,m-1}, 2r_0\right\}\right) \\
= &Pr\left(\text{no VoI in } \left[\sum_{i=1}^{m-2} \min\left\{l_{k,i}, 2r_0\right\} + 2r_0\right]\right) \times Pr\left(\exists \text{ VoI in } \min\left\{l_{k,m-1}, 2r_0\right\}\right),
\end{aligned} \tag{B.3}$$

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

Define two parameters $h_{k,i}$ and $H_{k,m}$ as follows:

$$h_{k,i} = \min\left\{l_{k,i}, 2r_0\right\}, i = 1, 2, ...m-1 \tag{B.4}$$

and

$$H_{k,m} = \sum_{i=1}^{m-2} h_{k,i} + 2r_0, m = 2, 3, ..., \tag{B.5}$$

with $H_{k,2} = 2r_0$. Because $l_{k,i}, i = 1, 2, ...m-1$ are i.i.d. random variables, it follows that $h_{k,i}, i = 1, 2, ...m-1$ are i.i.d. random variables. Denote by $\bar{f}_{h_{k,i}}(x)$ and $\bar{f}_{H_{k,m}}(x)$ the pdf of random variables $h_{k,i}$ and $H_{k,m}$ respectively. As an easy consequence of the total probability theorem, (B.3) can be calculated by:

$$\begin{aligned}
Pr(m_k = m) \approx &Pr\left(\text{no VoI in } H_{k,m}\right) \times Pr\left(\exists\text{VoI in } h_{k,m-1}\right) \\
= &\int_0^\infty e^{-p\rho x} \bar{f}_{H_{k,m}}(x)dx \times \int_0^\infty \left(1 - e^{-p\rho y}\right) \bar{f}_{h_{k,m-1}}(y)dy \\
= &E\left[e^{-p\rho H_{k,m}}\right] \times \left(1 - E\left[e^{-p\rho h_{k,m-1}}\right]\right).
\end{aligned} \tag{B.6}$$

From (B.6), we can see that when $m \geq 2$, the value of $Pr(m_k = m)$ is the product of two factors, $E\left[e^{-p\rho H_{k,m}}\right]$, and $\left(1 - E\left[e^{-p\rho h_{k,m-1}}\right]\right)$. Further note that $E\left[e^{-p\rho H_{k,m}}\right]$

and $E\left[e^{-p\rho h_{k,m-1}}\right]$ are in the form of the moment generating functions (MGF) of the random variables $H_{k,m}$ and $h_{k,m-1}$ respectively. For a random variable X, its MGF is defined as follows [99]:

$$M_X(t) \triangleq E[e^{tX}], \quad t \in \mathbb{R}. \tag{B.7}$$

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

$$\begin{aligned} Pr(m_k = m) &\approx \left(M_{H_{k,m}}(t) \cdot \left(1 - M_{h_{k,m-1}}(t)\right)\right)|_{t=-p\rho} \\ &= \left(\prod_{i=1}^{m-2} M_{h_{k,i}}(t) \cdot M_{2r_0}(t) \cdot \left(1 - M_{h_{k,m-1}}(t)\right)\right)|_{t=-p\rho} \\ &= \left(\left(M_{h_{k,i}}(t)\right)^{m-2} \cdot e^{2r_0 t} \cdot \left(1 - M_{h_{k,i}}(t)\right)\right)|_{t=-p\rho} \end{aligned} \tag{B.8}$$

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

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

$$\bar{f}_{h_{k,i}}(x) = \begin{cases} (1-p)\rho e^{-(1-p)\rho x} & x < 2r_0 \\ e^{-(1-p)\rho 2r_0}\delta(x - 2r_0) & x \geq 2r_0 \end{cases} \tag{B.9}$$

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

$$\delta(x) = \begin{cases} 1, & x = 0 \\ 0, & x \neq 0 \end{cases} \tag{B.10}$$

Using (B.9), the MGF of $h_{k,i}$ can be obtained as follows:

$$\begin{aligned}M_{h_{k,i}}(t) =&E[e^{t\cdot h_{k,i}}]\\ =&\int_0^{\infty} e^{tx}\bar{f}_{h_{k,i}}(x)dx\\ =&\int_0^{2r_0} e^{tx}(1-p)\rho e^{-(1-p)\rho x}dx + e^{t2r_0}e^{-(1-p)\rho 2r_0}\\ =&\frac{te^{(t-(1-p)\rho)2r_0}-(1-p)\rho}{t-(1-p)\rho}.\end{aligned} \tag{B.11}$$

