Communications and Networking for Connected Vehicles 2020

Lead Guest Editor: Li Zhu
Guest Editors: Richard Yu, Victor Leung, Hongwei Wang, and Changqing Luo

Communications and Networking for Connected Vehicles 2020

Communications and Networking for Connected Vehicles 2020

Lead Guest Editor: Li Zhu
Guest Editors: Richard Yu, Victor Leung, Hongwei Wang, and Changqing Luo

Reza Monir Vaghefi, USA
Juan F. Valenzuela-Valdés, Spain
Lorenzo Vangelista, Italy
Quoc-Tuan Vien, United Kingdom
Enrico M. Vitucci, Italy
Yingjie Wang, China
Huaqun Wang, China
Honggang Wang, USA
Ding Wang, China
Lifei Wei, China
Miaowen Wen, China
Dapeng Wu, China
Huaming Wu, China
liang wu, China
Ding Xu, China
Jie Yang, USA
Long Yang, China
YAN YAO, China
Qiang Ye, Canada
Ya-Ju Yu, Taiwan
Marat V. Yuldashev, P.O. Box 35 (Agora),
FIN-40014, Finland, Finland
Sherali Zeadally, USA
Jie Zhang, United Kingdom
Yin Zhang, China
Hong-Hai Zhang, USA
Yushu Zhang, China
Lei Zhang, Spain
Jiliang Zhang, United Kingdom
Wence Zhang, China
Xu Zheng, USA
Fuhui Zhou, USA
Meiling Zhu, United Kingdom
Zhengyu Zhu, China

Contents

Editorial

Communications and Networking for Connected Vehicles 2020

Li Zhu [1], **Richard Yu** [2], **Victor Leung** [3], **Hongwei Wang**,[1] and **Changqing Luo**[4]

[1]*State Key Lab. of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China*
[2]*Department of Systems and Computer Eng., Carleton University, Ottawa, Canada*
[3]*Department of Electrical and Computer Engineering University of British Columbia, Vancouver, Canada*
[4]*Virginia Commonwealth University, Richmond, USA*

Correspondence should be addressed to Li Zhu; zhulibjtu@gmail.com

With the new wave of urbanization, increasingly stringent emission standards, and intense pressure to improve the efficiency of private and public transportation, the development of more sustainable transportation systems becomes one of the fundamental social challenges of the next decade. Connected vehicles have been envisioned to provide enabling key technologies to enhance transportation efficiency, reduce incidents, improve safety, and mitigate traffic congestion impacts. The seamless integration and convergence of vehicular communication networks, information and transportation systems, and mobile devices and networks will face a number of technical, economic, and regulatory challenges. It is of paramount importance to (i) design vehicular communication systems that enable road users and other actors to exchange information in real-time with high reliability; (ii) enable pervasive sensing to monitor the status of vehicles and the surroundings; (iii) develop data analytics tools for processing large amounts of data generated by the connected vehicles; and (iv) develop middleware platforms for data management and sharing.

Within this context, we present a collection of high-quality research papers presenting the latest developments, current research challenges, and future directions in the use of control, communication, and emerging technologies for safer and more efficient communication and networking of connected vehicles.

Z. Wang and his research fellows aim to achieve efficient bandwidth aggregation of heterogeneous wireless networks in vehicle-to-ground communication. A network coding scheme called Delay Determined Group Size (DDGS) is proposed for adaptively adjusting the coding group according to the heterogeneous wireless network state. The mathematical analysis and process design of the DDGS scheme are discussed in detail. A large number of simulations proved that the DDGS scheme is significantly superior to other coding group determination schemes in terms of decoding time cost and bandwidth aggregation efficiency. DDGS is very suitable for heterogeneous networks with low latency, high bandwidth links, high latency, and low bandwidth links in vehicle networks.

As one of the important parts of an intelligent transportation system (ITS), the V2V communication system allows vehicles to exchange information about the surrounding traffic situation to improve safety, reduce traffic congestion, and provide a comfortable driving experience. The channel characteristics are of vital importance in the research, design, and deployment of the V2V communication system. L. Xiong et al. in Beijing Jiaotong University investigate the vehicle-to-vehicle (V2V) channel characteristics in peak hours at the 5.9 GHz band in two typical urban road scenarios, the urban straight road and the intersection. The channel characteristics, such as path loss, root mean square (RMS) delay spread, and angular spread, are derived from the ray-tracing (RT) simulations. Due to the low height of antennas at both the transmitter (Tx) and the receiver (Rx), the line of sight (LOS) between the Tx and the Rx will often be obstructed by other vehicles. Based on the RT simulation results, the shadowing loss is modelled by the multimodal Gaussian distribution, and path loss models in both LOS and non-LOS (NLOS) conditions are obtained. And the RMS delay spread in two scenarios can be modelled by the Weibull distribution. In addition, the deployment of

an antenna array is discussed based on the statistics distribution of the angular spread.

How to improve delay-sensitive traffic throughput is an open issue in vehicular communication networks, where a great number of vehicle to infrastructure (V2I) and vehicle to vehicle (V2V) links coexist. To address this issue, Y. Huang et al. in Chongqing University of Arts and Sciences propose a hybrid deep transfer learning scheme for allocating radio resources. Specifically, the traffic throughput maximization problem is first formulated by considering interchannel interference and statistical delay guarantee. The effective capacity theory is then applied to develop a power allocation scheme on each channel reused by a V2I and a V2V link. Thereafter, a deep transfer learning scheme is proposed to obtain the optimal channel assignment for each V2I and V2V link. Simulation results validate that the proposed scheme provides a close performance guarantee compared to a globally optimal scheme. Besides, the proposed scheme can guarantee lower delay violation probability than the schemes aiming to maximize the channel capacity.

Applying FD and EH techniques is inevitable for future wireless networks, especially for V2V communication systems, because of the big advantages of these techniques. Authors in Telecommunications University evaluate the ergodic capacity (EC) of full-duplex (FD) amplify-and-forward (AF) and decode-and-forward (DF) relay system with energy harvesting (EH) for vehicle-to-vehicle (V2V) communications. Unlike previous works on FD-EH systems, the case that both relay and destination are mobile vehicles while the source is a static base station is considered. The exact closed-form expressions of ECs of both AF and DF protocols of the considered FD-EH-V2V relay system over cascade (double) Rayleigh fading are derived mathematically. Numerical results show that the ECs in the case of the V2V communication system are reduced compared to those in the case of stationary nodes. Also, with a specific value of residual self-interference (RSI), the ECs of the considered FD-EH-V2V relay system can be higher or lower than those of half-duplex- (HD-) EH-V2V system, depending on the average transmission power of the source. Furthermore, when the transmission power of the source and RSI are fixed, the ECs of the considered system can achieve peak values by using optimal EH time duration. On the other hand, the ECs of both AF and DF protocols reach the capacity floors in the high signal-to-noise ratio (SNR) regime due to the RSI impact. Also, the effect of RSI dominates the impact of cascade Rayleigh fading in the high SNR regime. Finally, the analysis approach is validated through Monte Carlo simulations.

Content-centric networks (CCNs) have become a promising technology for relieving the increasing wireless traffic demands. J. Ren et al. in Donghua University explore the scaling performance of mobile content-centric networks based on the nonuniform spatial distribution of nodes, where each node moves around its own home point and requests the desired content according to a Zipf distribution. The authors assume each mobile node is equipped with a finite local cache, which is applied to cache contents follow-

ing a static cache allocation scheme. According to the nonuniform spatial distribution of cache-enabled nodes, the authors introduce two kinds of clustered models, i.e., the clustered grid model and the clustered random model. In each clustered model, we analyze throughput and delay performance when the number of nodes goes infinity by means of the proposed cell-partition scheduling scheme and the distributed multihop routing scheme. The authors show that the node mobility degree and the clustering behavior play the fundamental roles in the aforementioned asymptotic performance. Finally, the authors study the optimal cache allocation problem in the two kinds of clustered models. Their findings provide a guidance for developing the optimal caching scheme.

Communication-Based Train Control (CBTC) system is a critical signal system to ensure rail transit's safe operation. Compared with the train-ground CBTC system, the train control system based on train-to-train (T2T) communication has the advantages of fast response speed, simple structure, and low operating cost. As the core part of the train control system based on T2T communication, the reliability of the data communication system (DCS) is of great significance to ensure the train's safe and efficient operation. According to the T2T communication system requirements, H. Liang et al. in Beijing Jiaotong University adopt Long-Term Evolution-Unlicensed (LTE-U) technology to design the DCS structure and establishes the reliability model of the communication system based on Deterministic and Stochastic Petri Nets (DSPN). Based on testing the real line's communication performance parameters, the DSPN model is simulated and solved by Π-tool, and the reliability index of the system is obtained. The research results show that the LTE-U-based T2T communication system designed meets the train control system's needs for communication transmission. The proposed reliability evaluation method can complete the reliability modeling of train control DCS based on T2T communication. The research will provide a strong practical and theoretical basis for the design and optimization of train control DCS based on T2T communication.

Intelligent transportation system (ITS) is more and more crucial in the modern transportation field, such as the applications of autonomous vehicles, dynamic traffic light sequences, and automatic road enforcement. As the upcoming fifth-generation mobile network (5G) is entering the deployment phase, the idea of cellular vehicle-to-everything (C-V2X) is proposed. The same 5G networks, coming to mobile phones, will also allow vehicles to communicate wirelessly with each other. Hence, 3.5 GHz, as the main sub-6 GHz band licensed in 5G, is focused. Y. Zeng et al. in Guangdong Communications & Networks Institute conduct a comprehensive study on the channel characteristics for vehicle-to-infrastructure (V2I) link at 3.5 GHz frequency band through channel measurements and ray-tracing (RT) simulations. Firstly, the channel parameters of the V2I link are characterized based on the measurements, including power delay profile (PDP), path loss, root-mean-square (RMS) delay spread, and coherence bandwidth. Then, the measurement-validated RT simulator is utilized

to conduct the simulations in order to supplement other channel parameters, in terms of the Ricean K-factor, angular spreads, the cross-correlations of abovementioned parameters, and the autocorrelation of each parameter itself. This work is aimed at helping the researchers understand the channel characteristics of the V2I link at 3.5 GHz and support the link-level and system level design for future vehicular communications of 5G.

As one of the mainstream technologies of vehicle-to-everything (V2X) communication, Cellular-V2X (C-V2X) provides high-reliability and low-latency V2X communications. And with the development of mobile cellular systems, C-V2X is evolving from long-term evolution-V2X (LTE-V2X) to new radio-V2X (NR-V2X). However, C-V2X test specification has not been completely set in the industry. In order to promote the formulation of relevant standards and accelerate the implementation of industrialization, Y. Fan et al. in Beijing Jiaotong University conduct the field test and analysis based on LTE-V2X in the industrial park scenario. Firstly, key technologies of LTE-V2X are introduced. Then, the specific methods and contents of this test are proposed, which consists of functional and network performance tests to comprehensively evaluate the communication property of LTE-V2X. Static and dynamic tests are required in both line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios to evaluate network performance. Next, the test results verify that all functions are normal, and the performance evaluation indexes are appraised and analyzed. Finally, it summarizes the whole paper and puts forward the future work.

As an emerging type of Internet of Things (IoT), Internet of vehicles (IoV) denotes the vehicle network capable of supporting diverse types of intelligent services and has attracted great attention in the 5G era. K. Wang et al. in Xi'an University of Technology consider the multimedia content caching with multicast beamforming in IoV-based vehicular edge networks. First, the authors formulate a joint vehicle-to-vehicle- (V2V-) assisted clustering, caching, and multicasting optimization problem, to minimize the weighted sum of flow cost and power cost, subject to the quality-of-service (QoS) constraints for each multicast group. Then, with the two-timescale setup, the intractable and stochastic original problem is decoupled at separate timescales. More precisely, at the large timescale, the authors leverage the sample average approximation (SAA) technique to solve the joint V2V-assisted clustering and caching problem and then demonstrate the equivalence of optimal solutions between the original problem and its relaxed linear programming (LP) counterpart; and at the small timescale, the authors leverage the successive convex approximation (SCA) method to solve the nonconvex multicast beamforming problem, whereby a series of convex subproblems can be acquired, with the convergence also assured. Finally, simulations are conducted with different system parameters to show the effectiveness of the proposed algorithm, revealing that the network performance can benefit from not only the power saving from wireless multicast beamforming in vehicular networks but also the content caching among vehicles.

The diversified service requirements in vehicular networks have stimulated the investigation to develop suitable technologies to satisfy the demands of vehicles. In this context, network slicing has been considered as one of the most promising architectural techniques to cater to the various strict service requirements. However, the unpredictability of the service traffic of each slice caused by the complex communication environments leads to a weak utilization of the allocated slicing resources. Thus, Y. Cui et al. in Chongqing University of Posts and Telecommunications use Long Short-Term Memory- (LSTM-) based resource allocation to reduce the total system delay. Specially, the authors first formulated the radio resource allocation problem as a convex optimization problem to minimize system delay. Secondly, to further reduce delay, the authors design a Convolutional LSTM- (ConvLSTM-) based traffic prediction to predict traffic of complex slice services in vehicular networks, which is used in the resource allocation processing. And three types of traffic are considered, that is, SMS, phone, and web traffic. Finally, based on the predicted results, i.e., the traffic of each slice and user load distribution, the authors exploit the primal-dual interior-point method to explore the optimal slice weight of resources. Numerical results show that the average error rates of predicted SMS, phone, and web traffic are 25.0%, 12.4%, and 12.2%, respectively, and the total delay is significantly reduced, which verifies the accuracy of the traffic prediction and the effectiveness of the proposed strategy.

With the rapid developments of wireless communication and increasing number of connected vehicles, Vehicular Ad Hoc Networks (VANETs) enable cyberinteractions in the physical transportation system. Future networks require real-time control capability to support delay-sensitive application such as connected autonomous vehicles. In recent years, fog computing becomes an emerging technology to deal with the insufficiency in traditional cloud computing. Z. Wang et al. propose a fog-based distributed network control design toward connected and automated vehicle application. The proposed architecture combines VANETs with the new fog paradigm to enhance the connectivity and collaboration among distributed vehicles. A case study of connected cruise control (CCC) is introduced to demonstrate the efficiency of the proposed architecture and control design. Finally, the authors discuss some future research directions and open issues to be addressed.

A vehicular cloud (VC) can reduce latency and improve resource utilization of the Internet of vehicles by effectively using the underutilized computing resources of nearby vehicles. Although the task offloading of the VC enhances road safety and traffic management on the Internet of vehicles and meets the low-latency requirements for driving safety services on the Internet of vehicles business, there are still some key challenges such as the resource allocation mechanism of differentiated services (DiffServ) and task offloading mechanism of improving user experience. To address these issues, Y. Kang et al. in Chongqing University of Posts and Telecommunications study the task offloading and resource allocation strategy of the VC system where tasks generated by vehicles can be offloaded and executed cooperatively by

vehicles in VC. Specifically, the computing task is further divided into independent subtasks and executed in different vehicles in VC to maximize the offloading utility. Considering the mobility of vehicles, the deadline of tasks, and the limited computing resources, the authors propose the optimization problem of task offloading in the VC system in the cause of improved user experience. To characterize the difference in service requirements resulting from the diversity of tasks, a DiffServ model focusing on the pricing of a task is utilized. The initial pricing of a task is tailored by the characteristics of the task and the uniqueness of the network status. In this model, tasks are sorted and processed in order according to task pricing, so as to optimize resource allocation. Numerical results show that the proposed scheme can effectively increase the resource utilization and task completion ratio.

Mobile robot localization has attracted substantial consideration from the scientists during the last two decades. Mobile robot localization is the basics of successful navigation in a mobile network. Localization plays a key role to attain a high accuracy in mobile robot localization and robustness in vehicular localization. For this purpose, a mobile robot localization technique is evaluated to accomplish a high accuracy. This paper provides the performance evaluation of three localization techniques named Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), and Particle Filter (PF). I. Ullah et al. in Hohai University (HHU) propose three localization techniques. The performance of these three localization techniques is evaluated and analyzed while considering various aspects of localization. These aspects include localization coverage, time consumption, and velocity. The abovementioned localization techniques present a good accuracy and sound performance compared to other techniques.

In summary, it is essential that we continue to progress in the search for appropriate models that can adequately and faithfully improve the security and performance of Communications and Networking for Connected Vehicles. The progress reported in this special edition suggests that in the future, achieving these aims might be a distant prospect, but an attainable one.

Conflicts of Interest

The editors declare that they have no conflicts of interest regarding the publication of this special issue.

Funding

Li Zhu
Richard Yu
Victor Leung
Hongwei Wang
Changqing Luo

WILEY | Hindawi

Research Article

DDGS: A Network Coding Scheme for Dynamic Adaptation to Heterogeneous Vehicular Networks

Zongzheng Wang [iD]**, Ping Dong** [iD]**, Tao Zheng** [iD]**, and Hongke Zhang** [iD]

School of Electronic and Information Engineering, Beijing Jiaotong University, China

Correspondence should be addressed to Ping Dong; pdong@bjtu.edu.cn

The rapid development of the transportation industry has brought about the demand for massive data transmission. In order to make use of a large number of heterogeneous network resources in vehicular network, the research of applying network coding to multipath transmission has become a hot topic. Network coding can better solve the problems of packet reordering and low aggregation efficiency. The determination of coding scale is the key to network coding scheme. However, the existing research cannot adapt to the different characteristics of network resources in vehicular network, leading to larger decoding time cost and lower bandwidth aggregation efficiency. In this paper, we propose a network coding scheme called Delay Determined Group Size (DDGS), which could adaptively adjust the coding group according to the heterogeneous wireless networks state. The mathematical analysis and process design of the DDGS scheme are discussed in detail. Through a large number of simulations, we proved that the DDGS scheme is significantly superior to other coding group determination schemes in terms of decoding time cost and bandwidth aggregation efficiency.

1. Introduction

With a new wave of urbanization, the transportation industry is developing rapidly. A great quantity of vehicle services and applications emerges, creating a large number of application data to be transmitted to the ground [1]. As the development of wireless network technology, network device with multiple access interfaces has the ability to access heterogeneous wireless networks [2] (e.g., LTE, 5G, and satellite link) [3], which provide sufficient wireless channel resources for growing demands for high bandwidth vehicular applications [4] (e.g., HD video, online games, and live online). Based on the development of access technology, the multipath transmission scheme which uses multiple wireless interfaces simultaneously is widely studied in vehicle-to-ground communication [5, 6]. The aggregation of heterogeneous network resources is realized by distributing data on heterogeneous networks. In order to overcome the disorder of data packets caused by the heterogeneity of network resources in multipath transmission, scholars apply network coding [7, 8] to data scheduling process.

The encoding will encrypt [9] the data to a certain extent, and it brings a certain degree of security [10]. The core of network coding is to group packets and add redundancy. By encoding packets, the packets of the same group are equivalent to the decryption process. This feature eliminates the necessity for packets to arrive in strict order. To a certain extent, the reduction of communication efficiency caused by packet loss is also avoided. However, the addition of redundancy will cause the loss of bandwidth resources. Network coding is essentially a trade-off between link bandwidth and reliability. In order to efficiently utilize network coding schemes, the design of the coding scale is particularly important. We use group size (GS) to describe the number of data packets contained in a coding group. Designing appropriate GS according to the network status is the focus of this article.

The design of GS is related to the coding and decoding complexity, reception delay, bandwidth resource utilization, etc. The well-designed GS should be combined with the real-time state of the link to reasonably divide the data stream. Excessive GS design leads to greater computational complexity and delay, which affects the performance of

end-to-end transport protocol (TCP) [11]. Too small GS design will cause coding redundancy and reduce bandwidth utilization. The current research on the determination of group size has certain limitations in both method and theory. Most of the existing research on group size determination ignores the extremely different path characteristics in heterogeneous wireless networks. This is a prominent feature of vehicle-to-ground communication systems and has an important impact on the performance of network coding.

In order to make up for the deficiencies of current research, this paper proposes a Delay Determined Group Size (DDGS) scheme, which is aimed at achieving efficient bandwidth aggregation of heterogeneous wireless networks in vehicle-to-ground communication. This scheme can realize the dynamic adjustment of GS according to the real-time multipath characteristics. We treat the same round of data packets arriving at the receiving end as a batch [12, 13], which can be calculated as the difference in the number of packets sent between the fastest path and the slowest path in the transmission delay. In DDGS, the same batch of transmitted data packets is divided into fine-grained data packets with time scales on different links. A theoretically reasonable time segment division scheme can ensure a reasonable size of GS and achieve a reasonable decoding delay, thereby ensuring the continuity of the data stream at the receiving end.

At the sending end, before each group of data packets is encoded, the DDGS scheme dynamically senses the multilink delay difference [14] and converts the transmission data packets of each link within time Δt into a group. Then, the data packets are encoded by any k-row linearly independent encoding matrix, and k refers to the number of packets in the group. We define the coding rate k/n according to the packet loss rate of each link and send redundant packets on different links to compensate for the loss characteristics of the link. At the receiving end, we use the equivalent features of the same set of encoded data to decode, and the original data can be restored after receiving any k encoded data packets. We carry out the agreement of the coding matrix before data transmission at both ends of the communication and realize the information interaction required for decoding [15] through the header design containing the coding information of k and n. The vehicle network scenario is a typical heterogeneous wireless network scenario. In the vehicle network, the heterogeneous characteristics of various wireless networks have caused great differences in impact. DDGS scheme well supports the change in link delay difference caused by the heterogeneity and time-varying characteristics of wireless channels and compensates for the high reception and decoding delay and low bandwidth aggregation efficiency caused by the link delay difference.

Our contributions can be summarized as follows. First, we conducted a mathematical analysis on the formulation of coding groups in network coding. Second, we proposed a DDGS scheme based on mathematical analysis in heterogeneous network scenarios and designed the workflow of the scheme. DDGS scheme could achieve a smaller decoding time cost and higher bandwidth aggregation efficiency compared with the current research. Third, we did a large number of simulations to verify the superior performance of DDGS in the heterogeneous network scenario of vehicular network.

The structure of the paper is organized as follows. We discuss the related work and study the limitation of the existing research status of coding group size and coding rate in Section 2. In Section 3, we introduce the system architecture of DDGS. In Section 4, we analyze and introduce the DDGS scheme in detail. In Section 5, we take lots of tests and simulations to prove that our scheme is effective and reliable. In Section 6, we conclude the paper.

2. Related Work

In order to cope with the development of mobile transportation and the consequent demand for massive data communication [16], scholars focus on the research of the multipath transmission architecture and corresponding transmission schemes of vehicular networks [17]. In this section, we summarize the relevant research on multipath transmission in vehicular networks. This section will expand from two aspects. The first part is the research on the related protocols and architecture of multipath transmission. The second part is about the research progress of improving the packet loss and disorder in multipath transmission. Part of the main research-related direction is network coding.

2.1. Research on Protocol and Architecture of Multipath Transmission

2.1.1. Multipath Transmission Protocol. Research on multipath transmission protocols first focused on the middle layer of the network layer and the transport layer, such as HIP [18] and Shim6 [19]. The idea of these protocols is to introduce a mapping between the network layer and the transport layer to achieve network layer multipath implementation for upper-layer applications. However, due to the transparency of the transport layer, out-of-sequence packets caused by path differences will seriously affect throughput. Therefore, academia turned to the study of transport layer protocols. Multipath transmission support based on SCTP and TCP is the two main research directions.

SCTP was first specified in RFC2960 and redefined in RFC4960. Supporting multihosts is one of the most important features of the standard SCTP protocol, which can be used to establish multiple paths between two multihosts. However, only one path is allowed for data transmission, and the other paths can only be used as backup paths. There are many studies on the realization of multipath parallel transmission through modification. For example, Iyengar et al. [20] proposed the CMT-SCTP protocol, whose main idea is SFR (Split Fast Rete) scheme to manage each path separately, thereby realizing multipath transmission. Some other researches like WiMPSCTP [21] and cmpSCTP [22] also use similar ideas. However, the asymmetry of the path can easily cause disorder, and the resulting unnecessary retransmission and reduction of the congestion window are difficult to solve.

There is also a large amount of research foundation about the expansion of tcp multipath support. A series of TCPbased multipath transmission protocols are proposed, pTCP [23]

divides the transport layer function into central engine and TCP-v pipe, realizing multipath management of data packets through double sequence. mTCP [24] introduces virtual queues, and each end-to-end path uses subflows with independent congestion control mechanisms. Early previous research provided the basis for MPTCP, and MPTCP divides the transmission layer into MPTCP and TCP layers. The MPTCP layer is responsible for path management, packet scheduling, and congestion control. The TCP layer transmits data in the form of subflows and uses the sliding window mechanism for congestion control [25, 26]. To solve the problem that MPTCP throughput will be significantly reduced in the case of path heterogeneity. Schemes like ECF [27], BLEST [28], and BBP [29] are proposed. The main idea of these studies is to study the scheduling strategy of the sender and congestion control. Regardless of the middle layer protocol, SCTP, or TCP-related research, they all need terminal equipment to support the protocol stack, so it is difficult to deploy under existing conditions. Moreover, most of their application environments are static environments, and their performance is poor in the mobile heterogeneous environment of the vehicular networks. In contrast, our method is independent of the protocol stack and transparent to the transport-layer protocols.

2.1.2. Multipath Transmission Architecture. Research on multipath transmission architecture mainly focused on the network layer. There are two main ways to implement the architecture, NAT, and IP-in-IP encapsulation. Rodriguez et al. [30] introduced the MAR system, which acts as a NAT box. MAR can work in two modes: with or without the proxy server at the receiving end. When there is a proxy server, the mar strategy uses a packet-based scheduling scheme. In the opposite case, the scheduling scheme is flow-based.

Dong et al. [31] proposed a multipath network architecture for vehicular networks. It uses tunneling mechanisms through IP-in-IP encapsulation between two proxies. This method performs multipath scheduling of data packets based on the historical status data of wireless heterogeneous network links, thereby achieving bandwidth aggregation. This method has a poor response to packet loss on unreliable wireless links [32]. In this paper, we improved this architecture by introducing network coding to enhance the robustness of the architecture, and at the same time, designing dynamic coding schemes to improve communication performance.

2.2. Research on Network Coding in Multipath Transmission. Due to the mobile nature of vehicular networks, the status of wireless heterogeneous link resources is dynamically changing. The difference of paths will cause the inevitable out-of-sequence problem of multipath transmission. In order to solve this problem, scholars have introduced network coding strategies.

The general process of network coding is that at the sending end, the group size (GS) is appropriately formulated, and the data packets are grouped accordingly. The grouped data packets are redundantly coded through the coding matrix according to the coding strategy requirements. Then, the

encoded packets are sent through multiple links to the receiving end. After the receiver receives the same number of coded packets as the original packets, the receiver recovers the original packets by solving the linear equation through Gaussian elimination according to the extracted coding information. Through the analysis of the above process, we can conclude that GS parameters have an important impact on the efficiency of network coding to a certain extent. At present, the research on GS determination can be divided into three types: (1) fixed group size and coding rate, (2) feedback-based group size, (3) GS determination based on network state. [33, 34] adopt fixed group size and coding rate. These strategies cannot adapt to the dynamic changes of wireless heterogeneous network resources in vehicular networks, and it is difficult to achieve the best communication performance. [35–37] determines the group size based on the end-to-end feedback results. The disadvantage of feedback based is that it will introduce additional delay, which does not meet the low delay requirements of vehicular networks. The decision of the GS based on the real-time network parameters can make the network code better meet the needs of multipath transmission. MPTCP-PNC [13] and CMT-NC [12] fully consider the real-time status of the network, especially the difference in link delay. By using the same batch of packets as a group, packets of a group can arrive at the receiver within the same time round and avoid out-of-order packets between the two groups. These policies treat the same batch of packets as a group, and the packets of a group can arrive at the receiver within the same time round, thus, avoiding out-of-order packets between two groups. However, if only the same batch of data packets is used as a group, under the circumstances caused by the huge performance difference between heterogeneous networks, it will cause an excessively large group size. It greatly increases the computational complexity and decoding time of the receiver. We will perform mathematical analysis in Section 3.2 for further explanation. In our DDGS strategy, we divide the same batch of data packets into fine-grained groups for large delays. In our DDGS strategy, we divide the same batch of data packets into fine-grained groupings for large differences in link delays, thus, avoiding the above problems.

3. The System Architecture of DDGS

In this section, we will introduce the system architecture of DDGS from two aspects: network topology and system module. In order to facilitate the understanding of the formulas and diagrams in the article, we have summarized the symbols appearing in this article in Table 1.

3.1. The Network Topology of Multipath Transmission. The Delay Determined Group Size (DDGS) scheme we proposed is designed for the high bandwidth requirements in the vehicle-to-ground transmission scenario. In order to support multiple communication terminals on the vehicle and reduce the improvement of the protocol stack, we adopt the multipath transmission architecture of the network layer, as shown in Figure 1.

TABLE 1: Summary of symbols.

Symbols	Description
S	Link set between the sender and the receiver
i	Sequence number of a link
l	The link with the largest delay
m	The link with the minimum delay
b	Sequence number of a batch
N_b	Number of packets arriving in the same batch b
BW_i	Bandwidth of link i
RTT_i	Round-trip delay of link i
pe_i	Packet loss rate of link i
T_i	Delay of single packet on link i
q	Sequence number of a group
GSq	Group size of the group q
Mq	Number of redundant packets appended to group q
CMq	Coding matrix for group q
DMq	Decoding matrix for group q

At the sending end, the mobile router completes the aggregation of packets generated by device onboard and realizes grouping, coding, and distributing of packets. At the receiver side, the aggregation router decodes packets from heterogeneous networks and forward them to the server.

We use multiaccess router (MAR) and aggregation router (AR) as the two ends of multipath communication proxy on the vehicle and on the ground, respectively. As the gateway, the multiaccess router is responsible for aggregating the application traffic of users on board. The multiaccess router has multiple network interfaces and can access multiple heterogeneous channels. Traffic is transmitted in parallel to heterogeneous channels and converges at the aggregation router. Our DDGS scheme deploys in MAR and AR. At the MAR, DDGS dynamically groups data packets based on channel state parameters. Each group of data packets is encoded according to the packet loss rate of each channel and sent to the heterogeneous network. At the receiver side, the aggregation router decodes packets from heterogeneous networks and forward them to the server.

3.2. System Module of DDGS. In order to clearly describe the workflow of DDGS, we abstract the communication process as Figure 2. Sending end mainly consists of four modules, network measurement module, group size determination module, network coder, and scheduling module. The network measurement module is responsible for monitoring the status parameters of each link, including RTT, bandwidth, and packet loss rate. The RTT and packet loss rate are obtained by a modified "ping" program. We get these state data by sending probe packets regularly. The bandwidth is passively measured by calculating the payload combined with the probe packet interval. The group size determination module determines the size of the encoding group according to the real-time network status parameters, and the specific

calculation process will be introduced in Section 4. The general grouping process is shown in Figure 2. The DDGS algorithm divides the application layer packets into different batches according to the real-time status of the network, which is represented by packets of different colors in Figure 2. On this basis, the packets are divided into fixed time slices according to the link delay difference. In this way, the data packets in transmission can be divided into boxes as shown in the figure. The network coder dynamically selects the encoding matrix according to the group size and coding rate, and the encoded data packets are forwarded to each link through the scheduling module. At the receiving end, the encoded packets enter the receiving buffer. Decoder designs a decoding matrix to decode by extracting the encoding information of the packet header.

The encoding and decoding information interaction will be introduced in Section 5.

4. The Analysis and Principle of DDGS Scheme

When packets are transmitted through multiple paths, the problem of out-of-sequence of data packets caused by link delay is particularly prominent. Network coding makes the same group of data packets equivalent to decoding, so that the same group of data packets does not have to be in strict order when arriving at the receiving end. However, if the determination of the encoding group is unreasonable, when data packets belonging to different groups arrive, it will cause the buffer of the receiving end to be blocked. In the heterogeneous multilink scenario, there is a delay difference between links. During the link delay difference, the low-latency link will continue to receive data packets before the data packet reaches the receiving end through the high-latency link. We regard the packets sent in the transfer delay difference between the fastest path and the slowest path as the same batch b. [12, 13] use the number of packets in the same group as the code group size which can be expressed as N_b. In our paper, we added consideration of the influence of link bandwidth on N_b, and N_b can be mathematically formulated as.

$$N_b = 1 + \sum_{i \in S} \left\{ \frac{BW_i * (1 - pe_i)}{MTU} * \left[\left(\frac{RTT_l}{2} + \frac{MTU}{BW_l} \right) \cdot - \left(\frac{RTT_i}{2} + \frac{MTU}{BW_i} \right) \right] + 1 \right\}. \tag{1}$$

BW_i, RTT_i, and pe_i are measured in real-time by the network measurement module. $RTT_l/2 + MTU/BW_l$ represents the transmission delay of the packet on the path with the maximum delay. $RTT_i/2 + MTU/BW_i$ represents the transmission delay of the packet on the path i. $BW_i * (1 - pe_i)/MTU$ represents the number of transmitted packets per unit time of path i. Through the correction of the calculation, N_b is obtained. Taking N_b as the group size can prevent the data packets of different groups from arriving out of order to a certain extent. However, there is a big difference in the delay of vehicle heterogeneous links. This situation will lead to a large N_b value. We can calculate $N_b = 56$ through the test data in Table 2.
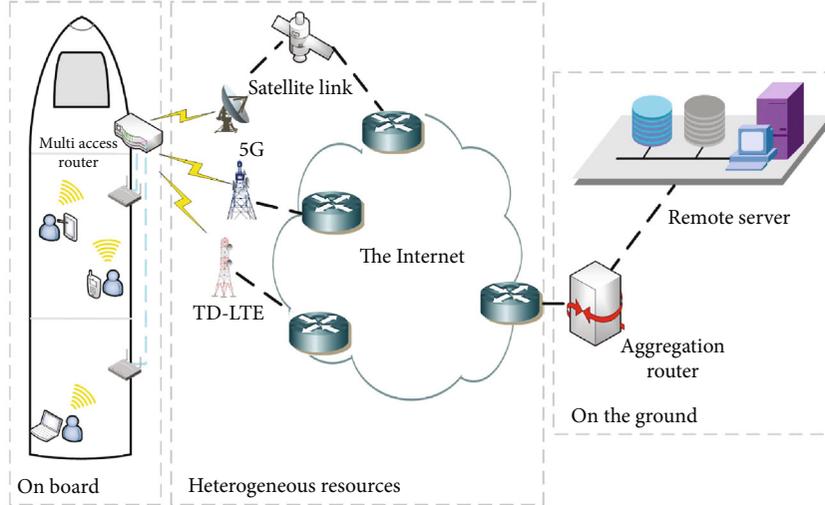
Figure 1: Multipath transmission architecture.
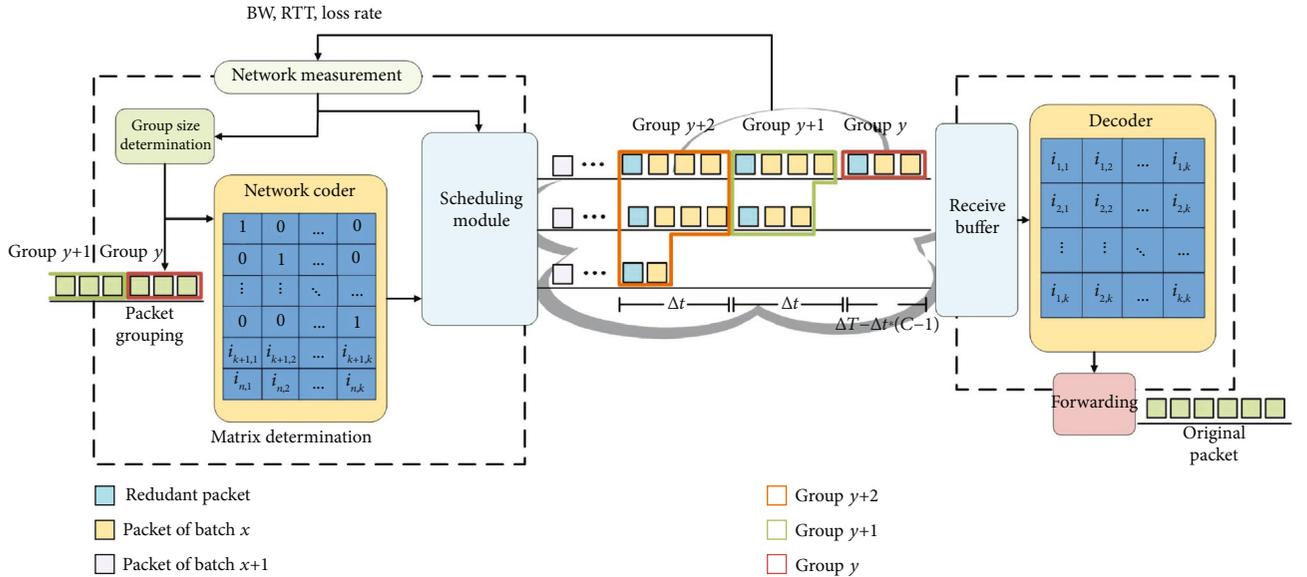
Figure 2: Workflow of DDGS scheme.

Table 2: Link configurations in simulations.

Link type	Delay	Bandwidth	Loss rate
TD-LTE	60 ms	5 Mbps	2%
Satellite link	180 ms	2 Mbps	5%

Using N_b as the number of coding groups makes the data transmission delay of this group equal to the delay of the longest path. What is more, when N_b is too large, it will increase the computational complexity of the encoding and decoding. In order to remedy the above problems, we adopt the DDGS scheme which could dynamically divide batches based on the performance of heterogeneous network resources. At the sender side, the parameter N_b indicates the size of the same Figure 2. Workflow of DDGS scheme batch of packets is sent through the parallel link at the same time. Through our pro-

posed DDGS strategy, we divide the batch of packets into fine-grained ones to avoid the problem caused by too large N_b value. In the next part, we will introduce the specific algorithm of DDGS.

4.1. The Grouping Principle of DDGS.

Combined with the real-time network state parameters measured by the network measurement module, we can calculate that the time consumption of one packet from sender to receiver on link i can be depicted as

$$T_i = \left(\frac{RTT_i}{2} + \frac{MTU}{BW_i} \right). \tag{2}$$

According to formula (2), we can get that the difference between the transmission delay of the data packet on link i

and the transmission delay on the path with the maximum delay is

$$\Delta T_i = \left(\frac{RTT_l}{2} + \frac{MTU}{BW_l} \right) - \left(\frac{RTT_i}{2} + \frac{MTU}{BW_i} \right). \qquad (3)$$

According to Section 4 part A, the packets transmitted within ΔT_i on the link i can arrive at the receiving end in the same batch as the packets transmitted simultaneously on the path with the highest delay. In DDGS, we choose Δt to further divide ΔT_i. In the actual deployment, we make Δt equal to the interval time of delay ACK feature of TCP (usually 40 ms). The advantage of this choice is that it can make full use of the end-to-end TCP characteristics, so that the decoded packets of the same group can be acknowledged in one ack. By calculating the transmission delay difference between the maximum delay path and the minimum delay path, we can get the number of groups C_b that this batch of packets can be divided into.

$$C_i = \left\lceil \frac{((RTT_l/2) + (MTU/BW_l)) - ((RTT_i/2) + (MTU/BW_i))}{\Delta t} \right\rceil. \qquad (4)$$

In formula (5), m refers to the link with minimum delay, l refers to the link with largest delay, and we use q to represent the group sequence, $q \in 1, 2, \cdots, C_b$. For any link i, the distribution of data packets in the same batch can be expressed as a sequence number starting from C_i to ending at C_b.

$$C_i = \left\lceil \frac{((RTT_l/2) + (MTU/BW_l)) - ((RTT_i/2) + (MTU/BW_i))}{\Delta t} \right\rceil. \qquad (5)$$

According to the group distribution analysis for link i, we can express the number of packets transmitted on link i with group sequence q as n_i^q.

$$n_i^q = \begin{cases} \dfrac{[\Delta Ti - \Delta t * (C_i - 1)] * BW_i}{MTU} & q = C_i, \\ \Delta T * \dfrac{BW_i}{MTU} & q < C_i. \end{cases} \qquad (6)$$

Then the group size of qth group transmitted in all paths can be expressed as

$$GSq = \sum_{i \in S} n_i^q. \qquad (7)$$

Next, we introduce the packet encoding process based on group.

4.2. Dynamic Coding and Encoding Based on GS. Due to the mobile characteristics and heterogeneity of vehicular networks, the link will cause serious packet loss and out-of-order data packet transmission. We alleviate the two major problems by encoding ordinary data packets to generate redundant data packets. According to the packet loss rate of each link, we can conclude that the number of additional redundant data packets required to ensure the completion of the decoding on each link is

$$r_i^q = \left\lceil \frac{pe_i}{1 - pe_i} * n_i^q \right\rceil. \qquad (8)$$

Adding the redundant data packets on each link to obtain the number of redundant data packets required for the code group q is

$$M_q = \sum_{i \in S} r_i^q. \qquad (9)$$

Based on the above analysis, we can calculate the coding ratio $r = GS_q/(GS_q + M_q)$ of the coding group q. We design a coding matrix that can be dynamically expanded so that the network coder module dynamically determines the coding matrix according to different values of r. The scalable coding matrix is designed as follows.

$$CM_q = \begin{bmatrix} I \\ RM \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 1 & 2 & \cdots & GS_q \\ GS_q & 1 & \cdots & GS_{q-1} \\ \vdots & \vdots & \ddots & \vdots \\ 2 & 3 & \cdots & 1 \end{bmatrix}. \qquad (10)$$

The coding matrix CM_q for group q has $GS_q + M_q$ rows and GS_q columns. The first GS_q rows of CM_q is a GS_q order identity matrix I, and the last M_q rows is a special form of the Toplitz matrix RM. [34] has proved that any GS_q rows of the constructed matrix are linearly independent. Encoding data packets by matrix CM_q ensures that the receiver can decode the original data packets after receiving any GS_q encoded packets. The construction prototype of coding matrix with different coding ratio r is negotiated by the sender and the receiver before communication. At the receiving end, after receiving any GS_q encoded packets, decoder module can trigger the decoding process. The decoding matrix DM_q for group q can be constructed by inverting the linear combination matrix of the received encoded packets.

Through the analysis of the above process, for realizing the interaction of decoding information, the encoded data packet only needs to carry two pieces of information, encoding ratio r and line number N which the encoding packet is linear combined through, then, the receiving end can realize the construction of the decoding matrix. We add these fields in the encoded packet header, as shown in Figure 3. The decoder module extracts the header information at the

receiving end and strips the header to decode. This design can effectively increase the payload of the data packet.

4.3. Packet Scheduling. According to formula (7), we can get the number of packets of group q transmitted on link i. Therefore, in the process of dispatching packets at the sending end, we can calculate the coding group of the batch of packets sent on each link in advance, so as to obtain the sequence of packets on the transmission path. The scheduling process is shown in Figure 2, in order to achieve the effect of the same batch of packets arriving at the receiving end at the same time, the data packets with larger coding group number are sent on the high delay link, and the coding groups are sent sequentially on the low delay link. The scheduling process is carried out according to the group number formulated by the DDGS scheme, which improves the receiving efficiency of the receiver. After completing a batch of data packet scheduling, in order to ensure real-time feedback of the link status, the scheduling module will trigger the update of GS_q and M_q, as well as coding matrix.

5. The Analysis and Principle of DDGS Scheme

In this section, the performance of DDGS is illustrated by simulation. We choose two main research methods to determine the coding group as a comparison. One is based on the single transmission process to determine the coding group size, abbreviated as BOST, and the other is the fixed coding rate, abbreviated as FIXED. For the research point of DDGS, we evaluate the impact of different network parameters on encoding and decoding delay and average throughput. We deploy the above three schemes on Network Simulation 3 (NS3) version 3.29 and dynamically adjust the link parameters for simulation.

5.1. Network Topology for Simulation. The simulation topology is shown in Figure 4. We use Node2 as MAR and Node3 as AR. We create three links between Node2 and Node3 to simulate a heterogeneous network scenario. The three links are used to simulate heterogeneous links with different delay, bandwidth, and packet loss rate in subsequent simulation. For different simulation objectives, we choose the link parameters which have the greatest impact on the proposed scheme as variables to verify the performance of the strategy. We create a virtual device tun for DDGS deployment. We design from a client-server program from Node1 to Node2 to perform one-way communication simulation. At Node2, data packets are grouped, encrypted, and forwarded to three links. At Node3, the received data packets are cached, decrypted, and forwarded to node4. The pseudocode on the sender and receiver is shown below.

5.2. Time Consumption of Decoding. In this section, we analyze the decoding time cost of the DDGS scheme. The design purpose of DDGS is to select a reasonable coding group size for link delay the difference in heterogeneous networks. Therefore, for the analysis of decoding time cost, we take link delay as an independent variable, and the delay settings of each link are shown in Table 3. By dynamically adjusting
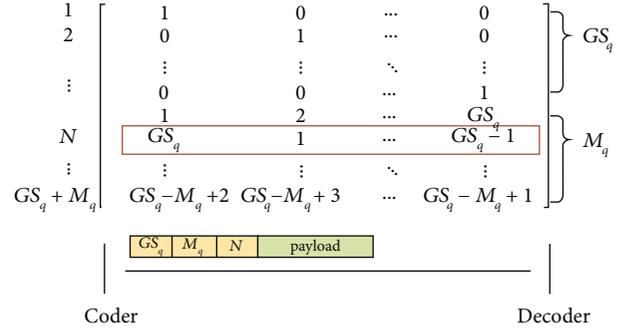
Figure 3: Decoding information interaction process.

the delay of different links, the performance of DDGS under different link delay combinations is displayed in Figure 5. We fix the delay of link A to 20 ms and adjust the delay of link B and C to achieve different link delay combinations. The abscissa represents the link delay of link C. The ordinates are the time required for the receiver to complete the decoding of the whole group of packets from the first packet of this group arriving in.

Through the analysis of Figure 5, we can see that when the three links have no delay difference or the delay difference is small, and the decoding time cost of the three schemes is similar. With the increase of delay difference, the time cost of the BOST scheme increases linearly due to the increase of the coding group, while the time cost of FIXED and DDGS schemes increases less. With the increase of the delay difference, the BOST scheme will lead to the excessive growth of the coding group, thus, increasing the waiting delay and decoding complexity of the receiver, resulting in the linear growth of the overhead. Although the DDGS scheme will adjust the coding group due to the increase of delay difference, the scheme ensures that the coding group will not grow excessively. The time cost is stable to a certain extent. Keeping the decoding cost time small is beneficial to reduce the end-to-end communication delay, which makes TCP communication avoid performance degradation caused by delay fluctuation. Although the FIXED scheme is slightly better than DDGS in decoding time cost, through subsequent analysis, we can see that DDGS can achieve higher throughput. According to the above analysis of simulation results, we can conclude that DDGS is suitable for delay fluctuation scenarios in heterogeneous networks.