Combing (B.8) and (B.11) and simplifying it, for $m \geq 2$, we have

$$Pr(m_k = m) \approx e^{-p\rho 2r_0}\cdot c_1^{m-2}(1-c_1) \tag{B.12}$$

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

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

$$\begin{aligned}F_{L_k}(x) =&Pr(L_k \leq x)\\ =&\sum_{m=1}^{\infty} Pr(R_c + \sum_{i=0}^{m_k-1} l_{k,i} \leq x | m_k = m)Pr(m_k = m)\\ =&\sum_{m=1}^{\infty} Pr(\sum_{i=0}^{m-1} l_{k,i} \leq x - R_c)Pr(m_k = m)\\ =&\sum_{m=1}^{\infty}\sum_{n=m}^{\infty} \frac{e^{-(1-p)\rho(x-R_c)}\left[(1-p)\rho(x-R_c)\right]^n}{n!}Pr(m_k = m)\end{aligned} \tag{B.13}$$

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

From (B.13) and with $f_{L_k}(x) = \frac{dF_{L_k}(x)}{dx}$, we have the pdf of L_k shown as (4.13), which completes the proof.

Appendix C

Proof of Theorem 4.5 in Chapter 4.2.3

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

(1). V2V communications:

Recall that under our optimum helper-VoI scheduling algorithm χ_{opt} for the V2V communications proposed in Theorem 4.2, for any two randomly chosen helper-VoI pairs at a randomly chosen time slot, the travel direction of the VoI in one pair is independent of the travel direction of the helper in the same pair, and is also independent of the travel direction of the VoI in the other pair. Therefore, at any randomly chosen time slot, the proportion of the helper-VoI pairs whose VoI is eastbound (westbound) is equal to the probability that the VoI of a randomly chosen helper-VoI pair is eastbound (westbound), denoted by P_{Ve} (P_{Vw}). Obviously, $P_{Ve}+P_{Vw}=1$. It follows that the maximum expected amount of data the eastbound and westbound VoIs can receive through V2V communications during time period $[0,t]$ from one cycle area, denoted by $D_{V2Ve}(t)$ and $D_{V2Vw}(t)$ respectively, can be calculated by $D_{V2Ve}(t) = P_{Ve} \cdot D_{V2V}(t)$ and $D_{V2Vw}(t) = P_{Vw} \cdot D_{V2V}(t)$, where $D_{V2V}(t)$ is the expected amount of data received by all the VoIs given in (4.3). Therefore, the capacity achieved by the eastbound and westbound VoIs through V2V communications from one cycle area are respectively:

$$\lim_{t\to\infty} \frac{D_{V2Ve}(t)}{t} = P_{Ve} \lim_{t\to\infty} \frac{D_{V2V}(t)}{t}, \tag{C.1}$$

and

$$\lim_{t\to\infty}\frac{D_{V2Vw}(t)}{t} = P_{Vw}\lim_{t\to\infty}\frac{D_{V2V}(t)}{t}. \tag{C.2}$$

In the following, we will calculate P_{Ve} and P_{V_w}. Suppose helper V_H is one of the simultaneous transmitters at a randomly chosen time slot. Noting that for a randomly chosen helper-VoI pair, the travel direction of its VoI is irrelevant to the travel direction of its helper, but only dependent on the original distribution of VoIs in each direction. Therefore, without loss of generality, we assume that helper V_H travels eastbound. Recall that we designate the east (right) direction as $+x$ direction. Here, designate the point to the $-x$ direction of helper V_H and at a distance r_0 to V_H as the origin of the coordinate system. Denote by z the location of the point from which the helper V_H starts to choose its receiver (VoI), i.e., the helper V_H chooses its receiver within road segment $[z, 2r_0]$. Therefore, according to the scheduling scheme χ_{opt}, z is equal to 0 if the VoI of the previous helper-VoI pair is not located within the coverage of helper V_H, otherwise $z > 0$. See Fig. C.1 for an illustration.

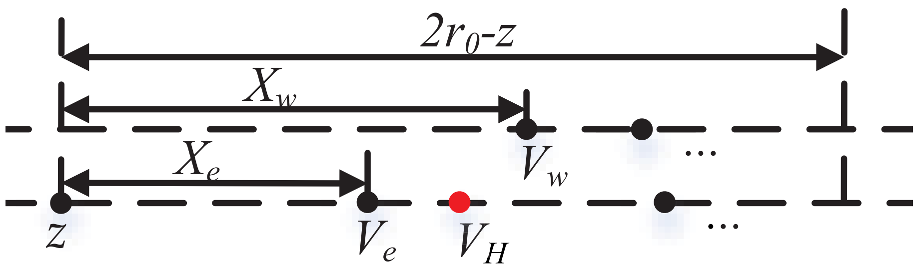


Figure C.1 : An illustration of the coordinate system, the location of the randomly chosen transmitter, and the left-most VoI from each direction that are located at the right of origin.

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

$$\begin{aligned}
P_{Ve} =& Pr(X_e \leq X_w | \text{there exists VoI in } [z, 2r_0]) \\
=& \frac{Pr(X_e \leq X_w, \text{there exists VoI in } [z, 2r_0])}{Pr(\text{there exists VoI in } [z, 2r_0])} \\
=& \frac{\int_0^{2r_0 - z} Pr(X_e \leq x) f_{X_w}(x) dx}{1 - e^{-p\rho(2r_0 - z)}} + \\
& \frac{\int_{2r_0 - z}^{\infty} Pr(X_e \leq 2r_0 - z) f_{X_w}(x) dx}{1 - e^{-p\rho(2r_0 - z)}} \\
=& \frac{\rho_1}{\rho_1 + \rho_2},
\end{aligned} \tag{C.3}$$

where the second step results by using Bayes' theorem; the third step is obtained by using total probability theorem. In addition, the result of (C.3) is irrelevant to the random variable z. Straightforwardly, we have:

$$P_{Vw} = 1 - P_{Ve} = \frac{\rho_2}{\rho_1 + \rho_2}. \tag{C.4}$$

Plugging (C.3) and (C.4) into (C.1) and (C.2) respectively, we can conclude that the expected amount of data received by the eastbound and westbound VoIs through V2V communications are proportional to their respective traffic densities.

(2). V2I communications:

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

$$\lim_{t\to\infty} \frac{D_{V2Ie}(t)}{t} = \frac{\rho_1}{\rho_1+\rho_2} \cdot \lim_{t\to\infty} \frac{D_{V2I}(t)}{t}, \tag{C.5}$$

and

$$\lim_{t\to\infty} \frac{D_{V2Iw}(t)}{t} = \frac{\rho_2}{\rho_1+\rho_2} \cdot \lim_{t\to\infty} \frac{D_{V2I}(t)}{t}. \tag{C.6}$$

where $D_{V2Ie}(t)$ and $D_{V2Iw}(t)$ are respectively the maximum expected amount of data the eastbound and westbound VoIs can receive through V2I communications during time period t from one cycle area.

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

Bibliography

[1] ([Online] Available: http://mobility.tamu.edu/ums/report/) 2015 Urban Mobility Scorecard, Texas A&M Transportation Institute.

[2] M. Alam, J. Ferreira and J. Fonseca, "Introduction to Intelligent Transportation Systems," *Intelligent Transportation Systems, Studies in Systems, Decision and Control*, Springer, Cham, vol 52. 2016.

[3] ([Online] Available: http://www.who.int/gho/publications/world_health_statistics/en/) World Health Organization, World Health Statistics 2017.

[4] J.R.Wootton, A.Garca-Ortiz, S.M.Amin, "Intelligent transportation systems: A global perspective," *Mathematical and Computer Modelling*, vol. 22, no. 4-7, pp. 259-268, 1995.

[5] T. Hasegawa, Traffic and Safety Sciences: Interdisciplinary Wisdom of IATSS, *The Japan Times, Ltd*, Mar. 2015.

[6] X. Yan, H. Zhang, C. Wu, "Research and Development of Intelligent Transportation Systems," *International Symposium on Distributed Computing and Applications to Business, Engineering* & *Science*, 2012.