5.3. Bandwidth Aggregation Efficiency. In this section, we will analyze the bandwidth aggregation performance of the DDGS scheme. First, we analyze the bandwidth aggregation performance of DDGS from the perspective of heterogeneous network delay. We set up two simulation environments, as shown in Table 4. In the first simulation, we set the three link delays to 10-30 ms fluctuations. This parameter setting represents the scenario with little difference in link delay for comparison. In the second simulation, the delay difference of the three links increases. This parameter setting indicates the heterogeneous network scenario with gradient delay difference, which is also the main scenario for DDGS. The simulation results are shown in Figure 6. The horizontal axis
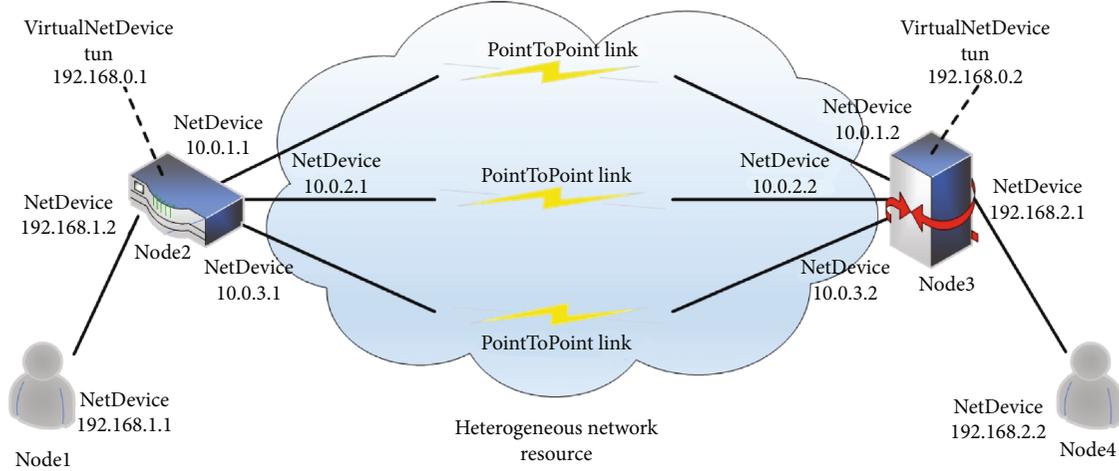
FIGURE 4: Simulation topology.

```
MAR sender.
1:   for RecvPacket() do
2:       if batch.pendingqueue=NULL then
3:           batch.pendingqueue=DetermineGS()
4:       end if
5:       Codingbuffer.q.queu=Grouping()
6:       if sizefo(Codingbuffer.q.queu)=Codingbuffer.q.k then
7:           ConstructCodingMartix()
8:           Encoding()
9:           Scheduling()
10:      end if
11:  end for
```

ALGORITHM 1

```
AR receiver.
1:   for RecvPacket() do
2:       HeaderAnalysis()
3:       recvbuffer.q.queue=Add2RecvBuffer()
4:       if sizefo(recvbuffer.q.queue)=recvbuffer.q.k then
5:           ConstructDecodingMartix()
6:           Decoding()
7:           Forwarding()
8:       end if
9:   end for
```

ALGORITHM 2

TABLE 3: Link configurations for decoding time cost.

Parameters	Link A	Link B	Link C
Delay	20 ms	20, 50, 80 ms	20, 50, 80, 110, 140 ms
Loss rate	0.05	0.05	0.05
Bandwidth	1 Mbps	1 Mbps	1 Mbps

represents simulation time, and the vertical axis represents real-time bandwidth statistics. It can be concluded that when the delay difference is small, the bandwidth aggregation performance of the three schemes is similar. When the link delay difference expands, FIXED and BOST have experienced significant performance degradation. In comparison, DDGS maintains better bandwidth aggregation performance.

Next, we will study the impact of link bandwidth on DDGS schemes. We set the delays of the three links to 20 ms, 80 ms, and 140 ms, respectively, to simulate the heterogeneous network scenario with large difference in link delays. We fixed the bandwidth of two links at 1 Mbps and adjusted the bandwidth of the other link from 0.1 Mbps to 2 Mbps to observe the impact of bandwidth changes of links with different delays on the performance of DDGS.

The simulation results are shown in Figure 7. From Figures 7(a) and 7(b), we can see that the aggregate bandwidth increases linearly with the increase of the bandwidth of the two links with a small delay. DDGS and BOST have a better convergence effect than the FIXED scheme (the slope of simulation results is larger than FIXED scheme), which is consistent with our assumption, because the FIXED scheme does not have the ability to adjust the size of the coding group according to the link delay difference, which will lead to the code group not suitable for the link state, resulting in excessive redundancy or coding groups out of order. The growth rate of Figure 7(a) is slightly larger than that of Figure 7(b). It can be seen that low delay link and high bandwidth link are more conducive to the overall performance improvement.

From Figure 7, we can see that the overall performance of DDGS is better than the other two schemes. The bandwidth aggregation performance of DDGS and BOST is relatively close. The difference between DDGS and BOST lies in the segmentation of the same batch of packets. The difference in bandwidth aggregation between DDGS and BOST is due to the aggregate bandwidth reduction caused by the receiving delay experienced by the receiver. Therefore, the bandwidth aggregation performance of the DDGS scheme is slightly better than that of the BOST scheme, but according to the

(a) Delay of link B is 20 ms

(b) Delay of link B is 50 ms

(c) Delay of link B is 80 ms
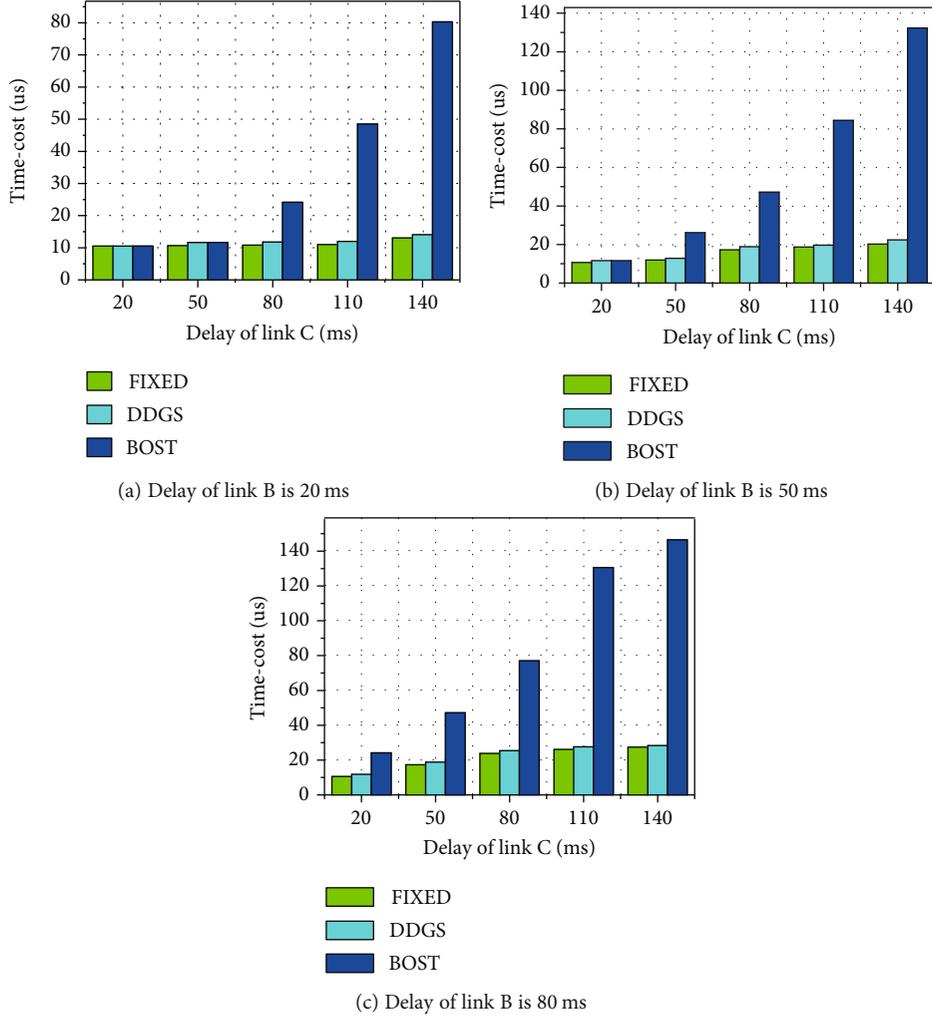
FIGURE 5: The time-cost of decoding processes at receiver for different schemes.

TABLE 4: Link configurations for bandwidth aggregation.

Parameters	Link A	Link B	Link C
Loss rate	0˜0.03	0˜0.03	0˜0.03
Bandwidth	0.5˜1.5 Mbps	0.5˜1.5 Mbps	0.5˜1.5 Mbps
Delay of simulation 1	10˜30 ms	10˜30 ms	10˜30 ms
Delay of simulation 2	10˜30 ms	70˜90 ms	130˜150 ms

analysis in Figure 5, DDGS is much better than BOST in receiving delay.

It is worth noting that in Figure 7(c), DDGS and BOST have a high-efficiency bandwidth aggregation effect when the bandwidth of the high-latency link c is small. When bandwidth of the link c increases to a certain value, the growth rate of the aggregate bandwidth decreases significantly. The reason for this is that the bandwidth of the low-latency link is insufficient to support the timely transmission of data packets on the low-latency link, resulting in queuing. The results show that the DDGS scheme has a better aggregation effect on low-latency, high-bandwidth links and high-latency, low-bandwidth links.

Next, we will study the impact of link loss rate on DDGS schemes. We maintain the link delay in the above simulation and set the bandwidth of the three links to 1 Mbps. We adjusted the packet loss rate of the three links from 0% to 20%.

From Figure 8, we can see that when the packet loss rate is less than 4%, the three schemes can achieve effective bandwidth aggregation. When the packet loss rate is greater than 4%, the bandwidth aggregation efficiency will decrease significantly. But DDGS can still guarantee a better aggregation effect than the other two schemes. When the link packet loss rate is less than 4%, the FIXED scheme uses a smaller group size, and redundant packets encoded to compensate for packet loss will greatly occupy effective bandwidth, resulting in a waste of bandwidth resources. When the packet loss rate increases, the performance of the BOST decreases more significantly than the other two schemes. This is because when the link delay difference is large, BOST will cause a large group size. In order to compensate for the link loss, each group of data needs to wait for more encoding redundant packets to arrive, resulting in a drop in throughput.

(a) Approximate delay of three links

(b) The delay of the three links is quite different
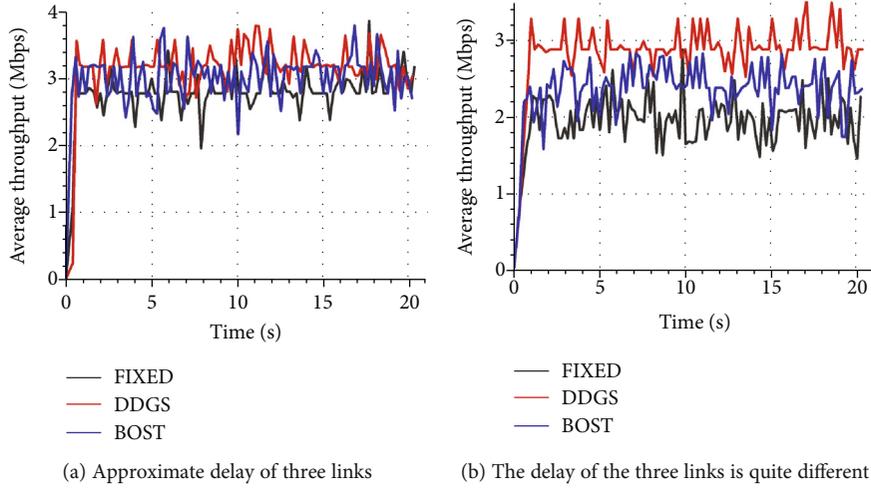
FIGURE 6: Bandwidth aggregation under different delay conditions.

(a) Link A as variable

(b) Link B as variable

(c) Link C as variable
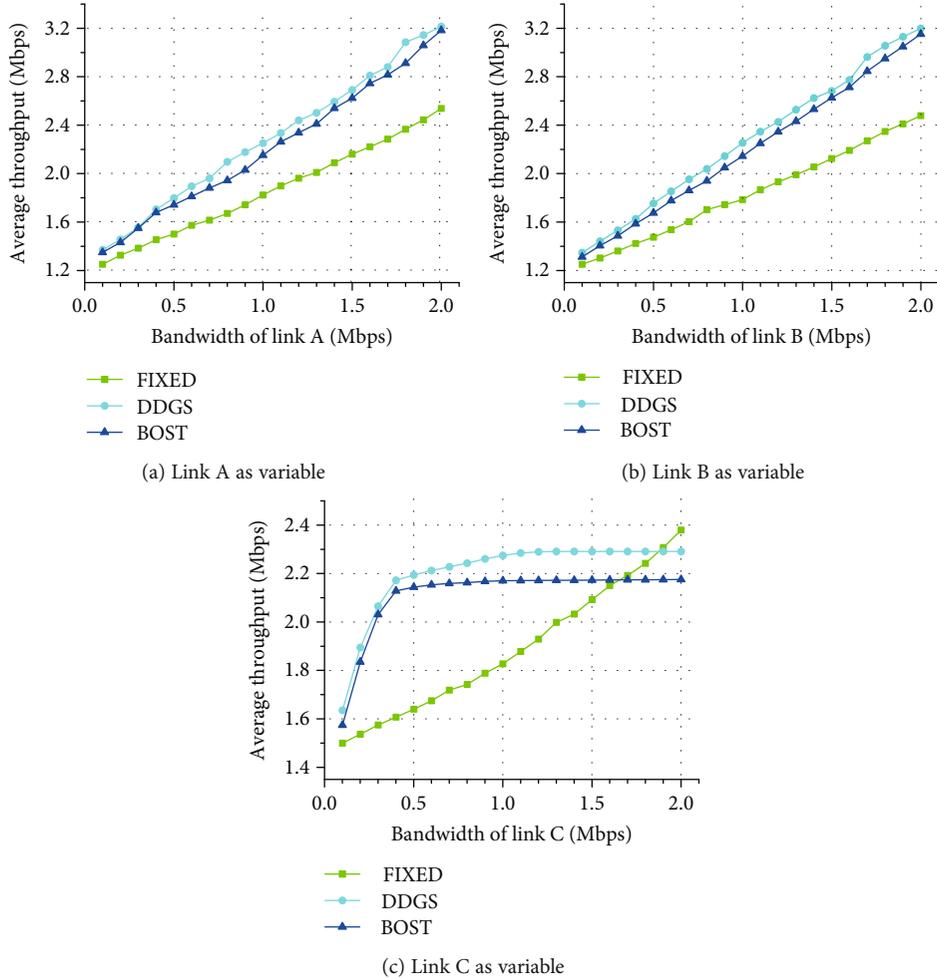
FIGURE 7: Bandwidth aggregation efficiency with link bandwidth fluctuation.

The following conclusions can be drawn from the above simulation results. (1) The DDGS scheme can guarantee a shorter packet decoding time than the BOST scheme, thereby ensuring data continuity and TCP friendliness. (2) The DDGS solution also has good bandwidth aggregation performance when the link delay difference is large. What is more, DDGS has a better aggregation effect for heterogeneous networks with low latency, high bandwidth links, high latency, and low bandwidth links in the vehicular network.
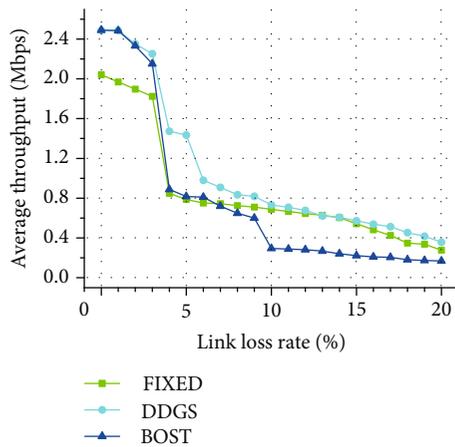
FIGURE 8: Bandwidth aggregation efficiency with loss rate fluctuation.

6. Conclusion

In this paper, aiming at solving the low efficiency of multi-path bandwidth aggregation caused by the heterogeneous characteristics of network resources in the vehicular network, we propose a DDGS scheme based on network coding. The scheme we proposed could adaptively adjust the coding group according to the heterogeneous wireless networks state, so as to make full use of the advantages of network coding. We discussed in detail the mathematical analysis and process design of the DDGS scheme. Finally, we did a lot of simulations, comprehensively considering a variety of network parameters to analyze DDGS. The results show that the DDGS scheme is significantly superior to other coding group determination schemes. DDGS is very suitable for heterogeneous networks with low latency, high bandwidth links, high latency, and low bandwidth links in vehicle networks.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] P. Dong, T. Zheng, X. Du, H. Zhang, and M. Guizani, "SVCC-HSR: providing secure vehicular cloud computing for intelligent high-speed rail," *IEEE Network*, vol. 32, no. 3, pp. 64–71, 2018.

[2] G. Liang, H. Yu, X. Guo, and Y. Qin, "Joint access selection and bandwidth allocation algorithm supporting user requirements and preferences in heterogeneous wireless networks," *IEEE Access*, vol. 7, pp. 23914–23929, 2019.

[3] J. Wu, B. Cheng, and M. Wang, "Improving multipath video transmission with raptor codes in heterogeneous wireless networks," *IEEE Transactions on Multimedia*, vol. 20, no. 2, pp. 457–472, 2018.

[4] X. Du, "QoS routing based on multi-class nodes for mobile ad hoc networks," *Ad Hoc Networks*, vol. 2, no. 3, pp. 241–254, 2004.

[5] L. Li, K. Xu, T. Li et al., "A measurement study on multi-path tcp with multiple cellular carriers on high speed rails," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, pp. 161–175, Budapest, Hungary, August 2018.

[6] Y. Zhang, P. Dong, S. Yu, H. Luo, T. Zheng, and H. Zhang, "An adaptive multipath algorithm to overcome the unpredictability of heterogeneous wireless networks for high-speed railway," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 11332–11344, 2018.

[7] R. Ahlswede, Ning Cai, S. Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[8] Z. Ding, K. K. Leung, D. L. Goeckel, and D. Towsley, "On the study of network coding with diversity," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1247–1259, 2009.

[9] Z. Guan, X. Liu, L. Wu et al., "Cross-lingual multi-keyword rank search with semantic extension over encrypted data," *Information Sciences*, vol. 514, pp. 523–540, 2020.

[10] L. Zhu, Y. Li, F. R. Yu, B. Ning, T. Tang, and X. Wang, "Cross-layer defense methods for jamming-resistant cbtc systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, pp. 1–13, 2020.

[11] J. K. Sundararajan, D. Shah, M. Médard, S. Jakubczak, M. Mitzenmacher, and J. Barros, "Network coding meets tcp: theory and implementation," *Proceedings of the IEEE*, vol. 99, no. 3, pp. 490–512, 2011.

[12] C. Xu, Z. Li, L. Zhong, H. Zhang, and G. M. Muntean, "Cmt-nc: improving the concurrent multipath transfer performance using network coding in wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1735–1751, 2016.

[13] C. Xu, P. Wang, C. Xiong, X. Wei, and G. M. Muntean, "Pipeline network coding-based multipath data transfer in heterogeneous wireless networks," *IEEE Transactions on Broadcasting*, vol. 63, no. 2, pp. 376–390, 2017.

[14] L. Zhu, Y. He, F. R. Yu, B. Ning, T. Tang, and N. Zhao, "Communication-based train control system performance optimization using deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10705–10717, 2017.

[15] L. Xue, Y. Yu, Y. Li, M. H. Au, X. du, and B. Yang, "Efficient attribute-based encryption with attribute revocation for assured data deletion," *Information Sciences*, vol. 479, pp. 640–650, 2019.

[16] L. Zhu, F. R. Yu, B. Ning, and T. Tang, "Cross-layer handoff design in mimo-enabled wlans for communication-based train control (cbtc) systems," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 719–728, 2012.

[17] M. Li, A. Lukyanenko, Z. Ou et al., "Multipath transmission for the internet: a survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2887–2925, 2016.

[18] R. Moskowitz, P. Nikander, P. Jokela et al., *Host Identity Protocol (hip) Architecture*, RFC 4423, Tech. Rep, 2006.

[19] M. Bagnulo and E. Nordmark, *Shim6: Level 3 Multihoming Shim Protocol for ipv6*, RFC 5533, Tech. Rep., 2009.

[20] J. R. Iyengar, P. D. Amer, and R. Stewart, "Concurrent multipath transfer using sctp multihoming over independent end-to-end paths," *IEEE/ACM Transactions on Networking*, vol. 14, no. 5, pp. 951–964, 2006.

[21] C.-M. Huang and C.-H. Tsai, "Wimp-sctp: multi-path transmission using stream control transmission protocol (sctp) in wireless networks," in *21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07)*, vol. 1, pp. 209–214, Niagara Falls, ON, Canada, 2007.

[22] J. Liao, J. Wang, and X. Zhu, "cmpsctp: an extension of sctp to support concurrent multi-path transfer," in *2008 IEEE International Conference on Communications*, pp. 5762–5766, Beijing, China, 2008.

[23] H.-Y. Hsieh and R. Sivakumar, "A transport layer approach for achieving aggregate bandwidths on multi-homed mobile hosts," *Wireless Networks*, vol. 11, no. 1-2, pp. 99–114, 2005.

[24] M. Zhang, J. Lai, A. Krishnamurthy, L. L. Peterson, and R. Y. Wang, "A transport layer approach for improving end-to-end performance and robustness using redundant paths," in *USENIX Annual Technical Conference, General Track*, pp. 99–112, Boston, MA, USA, 2004.

[25] A. Ford, C. Raiciu, M. Handley, S. Barre, and J. Iyengar, "Architectural guidelines for multipath tcp development," *IETF, Informational RFC*, vol. 6182, article 2070–1721, 2011.

[26] S. Barre, C. Paasch, and O. Bonaventure, "Multipath tcp: from theory to practice," in *International Conference on Research in Networking*, pp. 444–457, Springer, 2011.

[27] Y.-s. Lim, E. M. Nahum, D. Towsley, and R. J. Gibbens, "Ecf: an mptcp path scheduler to manage heterogeneous paths," in *Proceedings of the 13th International Conference on emerging Networking EXperiments and Technologies*, pp. 147–159, Incheon, Republic of Korea, 2017.

[28] S. Ferlin, O. Alay, O. Mehani, and R. Boreli, "Blest: blocking estimation-based mptcp scheduler for heterogeneous networks," in *2016 IFIP Networking Conference (IFIP Networking) and Workshops*, pp. 431–439, Vienna, Austria, May 2016.

[29] B.-H. Oh and J. Lee, "Feedback-based path failure detection and buffer blocking protection for mptcp," *IEEE/ACM Transactions on Networking*, vol. 24, no. 6, pp. 3450–3461, 2016.

[30] P. Rodriguez, R. Chakravorty, J. Chesterfield, I. Pratt, and S. Banerjee, "Mar: a commuter router infrastructure for the mobile internet," in *Proceedings of the 2nd international conference on Mobile systems, applications, and services - MobiSYS '04*, pp. 217–230, Boston, MA, USA, 2004.

[31] P. Dong, B. Song, H. Zhang, and X. du, "Improving onboard internet services for high-speed vehicles by multipath transmission in heterogeneous wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 9493–9507, 2016.

[32] X. Du and F. Lin, "Improving sensor network performance by deploying mobile sensors," in *PCCC 2005. 24th IEEE International Performance, Computing, and Communications Conference, 2005*, pp. 67–71, Phoenix, AZ, USA, April 2005.

[33] F. Gabriel, J. Acevedo, and F. H. P. Fitzek, "Network coding on wireless multipath for tactile internet with latency and resilience requirements," in *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Abu Dhabi, United Arab Emirates, December 2018.

[34] Y. Zhang, P. Dong, X. du, H. Luo, T. Zheng, and M. Guizani, "BNNC: improving performance of multipath transmission in heterogeneous vehicular networks," *IEEE Access*, vol. 7, pp. 158113–158125, 2019.

[35] Y. Cui, L. Wang, X. Wang, H. Wang, and Y. Wang, "Fmtcp: a fountain code-based multipath transmission control protocol," *IEEE/ACM Transactions on Networking*, vol. 23, no. 2, pp. 465–478, 2014.

[36] N. Arianpoo, I. Aydin, and V. C. M. Leung, "Network coding: a remedy for receiver buffer blocking in the concurrent multipath transfer of data over multi-hop wireless networks," in *2014 IEEE International Conference on Communications (ICC)*, pp. 3258–3263, Sydney, NSW, Australia, June 2014.

[37] J. Cloud and M. Medard, "Multi-path low delay network codes," in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–7, Washington, DC, USA, December 2016.

Research Article

Vehicle-to-Vehicle Channel Characterization Based on Ray-Tracing for Urban Road Scenarios

Lei Xiong [iD], Zhiyi Yao [iD], Haiyang Miao [iD], and Bo Ai [iD]

The State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, 100044 Beijing, China

Correspondence should be addressed to Lei Xiong; lxiong@bjtu.edu.cn

In this paper, the vehicle-to-vehicle (V2V) channel characteristics in peak hours at the 5.9 GHz band in two typical urban road scenarios, the urban straight road and the intersection, are investigated. The channel characteristics, such as path loss, root mean square (RMS) delay spread, and angular spread, are derived from the ray-tracing (RT) simulations. Due to the low height of antennas at both the transmitter (Tx) and the receiver (Rx), the line of sight (LOS) between the Tx and the Rx will often be obstructed by other vehicles. Based on the RT simulation results, the shadowing loss is modelled by the multimodal Gaussian distribution, and path loss models in both LOS and non-LOS (NLOS) conditions are obtained. And the RMS delay spread in two scenarios can be modelled by the Weibull distribution. In addition, the deployment of an antenna array is discussed based on the statistics distribution of the angular spread.

1. Introduction

As one of the important parts of an intelligent transportation system (ITS), the V2V communication system allows vehicles to exchange information about the surrounding traffic situation to improve safety, reduce traffic congestion, and provide a comfortable driving experience [1]. It is estimated by the U.S. Department of Transportation (DOT) that V2V communication can address up to 82% of all crashes, saving thousands of lives and billions of dollars. In recent years, V2V communication has attracted more and more attention and has achieved remarkable development.

The channel characteristics are of vital importance in the research, design, and deployment of the V2V communication system. Due to the low antenna height at both the Tx and the Rx, rapid time-varying environments, high mobility of vehicles, relatively short communication distance, diversity of scattering objects, etc., the V2V channel characteristics

are significantly different from those in the cellular networks [2].

Many measurement campaigns have been conducted to investigate the V2V channel characteristics over the past few years, and a number of V2V channel models have been proposed. Most campaigns are carried out between the vehicles on the straight road, either in the same direction [3–6] or in opposite [7, 8], and at intersections [9–11]. RT simulation technology is also used in [12], where the V2V channel characteristics of an urban intersection environment are analyzed, especially the power delay profile (PDP) and channel gain.

Many studies are carried out in LOS conditions, but some in NLOS conditions caused by large-size vehicles, such as buses, trucks, and vans. In [13], the shadowing effect caused by obstructing vehicles in the V2V channel has been addressed. In [14–16], based on the measurement in highways and urban environments, the shadow fading models

<div align="center">(a)</div>

<div align="center">(b)</div>
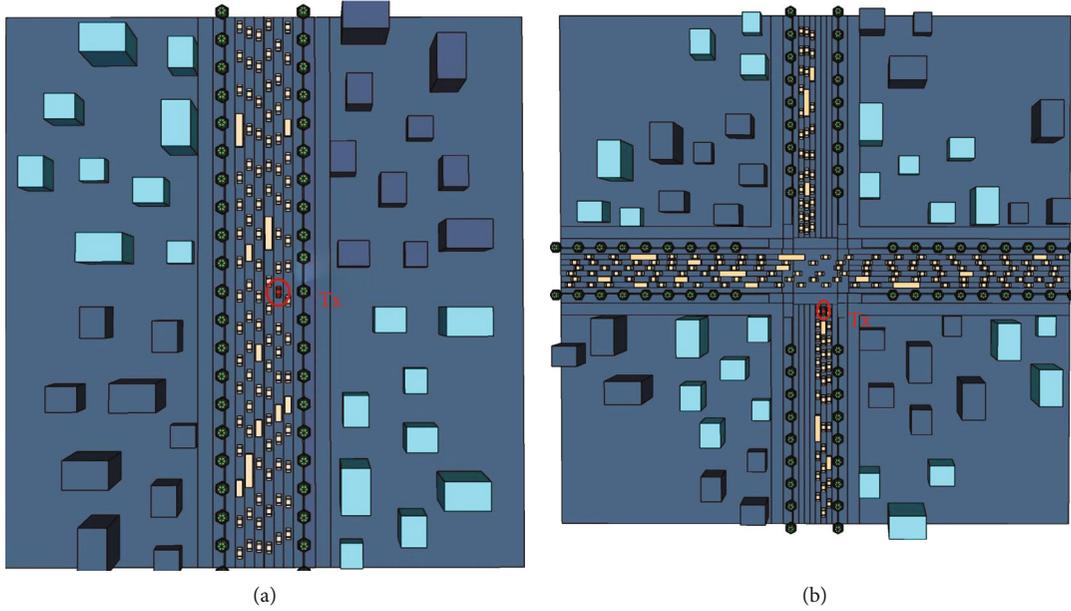
FIGURE 1: Straight road scenario and intersection scenario: (a) straight road; (b) intersection.

in LOS and NLOS conditions are established, and it is emphasized that the shadowing loss caused by the surrounding vehicles is about 10 dB. In [17], the V2V communication link is divided into three categories: LOS, LOS blocked by vehicles, and LOS blocked by stationary objects, in which the obstructing vehicles have a more significant impact on the V2V channel. In [14], according to the causes of LOS blockage (vehicles or buildings), the NLOS regions are divided into obstructed LOS (OLOS) and NLOS with log-normal distribution, respectively.

V2V channel characteristics are related to the environment (i.e., urban, suburban, highway, etc.), vehicle speed, and traffic density. During peak hours, urban traffic densities are usually higher and vehicle speeds are usually lower. Due to the mobility of Tx/Rx, the LOS will be blocked and appear in NLOS conditions, which means it is more urgent to investigate the channel characteristics in peak hours. Therefore, in this paper, the V2V channel characteristics of an urban straight road and intersection, e.g., path loss, RMS delay spread, and angular spread, in peak hours at the 5.9 GHz band are investigated based on RT. Channel characteristics in both LOS and NLOS conditions, the probability distribution of LOS in each scenario, and the distribution of shadowing loss and RMS delay spread are modelled. Finally, the RMS angular spread and antenna array deployment are discussed.

The rest of this paper is organized as follows. Section 2 presents the scenario models in urban roads and introduces the RT simulation technology. Section 3 investigates the channel characteristics in both LOS and NLOS conditions. It also includes the probability distribution of LOS in each scenario and the derivation of path loss and modeling of shadow fading in LOS and NLOS conditions. The deployment of an antenna array is also discussed. Conclusions are drawn in Section 4.

2. Urban Road Scenario and Ray-Tracing Simulation

2.1. Urban Road Modeling. The two typical V2V scenarios, urban straight road and intersection, are modelled in the paper. Multiple samples are established in these two scenarios, and Figure 1 takes one sample as an example in each scenario.

2.1.1. Straight Road. Six motor vehicle lanes and two nonmotor vehicle lanes are included. During peak hours, the speed of vehicles is relatively low (25 km/h), and the distance between adjacent vehicles is uniformly distributed between 4 and 10 m, where the distance between adjacent vehicles is defined as the distance from the rear of the front vehicle to the head of the adjacent rear vehicle on the same motor vehicle lane.

2.1.2. Intersection. The total length of the queues for the vehicles waiting for traffic lights from south to north is 50 m, and the distance between adjacent vehicles is uniformly distributed between 1 and 3 m. While in the same direction, for vehicles that are 50 to 100 m from the zebra crossing and driving into the queues, the distance between adjacent vehicles is uniformly distributed between 4 and 10 m. And the speed of the vehicles in the east-west direction is 25 km/h, and the distance between adjacent vehicles is uniformly distributed between 4 and 10 m, too.

2.2. Object Modeling. The surrounding objects in the V2V scenarios that will affect the radio propagation are considered, such as vehicles, buildings, and trees.

2.2.1. Vehicle Modeling. The vehicles in urban roads are divided into three types in terms of size, small-sized vehicles,
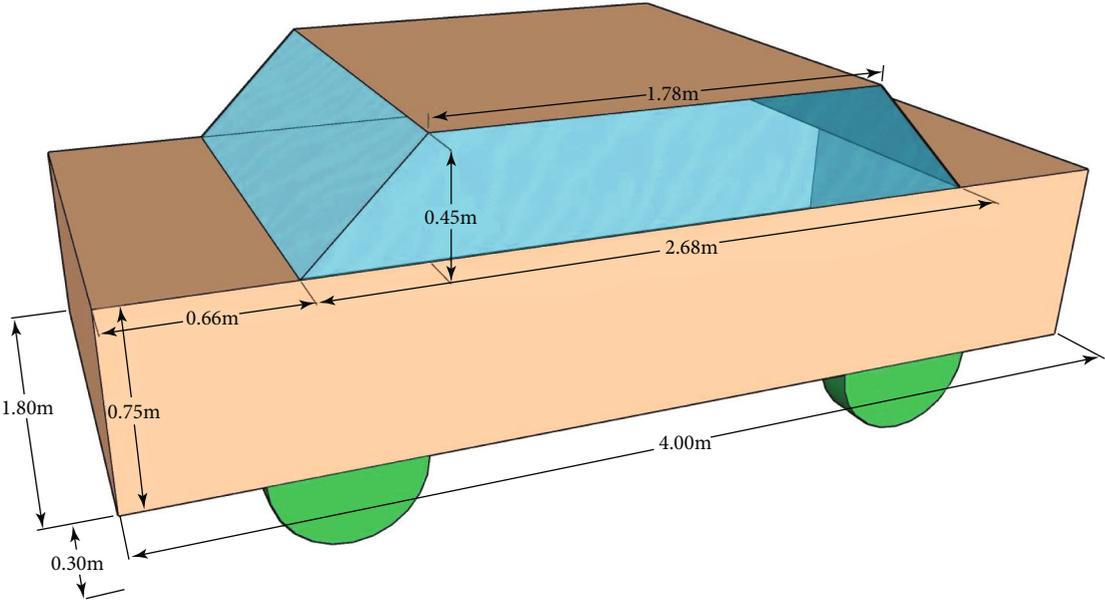
(a)
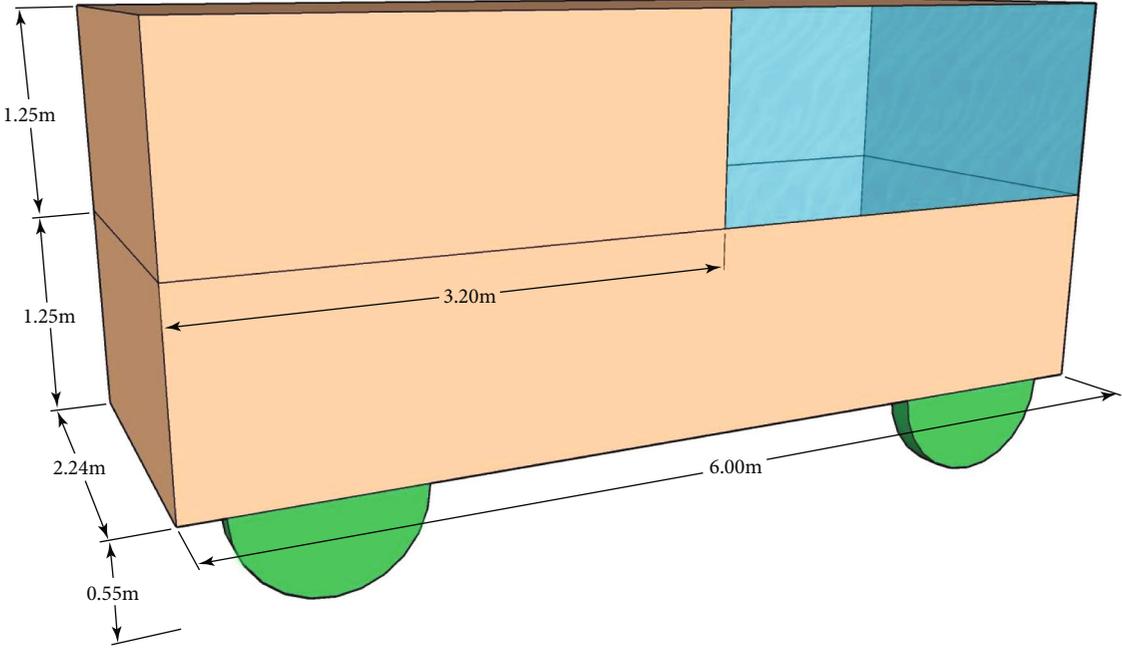
(b)

Figure 2: Continued.

(c)
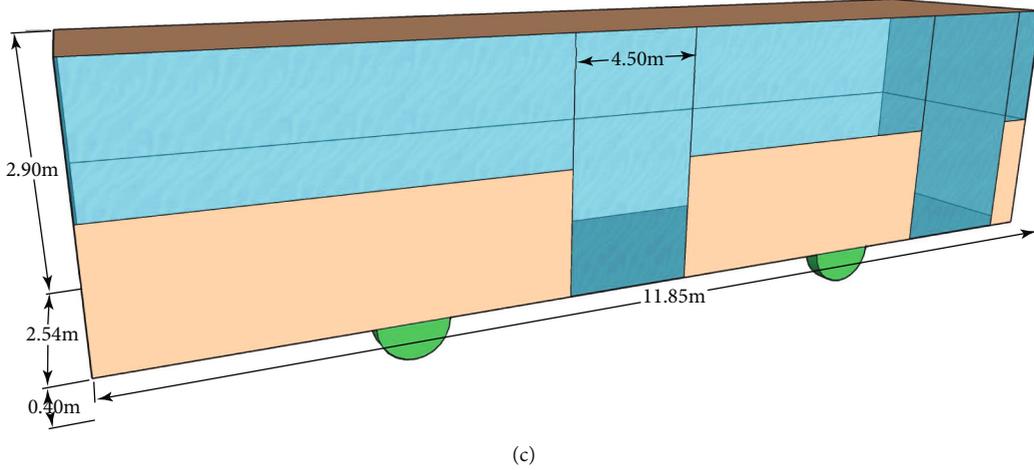
FIGURE 2: 3D model of vehicles: (a) car; (b) van; (c) bus.

medium-sized vehicles, and large vehicles, with the proportions of 89%, 7%, and 4%, respectively [18].

As shown in Figure 2, small-sized vehicles are modelled according to a 3-compartment car, medium-sized vehicles are modelled according to a van, and large-sized vehicles are modelled according to a bus.

2.2.2. Building Modeling. Urban buildings are modelled as low-rise, multistorey, and medium high-rise buildings, with the height of 7 m, 14 m and 23 m and the proportions of 20%, 50% and 30%, respectively [19], as shown in Figure 3.

And the distribution of buildings presents a commercial-residential-commercial trend. The buildings are arranged with a spacing of 13 m. The distance between a building and the sidewalk is 10 m.

2.2.3. Tree Modeling. The trees are 0.5 m away from the edge of the road with an interval of 10 m. The height of the trees is 10 m, in which the heights of the trunk and the crown are both 5 m, as shown in Figure 4. The trunk is modelled as a hexagonal prism, while the crown is modelled as an approximate sphere.

In addition, the electromagnetic parameters of all the materials in the scenario are listed in Table 1.

2.3. Ray-Tracing Simulation. The Tx/Rx antenna is placed on top of the car, with 0.05 m above the roof. For example, the omnidirectional antennas are used, as shown in Figure 5. And the simulation process is as follows.

2.3.1. Straight Road. A car in the middle lane is selected to simulate the V2V channel characteristics with other cars within 100 m, as shown in Figure 1(a).

2.3.2. Intersection. A car waiting for the traffic light near the zebra crossing in the north-south direction is selected, to simulate the V2V channel characteristics with other cars within around 100 m, as shown in Figure 1(b).

As a technique to predict radio propagation characteristics, RT is based on electromagnetic theory, geometrical optics (GO) theory, and uniform theory of diffraction

(UTD). It is assumed that the radio wave propagates in the plane wave, whose far-field propagation characteristics can be simplified to a ray model to describe all propagation paths from Tx to Rx, and the geometry and dielectric properties of scattering objects in the scenario need to be modelled.

The RT simulator employed in this study is developed by Beijing Jiaotong University. It is composed of a V2V RT simulator [20] and an UWB THz RT simulator [21]. Recently, this RT simulator is extended to a high-performance computing (HPC) cloud-based platform (CloudRT). More details of this platform can be found in [22] as well as via http://raytracer.cloud. This RT simulator has been validated by a large number of measurements from frequencies below the 6 GHz [1, 20] to 300 GHz band [21, 23]. With a Web browser, MATLAB, and SketchUp installed, users can prepare the needed models, configure/trigger simulation tasks, and download results.

RT can identify all possible rays between Tx and Rx, whose propagation modes include LOS, reflection, scattering, and diffraction. After a coherent superposition of all determined rays, the final output of the ray-based model is the time-variant CIR $\mathbf{h}(\tau, t) \in \mathbb{C}^{M_R \cdot M_T}$, which completely characterizes the frequency-selective channel for each Tx/Rx link for the M_T and M_R transmit and receive antennas, respectively. We can express the CIR for single-input single-output (SISO) transmission, i.e., $M_R = M_T = 1$, as

$$h(\tau, t) = \sum_{k=1}^{N(t)} a_k(t) e^{j(2\pi f \tau_k(t) + \varphi_k(t))} \delta(\tau - \tau_k) = \sum_{k=1}^{N(t)} \tilde{a}_k(t) \delta(\tau - \tau_k),$$

(1)

where the kth MPC is described by the amplitude $a_k(t)$, the delay $\tau_k(t)$, and the phase shift $\varphi_k(t)$ at time t. $N(t)$ and f denote the number of MPCs for each time instant and the carrier frequency, respectively. Based on the predicted CIR, additional characteristics like the PDP and RMS delay spread can be derived.

The simulation parameters are listed in Table 2, where the reflection order is defined as the maximum number of

(a)

(b)

(c)
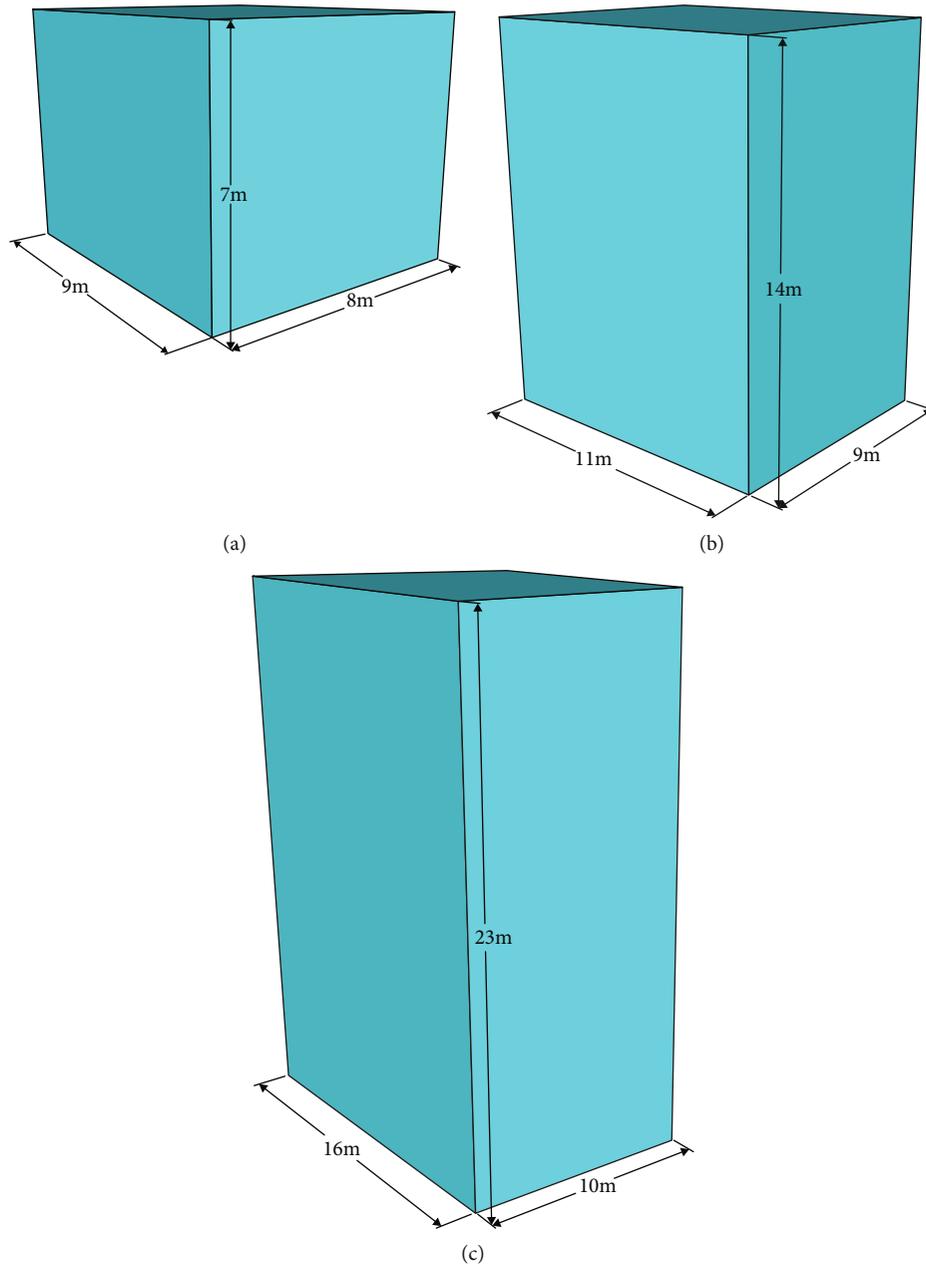
FIGURE 3: 3D model of buildings (take commercial buildings as an example): (a) low-rise buildings; (b) multistorey buildings; (c) medium high-rise buildings.

reflections that can be determined by the RT model based on the image method. Specular reflections are calculated recursively up to a desired order, but considering the complexity of the environment, only reflections up to order three are practical due to the limit of computational effort. In addition, the directive scattering (DS) model [24] which is widely used in optics is selected to calculate the power and direction of the scattering path in the platform.

3. Channel Characteristics

In this section, based on RT simulation technology, the channel characteristics in the straight road and intersection are discussed in peak hours.

3.1. LOS and NLOS. It can be seen from Figure 6(a), in the straight road scenario, as the Tx-Rx distance increases from 10 m to 90 m, the LOS probability falls from 100% to 40%.

As shown in Figure 6(b), the LOS probability in the intersection scenario has a similar trend. However, the LOS probability in different regions (from south to north, from north to south, and in east-west direction) is slightly different. For cars driving from north to south, if there are medium-sized or large-sized vehicles in the east-west direction in front of the Tx, LOS will be blocked. However, the distance between adjacent vehicles is relatively short in peak hours, which results in about 4-6 vehicles in the east-west direction ahead of the Tx. In addition, considering the proportion of vehicles (the probability of the appearance of medium-sized and
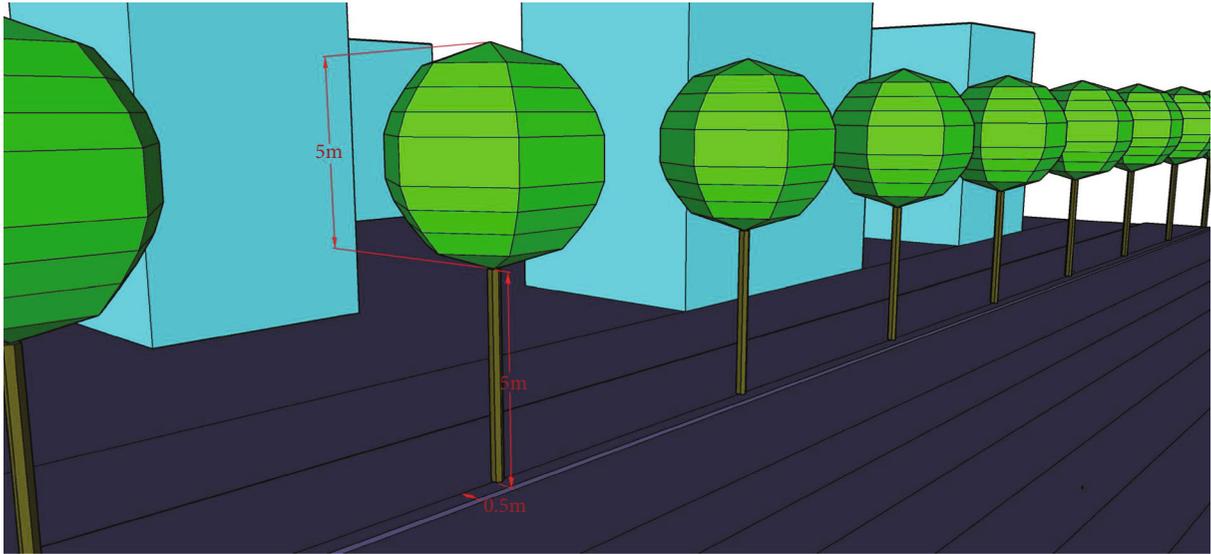
FIGURE 4: 3D model of trees.

TABLE 1: Material parameters.