[7] J. M. Sussman, Perspectives on Intelligent Transportation Systems (ITS), *Springer*, 2005.

[8] L. Figueiredo, I. Jesus, J. A. T. Machado, J. R. Ferreira, J. L. M. Carvalho," Towards the Development of Intelligent Transportation Systems," *IEEE Intelligent Transportation Systems Conference Proceedings*, 2001.

[9] O. Popescu, S. Sha-Mohammad, H. Abdel-Wahab, D. C. Popescu, and S. El-Tawab, "Automatic Incident Detection in Intelligent Transportation Systems Using Aggregation of Traffic Parameters Collected Through V2I Communications," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 2, pp. 64-75, Summer 2017.

[10] M. A. Javed, E. B. Hamida, "On the Interrelation of Security, QoS, and Safety in Cooperative ITS," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp 1943 - 1957, Jul. 2017.

[11] J. E. Siegel, D. C. Erb, and S. E. Sarma, "A Survey of the Connected Vehicle Landscape-Architectures, Enabling Technologies, Applications, and Development Areas," *IEEE Transactions on Intelligent Transportation Systems*, 2017.

[12] K. Zheng, Q. Zheng, et al., "Heterogeneous Vehicular Networking: A Survey on Architecture, Challenges, and Solutions," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2377-2396, Fourth Quarter, 2015.

[13] J. B. Kenney, "Dedicated Short-Range Communications (DSRC) Standards in the United States," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1162-1182, Jun. 2011.

[14] ETSI TR 101 607 (v1.1.1), Intelligent Transport Systems (ITS); Cooperative ITS (C-ITS); Release 1,, 2013.

[15] R. Molina-Masegosa, J. Gozalvez, "LTE-V for Sidelink 5G V2X Vehicular Communications: A New 5G Technology for Short-Range Vehicle-to-Everything Communications," *IEEE Vehicular Technology Magazine*, vol. 12, no. 4, pp. 30-39, Dec. 2017.

[16] IEEE Standard for Information TechnologyTelecommunications and information exchange between systems local andmetropolitan area networksSpecific

requirements Part 11:Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Std 802.11-2012 (Revision of IEEE Std 802.11-2007), pp. 12793, Mar 2012.

[17] ([Online] Available: https://static.tti.tamu.edu/tti.tamu.edu/documents/PRC-17-77-F.pdf) Connected Vehicle Infrastructure: Deployment and Funding Overview Final report, Texas A&M Transportation Institute.

[18] ([Online] Available:http://pressroom.toyota.com/releases)

[19] E. Ahmed, H. Gharavi, "Cooperative Vehicular Networking: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 996 - 1014, Mar. 2018.

[20] M. Khabazian, S. Aissa, and M. Mehmet-Ali, "Performance Modeling of Safety Messages Broadcast in Vehicular Ad Hoc Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 380-387, Mar. 2013.

[21] Z. Zhang, G. Mao, and B. D. O. Anderson, "Stochastic Characterization of Information Propagation Process in Vehicular Ad hoc Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 122-135, Feb. 2014.

[22] A. B. Reis, S. Sargento, F. Neves, and O. K. Tonguz, "Deploying Roadside Units in Sparse Vehicular Networks: What Really Works and What Does Not," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 6, pp. 2794-2806, Jul. 2014.

[23] H. Seo, K-D. Lee, et al., "LTE Evolution for Vehicle-to-Everything Services," *IEEE Communications Magazine*, vol. 54, no. 6, pp. 22-28, Jun. 2016.

[24] A. K. Ligo, J. M. Peha, "Cost-Effectiveness of Sharing Roadside Infrastructure for Internet of Vehicles," accepted by *IEEE Transactions on Intelligent Transportation Systems*, 2018.

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

[26] H. N. Dai, R. C. W. Wong, and H. Wang, "On Capacity and Delay of Multichannel Wireless Networks with Infrastructure Support," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1589 - 1604, Feb. 2017.

[27] M. Wang, et al., "Asymptotic Throughput Capacity Analysis of VANETs Exploiting Mobility Diversity," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 9, pp. 4187-4202, Sep. 2015.

[28] Y. Huang, M. Chen, et al., "Graph Theory Based Capacity Analysis for Vehicular Ad Hoc Networks," in *Proc. IEEE Globecom*, Dec. 2015, pp. 1 - 5

[29] G. Mao, Z. Lin, X. Ge, and Y. Yang, "Towards a Simple Relationship to Estimate the Capacity of Static and Mobile Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 3883-3895, Aug. 2013.

[30] Z. Lu, G. Qu, Z. Liu, "A Survey on Recent Advances in Vehicular Network Security, Trust, and Privacy," accepted by *IEEE Transactions on Intelligent Transportation Systems*, 2018.

[31] M. Azees, P. Vijayakumar, and L. J. Deborah, "Comprehensive survey on security services in vehicular ad-hoc networks," *IET Intelligent Transport Systems*, vol. 10, no. 6, pp. 379388, Aug. 2016.

[32] B. Brecht, D. Therriault, A. Weimerskirch, W. Whyte, V. Kumar, T. Hehn, R. Goudy, "A Security Credential Management System for V2X Communica-

tions," accepted by *IEEE Transactions on Intelligent Transportation Systems*, 2018.

[33] A. Rawat, S. Sharma, and R. Sushil, "VANET: Security attacks and its possible solutions," *Journal of Information and Operations Management*, vol. 3, no. 1, pp. 301-304, 2012.

[34] E. Fonseca and A. Festag, "A survey of existing approaches for secure ad hoc routing and their applicability to VANETs," *NEC Network Laboratories*, Heidelberg, Germany, NEC Tech. Rep. NLE-PR-2006-19, Version 1.1, Mar. 2006.

[35] J. Sun, C. Zhang, Y. Zhang, Y. Fang, "An Identity-Based Security System for User Privacy in Vehicular Ad Hoc Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 9, pp. 1227 - 1239, Sep. 2010.

[36] H. Tan, M. Ma, H. Labiod, A. Boudguiga, J. Zhang, and P. H. J. Chong, "A secure and authenticated key management protocol (SA-KMP) for vehicular networks," *IEEE Transactions Vehicular Technology*, vol. 65, no. 12, pp. 9570 - 9584, Dec. 2016.

[37] J. Petit, F. Schaub, M. Feiri, F. Kargl, "Pseudonym Schemes in Vehicular Networks: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 228-255, First Quarter, 2015.

[38] Q. Yang, H. Wang, "Toward trustworthy vehicular social networks," *IEEE Communications Magazine*, vol. 53, no. 8, pp, 42 - 47, Aug. 2015.

[39] S. Dietzel, R. Heijden, H. Decke and F. Kargl, "A flexible, subjective logic-based framework for misbehavior detection in V2V networks," in *Proceeding of IEEE WoWMoM*, 2014.

[40] Y. Du, M. Chowdhury, M. Rahman, K. Dey, A. Apon, A. Luckow, L. B. Ngo, "A Distributed Message Delivery Infrastructure for Connected Vehicle Technology Applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 787 - 801, Mar. 2018.

[41] S. Dietzel, J. Petit, G. Heijenk, and F. Kargl, "Graph-Based Metrics for Insider Attack Detection in VANET Multihop Data Dissemination Protocols," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 4, pp. 1505-1518, May. 2013.

[42] M. Raya, P. Papadimitratos, V. D. Gligor, and J. P. Hubaux, "On Data-Centric Trust Establishment in Ephemeral Ad Hoc Networks," in *IEEE INFOCOM*, 2008.

[43] Z. Huang, S. Ruj, M. A. Cavenaghi, M. Stojmenovic, and A. Nayak, "A social network approach to trust management in VANETs," *Peer-to-Peer Networking and Applications*, vol. 7, no. 3, pp. 229-242, 2014.

[44] K. Zaidi, M. B. Milojevic, V. Rakocevic, A. Nallanathan, and M. Rajarajan, "Host-Based Intrusion Detection for VANETs: A Statistical Approach to Rogue Node Detection," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6703-6714, Aug. 2014.

[45] J. Radak, B. Ducourthial, V. Cherfaoui, and S. Bonnet, "Detecting Road Events Using Distributed Data Fusion: Experimental Evaluation for the Icy Roads Case," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 184-194, Jan. 2016.