Object	Material	Relative permittivity	Loss tangent
Ground, sidewalk, and residential buildings	Concrete	1.06	0.65
Vehicle windows, commercial buildings	Tempered glass	0.0538	23.9211
Trunk	Wood	1.27	0.0038
Vehicle body, fence	Metal	1	10^7
Leaves	Matsuba	1.36	0.0441
Wheel	Rubber	2.168	0.038
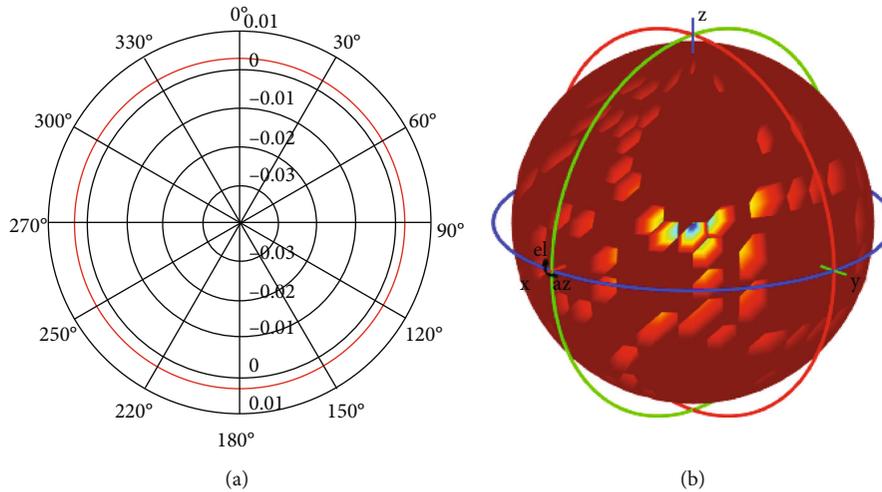
FIGURE 5: Omnidirectional antenna: (a) vertical cut; (b) 3D pattern.

large-sized vehicles appearing in urban roads is 11%), according to probability theory, the probability of more than one large- and medium-sized vehicle in these 4-6 vehicles, that is, the probability of LOS blockage, will be more than 40%, which means that the probability of LOS in that area will be less than 60%.

TABLE 2: Simulation parameters.

Parameter	Value
Frequency (GHz)	5.85-5.95
Resolution (MHz)	1
Reflection order	3
Scattering mode	Directive mode
Antenna (Tx, Rx)	Omnidirectional vertical polarization

(a)

(b)

South to north
East-west
North to south
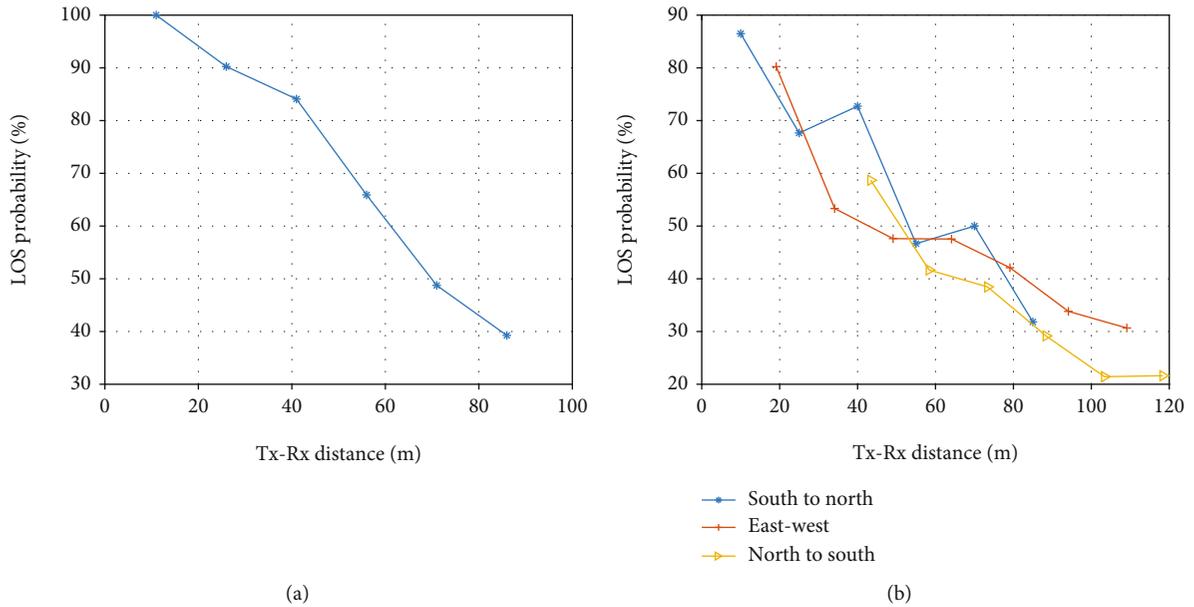
FIGURE 6: LOS probability at (a) straight road and (b) intersection.

(a)

(b)

(c)

(d)
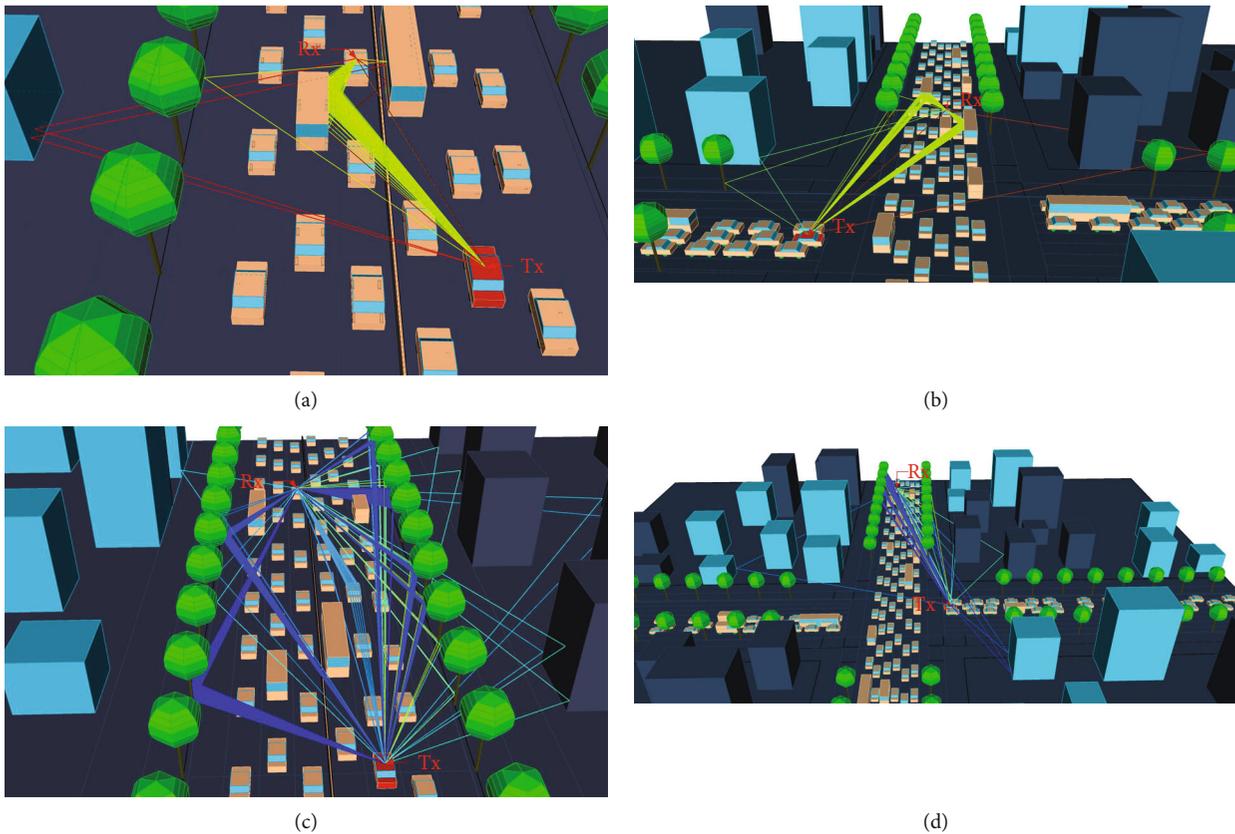
FIGURE 7: Propagation paths: (a) straight road scenario with LOS condition; (b) intersection scenario with LOS condition; (c) straight road scenario with NLOS condition; (d) intersection scenario with NLOS condition.

Figure 7 presents propagation paths in the straight road scenario and the intersection scenario in LOS and NLOS conditions.

In LOS conditions, there are a great number of reflections by the sides of the vehicle or the lower part of the side build-

ing. In NLOS conditions, the reflection rays, mainly from the sides of vehicles and buildings, if they exist, will dominate over a large number of scattering rays with relatively low power caused by the vehicles, buildings, and trees. Moreover,

(a)

(b)
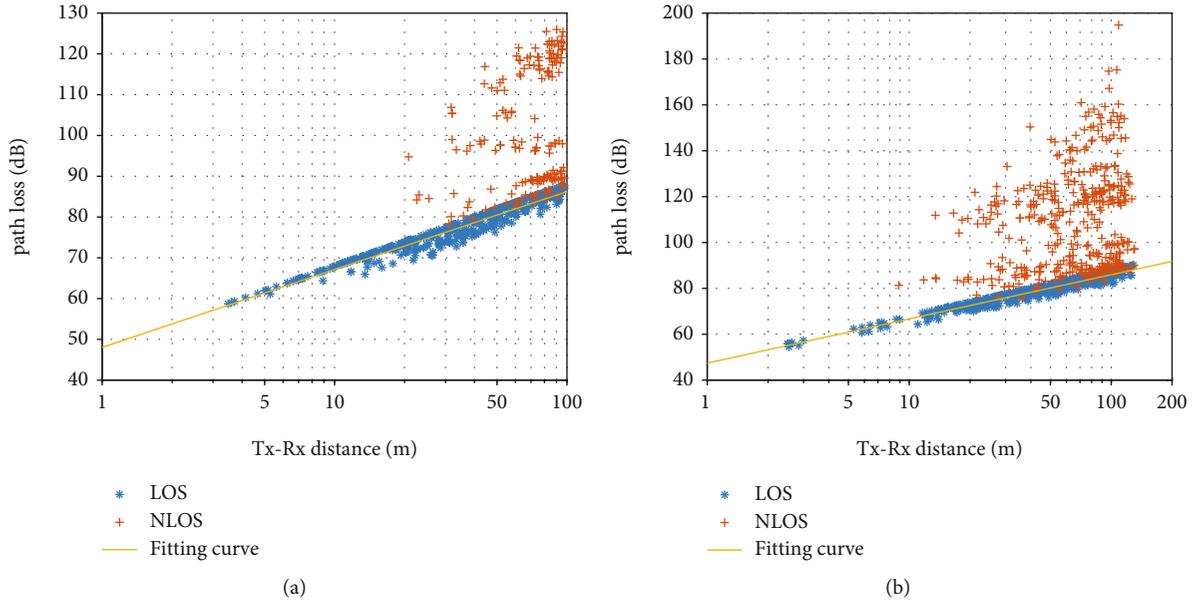
FIGURE 8: Path loss in terms of Tx-Rx distance at (a) straight road and (b) intersection.

TABLE 3: Parameters in path loss model.

Scenario	Parameter	Value
Straight road	n	1.937
	A (dB)	47.55
Intersection	n	1.927
	A (dB)	47.43

it can be seen that more reflecting objects (i.e., buildings and trees) exist near the Tx and Rx in the intersection scenario, so more multipath contributions (MPCs) can be provided when the Tx and the Rx are obstructed. According to ITU-R P.1411 [25], the radio propagation within 1 km is mainly affected by buildings and trees. Since most of the short-range wireless signal transmission is in the city and suburb, the influence of buildings is particularly important. The height of transmitters is near ground level, which leads to more interaction with objects in the close neighbor surroundings [26].

3.2. Path Loss. Figure 8 shows the path loss in the straight road scenario and the intersection scenario with LOS and NLOS conditions in terms of the Tx-Rx distance. The path loss in dB is roughly linear with the logarithmic distance in LOS conditions. In NLOS conditions, the shadowing loss caused by vehicle occlusion can exceed 20 dB even in the short Tx-Rx distance about 10 m, which is consistent with the conclusion in [13].

Considering that NLOS conditions will have a great impact on the path loss, in order to improve the accuracy, the path loss model, shown in (2), is used in LOS and NLOS conditions, independently, and the fitting parameters are shown in Table 3.

$$\mathrm{PL}(d) = A + 10n \log_{10}(d) + \begin{cases} L_a, & \text{LOS}, \\ L_b, & \text{NLOS}, \end{cases} \quad (2)$$

where A is the interception, d is the Tx-Rx distance (unit: m), n is the path loss exponent, and L_a and L_b are the shadowing loss in the LOS and NLOS conditions, respectively. Due to the obvious linear variation of the path loss in LOS conditions, both n and A are fitted in LOS conditions based on the minimum mean squared error (MMSE).

In the two scenarios, the waveguide effect occurs because of a set of waves reflected by the buildings on both sides of the road like valleys. Thus, the path loss exponent n in the straight road and the intersection are slightly smaller than those in the free space path loss (FSPL) model ($n = 2$). Moreover, the phenomenon ($n < 2$) has also been found in many other studies [4, 27–29]. And the difference of n and A in the two scenarios is very small.

The probability density function (PDF) of the shadowing loss and fitting results in the straight road scenario and the intersection scenario with LOS conditions is shown in Figure 9. It should be noted that the change trend of simulated data is approximately monotonically increasing, which is similar to special cases of Weibull distribution. Reference [30] points out that the Weibull distribution is a better fit over all empirical data than the Nakagami distribution in V2V channels. The PDF of the Weibull distribution is defined as

$$f(x, \lambda, k) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, \quad (3)$$

where $\lambda > 0$ denotes the scale parameter and $k > 0$ denotes the shape parameter. And the cumulative distribution
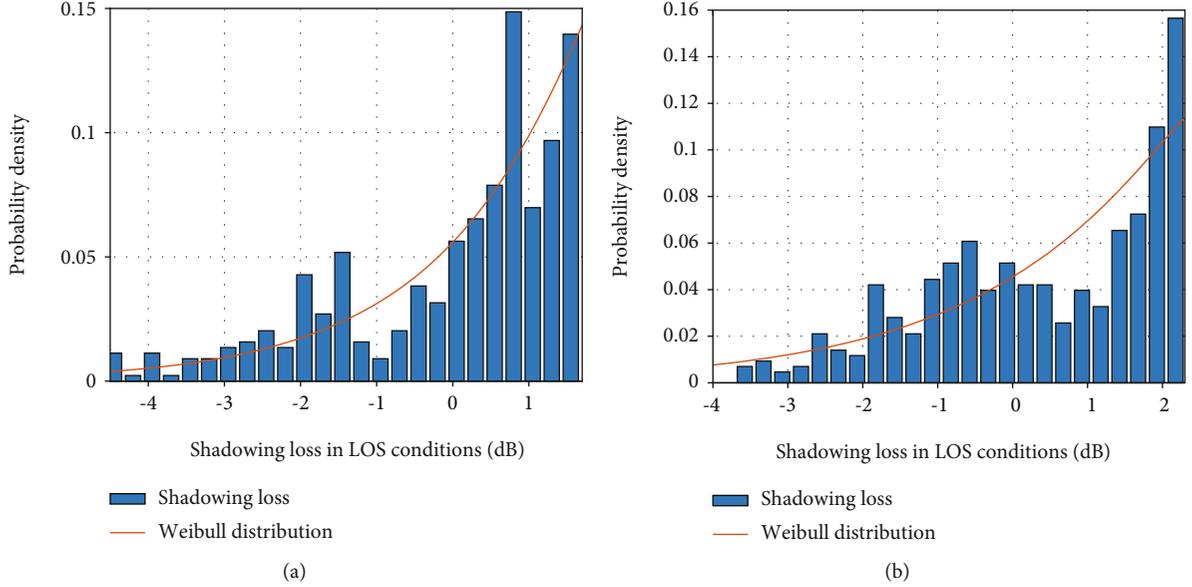
FIGURE 9: Shadowing loss distribution in LOS conditions: (a) straight road; (b) intersection.

TABLE 4: Parameters in shadowing loss model.

Scenario	Straight road	Intersection
λ	1.74	2.32
k	1.02	1.03
u_1	4.02	6.55
σ_1	3.00	8.51
u_2	16.94	37.04
σ_2	6.32	10.2
u_3	34.48	63.58
σ_3	5.08	7.22
a	0.60	0.46
b	0.07	0.38

function (CDF) is defined by

$$F(x, \lambda, k) = 1 - e^{-(x/\lambda)^k}. \tag{4}$$

In this paper, we need to modify the value of x by substituting the value of the maximum shadowing loss minus the shadowing loss; the fitting parameters of the Weibull distribution are presented in Table 4. The λ in the intersection is larger than that in the straight road, which shows that the probability of different shadowing loss values in the intersection with LOS conditions is more uniform. In the intersection, although the influence of multipaths at different locations is diversified, the effect of MPCs is more consistent in the distribution of the shadowing loss. The differences in the shadowing loss in LOS conditions are mainly attributed to the different participations of multipaths on the direct LOS path.

The PDFs of the shadowing loss and fitting results in the straight road scenario and the intersection scenario with

NLOS conditions are shown in Figure 10. Observed from data fitting, it is more in line with the multimodal Gaussian distribution, which is the superposition of three Gaussian distributions, corresponding to different compositions of multipath components, and can well describe the statistics of the shadowing loss, whether there is a straight road or a crossroad. Fitting parameters of the shadowing loss are shown in Table 4. The multimodal Gaussian distribution is defined as

$$
\begin{aligned}
f(x, u_1, \sigma_1, u_2, \sigma_2, u_3, \sigma_3, a, b) = & a \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x-u_1)^2}{2\sigma_1^2}\right) \\
& + b \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(x-u_2)^2}{2\sigma_2^2}\right) \\
& + (1-a-b) \frac{1}{\sqrt{2\pi}\sigma_3} \exp\left(-\frac{(x-u_3)^2}{2\sigma_3^2}\right),
\end{aligned}
\tag{5}
$$

where u_1, u_2, and u_3 are the mean values of the Gaussian distribution from left to right and σ_1, σ_2, and σ_3 are the corresponding standard deviations.

The mean value of the first Gaussian distribution of shadowing loss is 4.02 dB (straight road)/6.55 dB (intersection). This is mainly due to the existence of the single bounce reflections from the side of buildings and vehicles. Moreover, due to the larger value of a, the probability of the first Gaussian distribution is the largest. In [17, 31], it is reported that, in the absence of LOS, most of the power is received by single bounce reflections from physical objects. The measurement results in the urban road in [14] at the 5.6 GHz band show that the shadowing loss caused by obstructing vehicles follows the Gaussian distribution, and the mean value of the shadowing loss is about 7 dB, which is basically consistent
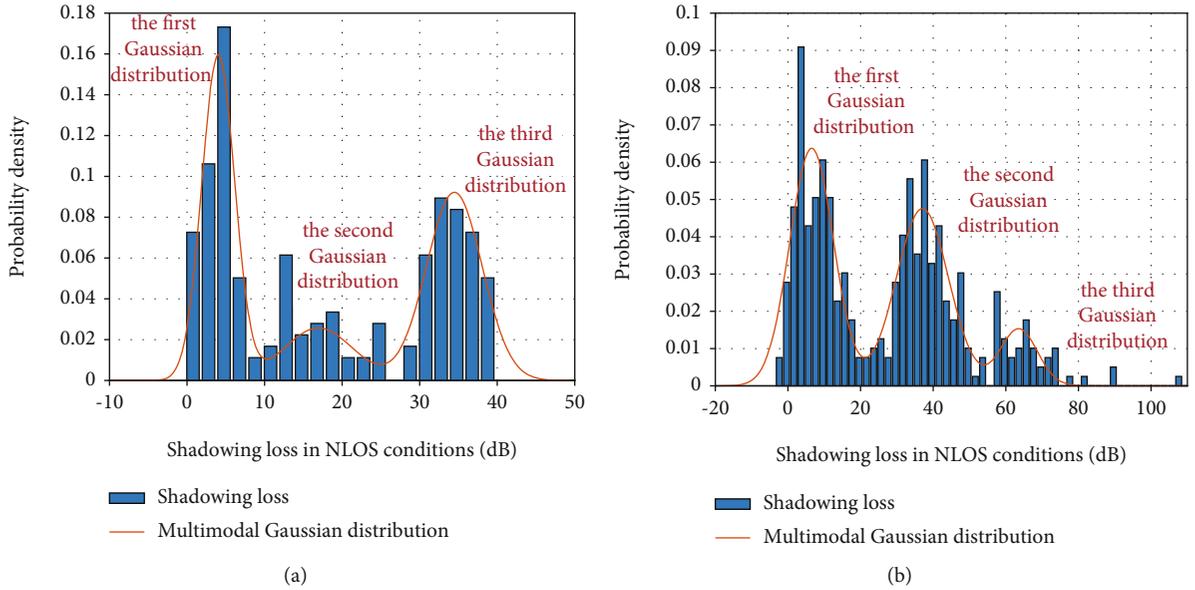
Figure 10: Shadowing loss distribution in NLOS conditions: (a) straight road; (b) intersection.
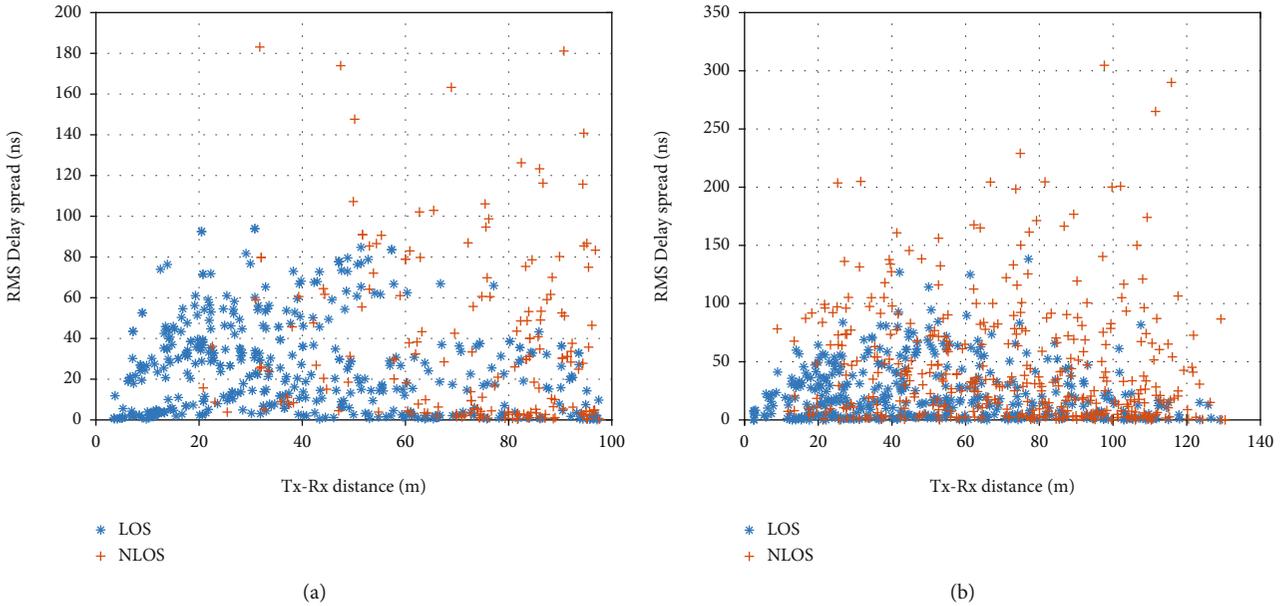
Figure 11: RMS delay spread in both LOS and NLOS conditions: (a) straight road; (b) intersection.

with the mean value of the first Gaussian distribution of the shadowing loss in this paper.

The mean value of the second Gaussian distribution of the shadowing loss in the straight road is 16.94 dB, with a low probability. It is mainly caused by multiple reflections from the vehicles.

The mean value of the second Gaussian distribution of the shadowing loss in the intersection and the third Gaussian distribution of the shadowing loss in the straight road is

34.48 dB (straight road)/37.06 dB (intersection). The rays are composed of a large number of scattering paths with a relatively low power from trees, buildings, and vehicles.

The mean value of the third Gaussian distribution of the shadowing loss in the intersection is 63.6 dB. It only appears when the Rx is driving in the east-west direction, extremely far away from the intersection and obstructed by many medium-sized or large-sized vehicles, resulting in only a small number of scattering paths existing, with lower power.
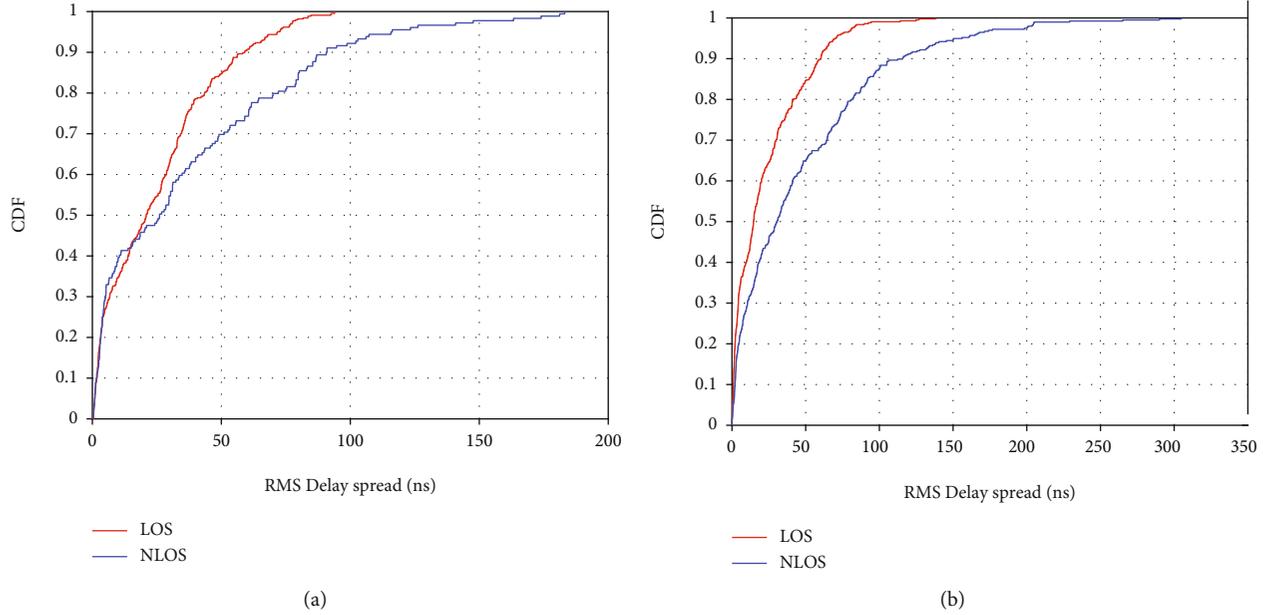
Figure 12: CDFs of RMS delay spread in both LOS and NLOS conditions: (a) straight road; (b) intersection.

Reference [9] points out that the path loss is high when the vehicles are far away from the intersection, with few physical objects actually providing propagation rays. In NLOS conditions, the strength of the received power is dependent on the availability of the reflection and scattering coming from buildings and trees, which may account for the main received power. In the absence of a large fraction of these reflections and scattering, the second Gaussian distribution or the third Gaussian distribution of the shadowing loss will appear.

3.3. RMS Delay Spread.
The RMS delay spread is defined as [32]

$$\tau_{\mathrm{rms}} = \sqrt{\frac{\sum_{n=1}^{N} P_n (\tau_n - \bar{\tau})^2}{\sum_{n=1}^{N} P_n}}, \tag{6}$$

where P_n and τ_n denote the power and the excess delay of the n-th ray, respectively; N is the number of rays; and the mean delay $\bar{\tau}$ is defined by

$$\bar{\tau} = \frac{\sum_{n=1}^{N} P_n \tau_n}{\sum_{n=1}^{N} P_n}. \tag{7}$$

Figure 11 shows the RMS delay spread in the straight road scenario and the intersection scenario with LOS and NLOS conditions. And the corresponding CDFs and PDFs are shown in Figures 12 and 13.

The median values of the RMS delay spread in the straight road are 20.86 ns (LOS) and 26.04 ns (NLOS), and 90% of the RMS delay spread are less than 58.24 ns (LOS) and 102.9 ns (NLOS). The median values of the RMS delay spread in the intersection are 20.86 ns (LOS) and 26.04 ns (NLOS), and 90% of the RMS delay spread are less than 58.24 ns (LOS) and 102.9 ns (NLOS).

The standard deviations of the RMS delay spread are 29.32 ns in the straight road and 41.66 ns in the intersection. The standard deviation in the intersection is larger than that in the straight road. The result illustrates that the surrounding scatterers in the intersection have a more significant impact on the V2V radio channel.

It can be observed that the RMS delay spread in LOS conditions is much smaller than that in NLOS conditions. The RMS delay spread in most locations is close to 0 ns, which indicates that MPCs with a large delay suffer greater attenuation.

Moreover, the RMS delay spread in the intersection is larger than that in the straight road. MPCs will be involved, and the Rx can receive more reflection and scattering paths from the surrounding buildings and trees, even from some distant buildings, as shown in Figures 6(b) and 6(d), which leads to a greater RMS delay spread.

Generally, the statistical characteristics of the RMS delay spread can be modelled as a Gaussian distribution [33, 34]. However, as can be seen from Figure 12, the Gaussian distribution is not applicable, so we use the Weibull distribution to model it. The results are presented in Table 5. A larger λ in the straight road indicates the extent to which the RMS delay spread variation is greater.

3.4. RMS Angular Spread.
The RMS angular spread is calculated as

$$\mathrm{AS} = \sqrt{-2 \ln \left(\left| \frac{\sum_{n=1}^{N} e^{(j\theta_n)} P_n}{\sum_{n=1}^{N} P_n} \right| \right)}, \tag{8}$$

where P_n and θ_n denote the power and angle (azimuth/elevation angle of arrival/departure) of the n-th ray, respectively, and N is the number of rays.

(a)

(b)
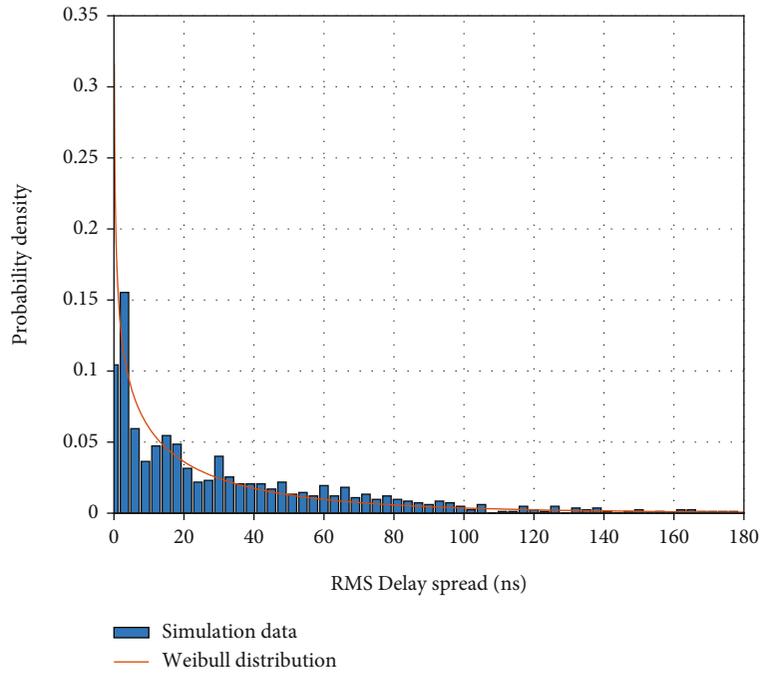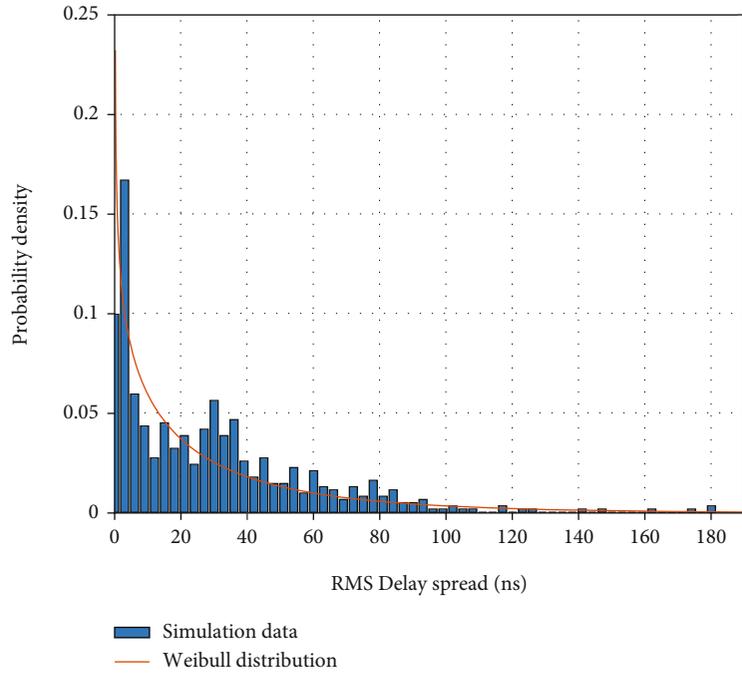
FIGURE 13: PDFs of RMS delay spread: (a) straight road; (b) intersection.

TABLE 5: Parameters in RMS delay spread.

Scenario	Straight road	Intersection
λ	27.55	26.16
k	0.80	0.74

In the V2V channel, the angle of arrival and the angle of departure are equivalent. Therefore, only the RMS angular spread of departure is discussed. Figure 14 shows the CDFs of the Azimuth Spread of Departure (ASD) and Elevation Spread of Departure (ESD) in the straight road and the intersection.

ESDs are almost $0°$ which indicate that the directions of the rays are almost parallel to the ground. The MPCs from the top of the buildings on both sides experience greater attenuation. The LOS path and the reflection paths from the buildings at the same height as the car contribute the main energy.

The median values of ASDs are $8.55°$ in the straight road and $8.68°$ in the intersection. 90% of ASDs are less than $32°$ in
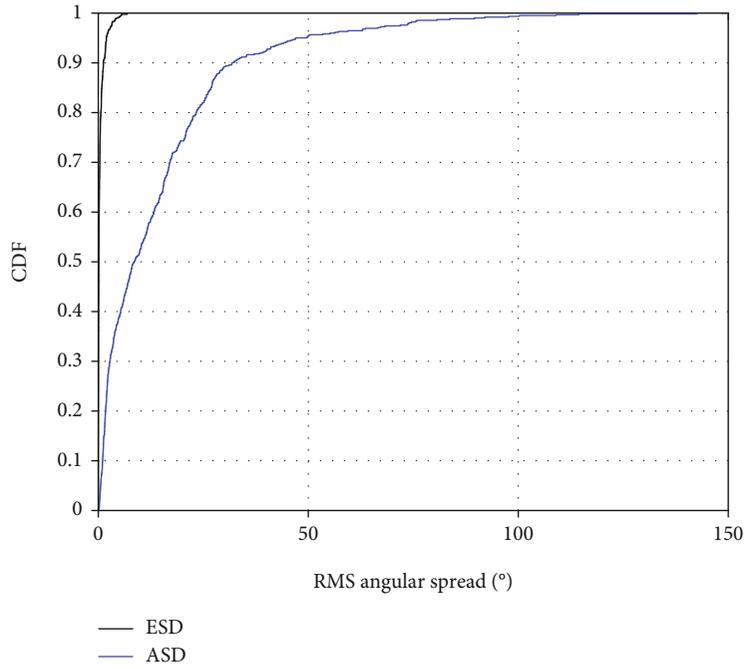
(a)
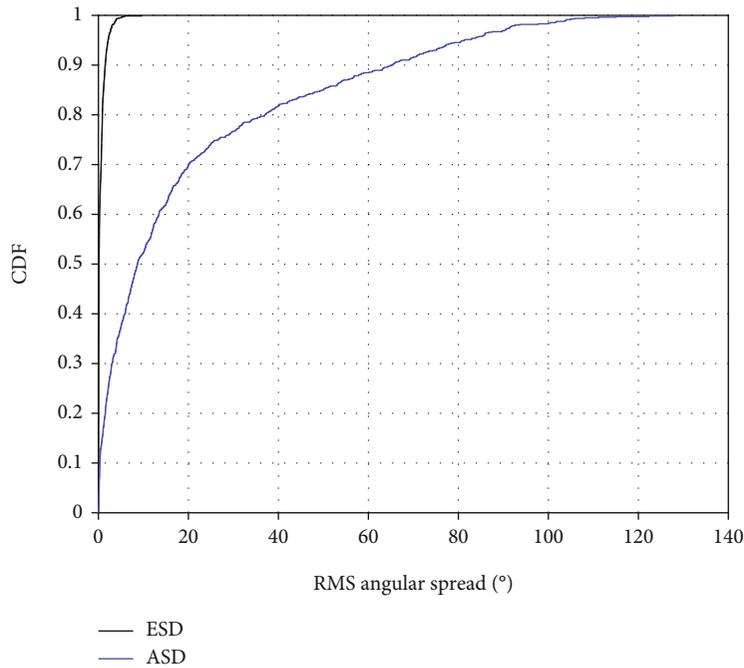
(b)

FIGURE 14: CDFs of RMS angular spread: (a) straight road; (b) intersection.

the straight road and 56° in the intersection. The ASDs in the intersection are larger than those in the straight road. The objects involved in the radio wave propagation in the straight road are arranged along both sides of the road, resulting in a relatively weak reflection and scattering paths.

The RMS angular spreads of the azimuth angle are larger than those of the elevation angle, illustrating that the low correlation and higher multiple-input multiple-output (MIMO)

gain can be improved when the antenna arrays are placed horizontally at the Tx locations.

4. Conclusion

Based on RT simulation technology, the V2V channel characteristics at the 5.9 GHz band are investigated. The two V2V scenarios, urban straight road and intersection,

including road models, vehicle models, building models, and tree models, are established. RT simulation results show that the LOS probability nearly linearly decreases with the Tx-Rx distance, and the shadowing loss caused by obstructing vehicles can exceed 20 dB even in the short Tx-Rx distance. The shadowing loss can be well characterized by the multimodal Gaussian distribution. Thus, the path loss models in the LOS and NLOS conditions are obtained. Furthermore, the distributions of the RMS delay spread in the two scenarios are obtained and are modelled by the Weibull distribution. Finally, the deployment of antenna arrays is discussed based on the statistics distribution of the angular spread.

In the past, the shadowing effect caused by obstructing vehicles is usually modelled as one Gaussian-correlated variable. However, different blocking levels, as we have observed, have different effects on shadow fading, which have largely been ignored. More importantly, we propose a multimodal Gaussian distribution to describe the shadowing loss, and the division of the multimode depends on the difference of multipath components caused by different blocking levels.

In the future work, based on RT simulation, we will carry out the research on V2V channel characteristics of specific scenarios and measurement campaigns to verify the RT simulation results and make the results more convincing.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

References

[1] T. Abbas, J. Nuckelt, T. Kurner, T. Zemen, C. F. Mecklenbrauker, and F. Tufvesson, "Simulation and measurement-based vehicle-to-vehicle channel characterization: accuracy and constraint analysis," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 7, pp. 3208–3218, 2015.

[2] Q. Wang, D. W. Matolak, and B. Ai, "Shadowing characterization for 5-GHz vehicle-to-vehicle channels," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 1855–1866, 2018.

[3] J. Kunisch and J. Pamp, "Wideband car-to-car radio channel measurements and model at 5.9 GHz," in *2008 IEEE 68th Vehicular Technology Conference*, pp. 1–5, Calgary, AB, Canada, 2008.

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

[5] I. Sen and D. W. Matolak, "Vehicle-vehicle channel models for the 5-GHz band," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 2, pp. 235–245, 2008.

[6] O. Renaudin, V. Kolmonen, P. Vainikainen, and C. Oestges, "Wideband MIMO car-to-car radio channel measurements at 5.3 GHz," in *2008 IEEE 68th Vehicular Technology Conference*, pp. 1–5, Calgary, AB, Canada, 2008.

[7] A. Paier, J. Karedal, N. Czink et al., "Characterization of vehicle-to-vehicle radio channels from measurements at 5.2 GHz," *Wireless Personal Communications*, vol. 50, no. 1, pp. 19–32, 2009.

[8] G. Acosta-Marum and M. Ingram, "Six time- and frequency-selective empirical channel models for vehicular wireless LANs," *IEEE Vehicular Technology Magazine*, vol. 2, no. 4, pp. 4–11, 2007.

[9] J. Karedal, F. Tufvesson, T. Abbas et al., "Radio channel measurements at street intersections for vehicle-to-vehicle applications," in *2010 IEEE 71st Vehicular Technology Conference*, pp. 1–5, Taipei, Taiwan, 2010.

[10] C. Schneider, M. Kaske, G. Sommerkorn et al., "Directional analysis of vehicle-to-vehicle propagation channels," in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Budapest, Hungary, 2011.

[11] I. Tan, W. Tang, K. Laberteaux, and A. Bahai, "Measurement and analysis of wireless channel impairments in DSRC vehicular communications," in *2008 IEEE International Conference on Communications*, pp. 4882–4888, Beijing, China, 2008.

[12] J. Nuckelt, T. Abbas, F. Tufvesson, C. Mecklenbrauker, L. Bernado, and T. Kurner, "Comparison of ray tracing and channel-sounder measurements for vehicular communications," in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Dresden, Germany, 2013.

[13] D. W. Matolak, "V2V communication channels: state of knowledge, new results, and what's next," in *International Workshop on Communication Technologies for Vehicles*, pp. 1–22, Villeneuve d'Ascq, France, 2013.

[14] T. Abbas, K. Sjöberg, J. Karedal, and F. Tufvesson, "A measurement based shadow fading model for vehicle-to-vehicle network simulations," *International Journal of Antennas and Propagation*, vol. 2015, Article ID 190607, 12 pages, 2015.

[15] P. Liu, D. W. Matolak, B. Ai, and R. Sun, "Path loss modeling for vehicle-to-vehicle communication on a slope," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 6, pp. 2954–2958, 2014.

[16] C. A. Gutiérrez, J. J. Jaime-Rodríguez, J. M. Luna-Rivera, D. U. Campos-Delgado, and J. Vázquez Castillo, "Modeling of non-WSSUS double-Rayleigh fading channels for vehicular communications," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 6394653, 15 pages, 2017.

[17] R. Meireles, M. Boban, P. Steenkiste, O. Tonguz, and J. Barros, "Experimental study on the impact of vehicular obstructions in VANETs," in *2010 IEEE Vehicular Networking Conference*, pp. 338–345, Jersey City, NJ, USA, 2010.

[18] J. Zhao, Y. Ji, and Y. Li, "Analysis on characteristics of road traffic flow in Tianjin," *Environmental Science Research*, vol. 32, no. 3, pp. 399–405, 2019.

[19] P. Zhang and X. Li, "Architectural types and spatial distribution characteristics of southern district of Qingdao," *Journal of Ecology*, vol. 33, no. 7, pp. 1882–1887, 2014.

[20] K. Guan, Z. Zhong, B. Ai, and T. Kürner, "Deterministic propagation modeling for the realistic high-speed railway environment," in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Dresden, Germany, 2013.

[21] S. Priebe and T. Kürner, "Stochastic modeling of THz indoor radio channels," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4445–4455, 2013.

[22] D. He, B. Ai, K. Guan, L. Wang, Z. Zhong, and T. Kürner, "The design and applications of high-performance ray-tracing simulation platform for 5G and beyond wireless communications: a tutorial," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 10–27, 2019.

[23] D. He, K. Guan, A. Fricke et al., "Stochastic channel modeling for kiosk applications in the terahertz band," *IEEE Transactions on Terahertz Science and Technology*, vol. 7, no. 5, pp. 502–513, 2017.

[24] S. Ju, S. H. A. Shah, M. A. Javed et al., "Scattering mechanisms and modeling for terahertz wireless communications," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, pp. 1–7, Shanghai, China, 2019.

[25] ITU-R. R. P.1411, "Propagation data and prediction methods for the planning of short-range outdoor radio communication systems and radio local area networks in the frequency range 300MHz and 100GHz," ITU, 2012.

[26] J. Lee, M. D. Kim, H. K. Chung, and J. Kim, "NLOS path loss model for low-height antenna links in high-rise urban street grid environments," *International Journal of Antennas and Propagation*, vol. 2015, Article ID 651438, 9 pages, 2015.

[27] A. F. Molisch, F. Tufvesson, J. Karedal, and C. Mecklenbrauker, "A survey on vehicle-to-vehicle propagation channels," *IEEE Wireless Communications*, vol. 16, no. 6, pp. 12–22, 2009.

[28] J. Karedal, N. Czink, A. Paier, F. Tufvesson, and A. F. Molisch, "Path loss modeling for vehicle-to-vehicle communications," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 1, pp. 323–328, 2011.

[29] H. Fernández, L. Rubio, J. Reig, V. M. Rodrigo-Peñarrocha, and A. Valero, "Path loss modeling for vehicular system performance and communication protocols evaluation," *Mobile Networks and Applications*, vol. 18, no. 6, pp. 755–765, 2013.

[30] Y. Li, B. Ai, X. Cheng, S. Lin, and Z. Zhong, "A TDL based non-WSSUS vehicle-to-vehicle channel model," *International Journal of Antennas and Propagation*, vol. 2013, Article ID 103461, 8 pages, 2013.

[31] T. Abbas, J. Karedal, F. Tufvesson, A. Paier, L. Bernado, and A. F. Molisch, "Directional analysis of vehicle-to-vehicle propagation channels," in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Budapest, Hungary, 2011.

[32] J. D. Parsons, *The Mobile Radio Propagation Channel*, Wiley, Hoboken, NJ, USA, 2nd edition, 2000.

[33] D. W. Matolak, I. Sen, W. Xiong, and N. T. Yaskoff, "5 GHz wireless channel characterization for vehicle to vehicle communications," in *MILCOM 2005 - 2005 IEEE Military Communications Conference*, pp. 3016–3022, Atlantic City, NJ, USA, 2005.

[34] A. F. Molisch, F. Tufvesson, J. Karedal, and C. Mecklenbrauker, "Propagation aspects of vehicle-to-vehicle communications - an overview," in *2009 IEEE Radio and Wireless Symposium*, pp. 179–182, San Diego, CA, USA, 2009.

WILEY | Hindawi

Research Article

Effective Capacity Maximization in beyond 5G Vehicular Networks: A Hybrid Deep Transfer Learning Method

Yi Huang [ID],[1,2] **Xinqiang Ma** [ID],[1,2] **Youyuan Liu** [ID],[1] **and Zhigang Yang** [ID][1]

[1]*Institute of Intelligent Computing and Visualization Based on Big Data, Chongqing University of Arts and Sciences, Chongqing, China*
[2]*College of Computer Science and Technology, Guizhou University, Guiyang, China*

Correspondence should be addressed to Yi Huang; cqhy@21cn.com and Xinqiang Ma; xinqma@zju.edu.cn

How to improve delay-sensitive traffic throughput is an open issue in vehicular communication networks, where a great number of vehicle to infrastructure (V2I) and vehicle to vehicle (V2V) links coexist. To address this issue, this paper proposes to employ a hybrid deep transfer learning scheme to allocate radio resources. Specifically, the traffic throughput maximization problem is first formulated by considering interchannel interference and statistical delay guarantee. The effective capacity theory is then applied to develop a power allocation scheme on each channel reused by a V2I and a V2V link. Thereafter, a deep transfer learning scheme is proposed to obtain the optimal channel assignment for each V2I and V2V link. Simulation results validate that the proposed scheme provides a close performance guarantee compared to a globally optimal scheme. Besides, the proposed scheme can guarantee lower delay violation probability than the schemes aiming to maximize the channel capacity.

1. Introduction

The rapid evolution of mobile communication technologies invites all human beings to the era of the Internet of Everything, where unprecedented changes will take place in all walks of life and have a profound impact on every single aspect of our daily interactions [1, 2]. Vehicular communications, widely regarded as a promising technology to enable intelligent transportation, autonomous driving, and even every potential application related to smart vehicles in beyond 5G networks have attracted extensive attention from both academia and industry [3]. Typically, the link types of vehicular communications include vehicle to infrastructure (V2I), vehicle to vehicle (V2V), and vehicle to everything (V2X) [4]. Worth noting is that different communication link types usually have to provide certain quality of service (QoS) guarantees [5]. For instance, an autonomous driving vehicle is expected to transmit its rough position information to the infrastructure to help the base station (BS) perceive the whole vehicular network. Besides, such vehicles are constantly exchanging various types of their instantaneous infor-

mation with adjacent vehicles to ensure transportation safety. Apparently, these two types of information should be transmitted with low delay, where the exchange of the instantaneous information between adjacent vehicles using V2V links is inherently more delay-sensitive than V2I communications. Moreover, since the spectrum resources are quite limited in existing cellular systems, how to effectively provide differentiated QoS guarantees for different traffic is a critical issue for vehicular communication networks [6].