[46] W. Zhu, D. Li, and W. Saad, "Multiple Vehicles Collaborative Data Download Protocol via Network Coding," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1607-1619, Apr. 2015.

[47] Y. Wang, J. Zheng, and N. Mitton, "Delivery Delay Analysis for Roadside Unit Deployment in Vehicular Ad Hoc Networks with Intermittent Connectivity," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 8591 - 8602, Oct. 2016.

[48] N. Lu, N. Zhang, et al., "Vehicles Meet Infrastructure: Toward Capacity-Cost Tradeoffs for Vehicular Access Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1266-1277, Sep. 2013.

[49] Z. Haibo, et al., "ChainCluster: Engineering a Cooperative Content Distribution Framework for Highway Vehicular Communications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no.6, pp. 2644-2657, Dec. 2014.

[50] B. Das, S. Misra, and U. Roy, "Coalition Formation for Cooperative Service-Based Message Sharing in Vehicular Ad Hoc Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 1, pp. 144-156, Jan. 2016.

[51] Y. Zhu, L. Song, S. Wu, H. Wang and C. Wang, "Cooperative Stepwise Relaying and Combining for Multihop Vehicular Wireless Communication," *IEEE Transactions on Vehicular Technology*, vol 64, no. 6, pp. 2663-2671, Jun. 2015.

[52] Q. Yan, M. Li, Z. Yang, W. Lou, and H. Zhai, "Throughput analysis of cooperative mobile content distribution in vehicular network using symbol level network coding," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 2, pp. 484-492, Feb. 2012.

[53] M. Li, Z. Yang, and W. Lou, "CodeOn: Cooperative Popular Content Distribution for Vehicular Networks using Symbol Level Network Coding," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 1, pp. 223-235, Jan. 2011.

[54] T. Wang, L. Song and Z. Han, "Coalitional Graph Games for Popular Content Distribution in Cognitive Radio VANETs," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 8, pp. 4010-4019. Oct. 2013.

[55] K. Liu, et al., "Cooperative Data Scheduling in Hybrid Vehicular Ad Hoc Networks: VANET as a Software Defined Network," *IEEE/ACM Transactions on Networking* , vol. 24, no. 3, pp. 1759 - 1773, Jun. 2016.

[56] K. Liu, J. K. Y. Ng, J. Wang, V. C. S. Lee, W. Wu, and S. H. Son, "Network-Coding-Assisted Data Dissemination via Cooperative Vehicle-to-Vehicle/-Infrastructure Communications," IEEE *Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, Jun. 2016.

[57] D. Zhang and C. Yeo, "Enabling efficient WiFi-based vehicular content distribution," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 3, pp. 479-492, Mar. 2013.

[58] Y. Li, X. Zhu, D. Jin and D. Wu, "Multiple Content Dissemination in Roadside-Unit-Aided Vehicular Opportunistic Networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 8, pp. 3947-3956, Oct. 2014.

[59] R. Kim, H. Lim, and B. Krishnamachari, "Prefetching-Based Data Dissemination in Vehicular Cloud Systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 1, pp. 292 - 306, Jan. 2016.

[60] G. Mauri, M. Gerla, et al., "Optimal Content Prefetching in NDN Vehicle-to-Infrastructure Scenario," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 3, pp. 2513 - 2525, Mar. 2017.

[61] K. Mershad, H. Artail, and M. Gerla, "We Can Deliver Messages to Far Vehicles," *IEEE Transactions on Intelligent Transportation Systems* vol. 13, no. 3, pp. 1099-1115, Sep. 2012.

[62] P. Si, Y. He, et al., "DaVe: Offloading Delay-Tolerant Data Traffic to Connected Vehicle Networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 3941 - 3953, Jun. 2016.

[63] Y. Wang, Y. Liu, et al., "Cooperative Store-Carry-Forward Scheme for Intermittently Connected Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 777 - 784, Jan. 2017.

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

[65] T. Yang, G. Mao, W. Zhang and X. Tao, "Transport Capacity of Distributed Wireless CSMA Networks," *IEEE Transactions on Wireless Communications*, Vol. 13, No. 10, pp. 5635 - 5647, Oct. 2014.

[66] M. Franceschetti, O. Dousse, D. N. C. 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, Mar. 2007.

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

[68] C. K. Chau, M. Chen, and S. C. Liew, "Capacity of large-scale csma wireless networks," *IEEE/ACM Transactions on Networking*, vol. 19, no. 3, pp. 893 - 906, Jun. 2011.

[69] P. Jacquet, S. Malik, B. Mans, and A. Silva, "On the throughput-delay tradeoff in georouting networks," in *IEEE INFOCOM*, pp. 765 - 773, 2012

[70] 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, Feb. 2009.

[71] N. Lu, X. S. Shen, "Scaling Laws for Throughput Capacity and Delay in Wireless Networks - A Survey," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 642-657, Second Quarter, 2014.

[72] M. Ni, J. Pan, L. Cai, J. Yu, H. Wu, Z. Zhong, "Interference-Based Capacity Analysis for Vehicular Ad Hoc Networks," *IEEE Communications Letters*, vol. 19, no. 4, pp. 621-624, Apr. 2015.

[73] H. Yang, W. Jin, "Instantaneous communication throughputs of vehicular ad hoc networks," *Transportation Research Part C: Emerging Technologies*, vol. 53, pp. 19-34, 2015.

[74] T. Gazdar, A. Rachedi, A. Benslimane, and A. Belghith, "A distributed advanced analytical trust model for VANETs," in *IEEE GLOBECOM* 2012.

[75] S. Ahmed, S. Al Rubeaai, and K. Tepe, "Novel Trust Framework for Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 9498 - 9511, Oct. 2017.

[76] N. Haddadou, A. Rachedi, and Y. Ghamri-Doudane, "A Job Market Signaling Scheme for Incentive and Trust Management in Vehicular Ad Hoc Networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 8, pp. 3657-3674, Aug. 2015.

[77] H. Sedjelmaci, and S. M. Senouci, "An accurate and efficient collaborative intrusion detection framework to secure vehicular networks," *Computers and Electrical Engineering*, vol. 43, pp. 33-47, 2015.

[78] S. K. Dhurandher, M. S. Obaidat, A. Jaiswal, A. Tiwari and A. Tyagi, "Vehicular Security Through Reputation and Plausibility Checks," *IEEE Systems Journal*, vol. 8, no. 2, pp. 384 - 394. Jun. 2014.

[79] W. Li, and H. Song, "ART: An Attack-Resistant Trust Management Scheme for Securing Vehicular Ad Hoc Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 960-969, Apr. 2016.

[80] K. Abboud and W. Zhuang, "tochastic Analysis of a Single-Hop Communication Link in Vehicular Ad Hoc Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2297-2307, Oct. 2014.

[81] W. Zhang, et al., "Multi-Hop Connectivity Probability in Infrastructure-Based Vehicular Networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 740-747, May. 2012.

[82] N. Wisitpongphan, B. Fan, P. Mudalige, V. Sadekar, and O. Tonguz, "Routing in Sparse Vehicular Ad Hoc Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 8, pp. 1538-1556, Oct. 2007.

[83] R. Mao and G. Mao, "Road traffic density estimation in vehicular networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2013.

[84] W. Wei, S. S. Liao, L. Xin, and J. S. Ren, "The Process of Information Propagation Along a Traffic Stream Through Intervehicle Communication," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 345-354, Feb. 2014.

[85] X. Ge, S. Tu, T. Han, Q. Li and G. Mao, "Energy Efficiency of Small Cell Backhaul Networks Based on Gauss-Markov Mobile Models," *IET Networks*, Vol. 4, No. 2, pp. 158-167, 2015.

[86] G. Mao, and B. D. O. Anderson, "Graph Theoretic Models and Tools for the Analysis of Dynamic Wireless Multihop Networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2009.

[87] W. Tan, W. Lau, O. Yue and T. Hui, "Analytical Models and Performance Evaluation of Drive-thru Internet Systems," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 1, pp. 207-222, Jan. 2011.

[88] J. Cheng, J. Cheng, M. Zhou, F. Liu, S. Gao, and C. Liu, "Routing in Internet of Vehicles: A Review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2339-2352, Oct. 2015.