Traffic throughput maximization usually serves as an objective to improve the overall spectrum efficiency of a given communication network [7]. However, as a vehicular network is generally required to provide different QoS guarantees for different traffic, traditional resource allocation schemes that aim to maximize the channel capacity may be no longer applicable. Considering the low delay constraint, actual resource optimization problems are usually more complex [8, 9]. Additionally, spectrum sharing is another potential solution to improve the spectrum efficiency of a communication system. In a vehicular network, V2I and V2V communications can reuse the same channel

to transmit data, which dramatically increases the number of access links for vehicular communications. Nevertheless, the introduced interference between the V2V and V2I links further complicates the performance analysis of vehicular networks. In summary, a fundamental challenge for vehicular communication networks is to design an efficient resource allocation scheme to maximize the network traffic throughput under diverse QoS requirements and various interference constraints.

To address this challenge, this paper focuses on a vehicular network where V2I and V2V links share the limited spectrum resource. Specifically, an interference model and a statistical delay model are both established, based on which a traffic throughput maximization problem is formulated aiming to acquire the optimal power allocation and spectrum sharing scheme. Subsequently, the optimization problem is decomposed into a power allocation subproblem for each pair of cellular user (CUE) and V2V user (VUE) and a spectrum sharing subproblem for the whole vehicular network. Firstly, the power allocation subproblem is solved analytically based on the effective capacity theory, with both the statistical information of small-scale channel fading and the instantaneous information of large-scale channel fading taken into account. Secondly, a supervised deep learning algorithm is proposed to solve the spectrum sharing problem. Moreover, to overcome the mismatch problem caused by the varying distribution of hidden network information and states, we propose a deep transfer learning algorithm to adapt fast to new scenarios and to achieve optimization under certain QoS requirements for vehicular networks. Simulation results validate the accuracy of our proposed learning schemes, and the performance analyses show that traditional channel capacity maximization schemes may incur a high delay violation probability for delay-sensitive traffic, which is systematically alleviated by our proposed learning schemes.

The contributions of this paper are summarized as follows:

(i) An analytical model is established to jointly consider the statistical delay guarantee and interchannel interference. Compared to the traditional model based on the average delay, our proposed analytical model better fits the context of beyond 5G networks, where the delay performance is commonly measured in a probabilistic dimension

(ii) A power allocation scheme is proposed to maximize the throughput of delay-sensitive traffic for a given CUE-VUE pair. The highlight of our proposed scheme lies in that its computation complexity only relates to the number of power levels, which enables its application on a real-world vehicular transmitter. In addition, the power allocation scheme can guarantee the traffic delay requirement for both CUE and VUE while other conventional schemes only guarantee either one of the vehicular links

(iii) A deep learning-based spectrum sharing scheme is proposed to quickly obtain the optimal channel reuse strategy. Based on the offline deep learning

algorithm, a deep transfer learning algorithm is further developed to deal with the mismatch problem commonly encountered in new scenarios, where the hidden information is dynamic and only few training samples can be obtained

The remainder of this paper is organized as follows. In Section 2, related work is introduced and discussed. In Section 3, the network model, interference model, and delay model are presented. Section 4 proposes the traffic maximization scheme, and Section 5 compares and discusses the simulation results. Finally, the paper is concluded in Section 6.

2. Related Work

In the literature, existing studies on throughput maximization for vehicular communications can be briefly summarized as follows.

In [10], a low complexity data routing policy was designed to maximize the data throughput from vehicles to roadside units. In [11], a data transmission and scheduling scheme was proposed to maximize the traffic throughput and reduce the resource contention for nonadjacent V2V communications. In [12], an information spread problem in vehicular networks with V2I and V2V links was formulated and solved, where the channel capacity of V2I links was maximized based on the Doppler effect. In [13], a coalition game model was introduced to optimize resource allocation and maximize the throughput of individual V2V links under a minimum V2I throughput requirement. In [14], a novel power allocation and spectrum sharing algorithm was proposed to optimize the throughput of V2I links while guaranteeing the minimum throughput requirement of V2V links. The abovementioned works [10–14] have designed novel resource optimization schemes for different scenarios. However, in these studies, the throughput performance was simply characterized by the Shannon channel capacity, and the transmission delay was not taken into account.

Since the transmission delay is a critical metric in vehicular communications, a significant number of researchers have paid close attention to the improvements of the delay performance. In the literature, the transmission delay was usually analyzed in the average and used as an indicator to calculate the tradeoff with other performance metrics based on the Lyapunov theory [15–18]. In [15], the TV white space bands were used to supplement the bandwidth for the computation offloading of vehicular terminals. The computation offloading and bandwidth allocation decisions were jointly optimized to balance the task delay and the cost of the TV white space bands. In [16], the extreme value theory and Lyapunov theory were employed to analyze the tail distribution of the age of information in a given vehicular network. A power control scheme was proposed to guarantee the mean delay requirement. In [17], a vehicle-centric approach was designed to optimize the node association and resource reallocation, by taking the additional latency caused by the overhead into account. In [18], the long-term time-averaged total
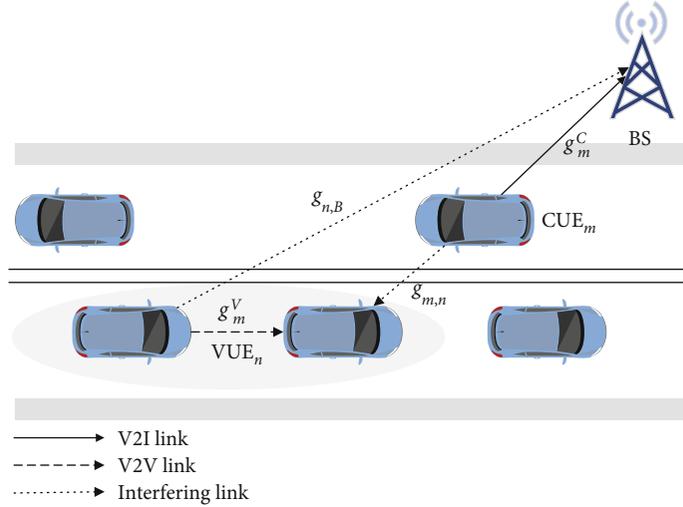
FIGURE 1: A typical vehicular network with CUEs and VUEs.

system capacity was maximized while satisfying the strict ultrareliable and low-latency requirements of vehicle communications. Generally, the mean delay is leveraged to characterize delay-tolerant traffic. However, there are many types of delay-sensitive traffic in vehicular networks where the positions of vehicles, wireless channel states, and traffic arrival intervals are all highly dynamic. Hence, the statistical delay guarantee is more useful for practical vehicular networks. In [19], the capacity of V2I links was maximized under a given delay and delay violation probability requirement, where the closed-form power allocation solution was derived for each V2I and V2V reuse link. However, [19] only considered the delay requirement of V2V links. To the best of our knowledge, how to provide the delay guarantee for both V2I and V2V links at the same time is still an open problem.

In addition, deep learning-based techniques are becoming more and more popular in wireless communications [20]. In [21], the authors integrated a convolutional neural network and a long short-term memory network to predict the channel state information. In [22], the authors constructed a feature learning framework for IoT applications to effectively classify data and detect anomaly events, using RBF-BP hybrid neural network. In [23], the deep learning assisted optimization methods for resource allocation in vehicular communications were introduced and compared. In [24], a multiagent reinforcement learning framework was proposed for the spectrum sharing in vehicular networks with V2I and V2V links. It is evident that deep learning is confirmed to be an effective tool for optimization in wireless communications. Hence, in this paper, we propose a hybrid deep transfer learning scheme to address the aforementioned problem in vehicular communications.

3. System Model

3.1. System Model. We consider a multivehicle single-cell network as depicted in Figure 1, where there are M vehicles as CUEs and $N(N \leq M)$ pairs of proximate vehicles as VUEs. The CUEs transmit information to the BS through V2I com-

munications with orthogonal channels, while the VUEs employ V2V communications to send and receive data through sharing the spectrum resource with CUEs. The total bandwidth of the considered network is B_{tot}. We assume that each CUE can only occupy one channel at a time, and a channel can only be allocated to one CUE. Hence, the channel bandwidth allocated to a CUE can be denoted as $B = B_{tot}/M$. In order to avoid the strong interference between V2I and V2V links, each VUE can only reuse one channel, and each channel can only be shared with one VUE. For notational expedience, we use $\mathcal{M} = \{1, 2, .., m, \cdots, M\}$ and $\mathcal{N} = \{1, 2, .., n, \cdots, N\}$ to denote the sets of CUEs and VUEs, respectively. In addition, all the CUEs and VUEs are equipped with a single antenna.

3.2. Communication Model. We denote the channel power gain from the mth CUE to the BS by $g_m^C = \varphi_m^C h_m^C$, where φ_m^C and h_m^C characterize the large-scale and small-scale fading components, respectively. The large-scale fading parameter can be further modeled as $\varphi_m^C = \phi \omega_m^C (l_m^C)^{-\alpha}$, where ϕ denotes the path loss constant, ω_m^C is the random log-normal shadowing parameter, l represents the distance between the CUE and the BS, and α is the power decay exponent. Similarly, we use $g_n^V = \varphi_n^V h_n^V$, $g_{n,B} = \varphi_{n,B} h_{n,B}$, and $g_{m,n} = \varphi_{m,n} h_{m,n}$ to represent the channel power gain of the nth VUE, the interference power gain from the nth VUE to the BS, and the interference power gain from the mth CUE to the nth VUE, respectively. In addition, due to the high mobility of vehicles and the varying delay requirement of different data traffic, it is impractical for the BS to always obtain the instantaneous small-scale fading information. However, the statistical information is easily accessible by the BS from the feedback of vehicles within hundreds of time slots. Hence, in this paper, we assume all the CUEs and VUEs undergo the small-scale Rayleigh fading. In other words, small-scale fading parameters h_m^C, h_n^V, $h_{n,B}$, and $h_{m,n}$ follow the independent and exponential distribution with unit mean in each time slot.

As the channel (interference) power gain is time-varying for both V2I and V2V links. The considered network should make a decision on the power management and spectrum sharing for all CUEs and VUEs when the statistical channel information changes. The transmission power of the mth CUE and the nth VUE is denoted by p_m^C and p_n^V, respectively. Besides, binary indicator $\tau_{m,n}$ is employed to characterize the channel reused by the mth CUE and the nth VUE, where $\tau_{m,n} = 1$ means the mth CUE and the nth VUE share the same channel. As a result, the signal to interference plus noise ratio (SINR) of the mth CUE and the nth VUE holds as

$$\gamma_m^C = \frac{p_m^C g_m^C}{N_0 B + \sum_{n=1}^{N} \tau_{m,n} p_n^V g_{n,B}}, \tag{1}$$

$$\gamma_n^V = \frac{p_n^V g_n^V}{N_0 B + \sum_{n=1}^{N} \tau_{m,n} p_m^C g_{m,n}}, \tag{2}$$

where N_0 denotes the power spectral density of background noise. According to the Shannon's Theorem, the channel capacity of the mth CUE and the nth VUE in each time slot can be obtained as

$$R_m^C = B \log_2 \left(1 + \gamma_m^C\right), \tag{3}$$

$$R_n^V = B \log_2 \left(1 + \gamma_n^V\right). \tag{4}$$

3.3. Performance Metric and Problem Formulation. Typically, a vehicle generates delay-sensitive traffic periodically and sends it to the BS or other vehicles in a V2X network, where the traffic is assumed to be infinitesimal. The corresponding cumulative arrivals during $(0, t]$ are denoted by $A_m^C(t)$ and $A_n^V(t)$. Similarly, the cumulative departures are denoted by $A_m^{C*}(t)$ and $A_n^{V*}(t)$. At t, the traffic delay can be obtained as

$$D_m^C(t) = \max \left\{ d : A_m^C(t) \geq A_m^{C*}(t + d) \right\},$$
$$D_n^V(t) = \max \left\{ d : A_n^V(t) \geq A_n^{V*}(t + d) \right\}. \tag{5}$$

In order to characterize the delay performance more intuitively, we model the delay metric according to the philosophy behind 5G ultrareliable low latency communications (uRLLC). Specifically, statistical delay characteristics are analyzed in this paper, as shown in

$$\Pr \left\{ D_m^C(t) > d_m^C \right\} \leq \varepsilon_m^C,$$
$$\Pr \left\{ D_n^V(t) > d_n^V \right\} \leq \varepsilon_n^V. \tag{6}$$

For the mth CUE, its statistical delay performance means the traffic delay exceeding threshold d_m^C should be controlled with probability ε_m^C, which also holds for the nth VUE. In a V2X network, a vehicle sustaining a higher traffic arrival rate under a specific delay requirement means that this vehicle is able to update its information to other vehicles or the infra-

structure more timely. Let λ_m^C ($m \in \mathcal{M}$) and λ_n^V ($n \in \mathcal{N}$) denote the maximum arrival rate (i.e., traffic throughput) sustained by the mth CUE and the n VUE under the delay requirement, respectively. We model the traffic throughput under the delay requirement as follows

$$\lambda_m^C = \max \left\{ \lambda : \Pr \left\{ D_m^C(t) > d_m^C \right\} \leq \varepsilon_m^C \right\},$$
$$\lambda_n^V = \max \left\{ \lambda : \mathrm{Pe} \left\{ D_n^V(t) > d_n^V \right\} \leq \varepsilon_n^V \right\}. \tag{7}$$

In this paper, we aim to maximize the traffic throughput for the considered network under diverse delay requirements through optimizing the power and spectrum allocation for each CUE and VUE. The optimization problem regarding to resource allocation can be formulated as

$$\mathbf{P1} \quad \max_{\{\tau_{m,n}\}, \{p_m^C\}, \{p_n^V\}} \sum_{m \in \mathcal{M}} \lambda_m^C + \sum_{n \in \mathcal{N}} \lambda_n^V$$

$$\text{s.t.} \quad C1 : \Pr \left\{ D_m^C(t) > d_m^C \right\} \leq \varepsilon_m^C \forall m \in \mathcal{M}$$

$$C2 : \Pr \left\{ D_n^V(t) > d_n^V \right\} \leq \varepsilon_n^V \forall n \in \mathcal{N}$$

$$C3 : 0 \leq p_m^C \leq p_{\max}^C \forall m \in \mathcal{M} \tag{8}$$

$$C4 : 0 \leq p_n^V \leq p_{\max}^V \forall n \in \mathcal{N}$$

$$C5 : \sum_{m \in \mathcal{M}} \tau_{m,n} \leq 1, \forall n \in \mathcal{N}$$

$$C6 : \sum_{n \in \mathcal{N}} \tau_{m,n} \leq 1, \forall m \in \mathcal{M}$$

In **P1**, C1 and C2 represent the delay constraints for CUEs and VUEs, respectively. C3 and C4 constrain the transmission power range of CUEs and VUEs, respectively. C5 and C6 are the spectrum sharing constraints ensuring that the channel of each CUE is reused by at most one VUE and each VUE reuses the channel of at most one CUE. Considering C5 and C6, **P1** is a mixed integer nonlinear programming (MINP) that cannot be solved by traditional convex optimization approaches. Moreover, due to the lack of tractable expressions to characterize C1 and C2, it is more challenging to solve **P1**, compared with other MINP problems, especially when the traffic throughput is modeled using the channel capacity without considering delay requirements. Therefore, we propose a hybrid deep transfer learning method to achieve the optimal resource allocation.

4. Joint Power Allocation and CUE-VUE Association Optimization

As the interference only exists in a channel that is reused by a CUE and a VUE, **P1** can be decomposed into a power allocation subproblem **P2** for a given CUE-VUE pair and CUE-VUE association subproblem **P3** for a given

power allocation. Specifically, **P2** and **P3** can be formulated as in

$$\textbf{P2} \quad \max_{p_m^C, p_n^V} \lambda_m^C + \lambda_n^V, \tag{9}$$
$$\text{s.t.} \quad C1 - C4$$

$$\textbf{P3} \quad \max_{\{\tau_{m,n}\}} \sum_{m \in \mathcal{M}} \lambda_m^C + \sum_{n \in \mathcal{N}} \lambda_n^V. \tag{10}$$
$$\text{s.t.} \quad C5 - C6$$

4.1. Power Allocation for CUE-VUE Pair. In order to solve **P2**, we need to first deduce the delay constraint for a given CUE-VUE pair. Without loss of generality, we randomly choose the mth CUE and the nth VUE as a pair for the subsequent analysis. According to the effective capacity theory, the delay violation probabilities for the mth CUE and the nth VUE can be obtained as

$$\Pr\left\{D_m^C(t) > d_m^C\right\} \le e^{-\theta_m^C \beta_m^C\left(\theta_m^C\right) d_m^C}, \tag{11}$$

$$\Pr\left\{D_n^V(t) > d_n^V\right\} \le e^{-\theta_n^V \beta_n^V\left(\theta_n^V\right) d_n^V}, \tag{12}$$

where β_m^C and β_n^V denote the effective capacity and θ_m^C and θ_m^C are the nonnegative QoS exponential parameters that can be further optimized. For a stable vehicular network, the effective capacity of CUE and VUE can be calculated as [25].

$$\beta_m^C\left(\theta_m^C\right) = -\frac{\ln \mathbb{E}\left[e^{-\theta_m^C R_m^C}\right]}{\theta_m^C} \ge \lambda_m^C, \tag{13}$$

$$\beta_n^V\left(\theta_n^V\right) = -\frac{\ln \mathbb{E}\left[e^{-\theta_n^V R_n^V}\right]}{\theta_n^V} \ge \lambda_n^V. \tag{14}$$

Combining (11) and (13), we have

$$\Pr\left\{D_m^C(t) > d_m^C\right\} \le \mathbb{E}\left[e^{-\theta_m^C R_m^C d_m^C}\right], \tag{15}$$

$$\Pr\left\{D_n^V(t) > d_n^V\right\} \le \mathbb{E}\left[e^{-\theta_n^V R_n^V d_n^V}\right]. \tag{16}$$

From (15), the delay violation probability of CUE can be improved by increasing θ_m^C. However, according to (13), the effective capacity of CUE decreases with θ_m^C, which implies that a low λ_m^C is guaranteed. Similar results can

be derived for VUE. Hence, **P2** can be transformed into the following feasible problem

$$\textbf{P4} \quad \max_{p_m^C, p_n^V, \theta_m^C, \theta_n^V} -\frac{\ln \mathbb{E}\left[e^{-\theta_m^C R_m^C}\right]}{\theta_m^C} - \frac{\ln \mathbb{E}\left[e^{-\theta_n^V R_n^V}\right]}{\theta_n^V}$$
$$\text{s.t.} \quad C1 : \mathbb{E}\left[e^{-\theta_m^C R_m^C d_m^C}\right] \le \varepsilon_m^C$$
$$C2 : \mathbb{E}\left[e^{-\theta_n^V R_n^V d_n^V}\right] \le \varepsilon_n^V \tag{17}$$
$$C3 : 0 \le p_m^C \le p_{\text{max}|}^C$$
$$C4 : 0 \le p_n^V \le p_{\text{max}}^V$$

To solve **P4**, the following theorem is derived.

Theorem 1. *If $\{p_m^{C*}, p_n^{V*}, \theta_m^{C*}, \theta_n^{V*}\}$ denotes the optimal solution for **P4**, the following equations must hold*

$$\mathbb{E}\left[e^{-\theta_m^{C*} R_m^C\left(p_m^{C*}, p_n^{V*}\right) d_m^C}\right] = \varepsilon_m^C, \tag{18}$$

$$\mathbb{E}\left[e^{-\theta_n^{V*} R_n^V\left(p_m^{C*}, p_n^{V*}\right) d_n^V}\right] = \varepsilon_n^V. \tag{19}$$

Proof. Firstly, we assume that for the optimal solution $\{p_m^{C*}, p_n^{V*}, \theta_m^{C*}, \theta_n^{V*}\}$,

$$\mathbb{E}\left[e^{-\theta_m^{C*} R_m^C\left(p_m^{C*}, p_n^{V*}\right) d_m^C}\right] < \varepsilon_m^C. \tag{20}$$

According to (13), the effective capacity of CUE, i.e., $\beta_m^C(\theta_m^C)$ is a continuously decreasing function in θ_m^C while the delay violation probability $\mathbb{E}[e^{-\theta_m^C R_m^C(p_m^C, p_n^V) d_m^C}]$ is also a continuously decreasing function in θ_m^C. As a result, there always exists $\widetilde{\theta}_m^C = \sigma \theta_m^{C*} < \theta_m^{C*}(\sigma \longrightarrow 1^-)$ meeting the delay constraint as

$$\mathbb{E}\left[e^{-\theta_m^{C*} R_m^C\left(p_m^{C*}, p_n^{V*}\right) d_m^C}\right] < \mathbb{E}\left[e^{-\widetilde{\theta}_m^{C*} R_m^C\left(p_m^{C*}, p_n^{V*}\right) d_m^C}\right] \le \varepsilon_m^C. \tag{21}$$

Also,

$$-\frac{\ln \mathbb{E}\left[e^{-\theta_m^{C*} R_m^C\left(p_m^{C*}, p_n^{V*}\right)}\right]}{\theta_m^{C*}} < -\frac{\ln \mathbb{E}\left[e^{\widetilde{\theta}_m^C R_m^C\left(p_m^{C*}, p_n^{V*}\right)}\right]}{\widetilde{\theta}_m^C}. \tag{22}$$

Hence, power allocation scheme $\{p_m^{C*}, p_n^{V*}, \widetilde{\theta}_m^C, \theta_n^{V*}\}$ guarantees a higher traffic throughput than $\{p_m^{C*}, p_n^{V*}, \theta_m^{C*}, \theta_n^{V*}\}$ under the delay and power constraints, which is a contradiction. On the other hand, we assume that for the optimal solution $\{p_m^{C*}, p_n^{V*}, \theta_m^{C*}, \theta_n^{V*}\}$, the following equation holds

$$\mathbb{E}\left[e^{-\theta_n^{V*} R_n^V\left(p_m^{C*}, p_n^{V*}\right) d_n^V}\right] < \varepsilon_n^V. \tag{23}$$

And a similar contradiction to the assumption can be observed. As a result, the optimal power allocation must meet C1 and C2 equally.

According to (18), the effective capacities of CUE and VUE can be further simplified as

$$
\begin{aligned}
\beta_m^C\left(\theta_m^C\right) &= -\frac{\ln \mathbb{E}\left[e^{-\theta_m^C R_m^C}\right]}{\theta_m^C} = \frac{\ln 1/\varepsilon_m^C}{d_m^C}\frac{1}{\theta_m^C}, \\
\beta_n^V\left(\theta_n^V\right) &= -\frac{\ln \mathbb{E}\left[e^{-\theta_n^V R_n^V}\right]}{\theta_n^V} = \frac{\ln 1/\varepsilon_n^V}{d_n^V}\frac{1}{\theta_n^V}.
\end{aligned}
\tag{24}
$$

Therefore, **P4** can be further transformed to

$$
\begin{aligned}
\textbf{P5} \quad &\max_{p_m^C, p_n^V} \Gamma_{m,n}\left(p_m^C, p_n^V\right) \triangleq \frac{\ln 1/\varepsilon_m^C}{d_m^C}\frac{1}{\theta_m^C} + \frac{\ln 1/\varepsilon_n^V}{d_n^V}\frac{1}{\theta_n^V} \\
\text{s.t.} \quad &C1: \mathbb{E}\left[e^{-\theta_m^C R_m^C d_m^C}\right] = \varepsilon_m^C \\
&C2: \mathbb{E}\left[e^{-\theta_n^V R_n^V d_n^V}\right] = \varepsilon_n^V \\
&C3: 0 \leq p_m^C \leq p_{\max}^C \\
&C4: 0 \leq p_n^V \leq p_{\max}^V
\end{aligned}
\tag{25}
$$

Note that θ_m^C and θ_n^V can be directly obtained according to C1 and C2 of (31) when transmission power $\{p_m^C, p_n^V\}$ is determined. And another theorem is proposed to further optimize the power allocation.

Theorem 2. *The optimal solution to* **P5** *always satisfies either* $p_m^{C*} = p_{\max}^C$ *or* $p_n^{V*} = p_{\max}^V$.

Proof. Firstly, we denote the optimal power allocation by $\{p_m^{C*}, p_n^{V*}\}$ and assume the following two expressions hold at the same time.

$$
\begin{aligned}
0 &\leq p_m^{C*} < p_{\max}^C, \\
0 &\leq p_n^{V*} < p_{\max}^V.
\end{aligned}
\tag{26}
$$

Additionally, let $p_m^{C*} < \tilde{p}_m^C = \xi p_m^{C*} < p_{\max}^C$ and $p_n^{V*} < \tilde{p}_n^V = \xi p_n^{V*} < p_{\max}^V$, where $\xi \longrightarrow 1^+$. According to (1), the following expressions hold

$$
\begin{aligned}
\gamma_m^C\left(\tilde{p}_m^C, \tilde{p}_n^V\right) &= \frac{\xi p_m^{C*} g_m^C}{N_0 B + \xi p_n^{V*} g_{n,B}} \\
&= \frac{p_m^{C*} g_m^C}{N_0 B/\xi + p_n^{V*} g_{n,B}} > \frac{p_m^{C*} g_m^C}{N_0 B + p_n^{V*} g_{n,B}} \\
&= \gamma_m^C\left(p_m^{C*}, p_n^{V*}\right),
\end{aligned}
$$

$$
\begin{aligned}
\gamma_n^V\left(\tilde{p}_m^C, \tilde{p}_n^V\right) &= \frac{\xi p_n^{V*} g_n^V}{N_0 B + \xi p_m^{C*} g_{m,n}} \\
&= \frac{p_n^{V*} g_n^V}{N_0 B/\xi + p_m^{C*} g_{m,n}} > \frac{p_n^{V*} g_n^V}{N_0 B + p_m^{C*} g_{m,n}} \\
&= \gamma_n^V\left(p_m^{C*}, p_n^{V*}\right).
\end{aligned}
\tag{27}
$$

According to (3), we have $R_m^C(\tilde{p}_m^C, \tilde{p}_n^V) > R_m^C(p_m^{C*}, p_n^{V*})$ and $R_n^V(\tilde{p}_m^C, \tilde{p}_n^V) > R_n^V(p_m^{C*}, p_n^{V*})$. From C1 and C2 in **P5**, θ_m^C is decreasing with R_m^C, and θ_n^V is decreasing with R_n^V, and therefore

$$
\begin{aligned}
&\frac{\ln 1/\varepsilon_m^C}{d_m^C}\frac{1}{\theta_m^C\left(\tilde{p}_m^C, \tilde{p}_n^V\right)} + \frac{\ln 1/\varepsilon_n^V}{d_n^V}\frac{1}{\theta_n^V\left(\tilde{p}_m^C, \tilde{p}_n^V\right)} \\
&> \frac{\ln 1/\varepsilon_m^C}{d_m^C}\frac{1}{\theta_m^C(p_m^{C*}, p_n^{V*})} + \frac{\ln 1/\varepsilon_n^V}{d_n^V}\frac{1}{\theta_n^V(p_m^{C*}, p_n^{V*})}.
\end{aligned}
\tag{28}
$$

Apparently, power allocation scheme $\{\tilde{p}_m^C, \tilde{p}_n^V\}$ guarantees a higher traffic throughput than $\{p_m^{C*}, p_n^{V*}\}$, under the delay and power constraints, which is also a contradiction to the assumption. Furthermore, for power allocation $\{p_m^C < p_{\max}^C, p_n^V < p_{\max}^V\}$, we can improve the traffic throughput under C1-C6 through increasing p_m^C and p_n^V with an equal proportion until one of them reaches the corresponding maximum. Consequently, Theorem 2 is proved.

From Theorem 2, **P5** is decomposed into two one-dimension optimization problems, i.e., optimizing p_n^V to maximize $\Gamma_{m,n}(p_{\max}^C, p_n^V)$ and optimizing p_m^C to maximize $\Gamma_{m,n}(p_m^C, p_{\max}^V)$. Thereafter, through comparing the corresponding optimal $\Gamma_{m,n}^*(p_{\max}^C, p_n^V)$ with $\Gamma_{m,n}^*(p_m^C, p_{\max}^V)$, we can choose the greater one as the optimal power allocation solution for **P5**. Note that in the procedure of solving the abovementioned two one-dimension optimization problems, the optimal θ_m^C and θ_n^V can be derived by using the bisection method on C1 and C2 in **P5**. Algorithm 1 summarizes how to ascertain the optimal power allocation for a given CUE-VUE pair. What should be highlighted is that, given power accuracy Δp, if C1 and C2 in **P5** can be solved analytically, the computation complexity of Algorithm 3 is $O(p_{\max}^C + p_{\max}^V/\Delta p)$. Otherwise, according to the bisection method, the computation complexity is $O(\log_2(\theta_{\max})p_{\max}^C + p_{\max}^V/\Delta p)$, where $\theta_{\max}$ denotes the maximum value of the QoS exponent. Typically, such a maximum value is small, and thus, the bisection method can converge rapidly.

4.2. Deep Learning-Based CUE-VUE Matching. After obtaining the optimal power allocation and the maximum traffic throughput under the delay constraints of each CUE-VUE pair, we further propose a supervised deep learning approach to solve **P3** in (10). In order to get the training label, we apply Hungarian algorithm to optimize the CUE-VUE matching, and the deep learning model should be properly trained to guarantee high accuracy.

---

1. **Initialize** statistical information of small-scale fading, locations of CUEs and VUEs, $p_m^C = 0$, $p_n^V = 0$, $p_m^{C*} = 0$, $p_n^{V*} = 0$, optimal traffic throughput $\Gamma_{m,n}^* = 0$, power accuracy Δp;
2. Fix $p_m^{C+} = p_{\max}^C$;
3. **while** $p_n^V \leq p_{\max}^V$ **do**
4.   Solve θ_m^C from C1 in **P5**.
5.   Solve θ_n^V from C2 in **P5**.
6.   Calculate $\Gamma_{m,n}$ according to (31).
7.   **if** $\Gamma_{m,n} > \Gamma_{m,n}^*$ **then**
8. $\Gamma_{m,n}^* = \Gamma_{m,n}, p_m^{C*} = p_{\max}^C, p_n^{V*} = p_n^V$;
9. $p_n^V = p_n^V + \Delta p$;
10.   **end if**
11. **end while**
12. Fix $p_n^{V+} = p_{\max}^V, p_m^C = 0$;
13. **while** $p_m^C \leq p_{\max}^C$ **do**
14.   Solve θ_m^C from C1 in **P5**.
15.   Solve θ_n^V from C2 in **P5**.
16.   Calculate $\Gamma_{m,n}$ according to (31).
17.   **if** $\Gamma_{m,n} > \Gamma_{m,n}^*$ **then**
18. $\Gamma_{m,n}^* = \Gamma_{m,n}, p_m^{C*} = p_m^C, p_n^{V*} = p_{\max}^V$;
19. $p_m^C = p_m^C + \Delta p$;
20.   **end if**
21. **end while**
22. **Output** $p_m^{C*}, p_n^{V*}, \Gamma_{m,n}^*$.

ALGORITHM 1: Optimal power allocation algorithm.

As the number of CUEs may be greater than that of VUEs, i.e., $M > N$, in this case, there are $(M - N)$ channels that are not reused by any VUE. In order to maximize the traffic throughput of those $(M - N)$ CUEs under their delay requirements, we introduce a set of virtual VUEs, denoted by $\mathcal{N}'$ and defined as

$$\mathcal{N}' = \begin{cases} \{N + 1, N + 2, \cdots, M\}, & \text{if } N < M \\ \varnothing, & \text{if } N = M \end{cases}. \tag{29}$$

Also, for a given CUE-VUE pair, if the VUE is a virtual VUE, we fix its transmission power as 0, and then, (31) can be transformed into a simple power optimization problem for a single CUE. It is easy to verify that the effective capacity of the CUE is an increasing function of transmission power p_m^C. Hence, the optimal power allocation for this CUE-VUE pair can be obtained as $\{p_{\max}^C, 0\}$, and the maximum traffic throughput can be calculated by

$$\Gamma_{m,n}^* = \frac{\ln 1/\varepsilon_m^C}{d_m^C} \frac{1}{\theta_m^C}, \tag{30}$$

where θ_m^C can be obtained by solving $\mathbb{E}[e^{-\theta_m^C R_m^C(p_{\max}^C, 0) d_m^C}] = \varepsilon_m^C$. Therefor, **P3** can be transformed into

$$\textbf{P6} \quad \max_{\{\tau_{m,n}\}} \sum_{m \in \mathcal{M}} \lambda_m^C + \sum_{n \in \mathcal{N}} \lambda_n^V = \sum_{m \in \mathcal{M}} \sum_{n \in \left\{ \mathcal{N} \cup \mathcal{N}' \right\}} \tau_{m,n} \Gamma_{m,n}^* \text{ s.t. C5}: \sum_{m \in \mathcal{M}} \tau_{m,n} \leq 1, \forall n \in \mathcal{N} \, \text{C6}: \sum_{n \in \mathcal{N}} \tau_{m,n} \leq 1, \forall m \in \mathcal{M}. \tag{31}$$

Note that **P6** is a bipartite matching problem [26] and can be effortlessly solved by the classical Hungarian algorithm. Specifically, the Hungarian algorithm is a sequential and combinatorial optimization algorithm first proposed to solve assignment problems [27]. The computation complexity of the Hungarian algorithm calculating the optimal CUE-VUE pair $\{\tau_{m,n}^*\}$ is $O(M^3)$, which is prohibitively high when the number of CUEs is large. Hence, we develop a supervised deep learning approach to solve **P6** after obtaining the offline labels from the Hungarian algorithm.

Firstly, we generate channel realizations with random positions of CUEs and VUEs. In each channel, 10^6 small-scale fading realizations are generated to solve C1 and C2 in **P5**. And then, we calculate the maximum traffic throughput $\Gamma_{m,n}^*$ sustained for each CUE-VUE pair and form a throughput matric as $\{\Gamma_{m,n}^*\}$. The total number of the training samples K^{train} is 5×10^4. In each sample, the traffic throughput for a CUE-VUE pair varies in a large value range and may be very different from those in other samples, which takes a long time to obtain the optimal training parameters. As a result, we normalize $\Gamma_{m,n}^*$ for each training sample as in

$$\Gamma_{m,n}^* \Longleftarrow \frac{\Gamma_{m,n}^* - \min\{\Gamma_{m,n}^*\}}{\max\{\Gamma_{m,n}^*\} - \min\{\Gamma_{m,n}^*\}}, m \in \mathcal{M}, n \in \mathcal{N} \cup \mathcal{N}'. \tag{32}$$

For each training sample, we can easily deduce corresponding label $\{\tau_{m,n}^*\}$. Note that there are M "1" and $M(M - 1)$ "0" elements in each $\{\tau_{m,n}^*\}$. This implies that each label is quite sparse and a latent poor training performance. Hence, we focus on the position of "1" elements for each $m \in \mathcal{M}$ and use a fixed number from 1 to M to represent it. For example, when "1" is spotted at the 5th column of the considered row, that row can then be characterized by 5. Therefore, the sparsity of the labels can be avoided.
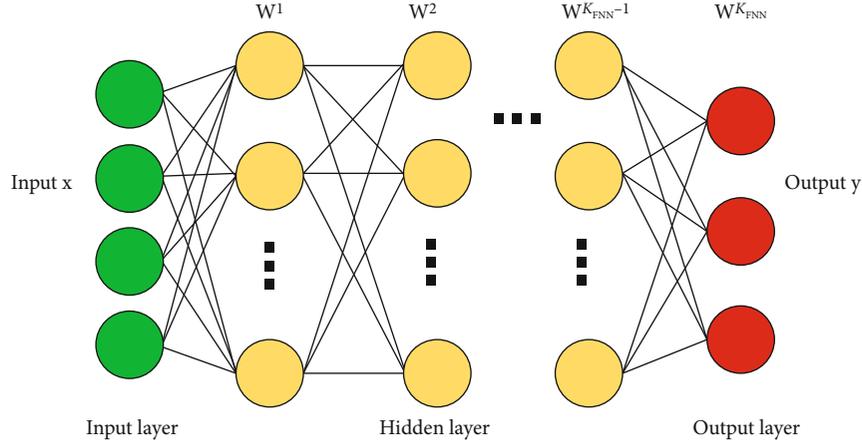
FIGURE 2: Fully connected neural network.

---

1. **Initialize** statistical information of small-scale fading, locations of CUEs and VUEs, the realization of large-scale fading;
2. Padding virtual VUEs to the considered network;
3. Obtain the maximum traffic throughput $\{\Gamma_{m,n}^*\}$ sustained for each CUE-VUE pair from Algorithm 1;
4. Calculate the optimal matching scheme $\{\tau_{m,n}^*\}$ as training labels;
5. Deal with the training samples $\{\Gamma_{m,n}^*\}$ according to (32);
6. Deal with the training labels $\{\tau_{m,n}^*\}$ to reduce the sparsity;
7. Train the FNN parameters with data samples until the loss function converges;
8. **Output** the optimal model.

ALGORITHM 2: Deep learning-based CUE-VUE matching.

In the model training stage, we construct a fully connected neural network (FNN) with $K_{\text{FNN}} = 5$ layers. In each layer, there are some neurons to be optimized and one activation function to introduce the nonlinear characteristics, as depicted in Figure 2. In detail, there are 1000 neurons in each middle (hidden) layer, and the ReLU activation function is employed. Besides, the input and output vectors of the kth layer are denoted by $\mathbf{x}^k$ and $\mathbf{y}^k$, and we have

$$
\begin{cases}
\mathbf{y}^k = \text{ReLU}\left(\mathbf{W}^k\mathbf{x}^k + \mathbf{b}^k\right), \\
\mathbf{x}^{k+1} = \mathbf{y}^k
\end{cases}
\tag{33}
$$

where $\mathbf{W}^k$ and $\mathbf{b}^k$ are the weight and bias vectors in the kth layer. In order to predict the CUE-VUE pair matching from 1 to M, the output layer has M outputs, i.e.,

$$
\mathbf{y}^{K_{\text{FNN}}} = \mathbf{W}^{K_{\text{FNN}}}\mathbf{x}^{K_{\text{FNN}}} + \mathbf{b}^{K_{\text{FNN}}}.
\tag{34}
$$

The training parameters of the FNN are initialized with Gaussian variables with zero mean and unit variance. In each epoch, a batch of training samples is randomly chosen from all training samples for parameter training. The loss function is defined as

$$
loss = \frac{1}{K^{\text{train}}} \sum_{i=1}^{K^{\text{train}}} \sum_{j=1}^{M} \left(y_{i,j} - y\wedge_{i,j}\right)^2,
\tag{35}
$$

where $y_{i,j}$ denotes the jth element of the label in the ith sample and $\widehat{y}_{i,j}$ denotes the training result. The training parameters can be optimized by the Adam algorithm to minimize the loss function [28]. The deep learning-based CUE-VUE pairing can be summarized by Algorithm 2.

4.3. Deep Transfer Learning for New Scenarios. As the FNN is trained offline, it works well only for a V2V network with the identical data distribution. However, in practical V2V networks, the traffic arrivals, channel fading, and positions of vehicles are highly dynamic and nonstationary. Since we only train the FNN with the maximum traffic throughput, hidden parameters such as the positions of vehicles and large scale channel fading are missing but have significant impacts on both the throughput matric and the optimal CUE-VUE matching. If the distribution of these parameters changes, the throughput matric will be affected, where the offline FNN will no longer perform well.

To address the mismatch problem, one potential approach is to retrain the FNN for each new scenario. However, it is hard to acquire enough training samples from a new scenario, and the traffic of both CUEs and VUEs is delay-sensitive. Hence, we resort to deep transfer learning to overcome the mismatch, especially when new training samples cannot be effectively obtained within a short time. The transfer learning framework is depicted in Figure 3. Specifically, we choose offline trained model $\{\mathbf{W}, \mathbf{b}\}$ as the initial parameter setting. In addition, fine-tuning is employed to adjust the
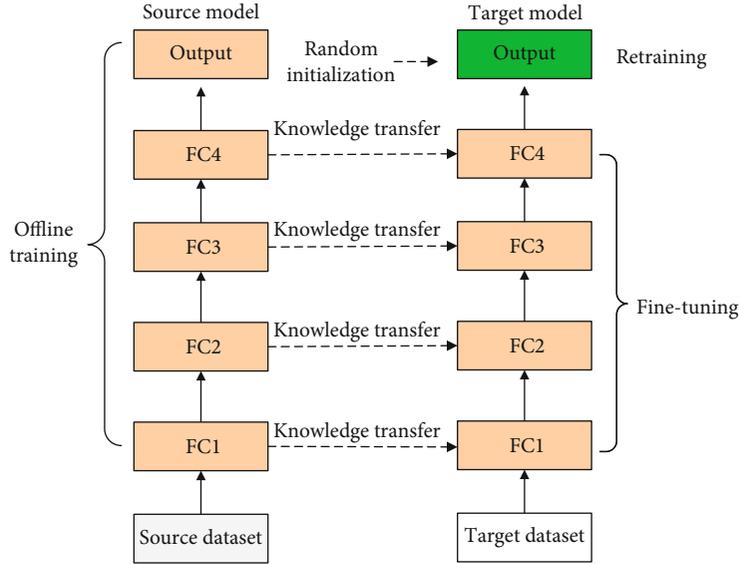
FIGURE 3: Deep transfer learning framework.

model parameters. Because the position of vehicles and the large-scale fading both depend on the velocities of vehicles, we need to fine-tune all the layers of the FNN.

5. Simulation Results

In this section, simulation results are presented and discussed. Network parameters and scenarios involved are set as follows, unless otherwise stated.

Similar with [14], we set up a simulation scenario with a 6-lane freeway (3 lanes in each direction) passing through a single cell, where the BS is located at the center of the roadside. The lane width is set to 4 m. The vehicles are randomly dropped according to a Poisson point process with density $2.5 \, s \times v$, where v (km/h) denotes the velocity of the vehicle. Then, we randomly choose CUEs and VUEs. Note that the CUE-VUE pair always includes two adjacent vehicles. In the simulation, the carrier frequency is set to 2 GHz, and the cell radius is set to 500 m. For the BS, the antenna height is set to 25 m, the antenna gain is set to 8 dBi, the receiver noise is set to 5 dB, and the distance to the freeway is set to 35 m. For each vehicle, the antenna height is set to 1.5 m, the antenna gain is set to 3 dBi, the receiver noise is set to 9 dB, and the velocity is set to 60 km/h. The numbers of CUEs and VUEs are both set as $M = N = 5$. The total bandwidth of the considered vehicular network B_{tot} is set to 10 MHz, and thus, the bandwidth for each CUE is $B = 2$ MHz. The delay requirement and the maximum tolerable violation probability for each CUE are set to $d_m^C = 1$ ms and $\varepsilon_m^C = 10^{-3}$. The delay requirement and the maximum tolerable violation probability for each VUE are set to $d_n^D = 1$ ms and $\varepsilon_n^D = 10^{-5}$. Also, the maximum transmission power of each CUE and VUE is set to $p_{max}^C = p_{max}^D = 20$ dBm. We simulate 2000 channel realizations and output the average result.

Figure 4 depicts the maximum traffic throughput sustained by a CUE-VUE pair under different vehicle velocities.
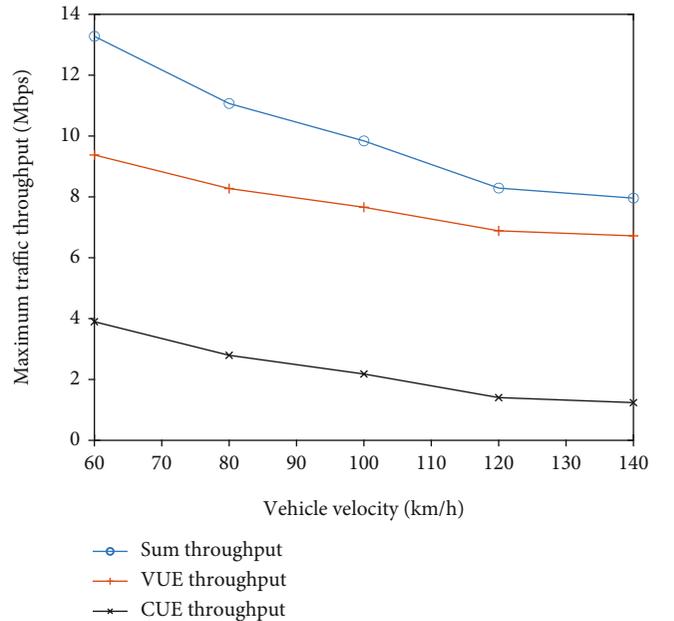
FIGURE 4: Traffic throughput for a CUE-VUE pair under given delay constraints.

It is shown that the throughput of both CUE and VUE decreases as the velocity increases. This is because the V2V distance increases with the vehicle velocity. As a result, the data transmission capability of VUEs degrades seriously due to the path loss. Hence, the VUE has to increase transmitting power to guarantee the low delay requirement, which introduces higher interchannel interference to the CUE. Though the communication distance from the CUE to the BS changes slightly, the high interference from the VUE still affects the traffic throughput of the CUE.

Figure 5 depicts how the loss function varies over training epochs of our proposed FNN tackling the channel
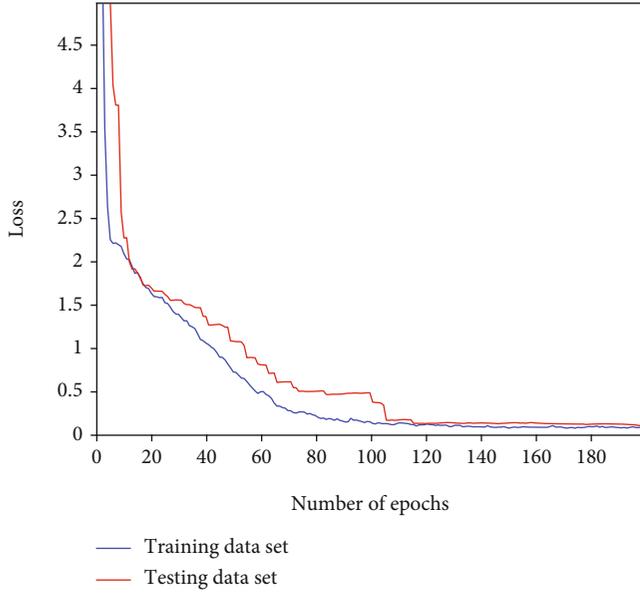
FIGURE 5: Loss function in training epochs.
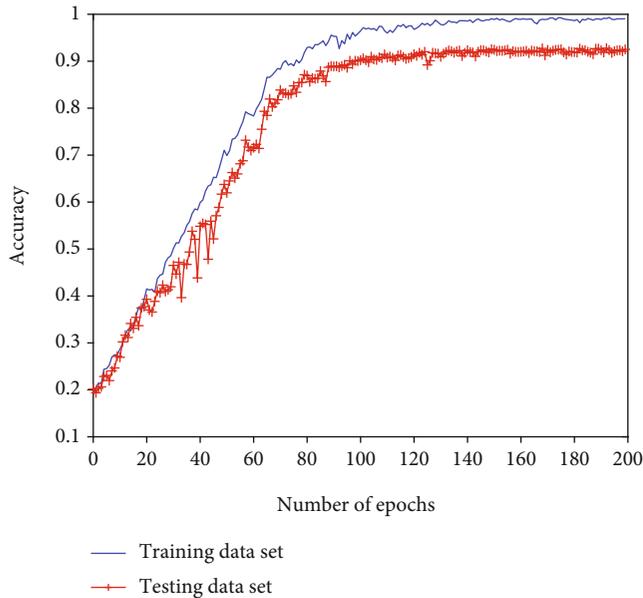
FIGURE 6: Accuracy in training epochs.

FIGURE 7: Network throughput under different schemes.
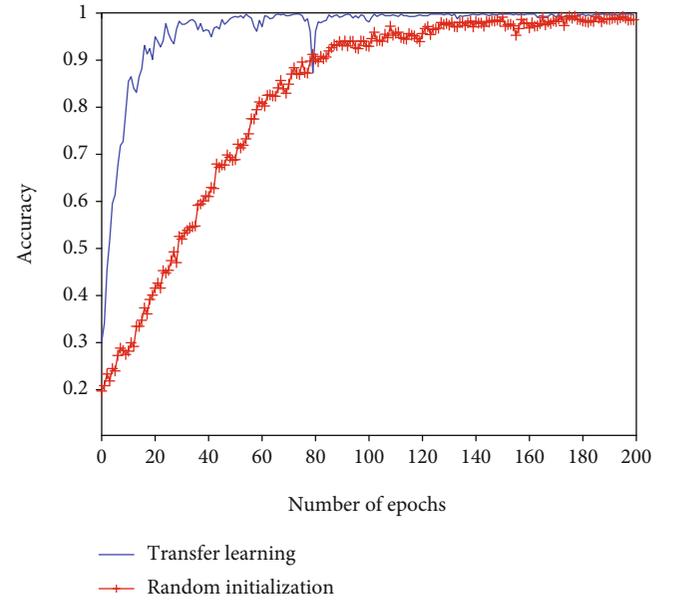
FIGURE 8: Accuracy of deep transfer learning.

assignment problem with offline data samples. It is observed that both training loss and testing loss converge within around 110 epochs. Note that we need to predict M integers whose values are from 1 to M, and consequently, the loss is low enough to guarantee the optimal channel assignment. In addition, the training accuracy is also presented in Figure 6, which verifies that our proposed FNN model is effective in solving the channel assignment problem with the testing accuracy above 90%. Hopefully, the model accuracy can be further improved if more training samples are fed into the FNN.

Figure 7 depicts the traffic throughput supported by different channel assignment schemes. Specifically, the global

optimal scheme calculated by the Hungarian algorithm has the complexity of $O(M^3)$. Our proposed FNN is a deep learning-based channel assignment scheme, where the channel capacity of the optimal scheme is leveraged to find the optimal channel assignment that maximizes the network throughput under given delay constraints. Obviously, the throughput prediction by our proposed FNN is close to that calculated by the Hungarian algorithm. Hence, we can use the trained model to predict traffic throughput rapidly while only incurring slight resource overheads. In addition, the

channel capacity provided by FNN is always higher than the corresponding traffic throughput while satisfying the delay constraint. In short, if the channel assignment aims for the channel capacity maximization, it will lead to a severe overestimation of the traffic throughput.

Figure 8 depicts the performance improvement by the deep transfer learning on the FNN training of new scenarios, where the distribution of hidden parameters varies. In the FNN training, the vehicle velocity is set to $v = 100$ km/h, whereas in the transfer learning process, we initialize the parameter as $v = 60$ km/h. It is verified that with the knowledge transfer, our proposed FNN model converges much faster than those with a random initialization. Hence, the proposed transfer learning scheme can rapidly retrain the channel assignment model for new sceneries and guarantee sufficiently high accuracy.

6. Conclusion

In this paper, a joint power allocation and spectrum sharing scheme was proposed to maximize the delay-sensitive traffic throughput for vehicular communications. Specifically, the interchannel interference model and traffic delay model were established, respectively, to derive the optimal power allocation for each CUE-VUE pair. Thereafter, a FNN was designed to deal with the channel assignment problem and speed up the allocation decision. Furthermore, a deep transfer learning scheme was proposed to leverage the offline knowledge to learn new scenarios where hidden parameters were unstable and training samples were insufficient. The effectiveness of the hybrid deep transfer learning scheme was also validated by extensive simulations. The results and analyses revealed that using the channel capacity to characterize the traffic throughput would incur a severe performance overestimation and degrade the traffic delay performance.

Data Availability

The simulation data used to support the findings of this study have not been made available because of the funding constraint.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] D. Wu, J. Yan, H. Wang, and R. Wang, "User-centric edge sharing mechanism in software-defined ultra-dense networks," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 7, pp. 1531–1541, 2020.

[2] R. Wang, H. Liu, H. Wang, Q. Yang, and D. Wu, "Distributed security architecture based on blockchain for connected health: architecture, challenges, and approaches," *IEEE Wireless Communications*, vol. 26, no. 6, pp. 30–36, 2019.

[3] K. Zheng, Q. Zheng, P. Chatzimisios, W. Xiang, and Y. Zhou, "Heterogeneous vehicular networking: a survey on architecture, challenges, and solutions," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 4, pp. 2377–2396, 2015.

[4] M. Hadded, P. Muhlethaler, A. Laouiti, R. Zagrouba, and L. A. Saidane, "TDMA-based mac protocols for vehicular ad hoc networks: a survey, qualitative analysis, and open research issues," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 4, pp. 2461–2492, 2015.

[5] D. Wu, J. Lyu, Z. Li, and R. Wang, "Mobility aware edge service migration strategy," *Journal of Communications*, vol. 41, no. 4, p. 1, 2020.

[6] J. Wang, K. Liu, K. Xiao, X. Wang, Q. Han, and V. C. S. Lee, "Delay-constrained routing via heterogeneous vehicular communications in software defined BusNet," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5957–5970, 2019.

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

[8] Z. Li, Y. Gao, P. Li, B. A. Salihu, L. Sang, and D. Yang, "Throughput analysis of an energy harvesting multichannel system under delay and energy storage constraints," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 7818–7832, 2017.

[9] D. Wu, Z. Zhang, S. Wu, J. Yang, and R. Wang, "Biologically inspired resource allocation for network slices in 5G-enabled Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9266–9279, 2019.

[10] M. Wang, H. Shan, T. H. Luan et al., "Asymptotic throughput capacity analysis of VANETs exploiting mobility diversity," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 9, pp. 4187–4202, 2015.

[11] D. M. Mughal, J. S. Kim, H. Lee, and M. Y. Chung, "Performance analysis of V2V communications: a novel scheduling assignment and data transmission scheme," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 7, pp. 7045–7056, 2019.

[12] Q. Wang, P. Fan, and K. B. Letaief, "On the joint V2I and V2V scheduling for cooperative VANETs with network coding," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 62–73, 2012.

[13] C. He, Q. Chen, C. Pan, X. Li, and F.-C. Zheng, "Resource allocation schemes based on coalition games for vehicular communications," *IEEE Communications Letters*, vol. 23, no. 12, pp. 2340–2343, 2019.

[14] L. Liang, G. Y. Li, and W. Xu, "Resource allocation for D2D-enabled vehicular communications," *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 3186–3197, 2017.

[15] J. Du, F. R. Yu, X. Chu, J. Feng, and G. Lu, "Computation offloading and resource allocation in vehicular networks based on dual-side cost minimization," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1079–1092, 2019.

[16] S. Hung, X. Zhang, A. Festag, K. C. Chen, and G. Fettweis, "Vehicle-centric network association in heterogeneous vehicle-to-vehicle networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5981–5996, 2019.

[17] M. K. Abdel-Aziz, S. Samarakoon, C. F. Liu, M. Bennis, and W. Saad, "Optimized age of information tail for ultra-reliable low-latency communications in vehicular networks," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1911–1924, 2020.

[18] Y. Chen, Y. Wang, M. Liu, J. Zhang, and L. Jiao, "Network slicing enabled resource management for service-oriented ultra-reliable and low-latency vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7847–7862, 2020.

[19] C. Guo, L. Liang, and G. Y. Li, "Resource allocation for low-latency vehicular communications: an effective capacity perspective," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 4, pp. 905–917, 2019.

[20] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: a survey," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 3, pp. 2224–2287, 2019.

[21] C. Luo, J. Ji, Q. Wang, X. Chen, and L. Pan, "Channel state information prediction for 5G wireless communications: a deep learning approach," IEEE Transactions on Network Science and Engineering, 2018.

[22] D. Wu, H. Shi, H. Wang, R. Wang, and H. Fang, "A feature-based learning system for Internet of Things applications," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1928–1937, 2019.

[23] L. Liang, H. Ye, G. Yu, and G. Y. Li, "Deep-learning-based wireless resource allocation with application to vehicular networks," *Proceedings of the IEEE*, vol. 108, no. 2, pp. 341–356, 2020.

[24] L. Liang, H. Ye, and G. Y. Li, "Spectrum sharing in vehicular networks based on multi-agent reinforcement learning," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2282–2292, 2019.

[25] Z. Li, Y. Jiang, Y. Gao, L. Sang, and D. Yang, "On buffer-constrained throughput of a wireless-powered communication system," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 2, pp. 283–297, 2019.

[26] S. Seng, C. Luo, X. Li, H. Zhang, and H. Ji, "User matching on blockchain for computation offloading in ultra-dense wireless networks," *NSE*, vol. 1–1, 2020.

[27] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[28] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *arXiv preprint arXiv*, vol. 1412, p. 6980, 2014.

Research Article

On the Capacity of Full-Duplex AF/DF Relay System with Energy Harvesting for Vehicle-to-Vehicle Communications

Ba Cao Nguyen (ID),[1] **Xuan Hung Le,**[1] **Van Duan Nguyen,**[1] **and Le The Dung** (ID)[2,3]

[1]*Telecommunications University, Khanh Hoa Province, Vietnam*
[2]*Division of Computational Physics, Institute for Computational Science, Ton Duc Thang University, Ho Chi Minh City, Vietnam*
[3]*Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam*

Correspondence should be addressed to Le The Dung; lethedung@tdtu.edu.vn

This paper studies the ergodic capacity (EC) of full-duplex (FD) amplify-and-forward (AF) and decode-and-forward (DF) relay system with energy harvesting (EH) for vehicle-to-vehicle (V2V) communications. Unlike previous works on FD-EH systems, we consider the case that both relay and destination are mobile vehicles while the source is a static base station. We mathematically derive the exact closed-form expressions of ECs of both AF and DF protocols of the considered FD-EH-V2V relay system over cascade (double) Rayleigh fading. Our numerical results show that the ECs in the case of the V2V communication system are reduced compared to those in the case of stationary nodes. Also, with a specific value of residual self-interference (RSI), the ECs of the considered FD-EH-V2V relay system can be higher or lower than those of half-duplex-(HD-) EH-V2V system, depending on the average transmission power of the source. Furthermore, when the transmission power of the source and RSI are fixed, the ECs of the considered system can achieve peak values by using optimal EH time duration. On the other hand, the ECs of both AF and DF protocols reach the capacity floors in the high signal-to-noise ratio (SNR) regime due to the RSI impact. Also, the effect of RSI dominates the impact of cascade Rayleigh fading in the high SNR regime. Finally, we validate our analysis approach through Monte-Carlo simulations.

1. Introduction

In the age of Industry 4.0, various new techniques have been fast developed to satisfy the requirements of capacity and energy consumption of the future wireless networks such as the fifth-generation (5G) and beyond (B5G) [1–3]. In addition, the emergence of trillions of Internet-of-Things (IoT) devices in the world requires devices to consume less energy and transmit data at a higher rate [3, 4]. Therefore, energy harvesting (EH) from radio frequency (RF) signals has been used to deal with these issues [4, 5]. Together with the traditional energy grid, EH can help to fulfill the energy requirements for different elements of 5G networks, including sensors in the IoT, mobile devices, heterogeneous networks (HetNets), relays in device-to-device (D2D) systems, and computing servers [3]. Additionally, the emergence of advanced materials and hardware designs helps realize the EH circuits for small portable consumer electronic devices

in the IoT. Furthermore, RF signals can be transmitted over the air all the time. Thus, EH from RF signals can provide stable energy for wireless devices that consume low power such as IoTs, sensors, and the remote area communication used in 5G and B5G systems [1, 3].

Meanwhile, full-duplex (FD) is a promising technique for achieving high spectral efficiency in wireless systems thanks to its capability to allocate the transmitted and received radio signals of a communication node on the same frequency and in the same time slot. Ideally, FD transmission increases the spectral efficiency twice compared to traditional half-duplex (HD) transmission. As a promising technology for next-generation (5G and B5G) wireless networks, FD wireless not only has the potential to double the spectrum efficiency in the physical layer but can also enhance the performance of wireless systems such as reducing feedback delay, end-to-end delay, and congestion, improving the network secrecy and efficiency and increasing the throughput and spectrum

usage flexibility [6, 7]. Overall, it is envisaged that FD communication technology can be adopted in the near future in a number of scenarios and applications, such as throughput enhancement in the sub-6 GHz band, supporting ultralow delay communication, small and dense cells [6, 8–10].

In the literature, various works have combined EH and FD techniques in a wireless communication system to solve the battery and spectrum efficiency issues [8, 10, 11]. The mathematical analysis and the experimental measurements have been applied for investigating the performance of FD-EH systems. Specifically, the mathematical analysis is used to derive expressions of outage probability (OP), symbol error rate (SER), bit error rate (BER), and ergodic capacity (EC) of the FD-EH systems [8, 10–13]. Based on these obtained mathematical expressions, the system behavior is analyzed under the effects of different parameters such as the residual self-interference (RSI) induced by FD transmission mode, the time switching ratio, and the channel conditions. It is found that optimal power allocation for FD transmission mode and optimal time switching ratio for EH can improve the system performance significantly [14–16]. Furthermore, the EC of the FD-EH system is generally higher than that of the HD-EH system for specific RSI values. Experimental measurements have also been widely used to evaluate the algorithms and solutions used to improve the performance of FD-EH systems [12, 17–19]. By applying the optimization problem in a nonconvex form, the FD-EH system's system power is minimized and better than that of the HD-EH system [19], and the sum rate and energy efficiency are maximized [18]. As a result, most of the works about FD-EH systems have analyzed RSI and other system parameters' impacts and then proposed solutions to reduce OP and SER and enhance EC and energy efficiency. Furthermore, exploiting FD-EH systems in different scenarios such as cognitive radio (CR), spatial modulation (SM), and cooperative communication has been widely carried out.

Recently, both FD and EH techniques are deployed in vehicle-to-vehicle (V2V) communication systems because of their various advantages in V2V environments. Thanks to FD and EH techniques, the V2V communication systems can reduce the delay time for signal transmission between vehicles and the power supply limitation problem [20, 21]. Hence, the FD-EH-V2V relay system can be applied for the intelligent transportation systems (ITS) and the road safety applications [8, 10, 20]. Specifically, the onboard unit was proposed in [20] to supply the energy for FD-V2V communications from the vehicle engine. Although this scheme could solve the energy issues, applying EH for V2V communication systems still becomes an inevitable trend thanks to many advantages of the EH technique. In recent reports, the OP, SER, and throughput of the FD-EH-V2V relay systems were obtained to investigate the system performance and evaluate the effects of several system parameters such as RSI, time switching ratio, and channel characteristics [8, 10]. In particular, papers [8, 10] derived the OP and SER expressions of FD-EH-V2V relay systems with AF and DF relaying protocols, respectively. However, the EC expressions of these FD-EH-V2V relay systems were not obtained. Meanwhile, we always want to get the lowest OP/SER and the highest EC

for wireless communication systems. Therefore, investigating OP/SER and ignoring EC when analyzing the performance of wireless communication systems may result in inaccurate conclusions on the system behaviors. Specifically, experiments and measurements indicate that the cascade (double) Rayleigh fading distribution best describes the channels between vehicle nodes [21, 22], while the traditional channels such as Nakagami, Rician, and Rayleigh cannot fully model the V2V communication channel. It is shown that, under the effects of RSI and cascade Rayleigh fading channel, the OP and SER of the FD-EH-V2V relay systems reach the error floor faster in the high SNR regime. Additionally, the cascade Rayleigh fading channel makes the derivation of closed-form expressions more difficult than Nakagami, Rician, and Rayleigh channels [8, 10, 22], leading to a lack of mathematical analysis of FD-EH-V2V relay systems over cascade Rayleigh fading channels.

As in the above discussions, the main benefit of FD transmission is high capacity compared with HD one. However, mathematical analysis of the EC of FD-EH-V2V relay systems under the effect of cascade Rayleigh fading has not been investigated yet. Meanwhile, EH and FD techniques in V2V communication systems are inevitable because these techniques can solve various issues in traditional V2V systems. Mainly, EH is an effective method for power supply in the case that the wireline power supply may not be deployed for moving vehicles. At the same time, the FD can improve the performance of safety applications for V2V systems. Therefore, it is required to mathematically analyze the ECs of the FD-EH-V2V relay system for both AF and DF protocols to evaluate the system behavior. This observation motivates us to perform a mathematical analysis of the ECs of the FD-EH-V2V relay system with AF/DF protocols over cascade Rayleigh fading channels. In our paper, we focused on analyzing the impacts of the cascade Rayleigh fading channel, RSI, and time switching ratio on the ECs of the FD-EH-V2V relay system with the AF/DF protocol by deriving the exact closed-form expressions of these attributes and comparing with those in the case of Rayleigh fading channel, perfect SIC, and HD system. So far, this is the first work deriving the EC expressions of the FD-EH-V2V relay system over cascade Rayleigh fading channels. It is noted that exploiting EH from RF signals provides a stable energy supply for low-power consumption networks such as IoTs, wireless sensor networks, extremely remote area communications used in 5G and B5G systems [4, 19], and vehicular networks [23]. Meanwhile, FD transmission mode can be exploited in various scenarios to support a set of safety applications in V2V systems because FD devices can transmit signals and sense the environment simultaneously, thus reducing the end-to-end delay of the systems [6]. Consequently, the combination of EH and FD in a V2V system can achieve many advantages such as solving the battery and spectrum efficiency issues. The main contributions of the paper are shortened as follows:

(i) We investigate an FD-EH-V2V relay system where the source is located at a fixed location while relay and destination move on the road. Besides, relay
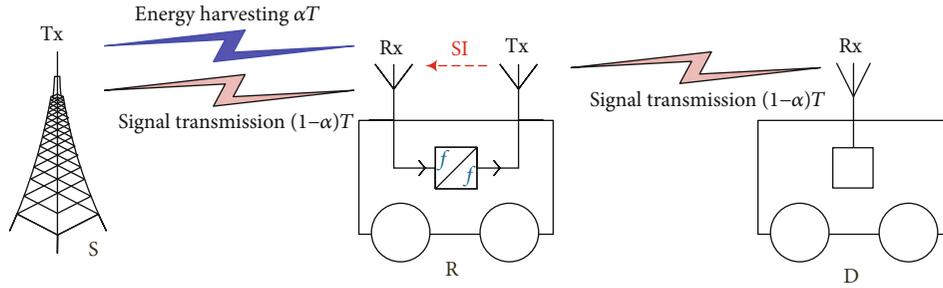
FIGURE 1: Block diagram of the considered FD-EH-V2V relay system.

harvests energy from the source while moving and exchanging signals. Furthermore, we consider both AF and DF protocols at the relay

(ii) We mathematically derive the exact closed-form expressions of the ECs of the considered FD-EH-V2V relay system for both AF and DF protocols under the influences of RSI and cascade Rayleigh fading, then validate these derived expressions by Monte-Carlo simulations. We also observe that cascade Rayleigh fading results in more difficulties deriving closed-form expression than traditional channels such as Rayleigh and Nakagami

(iii) We investigate the performance of the FD-EH-V2V relay system in different scenarios. Numerical results reveal that the cascade Rayleigh fading hurts the ECs of the considered system compared with the Rayleigh fading. Furthermore, the EC of the DF protocol is higher than that of AF one, and the ECs of the considered system are higher than those of the HD-EH-V2V relay system for certain RSI and SNR. On the other hand, when RSI and the average transmission power of the source are fixed, a suitable time switching ratio can be chosen to maximize the ECs of the considered system

The rest of the paper is organized as follows. Section 2 describes the system model of the considered FD-EH-V2V relay system with signal processing for both AF and DF protocols. Section 3 mathematically derives the EC expressions of the considered system. Section 4 provides numerical results and discussions. Finally, Section 5 draws some conclusions.

2. System Model

The considered FD-EH-V2V relay system consists of a static station (S) and two moving vehicles (R and D), as described in Figure 1. S and D are equipped with one antenna while R is equipped with two antennas, one for receiving and another for transmitting signals. S and D use traditional HD communication, while R employs FD communication. When R is moving on the road or restricted area, it is not easy to supply power to it; thus, R needs to harvest energy from RF signals for its operation. In particular, R is equipped with a suitable circuit that can harvest energy from the RF signal transmitted from S and then uses all the harvested energy

for signal transmission. In practice, R can use a shared antenna for both transmitting and receiving signals; however, exploiting a separate antenna can suppress self-interference (SI) better, especially in the propagation domain. Specifically, when a shared antenna is exploited at R, the isolation between its output and input may not be sufficient to satisfy the SIC requirements [24, 25]. Additionally, it is too difficult to apply the spatial suppression at R with a shared antenna. Meanwhile, various methods to suppress SI power with a separate antenna such as the usage of lossy materials, directional antennas, and spatial suppression can be easily deployed [25, 26]. Hence, R keeps antennas separately on the vehicle rooftop to perform this task at a far enough distance to make passive isolation remarkably efficient. In this paper, we also assume that R uses two separate antennas.

As can be seen from Figure 1, besides the signal processing as traditional AF/DF relay, the relay in the considered system has to deal with EH and SIC processes. In addition, various algorithms to suppress the SI are also exploited at the EH-FD relay. These operations lead to an increase in the architecture complexity and signal processing delay at the relay. Therefore, the considered FD-EH-V2V relay system is more complex than the traditional HD relay system without EH [27, 28]. On the other hand, the transmission delay was characterized in several papers such as [29–31]. Since the authors of [29–31] analyzed a two-hop HD relay system, the data transmitted from source to destination via relay needs two time slots, leading to a significant increase in the signal transmission delay. In contrast, there is only one time slot for transmitting data from source to destination in our work because of FD transmission mode. Thus, the delay is greatly reduced. Furthermore, we focus on the mathematical analysis of the ergodic capacity (EC) of the V2V relay system with two new techniques (EH and FD) and both amplify-and-forward (AF) and decode-and-forward (DF) relaying protocols by deriving the closed-form EC expressions. Then, we compare the ECs of the considered system with ECs of the traditional HD system or the system over Rayleigh fading to exhaustively investigate the impacts of cascade Rayleigh fading channel, RSI, and the benefits of FD transmission mode. In the near future, we will extend this work by analyzing the scenario where the delay is considered in FD-V2V relay systems.

The operation of the considered FD-EH-V2V relay system is illustrated in Figure 2. It consists of two stages. In the first stage, R harvests the energy from the RF signal transmitted from S in the time duration of αT, where $0 \leq \alpha \leq 1$ and
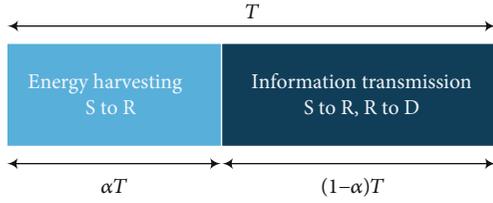
FIGURE 2: The operation of the considered FD-EH-V2V relay system with the TS protocol.

T are, respectively, the time switching ratio and the transmission block. In the second stage, S and R transmit signals to R and D, respectively, in the time duration of $(1 - \alpha)T$. Due to simultaneously transmitting and receiving at the same time and same frequency band, the received signal at R is distorted by the SI. Thus, all SIC techniques should be applied at R to mitigate this issue.

Also, the measurements and experiments in the literature such as [32, 33] indicated that the well-known channels such as Nakagami-m, Rayleigh, and Rician do not fit the R–D communication channel. It is because both R and D are moving vehicles. Instead, the cascade Rayleigh fading well characterizes the R–D communication channel in both fields of measurement and theoretical analysis [22, 32, 33]. Meanwhile, since S is static, the S–R communication link follows Rayleigh fading [32, 34, 35].

In the EH time duration αT, the harvested energy at R (denoted by E_h) is expressed as

$$E_h = \eta \alpha T P_S |h_{SR}|^2, \tag{1}$$

where $0 \leq \eta \leq 1$ is the energy conversion efficiency; P_S is the average transmission power of S; h_{SR} is the fading coefficient of the S–R channel.

Then, R uses all the harvested energy for signal transmission. Consequently, the transmission power of R is given by

$$P_R = \frac{\eta \alpha T P_S |h_{SR}|^2}{(1 - \alpha)T} = \frac{\eta \alpha P_S |h_{SR}|^2}{1 - \alpha}. \tag{2}$$

In the communication time, $(1 - \alpha)T$, S and R, respectively, transmit signals to R and D. Simultaneously transmitting and receiving signals of R causes SI from its transmitting antenna to its receiving antenna. The received signals at R is thus presented as

$$y_R = h_{SR} \sqrt{P_S} x_S + \tilde{h}_{RR} \sqrt{P_R} x_R + z_R, \tag{3}$$

where x_S and x_R are the transmitted signals at S and R, respectively; P_R is the transmission power of R given in (2); $\tilde{h}_{RR}$ is the fading coefficient of SI channel; z_R is the Gaussian noise at R with zero mean and variance of σ_R^2, i.e., $z_R \sim \mathcal{CN}(0, \sigma_R^2)$.

Since the distance between transmission and reception antennas of R is very small, especially for personal devices, R must apply all SIC techniques to suppress SI power before

performing further processes such as decoding and forwarding. The average SI power before SIC from (3) is calculated as

$$\mathbb{E}\left\{ \left| \tilde{h}_{RR} \right|^2 P_R \right\} = \frac{\eta \alpha P_S}{1 - \alpha} \mathbb{E}\left\{ \left| \tilde{h}_{RR} \right|^2 |h_{SR}|^2 \right\}, \tag{4}$$

where $\mathbb{E}\{.\}$ denotes the expectation operator.

First, in the propagation domain, R uses antenna directionality, isolation, and cross-polarization to reduce the SI power. Then, through circuits and algorithms in both analog and digital domains, the SI is further suppressed. In particular, since R knows its transmitted signal, it can subtract this signal from the received signals, especially by using digital cancellation. However, due to imperfect circuit hardware and SI channel estimation, the SI cannot be obliterated. Therefore, residual self-interference (RSI) still exists in the received signals of R. According to the measurements and experiments on the RSI of FD devices, the RSI after using all SIC techniques (denoted by I_R) follows complex Gaussian distribution with zero mean and variance of σ_{RSI}^2 [15, 16, 36–38], where σ_{RSI}^2 is expressed as

$$\sigma_{RSI}^2 = \frac{k \eta \alpha P_S}{1 - \alpha}, \tag{5}$$

where k is the RSI level at FD relay.

Now, the received signals at R becomes

$$y_R = h_{SR} \sqrt{P_S} x_S + I_R + z_R. \tag{6}$$

2.1. Amplify-and-Forward (AF). When R uses the AF protocol, it amplifies the received signals after SIC and forwards them to D. The transmitted signal at R is expressed as

$$x_R = G y_R, \tag{7}$$

where G is the relaying gain chosen so that $\mathbb{E}\{|x_R|^2\} = 1$, i.e.,

$$G = \sqrt{\frac{1}{|h_{SR}|^2 P_S + \sigma_{RSI}^2 + \sigma_R^2}}. \tag{8}$$

Then, R forwards signals to D. The received signal at D is presented as

$$y_D = h_{RD} \sqrt{P_R} x_R + z_D, \tag{9}$$

where h_{RD} is the fading coefficient of the R–D channel; z_D is the Gaussian noise at D, i.e., $z_D \sim \mathcal{CN}(0, \sigma_D^2)$.

Replacing (6), (7), and (8) into (9), we have

$$y_D = h_{RD} \sqrt{P_R} G \left(h_{SR} \sqrt{P_S} x_S + I_R + z_R \right) + z_D$$

$$= h_{SR} h_{RD} \sqrt{P_S} \sqrt{P_R} G x_S + h_{RD} \sqrt{P_R} G (I_R + z_R) + z_D. \tag{10}$$

From (10), the end-to-end signal-to-interference-plus-noise ratio (SINR) of the considered FD-EH-V2V relay system with the AF protocol (denoted by γ_{AF}) is computed as

$$\gamma_{AF} = \frac{|h_{SR}|^2 |h_{RD}|^2 P_S P_R G^2}{|h_{RD}|^2 P_R G^2 (\sigma_{RSI}^2 + \sigma_R^2) + \sigma_D^2}. \quad (11)$$

Substituting (2) and (8) into (11), γ_{AF} now becomes

$$\gamma_{AF} = \frac{|h_{SR}|^4 |h_{RD}|^2 \eta \alpha P_S^2}{|h_{SR}|^2 |h_{RD}|^2 \eta \alpha P_S (\sigma_{RSI}^2 + \sigma_R^2) + \sigma_D^2 (1-\alpha)(|h_{SR}|^2 P_S + \sigma_{RSI}^2 + \sigma_R^2)}. \quad (12)$$

2.2. Decode-and-Forward (DF). When R uses the DF protocol, it decodes the received signals after SIC, recodes the intended signals, and forwards them to D. The received signals at D is given in (9). Based on the received signals at R and D given as (6) and (9), the SINRs for decoding signals at R (denoted by γ_R) and D (denoted by γ_D) are, respectively, calculated as

$$\gamma_R = \frac{|h_{SR}|^2 P_S}{\sigma_{RSI}^2 + \sigma_R^2}, \quad (13)$$

$$\gamma_D = \frac{|h_{RD}|^2 P_R}{\sigma_D^2} = \frac{|h_{SR}|^2 |h_{RD}|^2 \eta \alpha P_S}{\sigma_D^2 (1-\alpha)}. \quad (14)$$

For the DF protocol, the end-to-end SINR (denoted by γ_{DF}) of the considered FD-EH-V2V relay system is given by

$$\gamma_{DF} = \min(\gamma_R, \gamma_D). \quad (15)$$

3. Performance Analysis

In this section, the ECs of the considered FD-EH-V2V relay system for both AF and DF protocols are obtained. Generally, the EC is computed as

$$\mathscr{C} = \mathbb{E}\{\log_2(1 + \gamma_{e2e})\} = \int_0^\infty \log_2(1 + \gamma_{e2e}) f_{\gamma_{e2e}}(\gamma) d\gamma, \quad (16)$$

where γ_{e2e} is the end-to-end SINR of the considered system given in (12) and (15) for AF and DF protocols, respectively; $f_{\gamma_{e2e}}(\gamma)$ is the probability density function (PDF) of γ_{e2e}.

After some mathematical manipulations, (16) becomes

$$\mathscr{C} = \frac{1}{\ln 2} \int_0^\infty \frac{1 - F_{\gamma_{e2e}}(x)}{1 + x} dx, \quad (17)$$

where $F_{\gamma_{e2e}}(x)$ is the cumulative distribution function (CDF) of γ_{e2e}.

From (17), the ECs of the considered FD-EH-V2V relay system for both AF and DF protocols are derived in the following theorem.

Theorem 1. *Under the impacts of RSI and the cascade Rayleigh fading channels, the ECs of the considered FD-EH-V2V*

relay system using AF (denoted by $\mathscr{C}_{AF}$) and DF (denoted by $\mathscr{C}_{DF}$) protocols are expressed as

$$\mathscr{C}_{AF} = \frac{\pi^2}{4MN \ln 2} \sum_{m=1}^M \sum_{n=1}^N \frac{\sqrt{(1 - \phi_m^2)(1 - \phi_n^2)}}{\Psi - \ln \nu}$$
$$\cdot \sqrt{\frac{\Phi \ln \nu (\ln u + \ln \nu - \Psi)}{\Psi \ln u (\ln u + \ln \nu)}} K_1$$
$$\cdot \left(\sqrt{\frac{\Phi \ln \nu (\ln u + \ln \nu - \Psi)}{\Psi \ln u (\ln u + \ln \nu)}} \right), \quad (18)$$

$$\mathscr{C}_{DF} = \frac{\pi^2}{4MN \ln 2} \sum_{m=1}^M \sum_{n=1}^N \frac{\sqrt{(1 - \phi_m^2)(1 - \phi_n^2)}}{\Psi - \ln \nu}$$
$$\cdot \sqrt{\frac{\Phi \ln \nu}{\Psi \ln u (\ln u + \ln \nu)}} K_1 \left(\sqrt{\frac{\Phi \ln \nu}{\Psi \ln u (\ln u + \ln \nu)}} \right), \quad (19)$$

where $\Psi = (\sigma_{RSI}^2 + \sigma_R^2)/P_S$; $\Phi = (4\sigma_D^2(1-\alpha))/\eta\alpha P_S$; $\phi_m = \cos(((2m-1)\pi)/2M)$; $u = 1/2(\phi_m + 1)$; $\phi_n = \cos(((2n-1)\pi)/2N)$; $\nu = 1/2(\phi_n + 1)$; *M and N are complexity-accuracy trade-off parameters* [39]; $K_1(.)$ *is the first order modified Bessel function of the second kind* [40].

Proof. To obtain ECs for both cases of AF and DF protocols, we have to derive the CDFs of γ_{AF} and γ_{DF} first and then replace them into (17).

For the AF protocol, the CDF of γ_{AF} (denoted by $F_{AF}(x)$) is computed as

$$F_{AF}(x) = \Pr\{\gamma_{AF} < x\}$$
$$= \Pr\left\{ \frac{|h_{SR}|^4 |h_{RD}|^2 \eta \alpha P_S^2}{|h_{SR}|^2 |h_{RD}|^2 \eta \alpha P_S (\sigma_{RSI}^2 + \sigma_R^2) + \sigma_D^2 (1-\alpha)(|h_{SR}|^2 P_S + \sigma_{RSI}^2 + \sigma_R^2)} < x \right\}$$
$$= \Pr\{|h_{SR}|^2 |h_{RD}|^2 \eta \alpha P_S [|h_{SR}|^2 P_S - x(\sigma_{RSI}^2 + \sigma_R^2)] < x\sigma_D^2 (1-\alpha)$$
$$\cdot (|h_{SR}|^2 P_S + \sigma_{RSI}^2 + \sigma_R^2)\}. \quad (20)$$

By changing variable, i.e., $|h_{SR}|^2 = y + (x(\sigma_{RSI}^2 + \sigma_R^2)/P_S)$, (20) becomes

$$F_{AF}(x) = \Pr\left\{ |h_{RD}|^2 < \frac{\sigma_D^2 (1-\alpha)[(\sigma_{RSI}^2 + \sigma_R^2)(x^2 + x) + P_S yx]}{\eta \alpha P_S y [P_S y + (\sigma_{RSI}^2 + \sigma_R^2)x]} \right\}$$
$$= 1 - \int_0^\infty \left[1 - F_{|h_{RD}|^2}\left(\frac{\sigma_D^2 (1-\alpha)[(\sigma_{RSI}^2 + \sigma_R^2)(x^2 + x) + P_S yx]}{\eta \alpha P_S y [P_S y + (\sigma_{RSI}^2 + \sigma_R^2)x]} \right) \right]$$
$$\cdot f_{|h_{SR}|^2}\left(y + \frac{x(\sigma_{RSI}^2 + \sigma_R^2)}{P_S} \right) dy. \quad (21)$$

To calculate (21), we begin with the CDF and PDF of the instantaneous channel gain, $|h_{SR}|^2$, that follows Rayleigh distribution, i.e.,

$$F_{|h_{SR}|^2}(x) = 1 - \exp\left(-\frac{x}{\Omega} \right), x \geq 0, \quad (22)$$

$$f_{|h_{\mathrm{SR}}|^2}(x) = \frac{1}{\Omega} \exp\left(-\frac{x}{\Omega}\right), x \geq 0, \qquad (23)$$

where $\Omega = \mathbb{E}\{|h|^2\}$ is the average channel gain.

To shorten the derived expressions, all the average channel gains are normalized, i.e., $\Omega = 1$. Therefore, (22) and (23) become

$$F_{|h_{\mathrm{SR}}|^2}(x) = 1 - \exp(-x), x \geq 0, \qquad (24)$$

$$f_{|h_{\mathrm{SR}}|^2}(x) = \exp(-x), x \geq 0. \qquad (25)$$

In addition, since the R–D channel is influenced by cascade Rayleigh fading, the CDF and PDF of $|h_{\mathrm{RD}}|^2$ are, respectively, expressed as [22, 34, 41]

$$F_{|h_{\mathrm{RD}}|^2}(x) = 1 - \sqrt{4x} K_1\left(\sqrt{4x}\right), \qquad (26)$$

$$f_{|h_{\mathrm{RD}}|^2}(x) = 2K_0\left(\sqrt{4x}\right), \qquad (27)$$

where $K_0(.)$ is the zero-order modified Bessel function of the second kind [40].

It should be better to know that the movements of both transmitter and receiver make the signal's amplitude and phase fluctuate and cause the Doppler shifts. As a result, the cascade Rayleigh fading channels are widely used in V2V communication systems such as in [22, 35, 42] to best describe the characteristics of V2V channels. Through mathematical analysis, previous works such as [22, 33, 35, 43] take these characteristics into account when deriving the CDF and PDF of the V2V communication channels. In other words, the CDF in (26) and the PDF in (27) implicitly reflect the movements of both transmitter and receiver in the V2V system.

Applying (26) and (25), (21) becomes

$$F_{\mathrm{AF}}(x) = 1 - \int_0^\infty \sqrt{\frac{4\sigma_{\mathrm{D}}^2(1-\alpha)\left[\left(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2\right)(x^2+x) + P_{\mathrm{S}}yx\right]}{\eta\alpha P_{\mathrm{S}}y\left[P_{\mathrm{S}}y + \left(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2\right)x\right]}}$$
$$\times K_1\left(\sqrt{\frac{4\sigma_{\mathrm{D}}^2(1-\alpha)\left[\left(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2\right)(x^2+x) + P_{\mathrm{S}}yx\right]}{\eta\alpha P_{\mathrm{S}}y\left[P_{\mathrm{S}}y + \left(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2\right)x\right]}}\right)$$
$$\cdot \exp\left(-y - \frac{x\left(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2\right)}{P_{\mathrm{S}}}\right) dy. \qquad (28)$$

Let $z = \exp(-y)$ be a new variable, we can rewrite (28) as

$$F_{\mathrm{AF}}(x) = 1 - \exp\left(-\frac{x\left(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2\right)}{P_{\mathrm{S}}}\right)\int_0^1$$
$$\cdot \sqrt{\frac{4\sigma_{\mathrm{D}}^2(1-\alpha)\left[\left(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2\right)(x^2+x) + P_{\mathrm{S}}x\ln(1/z)\right]}{\eta\alpha P_{\mathrm{S}}\left[P_{\mathrm{S}}\ln(1/z) + \left(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2\right)x\right]\ln(1/z)}}$$

$$\times K_1\left(\sqrt{\frac{4\sigma_{\mathrm{D}}^2(1-\alpha)\left[\left(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2\right)(x^2+x) + P_{\mathrm{S}}x\ln(1/z)\right]}{\eta\alpha P_{\mathrm{S}}\left[P_{\mathrm{S}}\ln(1/z) + \left(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2\right)x\right]\ln(1/z)}}\right)$$
$$\cdot dz = 1 - \exp(-\Psi x)\int_0^1 \sqrt{\frac{\Phi[\Psi(x^2+x) + x\ln(1/z)]}{(\ln(1/z) + \Psi x)\ln(1/z)}} K_1$$
$$\cdot \left(\sqrt{\frac{\Phi[\Psi(x^2+x) + x\ln(1/z)]}{(\ln(1/z) + \Psi x)\ln(1/z)}}\right) dz. \qquad (29)$$

Using [39] (Eq. (25.4.30)), (29) is computed as

$$\int_0^1 \sqrt{\frac{\Phi[\Psi(x^2+x) + x\ln(1/z)]}{(\ln(1/z) + \Psi x)\ln(1/z)}} K_1\left(\sqrt{\frac{\Phi[\Psi(x^2+x) + x\ln(1/z)]}{(\ln(1/z) + \Psi x)\ln(1/z)}}\right) dz$$
$$= \frac{\pi}{2M}\sum_{m=1}^M \sqrt{1-\phi_m^2}\sqrt{\frac{\Phi[\Psi(x^2+x) - x\ln u]}{-\ln u(-\ln u + \Psi x)}} K_1$$
$$\cdot \left(\sqrt{\frac{\Phi[\Psi(x^2+x) - x\ln u]}{-\ln u(-\ln u + \Psi x)}}\right), \qquad (30)$$

where M, ϕ_m, and u were defined in the theorem.

Then, $F_{\mathrm{AF}}(x)$ is expressed as

$$F_{\mathrm{AF}}(x) = 1 - \exp(-\Psi x)\frac{\pi}{2M}\sum_{m=1}^M \sqrt{1-\phi_m^2}\sqrt{\frac{\Phi[\Psi(x^2+x) - x\ln u]}{-\ln u(-\ln u + \Psi x)}} K_1$$
$$\cdot \left(\sqrt{\frac{\Phi[\Psi(x^2+x) - x\ln u]}{-\ln u(-\ln u + \Psi x)}}\right). \qquad (31)$$

Substituting $F_{\mathrm{AF}}(x)$ in (31) into (17) for calculating the EC corresponding to the AF protocol, we have

$$\mathscr{C}_{\mathrm{AF}} = \frac{1}{\ln 2}\int_0^\infty \frac{1}{1+x}\exp(-\Psi x)\frac{\pi}{2M}\sum_{m=1}^M \sqrt{1-\phi_m^2}$$
$$\cdot \sqrt{\frac{\Phi[\Psi(x^2+x) - x\ln u]}{-\ln u(-\ln u + \Psi x)}} K_1\left(\sqrt{\frac{\Phi[\Psi(x^2+x) - x\ln u]}{-\ln u(-\ln u + \Psi x)}}\right) dx. \qquad (32)$$

Let $t = \exp(-\Psi x)$ be a new variable, (32) can be rewritten as

$$\mathscr{C}_{\mathrm{AF}} = \frac{\pi}{2M\ln 2}\sum_{m=1}^M \sqrt{1-\phi_m^2}\int_0^1 \frac{1}{\Psi + \ln(1/t)}$$
$$\cdot \sqrt{\frac{\Phi\ln(1/t)(\ln(1/t) - \ln u + \Psi)}{-\Psi\ln u(\ln(1/t) - \ln u)}} K_1 \qquad (33)$$
$$\cdot \left(\sqrt{\frac{\Phi[\Psi(x^2+x) - x\ln u]}{-\Psi\ln u(\ln(1/t) - \ln u)}}\right) dt.$$

Applying [39] (Eq. (25.4.30)), the integral in (33) is calculated as

$$\int_0^1 \frac{1}{\Psi + \ln(1/t)} \sqrt{\frac{\Phi \ln(1/t)(\ln(1/t) - \ln u + \Psi)}{-\Psi \ln u(\ln(1/t) - \ln u)}} K_1$$
$$\cdot \left(\sqrt{\frac{\Phi \ln(1/t)(\ln(1/t) - \ln u + \Psi)}{-\Psi \ln u(\ln(1/t) - \ln u)}} \right) dt$$
$$= \frac{\pi}{2N} \sum_{n=1}^N \frac{\sqrt{1 - \phi_n^2}}{\Psi - \ln v} \sqrt{\frac{\Phi \ln v(\ln u + \ln u - \Psi)}{\Psi \ln u(\ln u + \ln v)}} K_1$$
$$\cdot \left(\sqrt{\frac{\Phi \ln v(\ln u + \ln u - \Psi)}{\Psi \ln u(\ln u + \ln v)}} \right), \tag{34}$$

where N, ϕ_n, and v were defined in the theorem.

Replacing (34) into (33), we obtain the EC of the considered system with the AF protocol as (18).

For the DF protocol, the CDF of γ_{DF} (denoted by $F_{\mathrm{DF}}(x)$) is computed as

$$F_{\mathrm{DF}}(x) = \Pr\{\gamma_{\mathrm{DF}} < x\} = \Pr\{\min(\gamma_{\mathrm{R}}, \gamma_{\mathrm{D}}) < x\}$$
$$= 1 - \Pr\{\gamma_{\mathrm{R}} > x, \gamma_{\mathrm{D}} > x\}. \tag{35}$$

Substituting γ_{R} and γ_{D} in (13) and (14) into (35), we have

$$F_{\mathrm{DF}}(x) = 1 - \Pr\left\{ \frac{|h_{\mathrm{SR}}|^2 P_{\mathrm{S}}}{\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2} > x, \frac{|h_{\mathrm{SR}}|^2 |h_{\mathrm{RD}}|^2 \eta \alpha P_{\mathrm{S}}}{\sigma_{\mathrm{D}}^2(1 - \alpha)} > x \right\}$$
$$= 1 - \Pr\left\{ |h_{\mathrm{SR}}|^2 > \frac{x(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2)}{P_{\mathrm{S}}}, |h_{\mathrm{RD}}|^2 > \frac{x\sigma_{\mathrm{D}}^2(1 - \alpha)}{|h_{\mathrm{SR}}|^2 \eta \alpha P_{\mathrm{S}}} \right\}$$
$$= 1 - \left(1 - \Pr\left\{ |v|^2 \leq \left. \frac{x\sigma_{\mathrm{D}}^2(1 - \alpha)}{|h_{\mathrm{SR}}|^2 \eta \alpha P_{\mathrm{S}}} \right|_{|h_{\mathrm{SR}}|^2 > \left(x\left(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2\right)/P_{\mathrm{S}}\right)} \right\} \right). \tag{36}$$

Using the conditional probability [44], (36) can be expressed as

$$F_{\mathrm{DF}}(x) = 1 - \int_0^\infty \left(1 - F_{|h_{\mathrm{RD}}|^2}\left(\frac{x\sigma_{\mathrm{D}}^2(1 - \alpha)}{|h_{\mathrm{SR}}|^2 \eta \alpha P_{\mathrm{S}}} \right) \right) f_{|h_{\mathrm{SR}}|^2}\left(y + \frac{x(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2)}{P_{\mathrm{S}}} \right) dy, \tag{37}$$

where $y = |h_{\mathrm{SR}}|^2 - (x(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2)/P_{\mathrm{S}})$.

Applying (25) and (26), (37) becomes

$$F_{\mathrm{DF}}(x) = 1 - \int_0^\infty \sqrt{\frac{4x\sigma_{\mathrm{D}}^2(1 - \alpha)}{\left(y + \left(x\left(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2\right)/P_{\mathrm{S}}\right)\right)\eta \alpha P_{\mathrm{S}}}} K_1$$
$$\cdot \left(\sqrt{\frac{4x\sigma_{\mathrm{D}}^2(1 - \alpha)}{\left(y + \left(x\left(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2\right)/P_{\mathrm{S}}\right)\right)\eta \alpha P_{\mathrm{S}}}} \right) \exp$$
$$\cdot \left(-y - \frac{x(\sigma_{\mathrm{RSI}}^2 + \sigma_{\mathrm{R}}^2)}{P_{\mathrm{S}}} \right) dy = 1 - \int_0^\infty \sqrt{\frac{\Phi x}{y + \Psi x}} K_1$$
$$\cdot \left(\sqrt{\frac{\Phi x}{y + \Psi x}} \right) \exp(-y - \Psi x) dy z$$

$$= 1 - \exp(-\Psi x) \int_0^\infty \sqrt{\frac{\Phi x}{y + \Psi x}} K_1\left(\sqrt{\frac{\Phi x}{y + \Psi x}} \right) \exp(-y) dy. \tag{38}$$

Applying all steps used for the AF protocol such as changing variable $z = \exp(-y)$ and using [39] (Eq. (25.4.30)), we can obtain the CDF of the considered system with the DF protocol as

$$F_{\mathrm{DF}}(x) = 1 - \frac{\pi}{2M} \exp(-\Psi x) \sum_{m=1}^M \sqrt{1 - \phi_m^2} \sqrt{\frac{\Phi x}{\Psi x - \ln u}} K_1$$
$$\cdot \left(\sqrt{\frac{\Phi x}{\Psi x - \ln u}} \right). \tag{39}$$

Then, substituting (39) into (17) and applying similar transformations used for the AF protocol, we obtain the EC of the considered system with the DF protocol as (19). The proof is complete.