[89] N. Lu, T. H. Luan, et al., "Bounds of Asymptotic Performance Limits of Social-Proximity Vehicular Networks," *IEEE/ACM Transactions on Networking*, vol. 22, no. 3, pp. 812-825, Jun. 2014.

[90] L. Zhu, C. Li, et al., "On Stochastic Analysis of Greedy Routing in Vehicular Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3353 - 3366, Dec. 2015.

[91] J. Ning, S. Singh, et al., "Forensic Analysis of Packet Losses in Wireless Networks," *IEEE/ACM Transactions on Networking*, vol. 24, no. 4, pp. 1975-1988, Aug. 2016.

[92] A. Le, A. S. Tehrani, et al., "Recovery of Packet Losses in Wireless Broadcast for Real-Time Applications," *IEEE/ACM Transactions on Networking*, vol. 25, no. 2, pp. 676-689, Apr. 2017.

[93] A. Ashish and P. R. Kumar, "Capacity bounds for ad hoc and hybrid wireless networks," *ACM SIGCOMM Computer Communications Review*, vol. 34, no. 3, pp. 71-81, Jul. 2004.

[94] P. Wang, G. Mao, Z. Li, X. Ge and B. D. O. Anderson, "Network Coding based Wireless Broadcast with Performance Guarantee," *IEEE Transactions on Wireless Communications*, Vol. 14, No. 1, pp. 532-544, Jan. 2015.

[95] R. G. Gallager, *Stochastic Processes: Theory for Applications.* Cambridge University Press, 2013.

[96] W. Feller, An Introduction to Probability Theory and Its Applications, vol. 2. New York, NY, USA: *Wiley*, 1971.

[97] Z. Zhang, G. Mao, T. Han and B. D. O. Anderson, "Cooperative Information Forwarding in Vehicular Networks Subject to Channel Randomness," in *Proceedings of IEEE ICC*, 2014.

[98] K. Zheng, F. Liu, Q. Zheng, W. Xiang and W. Wang, "A Graph-Based Cooperative Scheduling Scheme for Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 4, pp. 1450-1458, May. 2013.

[99] R. Nelson, Probability, Stochastic Processes, and Queueing Theory: The Mathematics of Computer Performance Modeling, New York: *Springer- Verlag*, 1995.

[100] F. Baccelli and B. Błaszczyszyn, *Stochastic Geometry and Wireless Networks Volume II Applications*, NOW: Foundations and Trends in Networking, 2010.

[101] J. Ma, F. Le, A. Russo, J. Lobo, "Detecting distributed signature-based intrusion: The case of multi-path routing attacks," in *IEEE Conference on Computer Communications (INFOCOM)*, 2015.

[102] J. Ponniah, Y. C. Hu and P. R. Kumar, "A Clean Slate Approach to Secure Ad Hoc Wireless Networking - Open Unsynchronized Networks," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 37 - 48, March, 2017.

[103] Y. Wang, H. Ieda, and F. Mannering, "Estimating Rear-End Accident Probabilities at Signalized Intersections: Occurrence-Mechanism Approach," *Journal of Transportation engineering*, vol. 129, no. 4, Jul. 2003.

[104] M. Kam, Q. Zhu and W. S. Gray, "Optimal data fusion of correlated local decisions in multiple sensor detection systems," *IEEE Transactions on Aerospace and Electronic Systems* , vol. 28, no. 3, pp. 916-920, Jul. 1992.

[105] Y. Zhu, Multisensor Decision and Estimation Fusion, United States, *Springer*, 2003.

[106] M. H. DeGroot, and M. J. Schervish, Probability and statistics, fourth edition, Boston, MA: Addison-Wesley, 2002.

[107] J. L. Gross, and J. Yellen, Graph theory and its applications, second edition, Boca Raton, Florida, USA, Chapman & Hall/CRC, 2006.

[108] D. R. Karger and C. Stein, "A new approach to the minimum cut problem," *Journal of the ACM (JACM)*, vol. 43, no. 4, pp. 601 - 640, Dec. 1996.

[109] A. V. Goldberg and R. E. Tarjan, "A new approach to the maximum-flow problem," *Journal of the Association for Computing Machinery* Mach., vol. 35, no. 4, pp. 921 - 940, Oct. 1988.