4. Numerical Results and Discussions

In this section, the mathematical expressions in the previous section are used to evaluate the ECs of the considered FD-EH-V2V relay system with AF/DF protocols. The effect of several system parameters such as the RSI level, the average transmission power, the time switching ratio, and the cascade Rayleigh fading is investigated to provide the system deployment guidelines in practice. All mathematical expressions are validated through Monte-Carlo simulations. In particular, we use the MATLAB simulator to obtain the simulation results. To generate a cascade Rayleigh fading channel, we generate two independent Rayleigh fading channels and then multiply these two channels. To realize the ECs of the HD-EH-V2V relay systems, we set the RSI level $k = 0$ in the obtained EC expressions of the FD-EH-V2V relay systems and divide the results by two because HD-EH-V2V relay systems need two time slots to transmit signal from S to D via R. In all scenarios, we set $\sigma_{\mathrm{R}}^2 = \sigma_{\mathrm{D}}^2 = \sigma^2$, the energy harvesting efficiency $\eta = 0.85$, and the complexity-accuracy trade-off parameters $M = N = 20$. In addition, the average transmission power is calculated as $\mathrm{SNR} = P_{\mathrm{S}}/\sigma^2$. Furthermore, ECs of the considered system are compared with those in the case of Rayleigh fading channel and HD transmission mode to demonstrate the benefits of the FD technique and the impacts of the RSI and cascade Rayleigh fading. For the sake of redoing the simulation easily by other researchers, we summarize the parameter settings for evaluating the system performance in Table 1.

Figure 3 illustrates the ECs of the considered FD-EH-V2V relay system versus the average SNR for different values of the RSI level, i.e., $k = -30, -20,$ and -10 dB. The time switching ratio is $\alpha = 0.5$. We use (18) and (19) in the theorem to plot the ECs of the considered FD-EH-V2V relay system with AF and DF protocols. In Figure 3, the ECs of the

TABLE 1: Parameter settings for evaluating the system performance.

Notation	Description	Fixed value	Varying range
SNR	Signal-to-noise ratio	40 dB	10, 20, 30, 50 dB; 0~50 dB
σ^2	Variance of Gaussian noise	1	None
η	Energy harvesting efficiency	0.85	None
α	Time switching ratio	0.5	0.1~0.9
k	RSI level	-20 dB	-10, -30 dB; 0~0.2
M, N	Trade-off parameters	20	None
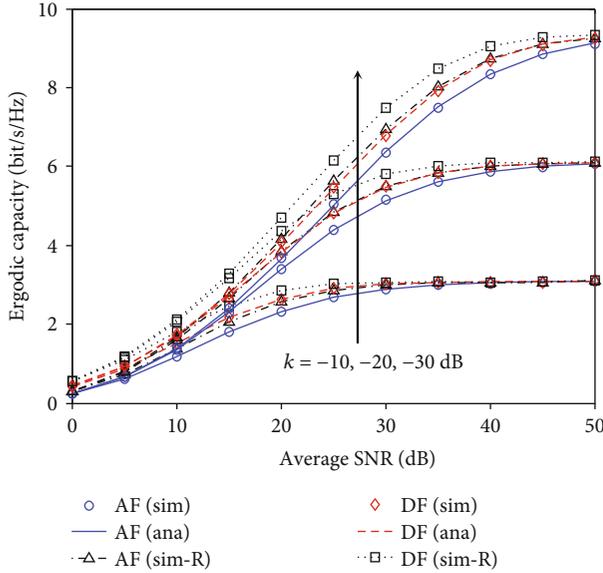
FIGURE 3: The ECs of the considered FD-EH-V2V relay system for different values of the RSI level, $k = -30, -20,$ and -10 dB; $\alpha = 0.5$.

FIGURE 4: The impact of RSI on the ECs of the considered FD-EH-V2V relay system for different average transmission power, SNR = 10, 20, 30, and 40 dB; $\alpha = 0.5$.

FD-EH system over Rayleigh fading channels are denoted by "Sim-R." As shown in Figure 3, the cascade Rayleigh fading reduces the ECs compared with the Rayleigh fading, especially in low SNR regime and low RSI level. Particularly, when RSI is low, i.e., $k = -30$ dB, the ECs of the considered system are 0.2 bit/s/Hz lower than those in the case of Rayleigh fading for both AF and DF protocols at SNR = 30 dB. However, all ECs approximate 9 bit/s/Hz at SNR = 50 dB. With higher RSI levels, i.e., $k = -20$ and -10 dB, all ECs reach the capacity ceilings at SNR = 40 dB and 50 dB for the case $k = -10$ dB and $k = -20$ dB, respectively. In these two cases, the ECs only reach 3.1 and 6 bit/s/Hz for $k = -10$ dB and $k = -20$ dB, respectively. These features indicate a strong impact of RSI on the ECs of the considered FD-EH-V2V relay system. On the other hand, the ECs with the DF protocol are always higher than those with the AF protocol. This result is reasonable because AF amplifies not only noise but also the RSI.

Figure 4 investigates the effect of the RSI on the ECs of the considered FD-EH-V2V relay system for different average transmission power, i.e., SNR = 10, 20, 30, and 40 dB. We also provide the ECs of the considered system with HD transmission mode to clearly show the impact of RSI and the benefit of FD transmission mode. It is evident from

Figure 4 that the effect of RSI is significant for high SNRs. Specifically, when self-interference is obliterated ($k = 0$), the ECs of the FD-EH-V2V relay system are two times higher than the ECs of the HD-EH-V2V relay system. When k increases, the ECs of the considered FD-EH-V2V relay system decrease. In particular, the ECs of the FD-EH-V2V relay system are always higher than ECs of the HD-EH-V2V relay system for low SNRs, i.e., SNR = 10 and 20 dB. However, for higher SNRs, i.e., SNR = 30 and 40 dB, the ECs of the FD-EH-V2V relay system are higher or lower than the ECs of the HD-EH-V2V relay system. For example, the ECs of the FD-EH-V2V relay system are higher than the ECs of the HD-EH-V2V relay system when k ranges from 0 to 0.07 and SNR = 30 dB. When $k > 0.07$, the ECs of the FD-EH-V2V relay system are lower than the ECs of the HD-EH-V2V relay system. With higher SNR, i.e., SNR = 40 dB, the ECs of the FD-EH-V2V relay system are only higher than those of the HD-EH-V2V relay system when $k < 0.03$. Therefore, depending on the measured and experimented RSI levels in practice, we can select FD or HD transmission mode to obtain high system capacity.
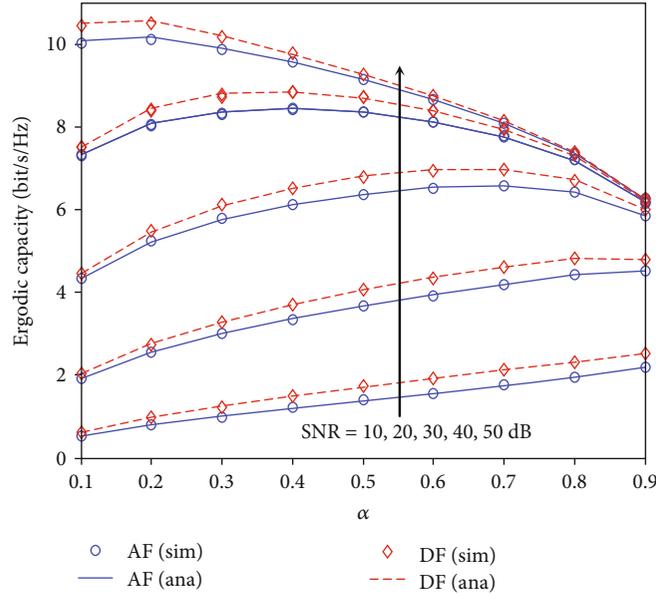
FIGURE 5: The ECs of the considered FD-EH-V2V relay system versus the time switching ratio α for different SNRs, $k = -30$ dB.
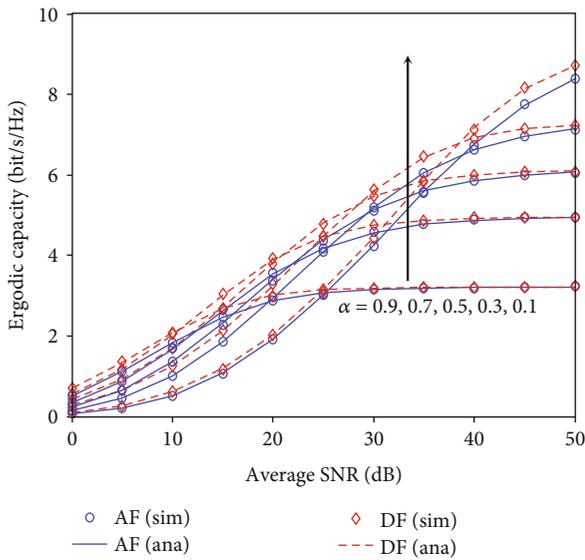
FIGURE 6: The ECs of the considered FD-EH-V2V relay system versus the average SNR for different values of time switching ratio, $k = -20$ dB.

FIGURE 7: The joint impacts of the RSI level k and the time switching ratio α on the ECs of the considered FD-EH-V2V relay system.

Figure 5 shows the ECs of the considered FD-EH-V2V relay system versus the time switching ratio α for different SNRs, e.g., SNR = 10, 20, 30, 40, and 50 dB. The RSI level is $k = -30$ dB. As observed in Figure 5, it is obvious that the ECs of the considered FD-EH-V2V relay system closely depend on α. When SNR is small, i.e., SNR = 10 and 20 dB, the ECs increase with α. It is because small SNR means low transmission power of S; thus, R needs more time for EH to have enough signal transmission power. However, when SNR is higher, i.e., SNR = 30, 40, and 50 dB, the ECs firstly increase and then decrease when α gets higher. Specifically, for SNR = 30 dB and SNR = 40 dB, ECs are maximal when $\alpha = 0.7$ and $\alpha = 0.4$, respectively, for both AF and DF proto-

cols. If the transmission power of S continues to increase, i.e., SNR = 50 dB, ECs are maximal when $\alpha = 0.2$. Therefore, based on the transmission power of S, the optimal α can be used to obtain the peak ECs of the considered FD-EH-V2V relay system.

To understand the relationship between the average transmission power of source and the time switching ratio α, we plot the ECs of the considered FD-EH-V2V relay system versus the average SNR for various time switching ratio, $\alpha = 0.1, 0.3, 0.5, 0.7, 0.9$, as shown in Figure 6. We can see that, in a low SNR regime, i.e., SNR < 20 dB, the EC of the case $\alpha = 0.7$ is the best while the EC of the case $\alpha = 0.1$ is the worst. However, for higher SNR range, 20 < SNR < 40 dB, the EC of the case $\alpha = 0.3$ is the best while the EC of the case $\alpha = 0.9$ is

the worst. If the average SNR still increases, i.e., SNR > 40 dB, the EC of the case $\alpha = 0.1$ is the best while the EC of the case $\alpha = 0.9$ is still the worst. To sum up, low SNR leads to a high α and vice versa.

The joint and cross impacts of the RSI level k and the time switching ratio α on the ECs of the considered FD-EH-V2V relay system are illustrated in Figure 7, where the average SNR is set as SNR = 40 dB. Similar to the ECs in Figure 5, with a certain value of k, there is an optimal α that maximizes ECs of AF/DF protocols. In particular, when $k = 0.01$, the ECs reach the maximum of 6.9 and 7.3 bit/s/Hz for AF and DF protocols, respectively, at $\alpha = 0.15$. However, with a higher value of k, the ECs are the highest when $\alpha = 0.1$. In other words, for k ranges from 0.02 to 0.2, ECs of AF/DF protocols decrease when α increases. Surprisingly, in the case of perfect SIC ($k = 0$), the ECs increase when α increases. As a result, besides reducing the ECs of the considered FD-EH-V2V relay system, the RSI also influences the optimal ECs. Therefore, for choosing an optimal α that maximizes the ECs of the considered system, it is necessary to know the RSI level and the average transmission power of the source before deploying this system in practice.

5. Conclusion

Applying FD and EH techniques are inevitable for future wireless networks, especially for V2V communication systems, because of the big advantages of these techniques. Therefore, in this paper, we evaluate the ergodic capacities of an FD-EH-V2V relay system with both AF and DF protocols under the impact of RSI and cascade Rayleigh fading channel. We mathematically derive the exact closed-form expressions of the EC of both AF and DF protocols. Numerical results reveal that the EC of the DF protocol is slightly higher than that of the AF one. Also, the cascade Rayleigh fading reduces the ECs of the considered system compared with the traditional Rayleigh fading. Furthermore, the ECs of the considered system are compared with those in the case of HD transmission mode to demonstrate the benefit of the FD technique. We also observe that, based on the RSI level and the source's average transmission power, we can choose an optimal time switching ratio to maximize the ECs of the considered FD-EH-V2V relay system.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Y. Zhao, "A survey of 6G wireless communications: emerging technologies," 2020, http://arxiv.org/abs/2004.08549.

[2] M. Babaei, U. Aygolu, M. Basaran, and L. Durak-Ata, "BER performance of full-duplex cognitive radio network with non-linear energy harvesting," *IEEE Transactions on Green Communications and Networking*, vol. 4, no. 2, pp. 448–460, 2020.

[3] Q.-V. Pham, F. Fang, V. N. Ha et al., "A survey of multi-access edge computing in 5G and beyond: fundamentals, technology integration, and state-of-the-art," *IEEE Access*, vol. 8, pp. 116974–117017, 2020.

[4] B. Clerckx, R. Zhang, R. Schober, D. W. K. Ng, D. I. Kim, and H. V. Poor, "Fundamentals of wireless information and power transfer: from RF energy harvester models to signal and system designs," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 1, pp. 4–33, 2018.

[5] H. H. M. Tam, H. D. Tuan, A. A. Nasir, T. Q. Duong, and H. V. Poor, "MIMO energy harvesting in full-duplex multi-user networks," *IEEE Transactionson Wireless Communications*, vol. 16, no. 5, pp. 3282–3297, 2017.

[6] A. H. Gazestani, S. A. Ghorashi, B. Mousavinasab, and M. Shikh-Bahaei, "A survey on implementation and applications of full duplex wireless communications," *Physical Communication*, vol. 34, pp. 121–134, 2019.

[7] Y. Deng, K. J. Kim, T. Q. Duong, M. Elkashlan, G. K. Karagiannidis, and A. Nallanathan, "Full-duplex spectrum sharing in cooperative single carrier systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 2, no. 1, pp. 68–82, 2016.

[8] B. C. Nguyen, X. N. Tran, and L. T. Dung, "On the performance of roadside unit-assisted energy harvesting full-duplex amplify-and-forward vehicle-to-vehicle relay systems," *AEU-International Journal of Electronics and Communications*, vol. 123, article 153289, 2020.

[9] N.-P. Nguyen, C. Kundu, H. Q. Ngo, T. Q. Duong, and B. Canberk, "Secure full-duplex small-cell networks in a spectrum sharing environment," *IEEE Access*, vol. 4, pp. 3087–3099, 2016.

[10] B. C. Nguyen, N. N. Thang, T. M. Hoang, and L. T. Dung, "Analysis of outage probability and throughput for energy harvesting full-duplex decode-and-forward vehicle-to-vehicle relay system," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 3539450, 10 pages, 2020.

[11] Y. Alsaba, C. Y. Leow, and S. K. A. Rahim, "Full-duplex cooperative non-orthogonal multiple access with beamforming and energy harvesting," *IEEE Access*, vol. 6, pp. 19726–19738, 2018.

[12] A. Koc, I. Altunbas, and E. Basar, "Two-way full-duplex spatial modulation systems with wireless powered AF relaying," *IEEE Wireless Communications Letters*, vol. 7, no. 3, pp. 444–447, 2018.

[13] C. Guo, L. Zhao, C. Feng, Z. Ding, and H.-H. Chen, "Energy harvesting enabled NOMA systems with full-duplex relaying," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 7, pp. 7179–7183, 2019.

[14] T. M. Hoang, V. Van Son, N. C. Dinh, and P. T. Hiep, "Optimizing duration of energy harvesting for downlink NOMA full-duplex over Nakagami-m fading channel," *AEU-International Journal of Electronics and Communications*, vol. 95, pp. 199–206, 2018.

[15] C. Li, Z. Chen, Y. Wang, Y. Yao, and B. Xia, "Outage analysis of the full-duplex decode-and-forward two-way relay system," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4073–4086, 2017.

[16] B. C. Nguyen, X. N. Tran, D. T. Tran, and L. T. Dung, "Full-duplex amplify-and-forward relay system with direct link:

performance analysis and optimization," *Physical Communication*, vol. 37, article 100888, 2019.

[17] X.-T. Doan, N.-P. Nguyen, C. Yin, D. B. Da Costa, and T. Q. Duong, "Cognitive full-duplex relay networks under the peak interference power constraint of multiple primary users," *EURASIP Journal on Wireless Communications and Networking*, vol. 2017, no. 1, 2017.

[18] V.-D. Nguyen, T. Q. Duong, H. D. Tuan, O.-S. Shin, and H. V. Poor, "Spectral and energy efficiencies in full-duplex wireless information and power transfer," *IEEE Transactions on Communications*, vol. 65, no. 5, pp. 2220–2233, 2017.

[19] D. H. Chen and Y. C. He, "Full-duplex secure communications in cellular networks with downlink wireless power transfer," *IEEE Transactions on Communications*, vol. 66, no. 1, pp. 265–277, 2018.

[20] C. Campolo, A. Molinaro, A. O. Berthet, and A. Vinel, "Full-duplex radios for vehicular communications," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 182–189, 2017.

[21] M. Yang, S.-W. Jeon, and D. K. Kim, "Interference management for in-band full-duplex vehicular access networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1820–1824, 2018.

[22] Y. Ai, M. Cheffena, A. Mathur, and H. Lei, "On physical layer security of double Rayleigh fading channels for vehicular communications," *IEEE Wireless Communications Letters*, vol. 7, no. 6, pp. 1038–1041, 2018.

[23] R. Atallah, M. Khabbaz, and C. Assi, "Energy harvesting in vehicular networks: a contemporary survey," *IEEE Wireless Communications*, vol. 23, no. 2, pp. 70–77, 2016.

[24] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: challenges and opportunities," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 9, pp. 1637–1652, 2014.

[25] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5983–5993, 2011.

[26] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 680–694, 2014.

[27] Q. N. Le, V. N. Q. Bao, and B. An, "Full-duplex distributed switch-and-stay energy harvesting selection relaying networks with imperfect CSI: design and outage analysis," *Journal of Communications and Networks*, vol. 20, no. 1, pp. 29–46, 2018.

[28] C. Zhong, H. A. Suraweera, G. Zheng, I. Krikidis, and Z. Zhang, "Wireless information and power transfer with full duplex relaying," *IEEE Transactions on Communications*, vol. 62, no. 10, pp. 3447–3461, 2014.

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

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

[31] D. Tian and V. C. Leung, "Analysis of broadcasting delays in vehicular ad hoc networks," *Wireless Communications and Mobile Computing*, vol. 11, no. 11, pp. 1433–1445, 2011.

[32] J. Salo, H. M. El-Sallabi, and P. Vainikainen, "Statistical analysis of the multiple scattering radio channel," *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 11, pp. 3114–3124, 2006.

[33] I. Kovacs, *Radio channel characterisation for private mobile radio systems - mobile-to-mobile radio link investigations: mobile-to-mobile radio link investigations*, Ph.D. dissertation, 2002.

[34] B. C. Nguyen, X. N. Tran, T. M. Hoang, and L. T. Dung, "Performance analysis of full-duplex vehicle-to-vehicle relay system over double-Rayleigh fading channels," *Mobile Networks and Applications*, vol. 25, no. 1, pp. 363–372, 2020.

[35] T. T. Duy, G. C. Alexandropoulos, V. T. Tung, V. N. Son, and T. Q. Duong, "Outage performance of cognitive cooperative networks with relay selection over double-Rayleigh fading channels," *IET Communications*, vol. 10, no. 1, pp. 57–64, 2016.

[36] B. C. Nguyen, T. M. Hoang, P. T. Tran, and T. N. Nguyen, "Outage probability of NOMA system with wireless power transfer at source and full-duplex relay," *AEU - International Journal of Electronics and Communications*, vol. 116, article 152957, 2020.

[37] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," in *SIGCOMM '13: Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, pp. 375–386, New York, NY, USA, August 2013.

[38] L. Van Nguyen, B. C. Nguyen, X. N. Tran, and L. T. Dung, "Transmit antenna selection for full-duplex spatial modulation multiple-input multiple-output system," *IEEE Systems Journal*, vol. 14, no. 4, pp. 4777–4785, 2020.

[39] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, vol. 9*, Dover, New York, NY, USA, 1972.

[40] A. Jeffrey and D. Zwillinger, *Table of Integrals, Series, and Products*, Academic Press, 2007.

[41] B. C. Nguyen, T. M. Hoang, and L. T. Dung, "Performance analysis of vehicle-to-vehicle communication with full-duplex amplify-and-forward relayover double-Rayleigh fading channels," *Vehicular Communications*, vol. 19, pp. 1–9, 2019.

[42] I. Z. Kovacs, P. C. F. Eggers, K. Olesen, and L. G. Petersen, "Investigations of outdoor-to-indoor mobile-to-mobile radio communication channels," in *Proceedings IEEE 56th Vehicular Technology Conference*, pp. 430–434, Vancouver, BC, Canada, September 2002.

[43] A. Pandey and S. Yadav, "Secrecy analysis of cooperative vehicular relaying networks over double-Rayleigh fading channels," *Wireless Personal Communications*, vol. 114, no. 3, pp. 2733–2753, 2020.

[44] A. Leon-Garcia and A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*, Pearson/Prentice Hall, UpperSaddle River, NJ, USA, 3rd edition, 2008.

WILEY | Hindawi

Research Article

Scaling Performance Analysis and Optimization Based on the Node Spatial Distribution in Mobile Content-Centric Networks

Jiajie Ren [iD],[1,2] Demin Li [iD],[1,2] Lei Zhang [iD],[1,2] and Guanglin Zhang [iD][1,2]

[1]*College of Information Science and Technology, Donghua University, 201620, China*
[2]*Engineering Research Center of Digitized Textile and Apparel Technology, Ministry of Education, 201620, China*

Correspondence should be addressed to Guanglin Zhang; glzhang@dhu.edu.cn

Content-centric networks (CCNs) have become a promising technology for relieving the increasing wireless traffic demands. In this paper, we explore the scaling performance of mobile content-centric networks based on the nonuniform spatial distribution of nodes, where each node moves around its own home point and requests the desired content according to a Zipf distribution. We assume each mobile node is equipped with a finite local cache, which is applied to cache contents following a static cache allocation scheme. According to the nonuniform spatial distribution of cache-enabled nodes, we introduce two kinds of clustered models, i.e., the clustered grid model and the clustered random model. In each clustered model, we analyze throughput and delay performance when the number of nodes goes infinity by means of the proposed cell-partition scheduling scheme and the distributed multihop routing scheme. We show that the node mobility degree and the clustering behavior play the fundamental roles in the aforementioned asymptotic performance. Finally, we study the optimal cache allocation problem in the two kinds of clustered models. Our findings provide a guidance for developing the optimal caching scheme. We further perform the numerical simulations to validate the theoretical scaling laws.

1. Introduction

During recent years, wireless traffic is undergoing explosively increase due to the subscribers' enormous data demands (such as video streaming). Content-centric networks (CCNs) [1] have emerged as a promising solution to deal with the increasing traffic, which shifts the traditional host-oriented communication pattern to the novel content-oriented communication pattern. In CCNs, nodes or user terminals are allowed to cache and forward contents based upon their names rather than the host addresses. This enables users request desired contents by local communications, without communicating with backhaul links to the core networks, which also reduces the delivery time of desired contents. In this context, as the number of users continually grows, the scaling performance of content-centric networks has attracted research interests, which is important to help us understand the scalability of CCNs.

In the pioneer work of Gupta and Kumar [2], they first study the scaling behavior of large-scale wireless ad hoc networks. In a static unit network consisting of n randomly distributed nodes, Gupta and Kumar [2] shows the asymptotic throughput of each node scales as $\Theta(1/\sqrt{n \log n})$ (Given two nonnegative functions $f(n)$ and $g(n)$: $f(n) = O(g(n))$ means there exists a constant c such that $f(n) \leq cg(n)$ for n large enough; $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$; $f(n) = \Theta(g(n))$ means both $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$; $f(n) = o(g(n))$ means $\lim_{n \to \infty} f(n)/g(n) = 0$; and $f(n) = \omega(g(n))$ means $\lim_{n \to \infty} g(n)/f(n) = 0$.), which indicates the poor scalability of wireless networks as the number of nodes increases. In [3], Franceschetti et al. apply the percolation theory to improve the asymptotic performance of wireless networks and per-node throughput is achieved as $\Theta(1/\sqrt{n})$. Grossglauser and Tse [4] first take the node mobility into consideration, and they propose a two-hop relaying policy to obtain the constant per-node throughput, which costs a vast transmission

delay. Subsequently, a series of researches focus on the scaling laws of various wireless ad hoc networks. Talak et al. [5] investigate the broadcast capacity and transmission delay in highly mobile wireless networks. In [6], Lin et al. investigate the optimal throughput-delay tradeoff under the i.i.d mobility model for mobile ad hoc networks. Jia et al. [7] introduce the correlated mobility into the analysis of throughput and delay. They find that strong correlation of node mobility results in poor asymptotic performance. The nature of different mobility models [6–9] affects the network performance remarkably. In addition, there are several technics to improve the performance of wireless networks, such as directional antennas [10–12], infrastructure support [13–15], secrecy analysis [16, 17], and reinforcement learning [18, 19].

In light of the asymptotic analysis of traditional wireless ad hoc networks, the throughput and delay scaling behaviors of large-scale cache-enabled content-centric networks have also received wide attention in recent years. In a *static* square network, Gitzenis et al. [20] formulate a joint optimization problem for content replication and transmission. They derive the minimum link capacity by utilizing a Zipf popularity law. In [21], by assuming that the content cached time in a node is finite, Azimdoost et al. investigate the throughput and delay of content-centric networks for the static grid network model and the static random network model, respectively. Jeon et al. [22] study the per-node throughput for wireless static device-to-device networks based on a decentralized caching scheme. In [23], Mahdian et al. analyze the scaling laws for pure static content-centric networks and heterogeneous content-centric networks, respectively. The authors further consider the optimal cache strategy for two kinds of networks. Zhang et al. [24] investigate the capacity of static hybrid content-centric wireless networks. In [25, 26], the authors study how the content popularity impact the network throughput. Contrast to the static networks, [27, 28] investigate the optimal throughput-delay tradeoff for *mobile* content-centric networks. Do et al. in [27] adopt a decoupling approach to achieve the optimal caching allocation in the hybrid mobile content-centric networks. In [29], Alfano et al. investigate the throughput and delay performance of mobile content-centric networks with limited cache space. They find that the stronger mobility degree of nodes results in poor network performance. In [30], Luo et al. introduce fast and slow mobility models into mobile content-centric networks according to different time scales. Then, the authors analyze the asymptotic performance with an arbitrary content popularity distribution.

The distribution of all nodes in the aforementioned literatures is uniform regardless of the traditional ad hoc networks or the content-centric networks. However, the node distribution or the mobility degree is nonuniform in the real world. For example, the density of nodes around hot spots or home points is relatively intensive. The mobile terminals are more likely to move around the airports or the tourist attractions. Motivated by the above considerations, the researchers begin to investigate the nonuniform wireless traffic [31–34]. Alfano et al. [31] first analyze the upper bounds to the per-node throughput of inhomogeneous static wireless networks, where the nodes are distributed following a shot-noise cox process. In [32, 33], Garetto et al. investigate the scaling behaviors of heterogeneous mobile ad hoc networks under the assumption that each node moves around a home point. The interesting results in [32, 33] show that the mobility degree directly affects the data exchange among the nodes. In [34], the authors study multicast capacity of heterogeneous sensor networks, jointly analyzing sensor energy efficiency. In contrast with [31–34], Zheng et al. [35] study the asymptotic performance of inhomogeneous static information-centric networks with infrastructure support. However, the mobility of nodes is of great importance in lots of application scenarios like vehicular ad hoc networks. It is meaningful to understand the impact of mobility on the scaling performance in the content-centric networks where node spatial distribution is nonuniform.

To this end, we investigate the scaling laws of mobile content-centric networks where node distribution is nonuniform in this paper. Instead of assuming each node moving the network area uniformly and independently, we consider that all nodes move around their corresponding home points, and they spend more time in the proximity of their home points than the network boundary. We assume each mobile node is equipped with a finite local cache and is able to store contents by using a static cache allocation scheme. Each cache-enabled node requests the desired content based on a Zipf popularity distribution. Moreover, we introduce two clustered models based on the spatial distribution of cache-enabled nodes, i.e., the clustered grid model and clustered random model. In each clustered model, we analyze the asymptotic performance according to the proposed scheduling and routing schemes. Finally, we study the optimal cache allocation problem of the two clustered models, respectively. We also perform the numerical simulations to validate the theoretical scaling laws. The main contributions of this paper are summarized as follows:

(i) Firstly, we construct a novel system model in the content-centric networks where each cache-enabled node moves around its corresponding home point. Based on the deployment of home point, we formalize the spatial distribution of nodes into the two clustered model

(ii) Secondly, we devise a cell-partition-based TDMA scheduling scheme to maximize the concurrent transmissions and develop a distributed multihop routing scheme. On this basis, we derive the asymptotic performance of the two kinds of clustered models and further establish the closed form of the throughput-delay tradeoff

(iii) Thirdly, we design the optimal cache allocation and investigate optimal throughput and delay performance by utilizing Lagrangian relaxation method under the assumption of a Zipf content popularity distribution. Moreover, massive numerical simulations are conducted to validate aforementioned theoretical results

The rest of the paper is organized as follows. In Section 2, we describe the system models and outline some definitions. In Section 3, we analyze the throughput and delay performance in both clustered grid model and clustered random model. In Section 4, we introduce the optimal cache allocation. Finally, we conclude the paper in Section 5.

2. System Models and Definitions

In this paper, we study a mobile content-centric network composed of n mobile nodes. We assume that n nodes move over a square region $\mathcal{O}$ of area n with the wrap-condition, to eliminate border effects. Note that, under aforementioned assumption, we adopt an extended network model that the node density over the square area remains constant as the number of mobile nodes increases. In the following subsections, we first describe the mobility model, content request model, and interference model. Then, we give some important definitions and notations used in this paper.

2.1. Mobility Model. In this paper, we adopt a bidimensional i.i.d mobility model for each mobile node, and time is divided into slots of equal duration. At the beginning of each slot, every node moves to a new location, which is independent of other nodes, and stays in the new location for the remaining duration of the slot. Let $X_i(t)$ denote the location of node i at time t, and $d_{ij} = \|X_i(t) - X_j(t)\|$ denotes the distance between node i and node j.

To characterize the spatial distribution of nodes, we assume that each node i is associated with a home point H_i, which is uniformly and independently selected over the square region $\mathcal{O}$. We consider that a node moves independently around its home point following a general ergodic process, which can be described by a rotationally invariant spatial distribution $\phi(d)$. Here, d denotes the Euclidean distance between the mobile node and its corresponding home point.

We further assume that $\phi(d)$ is an arbitrary nonincreasing function that decays as a power law of exponent δ, i.e., $\phi(d) \sim d^{-\delta}, \delta \geq 0$. We take function $s(d) = \min(1, d^{-\delta})$ and consider the following normalized probability density function (PDF) over the whole mobility area to avoid convergence problems in proximity of the home point.

$$\phi(d) = \frac{s(d)}{\int_{\mathcal{O}} s(d)} = \begin{cases} \Theta\left(s(d) n^{(\delta-2)/2}\right), & 0 \leq \delta < 2, \\ \Theta\left(\dfrac{s(d)}{\log n}\right), & \delta = 2, \\ \Theta(s(d)), & \delta > 2. \end{cases} \quad (1)$$

Exponent δ reflects the mobility degree of nodes, that is, the probability that each node moving to one point of the network area and stay at that point at a given time slot is not uniform. According to the PDF $\phi(d)$, it indicates that each node moves in a limited region, differing from the global mobility. In addition, the probability for the node moving in the proximity of its corresponding home point is larger than the probability that the node moves to the relatively far network area.

However, exponent δ just describes the individual mobility behavior and cannot reflect the mobile node density over the network area at a time slot. In real mobile world, the mobile node density is tightly related with the number of home points. Mobile nodes are more likely to move around the hotspots [36] or social spots [37]. For example, the density of mobile terminals (e.g., phones and ipads) is relatively dense in the office buildings or the tourist attractions while the density of mobile terminals in the suburb is reversely sparse due to the less home points. That is, the number of home points will affect the spatial distribution of mobile nodes. Motivated by this fact, clustering behavior has been found in [32, 33, 38] based on the long-term observations. In our work, we introduce the clustered model combined with the distribution of home points. First, we assume m clusters and each cluster has a middle point (the center of the cluster), which are distributed over the area $\mathcal{O}$ independently. Then, each home point randomly chooses one of the clusters with equal opportunity. Finally, the home points of the same cluster are belonged to a disk of radius R centered at the cluster middle point.

Considering the asymptotic performance, we assume that $m = n^\nu, 0 < \nu < 1$ and the cluster density over the whole network area $\rho_c = n^\nu/n = n^{\nu-1}$. The average distance between two cluster middle points is $d_c = n^{(1-\nu)/2}$. For the ease of analysis, we do not consider the cluster overlapping behavior. Hence, the cluster radius should be satisfied by $R = o(d_c)$.

2.2. Content Request Model. In the mobile content-centric networks, there are M distinct content objects of same size, where $M = n^\gamma, 0 < \gamma < 1$. We assume each mobile node is equipped with an equal-sized local cache, which can store K content objects (K is a positive constant). For the problem to be not trivial, we assume $K < M$, that is, each node has to decide which kinds of content to cache. We refer to a mobile node requesting a desired content $k, 1 \leq k \leq M$ at any time slot as the *requseter* of content k. We call a mobile node carrying a content k in its local cache the *holder* of content k.

In this paper, a caching scheme consists of two phases: content placement and content retrieve [22]. In the content placement phase, each node randomly and independently chooses contents to be stored in its local cache. Let $\mathcal{N}_k$ denote the set of nodes that cache content $k \in M$ in their local caches, where $N_k = |\mathcal{N}_k|$. Thus, the probability that the content k is cached by a mobile node in its local cache is N_k/n. In order to achieve a feasible cache allocation, the total cache constraint should be satisfied.

$$\sum_{k=1}^{M} N_k \leq nK. \quad (2)$$

In the content retrieve phase, each holder decides whether to deliver the requested content to the corresponding requester. During the retrieve phase, each node requests its desired content independently according to a Zipf

popularity distribution [39], i.e., the request probability p_k of content $k \in M$ is satisfied by

$$p_k = \frac{k^{-\alpha}}{H_\alpha(M)}, \tag{3}$$

where $\alpha > 0$ is the Zipf's law exponent and $H_\alpha(M) = \sum_{k=1}^{M} k^{-\alpha}$ is a normalization constant and is given by

$$H_\alpha(M) = \begin{cases} \Theta(1), & \alpha > 1, \\ \Theta(\log M), & \alpha = 1, \\ \Theta(M^{1-\alpha}), & \alpha < 1. \end{cases} \tag{4}$$

2.3. Interference Model. In this paper, to avoid multiuser simultaneous transmission interference, we adopt the protocol model in [2]. Moreover, we assume that the transmission range of mobile node i is T_i. If the content k is transmitted from mobile node i to node j successfully, then the following two conditions should be held:

(1) The distance between the transmitter i and the receiver j is no more than T_i, i.e.,

$$\left\| X_i(t) - X_j(t) \right\| \le T_i. \tag{5}$$

(2) Other transmitter l delivering different contents at the same time slot does not interfere the receiver j, i.e.,

$$\left\| X_l(t) - X_j(t) \right\| \ge (1 + \Delta) T_i. \tag{6}$$

Here, $\Delta > 0$ denotes a constant guard factor. $X_i(t)$ denotes the location of a mobile node at the time slot t. $\|\bullet\|$ denotes the Euclidean distance between the transmitter and the receiver. We further assume that each mobile node can deliver the contents at a constant rate W bits/sec.

2.4. Definitions

Definition 1. For a given scheduling and routing scheme, let $F(i, t)$ be the total number of bits of the requested contents received by mobile node i up to time t. We define the long-term throughput of mobile node i is

$$\liminf_{t \longrightarrow \infty} \frac{1}{t} F(i, t). \tag{7}$$

The average throughput over all nodes is given as

$$\frac{1}{n} \sum_{i=1}^{n} \liminf_{t \longrightarrow \infty} \frac{1}{t} F(i, t). \tag{8}$$

The throughput is defined as the expectation of the average throughput over all mobile nodes,

$$\lambda(n) = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \liminf_{t \longrightarrow \infty} \frac{1}{t} F(i, t) \right]. \tag{9}$$

Definition 2. The delay of a content retrieve process is the moment the interest packet leaves node i until the requested content arrives at node i from the closest holder. For a given scheduling and routing scheme, let $D(i, d)$ be the delay of the dth requested content of mobile node i. We define the long-term delay of mobile node i is

$$\limsup_{r \longrightarrow \infty} \frac{1}{r} \sum_{d=1}^{r} D(i, d). \tag{10}$$

The average delay over all nodes is given as

$$\frac{1}{n} \sum_{i=1}^{n} \limsup_{r \longrightarrow \infty} \frac{1}{r} \sum_{d=1}^{r} D(i, d). \tag{11}$$

The delay is defined as the expectation of the average delay over all mobile nodes.

$$D(n) = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \limsup_{r \longrightarrow \infty} \frac{1}{r} \sum_{d=1}^{r} D(i, d) \right]. \tag{12}$$

To facilitate the understanding, some important notations applied in this paper are listed in Table 1.

3. Throughput and Delay Analysis

For better understanding of content placement and retrieve process in the clustered random model, we first investigate the throughput and delay performance in the clustered grid model [33] in Section III-A, that is, the home points are placed regularly, which can be considered as a special instance of the distribution of the home points. In the real world, the home points can be deployed regularly, for example, the base stations (BSs) and the road side units (RSUs). In general, the BSs are deployed in the center of the cells, and the RSUs are deployed along the road (i.e., one RSU every 50 meters), which can be regarded as the examples of the grid clustered model. In Section III-B, we analyze the case of clustered random model. Specifically, when $\delta \le 2$ and n goes to infinity, the spatial distribution $\phi(d)$ tends to 0; we leave to the analysis of the network performance in this case for future work.

3.1. Clustered Grid Model. In the clustered grid model, we divide the whole network region into equal-sized cells of area $S_a = d_c^2$. The distribution of clusters should satisfy following two conditions:

(i) The middle point of each cluster is located at the center of each cell, and the distance between two adjacent cluster middle points is d_c

Symbols	Definitions
n	The number of mobile nodes, the area of network region $\mathcal{O}$
m	The number of clusters
M	The number of content objects
K	The size of a node's local cache
d_c	The average distance between two cluster middle points
ρ_c	The cluster density over the whole network area
S_a	The area of the cell in the clustered grid model
S_a'	The area of the cell in the clustered random model
p_k	The request probability of content k
T_i	The transmission radius of a mobile nodes
h	The maximum hops from the requester to the closest holder

(ii) The radii of all clusters equal to 0, i.e., the home points of the same cluster are gathered at their corresponding cluster's middle point (or the cell center).

Before introducing scheduling and routing schemes, we premise a necessary lemma that guarantees each cell has at least one mobile node at a time slot and the expected number of nodes in each cell is $\Theta(n^{1-\nu})$ which guarantees the network connectivity.

Lemma 3. *In the clustered grid model, assuming $\varepsilon = \Theta(\log\log n/\log n)$, for any $0 < \nu < 1 - \varepsilon$, each cell has at least one mobile node and the expected number of nodes in each cell S_a is $N(\bar{S}_a) = \Theta(n^{1-\nu})$ in a given slot with high probability.*

Proof. For an arbitrary cell, in a given slot, the probability that one node moves to any cell is d_c^2/n, i.e., $n^{-\nu}$. Hence, the probability p_c that each cell has at least one mobile node is

$$p_c = 1 - (1 - n^{-\nu})^n = 1 - e^{n \log (1 - n^{-\nu})} \geq 1 - e^{-n^{1-\nu}}, \quad (13)$$

where the last inequality follows that $\log (1 + x) \leq x$ for $x > -1$. Let $p_c' = 1 - e^{-n^{1-\nu}}$ and $p_c \geq p_c'$. By assuming $\varepsilon = \Theta(\log\log n/\log n)$, for any $0 < \nu < 1 - \varepsilon$, we have $n^{1-\nu} \geq \log n$. Note that when n goes infinity, ε tends to 0; the upper bound of ν is 1. Then, we have

$$p_c' = 1 - e^{-n^{1-\nu}} \geq 1 - e^{-\log n} = 1 - \frac{1}{n}, \quad (14)$$

which tends to 1 when n goes infinity. Hence, with high probability, each cell has at least one mobile node in a given slot.

Let $N(S_a)$ denote the number of nodes in the cell S_a, by definition:

$$N(S_a) = \sum_{i=1}^{n} I_{X_i \in S_a}, \quad (15)$$

where $I_{X_i \in S_a}$ is an i.i.d Bernoullian random variable, and $I_{X_i \in S_a} = 1$ if node i is in the cell S_a; otherwise, $I_{X_i \in S_a} = 0$. Hence,

$$E[N(S_a)] = E\left[\sum_{i=1}^{n} I_{X_i \in S_a}\right] = \sum_{i=1}^{n} E[I_{X_i \in S_a}] = n \cdot \frac{|S_a|}{n} = |S_a|. \quad (16)$$

We apply Chernoff bounds and have

$$P\left\{N(S_a) < \frac{1}{2}E[N(S_a)]\right\} \leq e^{-|S_a|/8},$$
$$P\{N(S_a) > 2E[N(S_a)]\} \leq e^{-|S_a|/3} < e^{-|S_a|/8}. \quad (17)$$

Then, letting $\varepsilon = \Theta(\log\log n/\log n)$, for any $0 < \nu < 1 - \varepsilon$, we have $|S_a| = n^{1-\nu} \geq 16 \log n$. Thus, we have

$$P\left\{\frac{1}{2}E[N(S_a)] \leq N(S_a) \leq 2E[N(S_a)]\right\}$$
$$\geq 1 - 2e^{-|S_a|/8} \geq 1 - 2e^{-2 \log n} = 1 - 2n^{-2}, \quad (18)$$

which tends to 1 when n goes infinity. Hence, we obtain the expected number of nodes in each cell is $\Theta(n^{1-\nu})$.

3.1.1. Scheduling Schemes. At a given slot, a scheduling scheme enables the contents retrieve between transmitter-receiver pairs not to be interfered. Based on the mobility model and the spatial distribution of mobile node, we adopt a cell-partition-based TDMA scheduling scheme to avoid the multiuser simultaneous transmission interference and to maximize the number of noninterfering transmission pairs at a time slot. Figure 1(a) illustrates a general real-world node distribution case, which shows mobile terminals are relatively dense near the home point, and there are fewer mobile terminals far away from home point. Figure 1(b) illustrates home points are deployed regularly and can be regarded as a clustered grid model. Figure 1(c) shows a cell-partition-based TDMA scheduling scheme corresponding to the time slot division, which guarantees multihop transmission in the clustered grid model.

We first divide each time slot into two half slots with equal length. Then, in the first half slot, we further divide the first half slot into several subslots with equal length. We partition each cell S_a into squares with same area S_e (S_e will be defined later). We call a square is *active* if a node in this square can transmit an interest packet or a content during the subslot. Based on the square partition, we assume that a node in a square can transmit an interest packet or a content to the node in the same square or the adjacent eight squares. We define the transmission radius of the node i as $T_i = \Theta(d_c)$. Hence, we obtain the following lemma according to the protocol model, which is essential to describe the noninterfering transmission process.
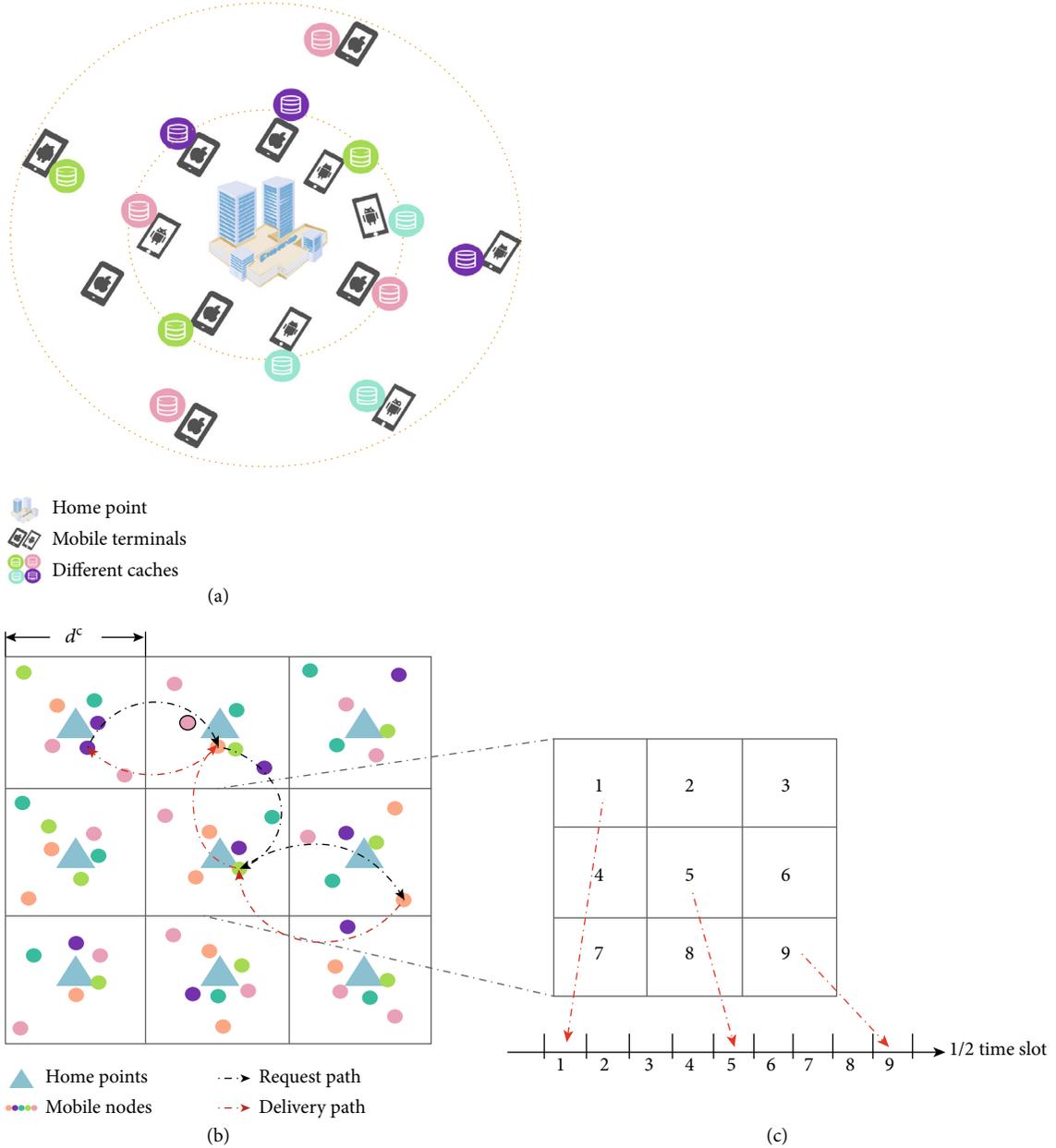
FIGURE 1: (a) A general real-world node distribution case. (b) Home points are deployed regularly and regarded as the clustered grid model. (c) A cell-partition-based TDMA scheduling scheme corresponding to the time slot division (the central cell is divided into 9 squares).

Lemma 4. *Each square can be active at most every $(1 + c_1)$ time subslots, and the achievable rate at the active square is $W/(1 + c_1)$.*

Proof. According to the interference model, we consider the distance between transmitter X_i and receiver X_j is bounded by $T_i = \Theta(d_c)$. If there is another simultaneously transmitting node X_l at the same time subslot and the transmission distance between X_l and X_j is less than $(2 + \Delta)T_i$, then X_l will cause the interference with X_j. Hence, the area of the total interference region is bounded by $(2(2 + \Delta)T_i + 3T_i)^2$. We obtain that each square has at most $c_1 = (2(2 + \Delta)T_i + 3T_i)^2$

$/T_i^2 = (2(2 + \Delta) + 3)^2$ interference neighbouring squares, which is a constant and independent of n.

Each square gets a noninterfering transmission opportunity and becomes active in every $(1 + c_1)$ time subslots. Since each transmitter-receiver pair can send W bits in a successful transmission, the achievable transmission rate at the active square is $W/(1 + c_1)$.

From Lemma 4, we can construct $(1 + c_1)$ subsets of regularly spaced, simultaneously transmitting squares without interference.

Next, we calculate the area of each square S_e. From Lemma 3, we obtain the average number of nodes in each cell

S_a is $\Theta(n^{1-\nu})$. We assume there exists point A in the cell that the distance between A and the home point is d. The mobility process of each node is independent at each slot, and the probability that the node moves to point A is $\Theta(d^{-\delta})$ at each slot according to the node distribution. Hence, when $\delta > 2$, the density of nodes at point A is $\Theta(n^{1-\nu}/d^{\delta})$ at each slot. Applying Lemma 4 in [33], we can obtain the maximum density of achievable transmitter-receiver pairs (T-R pairs) when $d = d_c/2$. Thus, the density of T-R pairs at any point of the network scales as $\phi_c = \Theta(n^{1-\nu}/d_c^{\delta}) = \Theta(n^{(1-\nu)(1-(\delta/2))})$. Therefore, the area of each square is equal to $1/\phi_c$, i.e., $S_e = \Theta(n^{-(1-\nu)(1-(\delta/2))})$.

At a given time slot, each cell can schedule the number of concurrent noninterfering transmissions χ is

$$\chi = \Theta\left(\frac{d_c^2}{(1+c_1)(1/\phi_c)}\right) = \Theta\left(n^{(1-\nu)(2-(\delta/2))}\right). \quad (19)$$

Since the area of each square is $\Theta(n^{-(1-\nu)(1-(\delta/2))})$, we can acquire that the probability that at least one node in a subcell is a constant by applying Lemma 3.

3.1.2. Routing Schemes. Considering the mobility model of nodes and the clustering behavior, we propose a multihop transmission scheme Π for the content transmission between the requester-holder pairs. We assume the content request and delivery process of each node is conducted in one time slot. According to the maximum hops (derived in the following section) in the clustered grid model, the value of the one time slot can be set as $\Omega(n^{\nu/2})$ to guarantee the multihop transmission within one time slot.

Π-1: according to the TDMA scheduling phase, each time slot is divided into two half slots with equal length.

Π-2: during the first half slot, we further divide the first half slot into several subslots with equal length. Each cell is divided into squares with the same area,

(1) if there is one requested content in the same square (or the same cluster), the requester i directly forwards an interest packet to the holder

(2) if there is no any requester-holder pair in the same square (or the same cluster), the requester i sends an interest packet to the relay node in the adjacent square at the beginning of subslot; multihop is used for all relay nodes until the interest packet reached the nearest holder. Furthermore, we assume that the first hop towards the adjacent subcell is on the horizon path (if possible) and then on the vertical path

Π-3: during the second half slot, we continue dividing the second half slot into several subslots with equal length,

(1)if there is one requested content in the same square (or the same cluster), the requested content will be directly forwarded back to the requester i. In this case, the requested content will be achieved in one hop

(2)if there is no any requester-holder pair in the same square (or the same cluster), the nearest holder receives the

interest packet in the first half slot, and then, the nearest holder in the second half slot sends the requested content back to its corresponding requester in the reverse direction

Lemma 5. *In the clustered grid model, for any node requesting content k, the probability $P(k)$ that an interest packet for content k is satisfied by one hop is* $\min\left(\Theta(N_k n^{(1-\nu)((\delta/2)-1)-1}), 1\right)$.

Proof. For an arbitrary square, the probability that there is at least one requested content k within the same square or the adjacent squares of the requester is $9S_e/n = 9n^{(1-\nu)((\delta/2)-1)}/n = 9n^{(1-\nu)((\delta/2)-1)-1}$. Thus,

$$P(k) = 1 - \left(1 - 9n^{(1-\nu)((\delta/2)-1)-1}\right)^{N_k} \le 9N_k n^{(1-\nu)((\delta/2)-1)-1}. \quad (20)$$

The inequality follows that $(1+x)^n \ge 1 + nx$ for any $x > -1$. If $9N_k n^{(1-\nu)((\delta/2)-1)-1} = o(1)$, then $P(k) = \Theta(N_k n^{(1-\nu)((\delta/2)-1)-1})$; otherwise, $P(k) = 1$.

In the clustered grid model, due to the transmission radius of each node is $\Theta(d_c)$, we can derive the maximum number of hops from a content requester to the closest holder is $h = \Theta(n^{\nu/2})$.

Lemma 6. *In clustered grid model, for any mobile node requesting content k, the average number of hops needed to transmit an interest packet along the path from the content requester to the closest content holder, denoted by $\mathbb{E}(H_k)$, is*

$$\mathbb{E}(H_k)$$
$$= \begin{cases} \Theta(1), & N_k = \Omega\left(n^{1-(1-\nu)((\delta/2)-1)}\right), \\ \Theta\left(\dfrac{1}{N_k n^{(1-\nu)((\delta/2)-1)-1}}\right), & N_k = o\left(n^{1-(1-\nu)((\delta/2)-1)}\right) \& N_k = \Omega\left(n^{(1/2)-(1-\nu)((\delta/2)-(3/2))}\right), \\ \Theta(n^{\nu/2}), & N_k = o\left(n^{(1/2)-(1-\nu)((\delta/2)-(3/2))}\right). \end{cases} \quad (21)$$

Proof. Let H_k denote the number of hops along the requesting path, and we have

$$\mathbb{E}(H_k) = \sum_{i=1}^{h-1} \left(i \cdot P(k)(1-P(k))^{i-1}\right) + h(1-P(k))^{h-1}$$
$$= \sum_{i=1}^{h} (1-P(k))^{i-1} = \frac{1-(1-P(k))^h}{P(k)}. \quad (22)$$

Based on Lemma 5 and the value of h, we consider three different conditions to further derive $\mathbb{E}(H_k)$.

(1)If $P(k) = \Theta(1)$, then $\mathbb{E}(H_k) = \Theta(1)$.

(2)If $P(k) = o(1)$ and $hP(k) = \Omega(1)$, $1 - (1-P(k))^h \ge 1 - e^{-hP(k)} = \Theta(1)$. Hence, we obtain that $\mathbb{E}(H_k) = \Theta(1/(P(k)))$.

(3)If $P(k) = o(1)$ and $hP(k) = o(1)$, applying the equivalent infinitesimal, we obtain that $1 - (1-P(k))^h \approx 1 - (1-hP(k)) \approx \Theta(hP(k))$. Hence, we get that $\mathbb{E}(H_k) = \Theta(h)$.

This completes the proof.

Theorem 7. *In the clustered grid model, the delay and throughput of each mobile node, denoted by $D(n)$ and $\lambda(n)$, are given by w.h.p.*

$$D(n) = \Theta\left(\sum_{m=1}^{M} p_k \mathbb{E}(H_k)\right), \qquad (23)$$

$$\lambda(n) = \Theta\left(\frac{n^{(1-\nu)(1-(\delta/2))}}{\sum_{m=1}^{M} p_k \mathbb{E}(H_k)}\right). \qquad (24)$$

Proof. First, we consider the delay performance. Based on the scheduling and routing scheme, each square can be active in every $1 + c_1$ time subslots, where c_1 is a constant. Thus, the time spent in each hop is a constant fraction of time. In Lemma 6, we obtain the average number of hops from a requester to the closest holder for content k is $\mathbb{E}(H_k)$. In addition, the requested content forwarded back to the requester takes the same route as its corresponding interest packet in reverse direction, the total number of hops for content k is $2\mathbb{E}(H_k)$. Taking the request probability p_k for the content k into consideration, we get the delay for all contents over the network as (6).

Next, we derive the throughput performance. Since there are $\Theta(n^\nu)$ cells over the network area, and the total number of bits that active squares in a cell can transmit is $(W/(1 + c_1))n^{(1-\nu)(2-(\delta/2))}$. Lemma 6 indicates the average number of hops for content k is $\mathbb{E}(H_k)$. Considering the request probability p_k for the content k, we obtain the average number of bits transmitted in the whole network is $n\lambda(n)\sum_{m=1}^{M} p_k \mathbb{E}(H_k)$. Then, we have

$$n\lambda(n) \sum_{m=1}^{M} p_k \mathbb{E}(H_k) \leq \frac{W}{(1 + c_1)} n^{(1-\nu)(2-(\delta/2))} \cdot n^\nu,$$

$$\lambda(n) \leq \frac{n^{(1-\nu)(1-(\delta/2))}}{\sum_{m=1}^{M} p_k \mathbb{E}(H_k)}. \qquad (25)$$

The achievable throughput of each mobile node is determined by the scheduling and routing scheme. Hence, we obtain the throughput performance as (24).

Corollary 8. *The throughput and delay tradeoff in the clustered grid model is given by*

$$\lambda(n) = \Theta\left(\frac{n^{(1-\nu)(1-(\delta/2))}}{D(n)}\right). \qquad (26)$$

3.2. Clustered Random Model. In this section, we analyze the throughput and delay of each node in the clustered random model, where clusters are distributed randomly and independently in the network, that is, the home points are deployed in a random manner.

We first introduce an important lemma given in [31]. This lemma indicates the appropriate partition of the whole network area can guarantee that the average number of clus-

ters in each cell is $\Theta(\log n)$, which makes the multihop transmission in the clustered random model possible.

Lemma 9. *Consider a set of m points independently distributed over a bidimensional domain $\mathcal{O}$ of area n, with density $\rho = m/n$. The domain $\mathcal{O}$ is partitioned by regular tessellations, and A_s denotes the tiles over the tessellations with area $|A_s| \geq 16(\log m/\rho)$, $\forall s$. Let $N(A_s)$ be the number of nodes falling within the tiles A_s. Then, uniformly over the tessellation, $N(A_s)$ is contained between $\rho|A_s|/2$ and $2\rho|A_s|$ with high probability, i.e.,*

$$\frac{\rho|A_s|}{2} < \inf_s N(A_s) \leq \sup_s N(A_s) < 2\rho|A_s|. \qquad (27)$$

This lemma can be proved by the Chernoff bound in [40]; we neglect the proof process for simplicity.

Considering Lemma 9 and the cluster density $\rho_c = n^{\nu-1}$, we partition the whole network into equal-sized cells with area $S_a' = 16(\log n^\nu/n^{\nu-1}) = \Theta(n^{1-\nu}\log n)$. Note that we choose the equal-sized cell area in the clustered random model in order to make sure that nodes in each cluster has the opportunity to retrieve the content; in fact, we can choose different size area only if $S_a' \geq \Theta(n^{1-\nu}\log n)$. Thus, we can obtain the expected number of clusters in each cell is $\Theta(\log n)$. We assume that the transmission radius of each mobile node in the clustered random model is $T_n = \Theta(n^{(1-\nu)/2}\sqrt{\log n})$, which guarantees the nodes can communicate with the nodes in their neighbor cells. Since the maximum distance between the requester and the closest holder is $\Theta(n^{1/2})$, the maximum number of hops from a content requester to the closest holder is $h = \Theta(n^{\nu/2}/\sqrt{\log n})$.

Based on the cell partition, we obtain the following lemma that guarantees the average number of nodes in every cluster is $\Theta(n^{1-\nu})$.

Lemma 10. *In the clustered random model, the average number of node in every cluster is $\Theta(n^{1-\nu})$.*

Proof. The proof process of this lemma is similar to that in Lemma 3 by applying he Chernoff bound.

The density of T-S pairs at any point of the network scales as $\phi'_c = \Theta(n^{1-\nu}/(n^{(\delta(1-\nu))/2}\log^{\delta/2}n)) = \Theta(n^{(1-\nu)(1-(\delta/2))}\log^{-(\delta/2)}n)$. We further partition each cell into squares with area $1/\phi'_c$, i.e., $S'_e = \Theta(n^{(1-\nu)((\delta/2)-1)}\log^{\delta/2}n)$. Hence, each active cell can schedule the number of concurrent noninterfering transmissions χ' is

$$\chi' = \Theta\left(\frac{S}{(1 + c_2)\left(1/\phi'_c\right)}\right) = \Theta\left(\frac{n^{1-\nu}\log n}{n^{(1-\nu)((\delta/2)-1)}\log^{\delta/2}n}\right)$$

$$= \Theta\left(n^{(1-\nu)(2-(\delta/2))}\log^{(1-(\delta/2))}n\right). \qquad (28)$$

Hence, a noninterference scheduling scheme in the clustered random model can be designed as following steps. First, we divide the whole network into equal-sized cells with area $\Theta(n^{1-\nu} \log n)$, which makes the transmission connectivity possible. Then, we partition each cell into equal-sized squares with area $\Theta(n^{(1-\nu)((\delta/2)-1)} \log^{\delta/2} n)$, which guarantees the content transmission free from interference. Figure 2 shows the cell partition and the content request process in the clustered random model. Finally, a routing scheme similar with that in the cluster grid model can be utilized to achieve the content multihop transmission at a given slot.

Lemma 11. *In the clustered random model, for any node requesting content k, the probability $P'(k)$ that an interest packet for content k is satisfied by one hop is* $\min (\Theta(N_k n^{(1-\nu)((\delta/2)-1)-1} \log^{\delta/2} n), 1)$.

Proof. For an arbitrary square, the probability that there is at least one requested content k is within the same square or the adjacent squares of the requester is $(9S_e'/n) = (9 n^{(1-\nu)((\delta/2)-1)} \log^{\delta/2} n)/n = 9n^{(1-\nu)((\delta/2)-1)-1} \log^{\delta/2} n$. Thus, we have

$$
\begin{aligned}
P'(k) &= 1 - \left(1 - 9n^{(1-\nu)((\delta/2)-1)-1} \log^{\delta/2} n\right)^{N_k} \\
&\leq 9N_k n^{(1-\nu)((\delta/2)-1)-1} \log^{\delta/2} n.
\end{aligned}
\tag{29}
$$

The inequality follows that $(1 + x)^n \geq 1 + nx$ for any $x > -1$. If $9N_k n^{(1-\nu)((\delta/2)-1)-1} \log^{\delta/2} n = o(1)$, then $P'(k) = \Theta(N_k n^{(1-\nu)((\delta/2)-1)-1} \log^{\delta/2} n)$. Otherwise, $P'(k) = 1$.

Lemma 12. *In the clustered random model, for any mobile node requesting content k, the average number of hops needed to transmit an interest packet along the path from the content requester to the closest content holder, denoted by $\mathbb{E}'(H_k)$, is*

$$
\mathbb{E}'(H_k) = \begin{cases}
\Theta(1), & N_k = \Omega\left(\dfrac{n^{1-(1-\nu)((\delta/2)-1)}}{\log^{\delta/2} n}\right), \\[4mm]
\Theta\left(\dfrac{n^{(1/2)-(1-\nu)((\delta/2)-(3/2))}}{N_k \log^{\delta/2} n}\right), & N_k = o\left(\dfrac{n^{1-(1-\nu)((\delta/2)-1)}}{\log^{\delta/2} n}\right) \& N_k = \Omega\left(\dfrac{n^{(1/2)-(1-\nu)((\delta/2)-(3/2))}}{\log^{\delta/2} n}\right), \\[4mm]
\Theta\left(\dfrac{n^{\nu/2}}{\sqrt{\log n}}\right), & N_k = o\left(\dfrac{n^{(1/2)-(1-\nu)((\delta/2)-(3/2))}}{\log^{\delta/2} n}\right).
\end{cases}
\tag{30}
$$

Proof. According to Lemma 6, we have

$$
\mathbb{E}'(H_k) = \frac{1 - \left(1 - P'(k)\right)^h}{P'(k)}.
\tag{31}
$$

Based on the value of h and $P'(k)$, we consider three different conditions to further derive $\mathbb{E}'(H_k)$.

(1) If $P'(k) = \Theta(1)$, then $\mathbb{E}'(H_k) = 1$

(2) If $P'(k) = o(1)$ and $hP'(k) = \Omega(1)$, $1 - (1 - P'(k))^h \geq 1 - e^{-hP'(k)} = \Theta(1)$. Hence, we obtain that $\mathbb{E}'(H_k) = \Theta(1/(P'(k)))=\Theta(n^{(1/2)-(1-\nu)((\delta/2)-(3/2))}/N_k \log^{\delta/2} n)$.

(3) If $P'(k) = o(1)$ and $hP'(k) = o(1)$, applying the equivalent infinitesimal, we obtain that $1 - (1 - P'(k))^h \approx 1 - e^{-hP'(k)} \approx \Theta(hP'(k))$. Hence, we get that $\mathbb{E}'(H_k) = \Theta(h) =\Theta(n^{\nu/2}/\sqrt{\log n})$.

This completes the proof.

Theorem 13. *In the clustered random model, the delay and throughput of each mobile node, denoted by $D(n)$ and $\lambda(n)$, are given by w.h.p.*

$$
\begin{aligned}
D(n) &= \Theta\left(\sum_{m=1}^{M} p_k \mathbb{E}'(H_k)\right), \\[3mm]
\lambda(n) &= \Theta\left(\frac{n^{(1-\nu)(1-(\delta/2))} \log^{-\delta/2} n}{\sum_{m=1}^{M} p_k \mathbb{E}'(H_k)}\right).
\end{aligned}
\tag{32}
$$

Proof. The proof of delay performance is similar to that of Theorem 7, and we do not repeat it for simplicity.

We analyze the throughput performance of each node. Based on the cell partition of the network, we can obtain the total number of cells is $\Theta(n^\nu/\log n)$. Considering the scheduling scheme and routing scheme, the total number of bits that active square in a cell can transmit is $(W/(1 + c_2)) n^{(1-\nu)(2-(\delta/2))} \log^{(1-(\delta/2))} n$. Moreover, Lemma 12 indicates the average number of hops for content k is $\mathbb{E}'(H_k)$. Since the request probability of the content k is p_k, the average number of bits transmitted in the whole network can be calculated as $n\lambda(n)\sum_{m=1}^{M} p_k \mathbb{E}'(H_k)$. The rest proof is similar to the Theorem 7; we neglect it here for simplicity.

From Theorem 13, we can conclude that the difference of throughput and delay performance between the clustered random model and the clustered grid model is the logarithm factor and the average transmission hops.
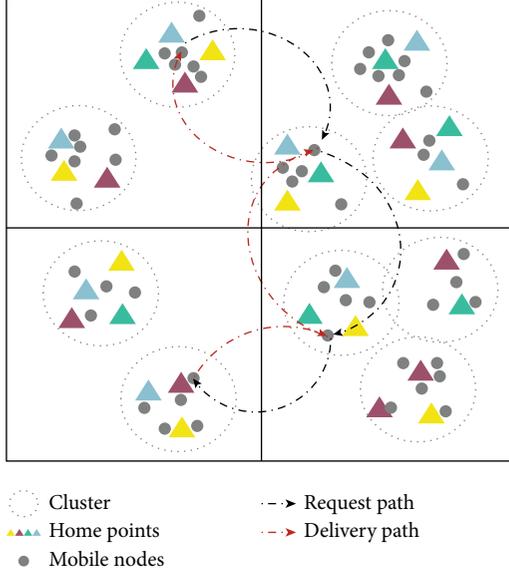
FIGURE 2: The cell partition and the content request process in the clustered random model.

Corollary 14. *The throughput and delay tradeoff in the clustered random model is given by*

$$\lambda(n) = \Theta\left(\frac{n^{(1-\nu)(1-(\delta/2))}}{D(n)\log^{\delta/2}n}\right),\tag{33}$$

which differs from Corollary 8 by a factor $\log^{-\delta/2}n$.

4. Optimal Cache Allocation

In this section, we analyze the optimal throughput and delay performance of the mobile content-centric network with respect to optimal cache allocation strategy. To achieve the optimal goal, we need to select appropriate $\{N_k\}_{k=1}^{M}$ based on the cache constraints for the clustered grid model and clustered random model, respectively.

4.1. Clustered Grid Model. From Corollary 8, it indicates that minimizing the transmission delay is equivalent to maximizing the throughput performance. In order to achieve the minimum delay, we formulate the following optimization problem:

$$
\begin{aligned}
\underset{N_k}{\text{minimize}} \quad & \sum_{k=1}^{M} p_k \mathbb{E}(H_k) \\
\text{subject to} \quad & \sum_{k=1}^{M} N_k \le nK, \\
& 1 \le N_k \le n^{1-(1-\nu)((\delta/2)-1)},
\end{aligned}
\tag{34}
$$

where $\mathbb{E}(H_k)$ is given by

$$
\mathbb{E}(H_k)
= \begin{cases}
\Theta(1), & N_k = \Omega\left(n^{1-(1-\nu)((\delta/2)-1)}\right), \\
\Theta\left(\dfrac{1}{N_k n^{(1-\nu)((\delta/2)-1)-1}}\right), & N_k = o\left(n^{1-(1-\nu)((\delta/2)-1)}\right) \& N_k = \Omega\left(n^{(1/2)-(1-\nu)((\delta/2)-(3/2))}\right), \\
\Theta(n^{\nu/2}), & N_k = o\left(n^{(1/2)-(1-\nu)((\delta/2)-(3/2))}\right).
\end{cases}
\tag{35}
$$

The first constraint in (34) comes from the feasible cache allocation in (2), and the second constraint guarantees that for different types of contents, there is at most one copy content in each square. Note that the second derivatives of the objective function in (34) is always positive, which can be considered as a strictly convex optimization problem. Hence, we apply Lagrangian relaxation method to find the unique optimal solution.

Based on the Zipf distribution law, it shows that p_k decreases as k increases and so is N_k. For the convenient and tractable analysis, we first define three sets K_1, K_2, and K_3 according to the size of N_k, respectively. That is,

(i) let $K_1 = \{1, 2, \cdots, k_1 - 1\}$ be the set of contents such that $N_k = n^{1-(1-\nu)((\delta/2)-1)}$

(ii) let $K_2 = \{k_1, k_1 + 1, \cdots, k_2 - 1\}$ be the set of contents such that $n^{(1/2)-(1-\nu)((\delta/2)-(3/2))} < N_k < n^{1-(1-\nu)((\delta/2)-1)}$

(iii) let $K_3 = \{k_2, k_2 + 1, \cdots, M\}$ be the set of contents such that $N_k = 1$

Next, we take the Lagrangian multiplier $\lambda \in R^+$ for the first constraint in (34) and combine with the second constraint; the necessary conditions for the minimal $D(n)$ are given as

$$
\frac{\partial D(n)}{\partial N_k} = -\frac{p_k}{N_k^2 n^{(1-\nu)((\delta/2)-1)-1}}
\begin{cases}
\le -\lambda, & \forall k \in K_1, \\
= -\lambda, & \forall k \in K_2, \\
\ge -\lambda, & \forall k \in K_3.
\end{cases}
\tag{36}
$$

When $k \in K_2$, we obtain

$$N_k = \frac{p_k^{1/2}}{\lambda^{1/2} n^{(1-\nu)((\delta/4)-(1/2))-(1/2)}}.\tag{37}$$

By adding up N_k in (37) for $k \in K_2$, we have

$$\lambda^{1/2} = \frac{n^{(1/2)-(1-\nu)((\delta/4)-(1/2))}\sum_{k=k_1}^{k_2-1} p_k^{1/2}}{\sum_{k=k_1}^{k_2-1} N_k}.\tag{38}$$

Combined (37) and (38), we obtain

$$N_k = \frac{p_k^{1/2}}{\sum_{k=k_1}^{k_2-1} p_k^{1/2}} \sum_{k=k_1}^{k_2-1} N_k = \frac{p_k^{1/2}}{\sum_{k=k_1}^{k_2-1} p_k^{1/2}} nK',\tag{39}$$

where $K' = K - ((k_1 - 1)/n^{(1-\nu)((\delta/2)-1)}) - ((M - k_2 + 1)/n)$. Therefore, we obtain the optimal number of content k, denoted by $N_k^{a\circ}$ (In this paper, we apply symbol a° to denote the optimal value.), i.e.,

$$
N_k^{\text{a}\circ} = \begin{cases} n^{1-(1-v)((\delta/2)-1)}, & \forall k \in K_1, \\ \dfrac{p_k^{1/2}}{\sum_{k=k_1}^{k_2-1} p_k^{1/2}} nK', & \forall k \in K_2, \\ 1, & \forall k \in K_3. \end{cases} \tag{40}
$$

By substituting (40) into (34), we achieve the minimum delay, denoted by $D(n)^{\text{a}\circ}$.

$$
D(n)^{\text{a}\circ} = \Theta\left(\sum_{k=1}^{k_1-1} p_k + \frac{\left(\sum_{k=k_1}^{k_2-1} p_k^{1/2}\right)^2}{n^{(1-v)((\delta/2)-1)}K'} + \sum_{k=k_2}^{M} n^{v/2} p_k \right). \tag{41}
$$

Lemma 15. *In the clustered grid model, for $n \longrightarrow \infty$, the values of k_1 and k_2 is given as follows*

$$
k_1 = \begin{cases} \Theta\left(n^{(1-v)((\delta/2)-1)} \right), & \alpha > 2, \\ \Theta\left(\dfrac{n^{(1-v)((\delta/2)-1)}}{\log M} \right), & \alpha = 2, \\ \min\left\{ M, \Theta\left(\dfrac{n^{(2/\alpha)(1-v)((\delta/2)-1)}}{M^{(2/\alpha)-1}} \right) \right\}, & 0 < \alpha < 2. \end{cases}
$$

$$
k_2 = \begin{cases} \min\left\{ M+1, \Theta\left(n^{(\alpha-2)/\alpha(1-v)((\delta/2)-1)+(2/\alpha)} \right) \right\}, & \alpha > 2, \\ M+1, & \alpha \leq 2. \end{cases} \tag{42}
$$

Proof. By applying the condition $\{k_1, k_2 - 1\} \in K_2$, we derive that

$$
\frac{k_1}{k_2} \simeq n^{(2/\alpha)(1-v)((\delta/2)-1)-(2/\alpha)}. \tag{43}
$$

When $n \longrightarrow \infty$, we obtain that $K' \longrightarrow K - ((k_1 - 1)/n^{(1-v)((\delta/2)-1)})$. Since $k_1 \in K_2$, according to (40), we have $N_{k_1} < n^{1-(1-v)((\delta/2)-1)}$. Hence, it follows that

$$
n^{(1-v)((\delta/2)-1)}K' < k_1^{\alpha/2}[H_{\alpha/2}(k_2-1) - H_{\alpha/2}(k_1-1)]. \tag{44}
$$

From the fact that k_1 is the smallest index in set K_2 such that $N_{k_1} < n^{1-(1-v)((\delta/2)-1)}$, we decrease k_1 by one, which will result in $N_{k_1-1} \geq n^{1-(1-v)((\delta/2)-1)}$. Hence, we obtain that

$$
n^{(1-v)((\delta/2)-1)}K' \geq (k_1-1)^{\alpha/2}[H_{\alpha/2}(k_2-1) - H_{\alpha/2}(k_1-2)]. \tag{45}
$$

Combining (44) and (45), for $k_1 > 1$, we obtain the approximation value of k_1 scales as

$$
n^{(1-v)((\delta/2)-1)}K' \simeq (k_1-1)^{\alpha/2}[H_{\alpha/2}(k_2-1) - H_{\alpha/2}(k_1-1)]. \tag{46}
$$

Similarly, we can derive the approximation value of k_2. We know that k_2 is the smallest index in set K_3 such that $N_{k_2} < n^{(1/2)-(1-v)((\delta/2)-(3/2))}$, which result in

$$
n^{(1/2)+(1-v)((\delta/2)-(3/2))}K' < k_2^{\alpha/2}[H_{\alpha/2}(k_2) - H_{\alpha/2}(k_1-1)], \tag{47}
$$

decreasing k_2 by one, we have $N_{k_2-1} \geq n^{(1/2)-(1-v)((\delta/2)-(3/2))}$, which leads to

$$
n^{(1/2)+(1-v)((\delta/2)-(3/2))}K' \geq (k_2-1)^{\alpha/2}[H_{\alpha/2}(k_2-1) - H_{\alpha/2}(k_1-1)]. \tag{48}
$$

Therefore, combining (47) and (48), we obtain the approximation value of k_2 scales as

$$
n^{(1/2)+(1-v)((\delta/2)-(3/2))}K' \simeq (k_2-1)^{\alpha/2}[H_{\alpha/2}(k_2-1) - H_{\alpha/2}(k_1-1)]. \tag{49}
$$

The value of k_1 and k_2 is determined by the different value of α. Therefore, we have following three conditions.

(1)For $\alpha > 2$: We can simplify (46) as

$$
n^{(1-v)((\delta/2)-1)}\left[K - \frac{(k_1-1)}{n^{(1-v)((\delta/2)-1)}} \right] \simeq (k_1-1)^{\alpha/2}\frac{-(k_1-1)^{1-(\alpha/2)}}{1-(\alpha/2)}, \tag{50}
$$

which results in

$$
k_1 \simeq 1 + \frac{\alpha-2}{2}n^{(1-v)((\delta/2)-1)}K. \tag{51}
$$

According to (43), k_2 can be obtained by

$$
k_2 \simeq \frac{\alpha-2}{2}n^{((\alpha-2)/\alpha)(1-v)((\delta/2)-1)+(2/\alpha)}K. \tag{52}
$$

(2)For $\alpha = 2$: We assume $k_2 \leq M$, and (49) can be simplified as

$$
n^{(1/2)+(1-v)((\delta/2)-(3/2))}\left[K - \frac{(k_1-1)}{n^{(1-v)((\delta/2)-1)}} \right] \simeq (k_2-1)\log k_2, \tag{53}
$$

which follows that

$$
(k_2-1)\left[1 + \frac{n^{(1-v)((\delta/2)-(3/2))-(1/2)}}{\log k_2} \right] \simeq \frac{n^{(1/2)+(1-v)((\delta/2)-(3/2))}}{\log k_2}K, \tag{54}
$$

resulting in $k_2 \simeq nK$. This contradicts $k_2 \leq M = n^\gamma, 0 < \gamma < 1$. Thus, we obtain $k_2 = M + 1$.

Assuming $k_1 > 1$, we can simplify (46) as

$$
n^{(1-v)((\delta/2)-1)}\left[K - \frac{(k_1-1)}{n^{(1-v)((\delta/2)-1)}} \right] \simeq (k_1-1)\log k_2, \tag{55}
$$

by using (43), we obtain $k_1 = \Theta(n^{(1-v)((\delta/2)-1)}/\log M)$.

(3)For $0 < \alpha < 2$: Assuming $1 \le k_1 < k_2 \le M$, we can simplify (49) as

$$n^{(1/2)+(1-\nu)((\delta/2)-(3/2))}\left[K - \frac{(k_1-1)}{n^{(1-\nu)((\delta/2)-1)}}\right] \simeq (k_2-1)^{\alpha/2}\frac{(k_2-1)^{1-(\alpha/2)}}{1-(\alpha/2)},$$
(56)

which results in

$$k_2 \simeq n^{(2-\alpha/\alpha)(1-\nu)(1-(\delta/2))+(2/\alpha)}K.$$
(57)

This contradicts $k_2 \le M$. Hence, we have $k_2 = M + 1$.
By simplifying (46), we have

$$n^{(1-\nu)((\delta/2)-1)}\left[K - \frac{(k_1-1)}{n^{(1-\nu)((\delta/2)-1)}}\right] \simeq (k_1-1)^{\alpha/2}M^{1-(\alpha/2)},$$
(58)

resulting in $k_1 = \Theta(n^{(2/\alpha)(1-\nu)((\delta/2)-1)}/M^{(2/\alpha)-1})$.

Theorem 16. *In the clustered grid model, letting $\omega = (1-\nu)((\delta/2)-1)$, according to the Zipf distribution, the optimal delay and throughput are given by w.h.p.*

$$D(n)^{a\circ} = \begin{cases} \Theta(1), & \alpha > 2, \\[2mm] \Theta\left(\dfrac{\log^2 M}{n^\omega}\right), & \alpha = 2, \\[2mm] \Theta\left(\dfrac{M^{2-\alpha}}{n^\omega}\right), & 1 < \alpha < 2, \\[2mm] \Theta\left(\dfrac{M}{n^\omega \log M}\right), & \alpha = 1, \\[2mm] \Theta\left(\dfrac{M}{n^\omega}\right), & 0 < \alpha < 1, \end{cases}$$
(59)

$$\lambda(n)^{a\circ} = \begin{cases} \Theta(n^{-\omega}), & \alpha > 2, \\[2mm] \Theta\left(\dfrac{1}{\log^2 M}\right), & \alpha = 2, \\[2mm] \Theta\left(\dfrac{1}{M^{2-\alpha}}\right), & 1 < \alpha < 2, \\[2mm] \Theta\left(\dfrac{\log M}{M}\right), & \alpha = 1, \\[2mm] \Theta\left(\dfrac{1}{M}\right), & 0 < \alpha < 1. \end{cases}$$

Proof. We first prove the optimal delay based on the values of k_1 and k_2. Then, the optimal throughput can be easily derived by the tradeoff relation given by (26). Substituting the value of p_k into (41), we have

$$D(n)^{a\circ} = \Theta\left(\frac{H_\alpha(k_1-1)}{H_\alpha(M)} + \frac{[H_{\alpha/2}(k_2-1)-H_{\alpha/2}(k_1-1)]^2}{n^{(1-\nu)(\delta/2-1)}K'H_\alpha(M)} + n^{\nu/2}\frac{H_\alpha(M)-H_\alpha(k_2-1)}{H_\alpha(M)}\right).$$
(60)

We divide the RHS of (60) into three terms, which are denoted by D_1, D_2, and D_3, respectively. Then, we have following five conditions according to the values of α, k_1, and k_2.

(1)For $\alpha > 2$: We have $D_1 = \Theta(1)$, and $D_2 = o(1)$. If $k_2 = M + 1$, we have $D_3 = 0$, else if $k_2 = \Theta(n^{((\alpha-2)/\alpha)(1-\nu)((\delta/2)-1)+(2/\alpha)})$, we have $D_3 = o(1)$, when $n \longrightarrow \infty$. Hence, we have $D^{a\circ} = \Theta(1)$.

(2)For $\alpha = 2$: We have $D_1 = \Theta(1)$, and $D_2 = \Theta((\log^2 M)/n^{(1-\nu)((\delta/2)-1)})$. Since $k_2 = M + 1$, we have we have $D_3 = 0$. Hence, we have $D^{a\circ} = \Theta((\log^2 M)/n^{(1-\nu)((\delta/2)-1)})$.

(3)For $1 < \alpha < 2$: We have $D_1 = \Theta(1)$, and $D_2 = \Theta(M^{2-\alpha}/n^{(1-\nu)((\delta/2)-1)})$. D_3 is same with the condition (2), i.e., $D_3 = 0$. Hence, we have $D^{a\circ} = \Theta(M^{2-\alpha}/n^{(1-\nu)((\delta/2)-1)})$.

(4)For $\alpha = 1$: We have $D_1 = \Theta(1)$, and $D_2 = \Theta(M/n^{(1-\nu)((\delta/2)-1)}\log M)$. D_3 is same with the condition (2), i.e., $D_3 = 0$. Hence, we have $D^{a\circ} = \Theta(M/n^{(1-\nu)((\delta/2)-1)}\log M)$.

(5)For $0 < \alpha < 1$: We have $D_1 = \Theta(1)$, and $D_2 = \Theta(M/n^{(1-\nu)((\delta/2)-1)})$. D_3 is same with condition (2), i.e., $D_3 = 0$. Hence, we have $D^{a\circ} = \Theta(M/n^{(1-\nu)((\delta/2)-1)})$.

In Figure 3(a), we have plotted the optimal delay performance of the clustered grid model for different values of α according to the Theorem 16. Similarly, the optimal throughput results for various values of α are plotted in Figure 3(b). We adopt $\nu = 0.5$ and $\gamma = 0.8$ for both figures. We observe that the delay curves appear the ascending tendency while the throughput curves present descending tendency as the number of mobile nodes n increases. We further find that when the number of nodes is fixed, the optimal delay shows a decreasing trend while the optimal throughput increases as α increases. From the simulations, we can conclude that the most popular contents are mainly cached and transmitted by mobile nodes when α is large, which reduces the content transmission time and increases the number of simultaneous transmissions. That is, the advantage of caching is more obvious as α increases.

4.2. Clustered Random Model. In the clustered random model, we first formulate the optimal delay problem according to Theorem 13, and then, we design the optimal cache allocation strategy to achieve the minimum delay. Combining Theorem 13 and the cache constraint, the objective function of minimum delay is given by

$$\begin{aligned} \underset{N_k}{\text{minimize}} \quad & \sum_{k=1}^{M} p_k \mathbb{E}'(H_k) \\ \text{subject to} \quad & \sum_{k=1}^{M} N_k \le nK, \\ & 1 \le N_k \le n^{1-(1-\nu)((\delta/2)-1)}\log^{-(\delta/2)}n, \end{aligned}$$
(61)

(a) Delay
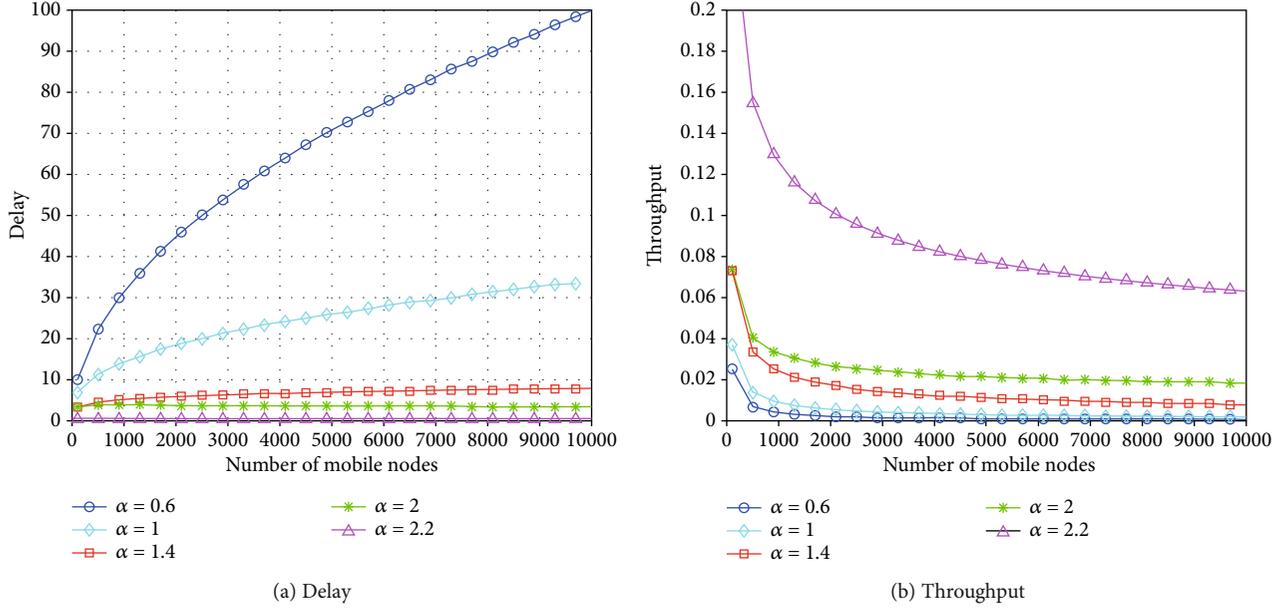
(b) Throughput

FIGURE 3: Delay and throughput performance of the clustered grid model for $\nu = 0.5$ and $\gamma = 0.8$.

(a) Delay
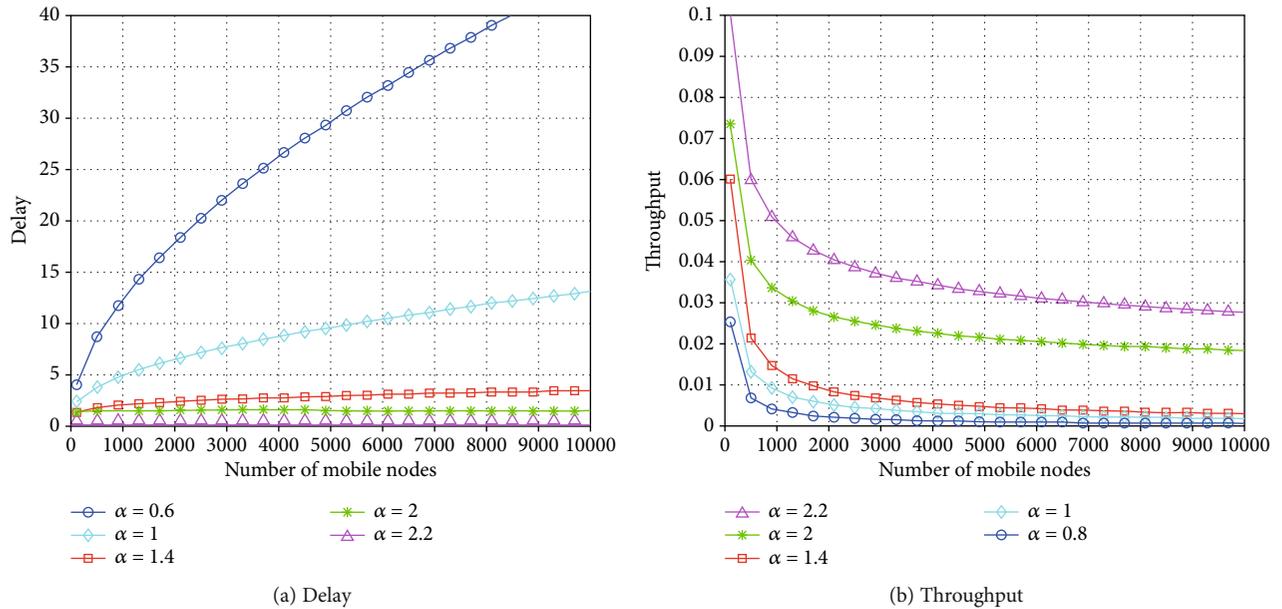
(b) Throughput

FIGURE 4: Delay and throughput performance of the clustered random model for $\nu = 0.5$ and $\gamma = 0.8$.

where $\mathbb{E}'(H_k)$ is given by

$$\mathbb{E}'(H_k)$$
$$= \begin{cases} \Theta(1), & N_k = \Omega\left(\dfrac{n^{1-(1-\nu)((\delta/2)-1)}}{\log^{\delta/2} n}\right), \\[3ex] \Theta\left(\dfrac{n^{1-(1-\nu)((\delta/2)-1)}}{N_k \log^{\delta/2} n}\right), & N_k = o\left(\dfrac{n^{1-(1-\nu)((\delta/2)-1)}}{\log^{\delta/2} n}\right) \& N_k = \Omega\left(\dfrac{n^{(1/2)-(1-\nu)((\delta/2)-(3/2))}}{\log^{\delta/2} n}\right), \\[3ex] \Theta\left(\dfrac{n^{\nu/2}}{\sqrt{\log n}}\right), & N_k = o\left(\dfrac{n^{(1/2)-(1-\nu)((\delta/2)-(3/2))}}{\log^{\delta/2} n}\right). \end{cases}$$

$$(62)$$

The optimization process is similar with the clustered grid model. Hence, we give a brief description.

We first define three sets K'_1, K'_2, and K'_3 according to the size of N_k, respectively. Let $K'_1 = \{1, 2, \cdots, k_1 - 1\}$ be the set of contents such that $N_k = n^{1-(1-\nu)((\delta/2)-1)}/\log^{\delta/2} n$, let $K'_2 = \{k_1, k_1 + 1, \cdots, k_2 - 1\}$ be the set of contents such that $N_k \in [n^{(1/2)-(1-\nu)((\delta/2)-(3/2))}/\log^{\delta/2} n, n^{1-(1-\nu)((\delta/2)-1)}/\log^{\delta/2} n)$. And let $K'_3 = \{k_2, k_2 + 1, \cdots, M\}$ be the set of contents such that $N_k = 1$.

Then, we take the Lagrangian multiplier $\lambda \in R^+$ for the first constraint in (61) and combine with the second

(a) Delay

(b) Throughput
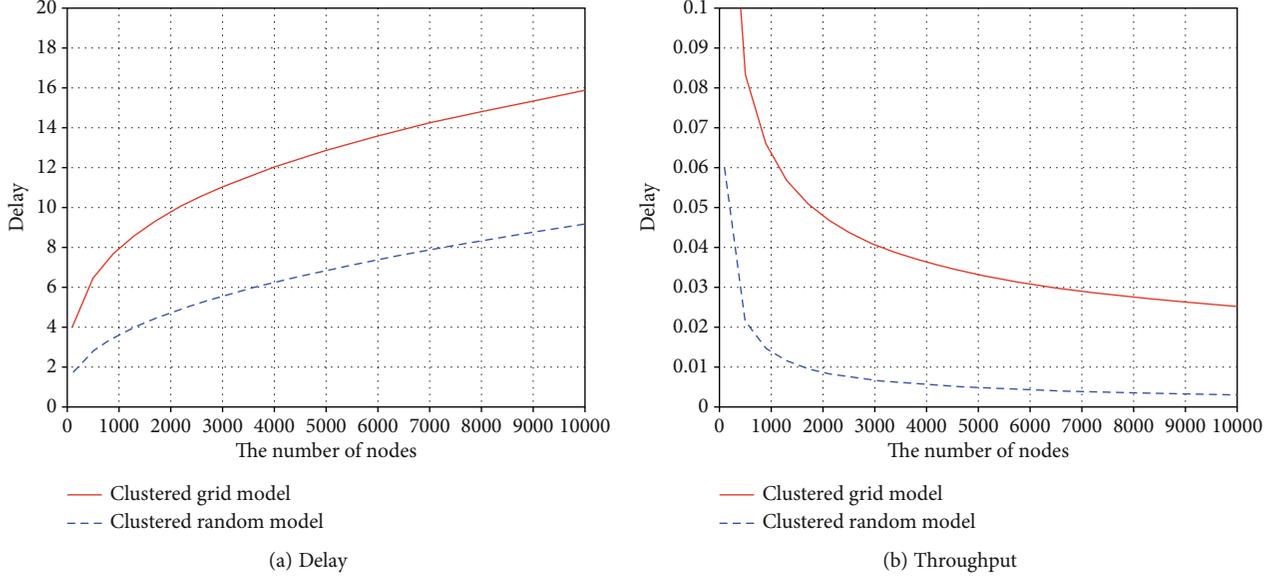
FIGURE 5: Comparisons of scaling performance for the clustered grid model and the clustered random model when $\alpha = 1.5$.

constraint; the necessary conditions for the minimal $D(n)$ are given as:

$$\frac{\partial D(n)}{\partial N_k} = -\frac{p_k n^{1-(1-v)((\delta/2)-1)}}{N_k^2 \log^{\delta/2} n} \begin{cases} \leq -\lambda, & \forall k \in K'_1, \\ = -\lambda, & \forall k \in K'_2, \\ \geq -\lambda, & \forall k \in K'_3. \end{cases} \quad (63)$$

Next, we apply the similar calculation method to obtain the values of $N_k^{a\circ}$. Hence, the minimum delay in the clustered random model is given by

$$D(n)^{a\circ} = \Theta \left(\sum_{k=1}^{k_1-1} p_k + \frac{\left(\sum_{k=k_1}^{k_2-1} p_k^{1/2} \right)^2}{n^{(1-v)((\delta/2)-1)} K' \log^{\delta/2} n} + \sum_{k=k_2}^{M} \frac{n^{v/2} p_k}{\sqrt{\log n}} \right). \quad (64)$$

Before obtaining the minimum delay $D(n)^{a\circ}$, we need to estimate the value of k_1 and k_2 by the following lemma.

Lemma 17. *In the clustered random model, for $n \longrightarrow \infty$, the values of k_1 and k_2 is given as follows*

$$k_1 = \begin{cases} \Theta \left(n^{(1-v)((\delta/2)-1)} \log^{(\delta/2)} n \right), & \alpha > 2, \\[2mm] \Theta \left(\dfrac{n^{(1-v)((\delta/2)-1)} \log^{\delta/2} n}{\log M} \right), & \alpha = 2, \\[3mm] \min \left\{ \Theta \left(n^{(1-v)((\delta/2)-1)} \log^{\delta/2} n \right), \Theta \left(\dfrac{M n^{(2/\alpha)(1-v)((\delta/2)-1)-(2/\alpha)}}{\log^{-\delta/2}} \right) \right\}, & 0 < \alpha < 2, \end{cases}$$

$$k_2 = \begin{cases} \min \left\{ M+1, \Theta \left(n^{(\alpha-2)/\alpha(1-v)((\delta/2)-1)+(2/\alpha)} \log^{\delta(\alpha-2)/2\alpha} n \right) \right\}, & \alpha > 2, \\[2mm] M+1, & \alpha = 2, \\[2mm] \min \left\{ M+1, \Theta \left(\dfrac{n^{(2-\alpha)/\alpha(1-v)(1-(\delta/2))+(2/\alpha)}}{\log^{\delta(\alpha-2)/2\alpha} n} \right) \right\}, & 0 < \alpha < 2. \end{cases} \quad (65)$$

Proof. The proof approach of this lemma is similar to that in Lemma 15. We omit it for simplicity.

Combining Lemma 17 and the different values of α, the optimimal delay and throughput in the clustered random model are given as follows:

Theorem 18. *In the clustered random model, letting $\omega = (1 - \nu)((\delta/2) - 1)$, according to the Zipf distribution, the optimal delay and throughput are given by w.h.p.*

$$
D(n)^{a_\circ} =
\begin{cases}
\Theta(1), & \alpha > 2, \\[2ex]
\Theta\left(\dfrac{\log^2 M}{n^\omega \log^{\delta/2} n}\right), & \alpha = 2, \\[2ex]
\min\left\{\Theta\left(\dfrac{M^{2-\alpha}}{n^\omega \log^{\delta/2} n}\right), \Theta\left(\dfrac{n^{-\left((\alpha^2-3\alpha+4)/\alpha\right)\omega + ((4-2\alpha)/\alpha)}}{\log^{\left(\delta((2-\alpha)^2+\alpha)\right)/2\alpha} n}\right)\right\}, & 1 < \alpha < 2, \\[2ex]
\min\left\{\Theta\left(\dfrac{M}{n^\omega \log^{\delta/2} n \log M}\right), \Theta\left(\dfrac{n^{2(1-\nu)(1-(\delta/2))+2}}{\log M \log^\delta n}\right)\right\}, & \alpha = 1, \\[2ex]
\min\left\{\Theta\left(\dfrac{M}{n^\omega \log^{\delta/2} n}\right), \Theta\left(\dfrac{n^{-\left((\alpha^2-3\alpha+4)/\alpha\right)\omega + ((4-2\alpha)/\alpha)}}{M^{1-\alpha} \log^{\left(\delta((2-\alpha)^2+\alpha)\right)/2\alpha} n}\right)\right\}, & 0 < \alpha < 1,
\end{cases}
$$

$$
\lambda(n)^{a_\circ} =
\begin{cases}
\Theta\left(\dfrac{n^{-\omega}}{\log^{\delta/2} n}\right), & \alpha > 2, \\[2ex]
\Theta\left(\dfrac{1}{\log^2 M}\right), & \alpha = 2, \\[2ex]
\max\left\{\Theta\left(\dfrac{1}{M^{2-\alpha}}\right), \Theta\left(\dfrac{\log^{\left(\delta(2-\alpha)^2\right)/2\alpha} n}{n^{-((\alpha^2-4\alpha+4)/\alpha)\omega + ((4-2\alpha)/\alpha)}}\right)\right\}, & 1 < \alpha < 2, \\[2ex]
\max\left\{\Theta\left(\dfrac{\log M}{M}\right), \Theta\left(\dfrac{\log M \log^{\delta/2} n}{n^{(1-\nu)(1-(\delta/2))+2}}\right)\right\}, & \alpha = 1, \\[2ex]
\max\left\{\Theta\left(\dfrac{1}{M}\right), \Theta\left(\dfrac{M^{1-\alpha} \log^{\left(\delta(2-\alpha)^2\right)/2\alpha} n}{n^{-((\alpha^2-4\alpha+4)/\alpha)\omega + ((4-2\alpha)/\alpha)}}\right)\right\}, & 0 < \alpha < 1.
\end{cases}
$$

$$(66)$$

Proof. The proof of this theorem is similar to that in Theorem 16. Hence, we neglect it for simplicity.

In Figure 4, we depict the optimal delay and throughput performance of the clustered random model for different values of α. We also adopt $\nu = 0.5$ and $\gamma = 0.8$ for both figures. In particular, for $\alpha < 2$, we choose the smaller value of the two delay and we choose the larger value of the two throughput. In Figure 5, we plot the comparison results of the scaling performance of the different clustering models for $\alpha = 1.5$. We observe that the delay performance is better in the clustered random model than that in the clustered grid model while the throughput degrades in the clustered random model. This can be explained that the transmission range in the clustered random model is larger than that in the clustered grid model, which leads to the decreased transmission hops in the clustered random model accordingly. On the contrary, the larger transmission range covers more network areas, which results in the decrease concurrent transmissions.

5. Conclusions and Future Works

In this paper, we analyzed the throughput and delay performance of mobile content-centric networks, where the node spatial distribution of nodes is not uniform. We adopted a cell-partition TDMA scheduling scheme and proposed a distributed multihop routing scheme to achieve the throughput-delay tradeoff in the clustered grid model and clustered random model, respectively. Moreover, according to the content popularity distribution, we applied Lagrangian relaxation method to optimize the cache allocation in two kinds of clustered models. Finally, we obtained the optimal throughput and delay in the mobile content-centric networks. Our theoretical results were validated by the numerical simulations.

There are several problems left for future research, for example, the scaling laws and the caching optimization for mobile content-centric networks with infrastructures support. In addition, the multicast capacity of nonuniform mobile content-centric networks has not been studied. Finally, it is meaningful to optimize cache allocation strategy when the distribution of contents is unknown.

Data Availability

The data used to support the findings of this study are available from the authors upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in *Proc. 5th Int Conf. Emerg. Netw. Experim. Technol*, pp. 1–12, Rome, Italy, 2009.

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

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

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

[5] R. Talak, S. Karaman, and E. Modiano, "Capacity and delay dcaling for broadcast transmission in highly mobile wireless networks," in *IEEE Trans*, Mobile Comput, 2019.

[6] X. Lin and N. B. Shroff, "The fundamental capacity-delay tradeoff in large mobile ad hoc networks," in *Proc. 3rd Annu. Medit. Ad Hoc Netw. Workshop*, pp. 1–12, Bodrum, Turkey, 2004.

[7] R. Jia, F. Yang, S. Yao et al., "Optimal capacity-delay tradeoff in MANETs with correlation of node mobility," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1772–1785, 2017.

[8] K. Guan, G. Li, T. Kurner et al., "On millimeter wave and THz mobile radio channel for smart rail mobility," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 5658–5674, 2017.

[9] K. Guan, D. He, B. Ai et al., "5-GHz obstructed vehicle-to-vehicle channel characterization for internet of intelligent vehicles," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 100–110, 2019.

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

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

[12] J. Ren, G. Zhang, and D. Li, "Multicast capacity for VANETs with directional antenna and delay constraint under random walk mobility model," *IEEE Access*, vol. 5, pp. 3958–3970, 2017.

[13] R. Hou, Y. Cheng, J. Li, M. Sheng, and K. S. Lui, "Capacity of hybrid wireless networks with long-range social contacts behavior," *IEEE/ACM Transactions on Networking*, vol. 25, no. 2, pp. 834–848, 2017.

[14] X. Wang, Z. Wang, and Q. Liang, "Outage throughput capacity of hybrid wireless networks over fading channels," *IEEE Access*, vol. 8, pp. 867–875, 2020.

[15] L. Zhu, F. R. Yu, B. Ning, and T. Tang, "Cross-layer handoff design in MIMO-enabled WLANs for communication-based train control (CBTC) systems," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 719–728, 2012.

[16] J. Zhang, L. Fu, and X. Wang, "Asymptotic analysis on secrecy capacity in large-scale wireless networks," *IEEE/ACM Transactions on Networking*, vol. 22, no. 1, pp. 66–79, 2014.

[17] L. Zhu, Y. Li, F. R. Yu, B. Ning, T. Tang, and X. Wang, "Cross-layer defense methods for jamming-resistant CBTC systems," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2020.

[18] J. Song, M. Sheng, T. Quek, C. Xu, and X. Wang, "Learning-based content caching and sharing for wireless networks," *IEEE Transactions on Communications*, vol. 65, no. 10, pp. 4309–4324, 2017.

[19] L. Zhu, Y. He, F. R. Yu, B. Ning, T. Tang, and N. Zhao, "Communication-based train control system performance optimization using deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10705–10717, 2017.

[20] S. Gitzenis, G. S. Paschos, and L. Tassiulas, "Asymptotic laws for joint content replication and delivery in wireless networks," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2760–2776, 2013.

[21] B. Azimdoost, C. Westphal, and H. R. Sadjadpour, "Fundamental limits on throughput capacity in information-centric networks," *IEEE Transactions on Communications*, vol. 64, no. 12, pp. 5037–5049, 2016.

[22] S. W. Jeon, S. N. Hong, M. Ji, G. Caire, and A. F. Molisch, "Wireless multihop device-to-device caching networks," *IEEE Transactions on Information Theory*, vol. 63, no. 3, pp. 1662–1676, 2017.

[23] M. Mahdian and E. M. Yeh, "Throughput and delay scaling of content-centric ad hoc and heterogeneous wireless networks," *IEEE/ACM Transactions on Networking*, vol. 25, no. 5, pp. 3030–3043, 2017.

[24] G. Zhang, J. Liu, J. Ren, L. Wang, and J. Zhang, "Capacity of content-centric hybrid wireless networks," *IEEE Access*, vol. 5, pp. 1449–1459, 2017.

[25] M.-C. Lee, M. Ji, A. F. Molisch, and N. Sastry, "Throughput-outage analysis and evaluation of cache-aided D2D networks with measured popularity distributions," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5316–5332, 2019.

[26] L. Qiu and G. Cao, "Popularity-aware caching increases the capacity of wireless networks," *IEEE Transactions on Mobile Computing*, vol. 19, no. 1, pp. 173–187, 2020.

[27] T. A. Do, S. W. Jeon, and W. Y. Shin, "How to cache in mobile hybrid IoT networks?," *IEEE Access*, vol. 7, pp. 27814–27828, 2019.

[28] A. Malik, S. H. Lim, and W.-Y. Shin, "On the effects of sub-packetization in content-centric mobile networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 8, pp. 1721–1736, 2018.

[29] G. Alfano, M. Garetto, and E. Leonardi, "Content-centric wireless networks with limited buffers: when mobility hurts," *IEEE/ACM Transactions on Networking*, vol. 24, no. 1, pp. 299–311, 2016.

[30] J. Luo, J. Zhang, Y. Cui, L. Yu, and X. Wang, "Asymptotic analysis on content placement and retrieval in MANETs," *IEEE/ACM Transactions on Networking*, vol. 25, no. 2, pp. 1103–1118, 2017.

[31] G. Alfano, M. Garetto, and E. Leonardi, "Capacity scaling of wireless networks with inhomogeneous node density: upper bounds," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 7, pp. 1147–1157, 2009.

[32] M. Gatetto, P. Giaccone, and E. Leonardi, "Capacity scaling in ad hoc networks with heterogeneous mobile nodes: the super-critical regime," *IEEE/ACM Transactions on Networking*, vol. 17, no. 5, pp. 1522–1535, 2009.

[33] M. Gatetto, P. Giaccone, and E. Leonardi, "Capacity scaling in ad hoc networks with heterogeneous mobile nodes: the sub-critical regime," *IEEE/ACM Transactions on Networking*, vol. 17, no. 6, pp. 1888–1901, 2009.

[34] X. Liu, L. Fu, J. Wang, X. Wang, and G. Chen, "Multicast scaling of capacity and energy efficiency in heterogeneous wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 15, no. 3, pp. 1–32, 2019.

[35] K. Zheng, Y. Cui, X. Liu, X. Wang, X. Jiang, and J. Tian, "Asymptotic analysis of inhomogeneous information-centric wireless networks with infrastructure support," *The IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 5245–5259, 2018.

[36] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," in *Proc. IEEE INFOCOM*, pp. 1–13, Barcelona, Spain, 2006.

[37] N. Lu, T. H. Luan, M. Wang, X. Shen, and F. Bai, "Bounds of asymptotic performance limits of social-proximity vehicular networks," *IEEE/ACM Transactions on Networking*, vol. 22, no. 3, pp. 812–825, 2014.

[38] T. Wang, X. Cao, and S. Wang, "Self-adaptive clustering and load-bandwidth management for uplink enhancement in heterogeneous vehicular networks," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5607–5617, 2019.

[39] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: evidence and implications," in *Proc. IEEE INFOCOM*, pp. 126–134, New York, USA, 1999.

[40] R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge Univ. Press, Cambridge, U. K., 1995.

WILEY | Hindawi

Research Article

LTE-U based Train to Train Communication System in CBTC: System Desin and Reliability Analysis

Hao Liang [ID],[1] **Hongli Zhao** [ID],[2] **Shuo Wang** [ID],[1] and **Yong Zhang** [ID][1]

[1]*School of Electronics and Information, Beijing Jiaotong University, 100044 Beijing, China*
[2]*State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, 100044 Beijing, China*

Correspondence should be addressed to Hongli Zhao; hlzhao@bjtu.edu.cn

Communication-Based Train Control (CBTC) system is a critical signal system to ensure rail transit's safe operation. Compared with the train-ground CBTC system, the train control system based on train-to-train (T2T) communication has the advantages of fast response speed, simple structure, and low operating cost. As the core part of the train control system based on T2T communication, the reliability of the data communication system (DCS) is of great significance to ensure the train's safe and efficient operation. According to the T2T communication system requirements, this paper adopts Long-Term Evolution-Unlicensed (LTE-U) technology to design the DCS structure and establishes the reliability model of the communication system based on Deterministic and Stochastic Petri Nets (DSPN). Based on testing the real line's communication performance parameters, the DSPN model is simulated and solved by π-tool, and the reliability index of the system is obtained. The research results show that the LTE-U-based T2T communication system designed in this paper meets the train control system's needs for communication transmission. This paper's reliability evaluation method can complete the reliability modeling of train control DCS based on T2T communication. The research in this paper will provide a strong practical and theoretical basis for the design and optimization of train control DCS based on T2T communication.

1. Introduction

Today, the CBTC system has been widely used in urban rail transit systems [1]. The existing CBTC system is technically very mature. However, there are still certain shortcomings: the system structure is involved, which leads to too many interfaces between subsystems; the information transmission process is cumbersome, and the on-board equipment cannot directly obtain relevant information, which limits the increase in train speed; the system consists of more equipment, resulting in higher system construction and operating costs, which is not conducive to the large-scale development of urban rail transit [2].

As an improvement of the train control system based on train-to-ground communication, the train control system based on T2T communication has many advantages: simple structure, short transmission time, strong system flexibility, and low operating cost. Therefore, train control systems based on T2T communication have become an important

development direction. Alstom implemented a streamlined CBTC system on line 1 of Lille, France, and first proposed a train control system based on T2T communication. There is no operating line of the train control system based on T2T communication in China, but there have been many related studies. Lin and others introduced T2T communication technology based on the existing train control system to achieve train collision protection, designed the overall architecture of the system, and analyzed the simplified system through the reliability block diagram [3]. Guan et al. conducted a series of detections on the T2T communication channel and investigated the influence of train operation on the communication channel [4]. Liu compared the new train control system with the traditional CBTC system in terms of structure, performance, maintenance, and backup mode and obtained the system [5]. These research results provide basic theoretical knowledge for the design of the train control DCS based on T2T communication. However, the existing research focuses on the performance of the communication

system, and there has not yet been comprehensive systematic research on the reliability of T2T communication.

The train control DCS based on T2T communication reduces the system coupling. It simplifies the train communication process, thereby significantly reducing the DCS complexity and improving the system's efficiency. Also, the wireless communication technology used in the DCS has a more significant impact on the data transmission process between train control subsystems related to the quality of communication. Compared with wireless local area network (WLAN) technology, LTE technology based on the 1.8 GHz licensed frequency band has strong anti-interference ability, complete mobility management function, high transmission rate, and low system delay. Now, it is successfully applied to urban rail transit [6]. However, with the advancement of smart urban rail applications, the contradiction between insufficient authorized spectrum resources and urban rail data transmission requirements is becoming more dangerous. As a supplementary technology to LTE, LTE-U aggregates authorized frequency band and unlicensed frequency band resources and has many advantages in licensed frequency bands. Simultaneously, the use of unlicensed frequency band resources can effectively alleviate the pressure of tight spectrum resources to meet the future development needs of urban rail transit.

LTE-U technology was first proposed at the 3GPP RAN62 summit in 2013, and 3GPP introduced it as a critical technology in LTE R13 [7]. Qing analyzed the technical advantages of LTE-U in applying urban rail transit through the professional research of LTE-U [8]. Bai researched the application of LTE-U in the urban rail transit vehicle-ground communication system, and its wireless resource scheduling management and mobility management algorithms [9]. At present, the research on the LTE-U DCS of urban rail transit is in its infancy, so it is of considerable significance to study the design and reliability of train DCS based on LTE-U.

The reliability of the DCS is of great significance for ensuring the safe and efficient operation of trains. The existing set of sophisticated research methods for the reliability research of the train control DCS based on train-to-ground communication is not entirely suitable for the reliability research of the train control system based on T2T communication. Besides, the existing model parameters in the reliability research of the DCS are almost all hypothetical, and the system performance parameters are not obtained from the real train control DCS. The main contributions of this paper are as follows:

(i) According to the requirements of the train control system based on T2T communication, using LTE-U technology, a train control DCS based on T2T communication is designed

(ii) According to the technical characteristics of LTE-U, and the unique needs of train control, the reliability of train control DCS based on T2T communication is fully defined. In order to evaluate the reliability of the system, this paper focuses on the analysis

of typical communication scenarios for T2T communication

(iii) Use DSPN to model T2T communication scenarios during train communication. Through the communication performance parameter test in the real line environment, the reliability of the T2T communication scene is obtained by using the π-tool simulation, and the influence of the communication limit time on the reliability of the communication system is studied

The rest of this paper is organized as follows. Section 2 introduces the design of the T2T communication system based on LTE-U. Section 3 proposes the reliability definition of train control DCS and analyzes and models typical scenarios. In Section 4, reliable performance parameters are obtained by testing the communication link of the actual line. Section 5 examines the reliability of the train control DCS based on the exact line performance parameters. Section 6 summarizes the paper.

2. The Design of Train Control DCS

2.1. LTE-U Wireless Data Communication Technology. In the early days, the DCS of CBTC used WLAN technology based on the IEEE802.11 series protocols to carry out two-way information transmission between train and ground [10]. Although the data wireless communication system based on WLAN can meet traditional urban rail transit requirements for availability, reliability, and safety, it can no longer meet the rapid development of urban rail transit operation requirements in the future. Its limitations mainly include limited operating speed and coverage. The distance is short, the link design is complicated, and wireless interference is dangerous. When multiple services are concurrent, it is impossible to schedule resources according to the priority [11].

LTE-M is an LTE system based on LTE wireless communication technology and customized according to the business needs of urban rail transit [12]. The LTE-M system fully considers the reliability and real-time requirements of urban rail transit business and combines the advantages of the LTE network to form a systematic and standardized solution, which makes up for the disadvantage of WLAN technology carrying CBTC service, and provides innovative technical support for the safe operation of urban rail transit. However, at present, LTE-M systems occupy the proprietary frequency band of 1785 MHz-1805 MHz [13] and are only allowed to use 20 MHz bandwidth resources at most. At the same time, CBTC services adopt dual-network redundancy mode, and their single-network services usually need 5 MHz bandwidth, so integrated bearer service networks often use 15 MHz bandwidth resources at most. In addition, if the frequency transfer method is used to solve the interference problem of the public-private network, the wireless resources available in the urban rail transit system will be further reduced, and the comprehensive carrying service network is easy to block the network capacity. Reduce the performance of the communication system. Reduce the

performance of the communication system. Therefore, new technology is needed to solve the problem of lack of spectrum resources to meet the growing needs of urban rail transit.

LTE-U aggregates authorized band and unlicensed band resources through carrier aggregation (CA) technology and introduces it into urban rail transit DCS to supplement the limited spectrum resources of the LTE-M system. While the authorized band provides reliable and stable transmission, an unlicensed band provides high throughput transmission to meet the future development needs of urban rail transit. LTE-U inherits many advanced technologies of the LTE system and introduces its unique channel sharing strategy. LTE-U is the same as LTE-M in network architecture. Still, the frequency band of the RF work unit is different, including Evolved Packet Core (EPC), Base Band Unit (BBU), Radio Remote Unit (RRU), and User Equipment (UE).

At present, there are mainly three bands of unauthorized spectrum available for LTE-U, namely, the 2.4 GHz group for the industry, science, and medicine, the 5 GHz band for unlicensed international information facilities, and the newly proposed 28~60 GHz millimeter-wave band [14]. Therefore, in the application of urban rail, LTE-U works in the 5 GHz group, which will have abundant unauthorized spectrum resources and have a good application prospect. In order to realize the integrated service bearer with a high transmission rate and real-time requirements in the high-speed operation scenario, LTE technology is introduced into the 5.8 GHz open frequency band, that is, LTE-U can achieve full high-speed data transmission of nonsecure information with higher security and environmental adaptability based on more precious frequency resources and higher spectrum efficiency [15].

As a new technology supported by 3GPP and MulteFire alliance, LTE-U technology has obvious technical advantages in urban rail transit applications, such as excellent mobility and handoff performance, small transmission delay, strong anti-interference ability, large coverage radius, good QoS guarantee, high security, and high reliability. Therefore, the application prospect of LTE-U technology in urban rail transit is broad.

2.2. Train Control System Based on T2T Communication. The train control system based on T2T communication simplifies the structure of train control system based on train-ground communication, reduces the ground equipment Zone Controller (ZC) and Computer Interlocking (CI), integrates its functions into on-board equipment, and adds Train Management Unit (TMU) and Object Controller (OC). The basic structure is shown in Figure 1.

In the process of train operation, the train runs according to the operation plan issued by the Automatic Train Supervision (ATS). When it is about to enter the next section of the line, the train actively sends a communication request to the TMU of the section, registers the identity information, and queries other train information in the line. TMU stores the information on the car and responds to the query of the train. According to the identity information obtained, the train establishes the communication between train and train and requests the location information of other trains. The train

uses the speed measurement and positioning equipment and the transponder to determine the position of the train on the line, map it to the electronic map, and judge the logical relationship between the front and rear trains.

The on-board equipment calculates the endpoint of the Movement Authority (MA) through the interlocking module in the Vehicle On-Board Controller (VOBC) according to the driving plan, the status of the signal equipment beside the track, and the position information of the front train on the line. The Automatic Train Protection (ATP) module will calculate the train speed protection curve based on the MA, line information, temporary speed limit, and other information and update it in real time. Automatic Train Operation (ATO) completes the functions of train traction, speed adjustment, and automatic return under the supervision of ATP. At the same time, the train will check the status of the associated signal equipment in the route and request the control of the trackside equipment from the OC. OC responds to the query, returns the state of the device, and transfers control rights. When the train is about to clear from the line section, the train actively disconnects the communication connection with other trains on the line and cancels the train information in the TMU.

2.3. The Design of the T2T Communication System Based on LTE-U. Based on the network structure of the LTE-U communication system, combined with the structure of the train control system based on T2T communication, a train control DCS based on T2T communication using LTE-U communication technology is designed. The system structure is shown in Figure 2. ATS interconnects with the core network through the backbone network, and the core network is connected with the base station on the line. Trains within the signal coverage of the base station can communicate with ATS through wireless communication.

The specific functions of the network elements in the LTE-U system are as follows:

(1) Core network EPC: EPC is a full IP packet core network, and all services are switched in the Packet Switch (PS) domain. EPC core network includes Mobility Management Entity (MME), Serving Gateway (SG), Packet Data Node Gateway (PGW), Home Subscriber Server (HSS), and other network elements. The primary function of MME is to control signaling transmission, authenticate and page the terminal, and realize mobility management such as handover and roaming. SGW routes and forwards packet data in the system. The functions of PGW include checking, filtering and filtering data, and interworking with networks outside the LTE system. HSS stores and manages user data and authentication data, including location information of mobile users

(2) Wireless access network E-UTRAN: wireless access network E-UTRAN is composed of a plurality of evolved NodeB (eNodeB). The communication network is connected wirelessly by eNodeB users. eNodeB has the functions of routing, establishing or
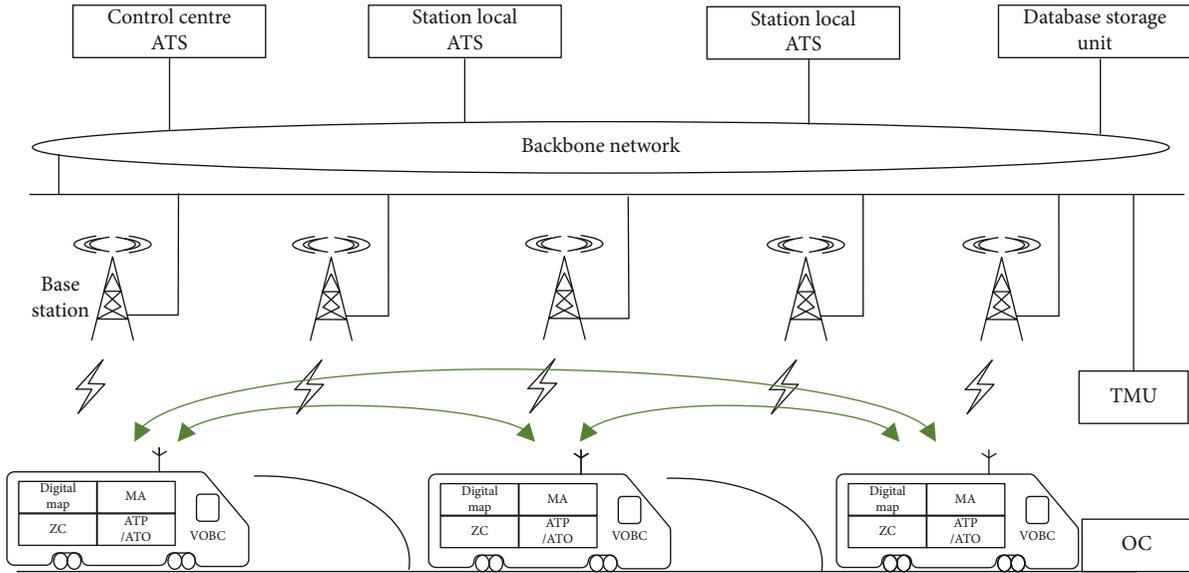
FIGURE 1: Basic structure diagram of the train control system based on T2T communication.
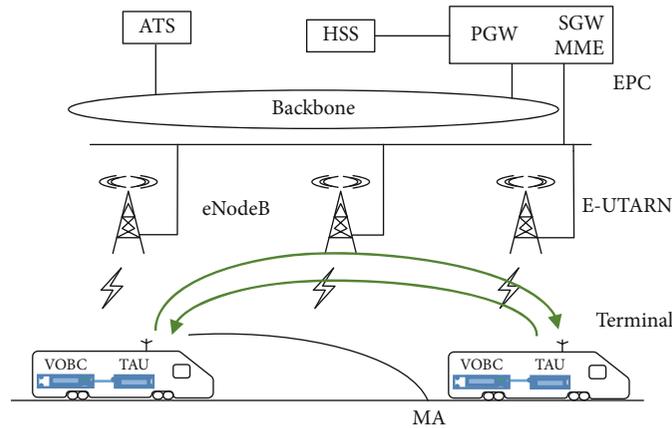
FIGURE 2: Train control DCS based on T2T communication.

releasing wireless links, mobility management, scheduling and allocation of wireless resources, bearer control, etc.

(3) Terminal: train installs train access unit (TAU) as the service data transmission equipment to realize wireless communication. The TAU is connected with the functional modules on the train to send the information that the train needs to upload through the antenna, and then, the information is transmitted to the ground equipment or other trains. At the same time, the TAU receives the information forwarded by the base station through the antenna and sends it to the train equipment after processing

3. Reliability Definition and Scenario Models of Train Control DCS

3.1. Data Transmission Flows of Train Control DCS Based on T2T Communication. Because the structure of train control

DCS based on LTE-U technology is different from that based on WLAN technology, the data communication flow is different from the previous communication system, so it is necessary to analyze the information transmission flow between train and train. As shown in Figure 3, the solid line is the data flow, and the dashed line is the signaling flow.

Curve 1 is the signaling flow channel after the train is electrified. When the train starts, TAU searches for cell signals, selects the cell, and triggers the random access process, ensuring communication links between terminals and networks. TAU sends a connection establishment request signal to the base station, allocating the radio resources for it. Simultaneously, through the S1-MME port, the base station sends the attachment request to the MME. After completing the TAU authentication, the MME sends the attached signal to the base station, and the base station sends the TAU capability information to the MME. Finally, the base station sends a security activation message, and after TAU activation, it sends configuration completion and attachment completion messages to the base station and MME.

Curve 2 is the data flow channel of train communication under different base stations. The transmitting terminal sends messages to the base station through the empty port, and the base station forwards the messages to the SGW through the S1-U port. After the SGW inquiry, the messages are forwarded to the base station in the cell where the receiving terminal TAU is located and then to the receiving terminal TAU.

When the train reaches the overlap of base station signals, the service area has to be changed. Curve 3 is the channel of signaling flow based on X2 port switching. First, the source station sends measurement control down to TAU, which feeds back measurement reports. The source base station decides to switch based on the measurement report and then sends the switching request to the destination base station. After receiving the acknowledgment signal, the source base station sends configuration information to TAU. After the TAU configuration is completed, the source base station begins switching and sends the connection success information to the destination base station. The base station requests MME to update the data channel. TAU releases the cell resources under the source base station to complete the switch between stations.

3.2. Reliability Definition of Train Control DCS. The reliability of train control DCS affects the probability of a successful exchange of train information, so it is vital to use reasonable methods to evaluate its reliability. Based on the communication performance parameters, this paper defines the reliability of DCS from two links.

The first link is the reliability of data transmission. The reliability of data transmission is defined as the probability that the available information reaches the receiving end within a limited time. Both transmission delay and continuous packet loss affect the reliability of data transmission. According to the definition of reliability, as shown in Figure 4, (a) the packet delay $t1$ is greater than the limit time T; then, the transmission of packet 1 is unreliable. Otherwise, the transfer of packet 2 is reliable; (b) when a continuous packet loss occurs, the duration of no data reception at the receiving end can be calculated according to the packet sending rate. If the term is greater than T, the data transmission is not reliable at this time.

The second link is the reliability of the communication scenario. During the train operation, communication processes such as data retransmission, switching between stations, and interrupt reconnection are required. Therefore, it is stipulated that the train needs to complete the communication process within the communication limit time. After the communication limit time is exceeded, the train will be in an unsafe operation state, and the reliability of the communication scenario is the probability that the train is in a safe state.

Since the existing reliability studies are mostly based on assumed parameters, in order to make the data communication system reliability research results more in line with the actual situation, the parameters of the model in this paper are derived from actual communication performance test results or the requirements of the communication system in
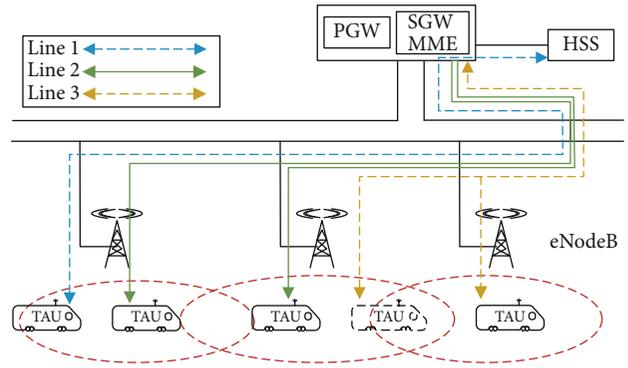
FIGURE 3: Signaling transmission flow and data transmission flow.

the urban rail transit communication specification, and the analysis results have higher practical significance.

3.3. Typical Scenario Analysis of T2T Communication. In the train control system based on T2T communication, the communication between trains is the primary demand, among which the communication scene of redundancy train access unit and roaming switching communication scene are typical communication scenes in the process of train operation.

3.3.1. Communication Scenarios of Redundancy Train Access Units. T2T communication is an important communication link in the process of train operation. However, the channel environment changes caused by geographical location, weather conditions, and other factors will reduce the communication quality, affect the real time and availability of the data received by the train, and then affect the regular operation of the train and the passenger's riding experience. In the existing research, it is a standard method to improve communication reliability using equipment redundancy. Therefore, the communication scenario of the redundant train access unit is typical in the T2T communication scenario.

As shown in Figure 5, two sets of train access units and on-board antennas are, respectively, configured at the head and tail ends of the train, and the two sets of communication equipment work independently and do not interfere with each other. Therefore, only one set of train access units can successfully receive reliable data within the specified time to ensure the train's regular operation. After receiving the information, the two sets of train access units process the information and obtain the available information required by on-board equipment, including train position information and obstacle information. Simultaneously, if the trackside equipment status and route information are obtained, the logic module of on-board equipment can calculate according to the relevant information to obtain the train operation permission terminal point. The emergency braking curve and automatic driving curve are calculated by the ATP and ATO logic module of on-board equipment, and the train is in normal operation.

3.3.2. Communication Scenarios of On-Board Terminal Handoff. Train cross-line operation is an inevitable requirement for the development of urban rail transit. It is necessary

(a) Unreliable transmission caused by delay

(b) Unreliable transmission caused by packet loss
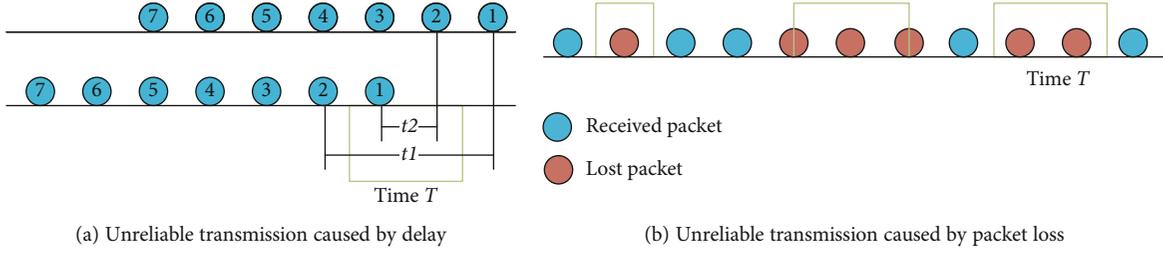
FIGURE 4: The definition of data transmission reliability.
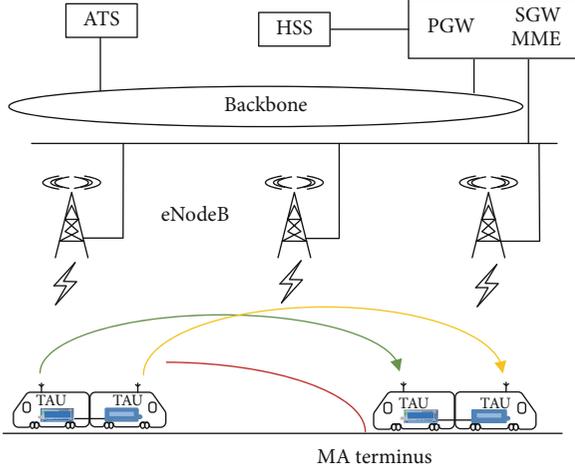
FIGURE 5: Communication scenario of redundancy train access units.

to ensure that the train can achieve on-board terminal roaming, realize interconnection, and save train resources. When the train is in the process of roaming switching, other trains on the line still need to communicate with each other. The roaming switching process has a more significant impact on the communication delay.

As shown in Figure 6, on the existing operation line, one set of communication equipment is used on the front line due to the construction planning. In contrast, another set of communication equipment is used on the follow-up line. When the train runs to the junction of two sets of core network equipment, it needs to carry out roaming switching.

As shown in Figure 7, when the train is roaming switching, the information sent by the ground equipment and other trains will be forwarded to the destination core network through the source core network and then to the base station of the switching train. After the roaming switching is completed, the train will obtain the information of the ground equipment and other trains on the line through the forwarding of the destination core network, which will increase large transmission delay. When the train crosses a line, the running interval of the line will change, and the train speed will be limited. To improve operation efficiency, it is necessary to quickly adjust the relevant parameters of train operation, so it is necessary to ensure the communication between trains.

3.4. DSPN Models of T2T Communication Scenarios

3.4.1. Deterministic and Stochastic Petri Net.

The train control data DCS is a complex dynamic system, and its data communication has large randomness and high correlation with time. Therefore, this paper uses Deterministic and Stochastic Petri Net (DSPN) to model the train control DCS based on its data transmission characteristics.

A system of DSPN can be defined as a decimal group: $\Sigma = (S, T, I, O, H, Y, K, \lambda, W, M_0)$. The store holds tokens, representing a network element of the system or a step in the process. In a DSPN, the maximum number of tokens that a store can hold can be set, the number of elements in the set K for that store. Therefore, the state of the Petri net can be defined by the number of tokens in each store, as

$$m = (\#(p1), \#(p2), \cdots, \#(pk)), \tag{1}$$

where m is also called the identification vector of the Petri net. Each element is the number of tokens in the store p, and it must satisfy $M(p) \leq K(p)$. The set of m is the set of system states.

In Petri nets, the triggering of changes represents events in the actual system, causing the system to transition from one state to another. Moreover, in stochastic Petri nets, transitions have implementation delays, and implementation delays are associated with a particular distribution, and the set of transition implementation rates is λ. In DSPN, according to different implementation delays, there are three changes:

(1) The implementation delay is zero. Such a transition is called an immediate transition, indicating a controlling role or a logical choice that does not require processing time

(2) The implementation delay is a random variable. This kind of vicissitude is one kind of time vicissitude, which is called exponential vicissitude. It indicates that the event takes a certain amount of time, and the distribution of time follows an exponential distribution: $F(t) = 1 - e^{-\lambda t} \lambda > 0$

(3) The implementation delay is a definite value. This kind of transition is called fixed-time transition, which means the implementation delay of the transition corresponds to the processing time in the real system
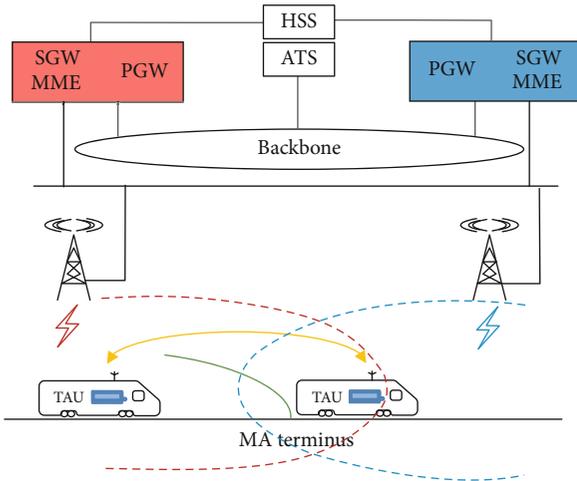
FIGURE 6: Train roaming communication scenario.

Besides, the Petri net visual simulation tool π-tool is used to establish a communication model and simulate [16–18]. The π-tool can realize the establishment of many kinds of Petri net models. It is a perfect modeling and simulation tool and has many advantages:

(i) Model building: the implementation rate of transitions in the software can be set to a variety of common distributions. For immediate transitions, the weight and priority of the transitions can be set. Ability to build structured models

(ii) Model analysis: the Petri net model can be deadlocked to detect system resource allocation and consumption. The reachability graph can be automatically generated according to the model to determine the state space and state transition of the system. Using Monte Carlo simulation analysis, the Petri net model with any transition can be analyzed to obtain the occupancy rate of the store, and a certain transition can be specifically observed to determine the time when it is first triggered

In the π-tool graphical interface, each element of the DSPN is graphically represented, as shown in Table 1.

3.4.2. Models of Communication Scenarios for Redundancy Train Access Units.

The model is shown in Figure 8. First, the train is in the typical running state *run*, and then triggers transition *GenMsg* to start data transmission. At the same time, triggering transition *trans*, timer *timer* is started. This timer is the communication limit time for the train to complete an individual link. After that, the token is transferred to the store *train0*, which means that the information is transmitted to another train, triggering the transition *eNodeB*, and the token is transferred to the store *train*, which means that the data is transmitted to the base station. At this time, there are two situations. The train is in the process of switching between stations or interrupting the reconnection process. The transition *handover* indicates that the train is performing the switching process. The stores *HOX2-TAU1.Msg* and *HOX2TAU2.Msg* are the switching processes required by the two train access units. The transition interrupt means that the train's reconnection process is performed, and the stores *AttachTAU1.connection* and *AttachTAU2.connection* are the reconnection process required by the two train access units. This paper considers the most unfavorable situation: the handover and reconnection processes need to start from the initial signaling.

In this communication scenario, if both train access units need to switch between stations, and in the case where one train access unit switches successfully, the switching time of the other train access unit is no longer limited. The method for judging whether the reconnection of the two train access units is successful is the same as the handover. Only one of the train access units completes the reconnection process within the communication limit can be regarded as a successful reconnection. Taking the first train access unit successfully switching or reconnecting as an example, the switching or reconnecting submodel transfers the token to store *success*. At this time, through the limiting effect of the test arc, stores *AttachTUA2.fail* and *HOX2TAU2.fail* will obtain the cause the token generated by the transition *fail-TAU2* that triggers the second train access unit to switch or reconnect the token in the submodel, and the model can start the next information transmission. However, if the timer has reached the communication limit time during the handover or reconnection, it is judged that the handover or reconnection is unsuccessful. At this time, all tokens in the submodel are cleared, and the system state changes from *normal* to *danger* until the next time the train successfully completes the handover or reconnection within the specified time; the system status changes to *normal*.

3.4.3. Models of Communication Scenarios of On-Board Terminal Handoff.

Figure 9 shows the DSPN model. This is the top-level model. There are three bottom-level submodels under the top-level model, namely, the *HOS1*, *Transmission*, and *TAUpd* submodel. When the train runs, it generates information that needs to be sent to other trains, that is, the token is transferred from the store *run* to *Msg*. At this time, the transition *trans* is triggered, and the token is transferred to the timer *timer*, which limits the time of the communication process. At the same time, a token is transferred to the store *train0* and data transmission begins. The transition *eNodeB* trigger indicates that the data has been transmitted to the base station. In the most unfavorable situation, another train needs to complete all interactive processes of roaming switching.

First, the train performs the cross-core network switching process, and the token is transferred to the submodel *HOS1* through the associated store *HOS1.handover*, and the store *PN.HOsucces* representing the completion of the switch in the submodel is associated with the *HOsuccess* store in the top model. After the submodel process is completed, the token is transferred to *HOsuccess*. Afterwards, the data is forwarded. The data is forwarded from the source core network to another core network. The store *Transmission.data* is associated with the store *data* in the submodel *Transmission*. In the submodel, after completing the process of data

FIGURE 7: Data forwarding process.

TABLE 1: Graphic elements and their meanings of DSPN.

Element	Symbol	Meaning
Store	◯	The store in the hierarchical model
	◯	The store associated with the previous or next level
Token	•	Resources in the store and their quantity
Transition	▯	Exponential transition
	▮	Fixed time transition
	▮	Immediate transition
Normal arc	⟶	Directed arcs, input and output functions between locations and transitions
Suppression arc	----○	Restrictions on transition, when there are a number of tokens corresponding to the arc weight in the store onnected to the transition use prohibition arc, the transition prohibition triggers
Test arc	----▸	Restrictions on the transition, when there are a number of tokens corresponding to the arc weight in the store connected to the transition test arc, the transition is allowed to trigger

FIGURE 8: Models of communication scenarios for redundancy train access unit.
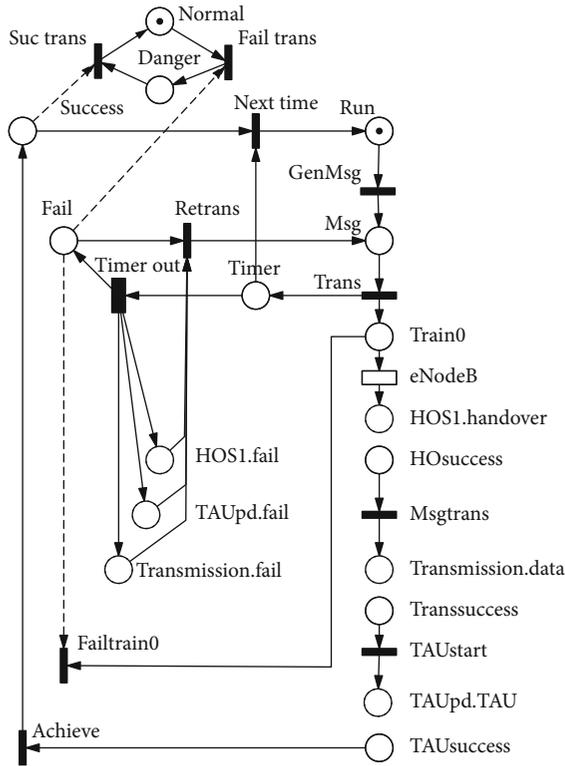
FIGURE 9: Models of communication scenarios of on-board terminal handoff.

transmission and retransmission, the token enters the store *transsuccess* in the top model through the associated store *PN.transsuccess*, indicating that the data forwarding process has been completed. Finally, by triggering the transition *TAUstart*, the tracking area update process begins. Similarly, the token enters the store *TAU* in the *TAUpd* submodel through the associated store *TAUpd.TAU*. After the token completes the tracking area update process in the signaling interaction sequence in the submodel, the token is transferred to the *TAUsuccess* store in the top model through the associated store *PN.TAUsuccess*. If the cross-core network handover process, data forwarding process, and tracking area update process are completed within the communication limit time, the next round of information transmission can begin. However, if the train fails to complete the roaming handover process when the timer reaches the communication limit time, indicating that the cross-core network handover fails, the top model triggers the *timeout* transition to cause orders to appear in the associated stores *HO21.fail*, *TAUpd.fail*, and *Transmission.fail* of the three submodels Card and then empty the token in the submodel. At the same time, the system state changes from *normal* to *danger* until the next successful roaming switching process; the system state changes to *normal*.

4. Communication Performance Test of Train Control Data Link

The transmission performance of the train control data link is very critical and affects the reliability of communication.

The existing reliability studies are mostly based on assumed performance parameters. Therefore, to make the research results of DCS reliability more in line with the actual situation, the model parameters in this paper need to be derived from actual communication performance test results or urban rail transit communications, the requirements for the communication system in the specification; such analysis results have higher practical significance. To determine the communication quality of the wireless link and obtain the transmission performance of the link during train operation, this paper is based on the actual operating line of Beijing metro line 15 to test the communication performance parameters of the link. The test section of the operating line is the 2 stations 1 section upstream section (including the station line) from Maquanying to Sunhe, a total of 3.3 km, of which the 2.3 km section east of Maquanying Station is an underground tunnel, which is immediately adjacent to the viaduct, with a length of 1 kilometer to Sunhe Station, as shown in Figure 10.

The LTE-U system test in the test section of line 15 mainly includes three parts: communication room center equipment (EPC/switch, CCTV/PIS ground equipment), trackside base station equipment (8 sets), and on-board equipment (TAU/switch). Test networking is shown in Figure 11. Among them, base station 1 to base station 5 are covered by antennas with a distance of 400~450 m, one side of base station 6 is covered by an antenna, a slotted waveguide covers the other side, and base station 7 and base station 8 are covered by the slotted waveguide.

First, test the data transmission delay and handover delay when the train is running. During the train operation, the train is within the signal coverage of a base station for a specified period. At this stage, the data transmission delay is only determined by the network environment and channel quality. When the train travels within the signal overlap range of the two base stations, the handover mechanism, handover between stations is required, and handover delay occurs. When the train is running on the line, seven handovers occur in a single journey, and seven handover delay data can be obtained. To obtain more handover delay data, the train runs back and forth between stations, and the handover delay is measured multiple times.

Figure 12 shows the delay and handover delay results obtained in a one-way test. At certain moments, the time delay will be significantly increased. The test time in the figure can be divided into eight stages, with a higher delay as the demarcation point. Each stage indicates that the train is within the signal coverage of an individual base station, corresponding to eight stages. For the signal coverage of each base station, a higher delay indicates that the train is at the junction of the base station's signal coverage. Therefore, there are seven handover delays in the figure. In the subsequent analysis, the delay and handover delay in all test data must be separately counted.

In order to avoid mutual influence when measuring the delay and packet loss at the same time, the delay and packet loss tests are carried out separately. To ensure sufficient data volume, the train runs multiple times to test and obtain packet loss data. If the software operation is interrupted, it
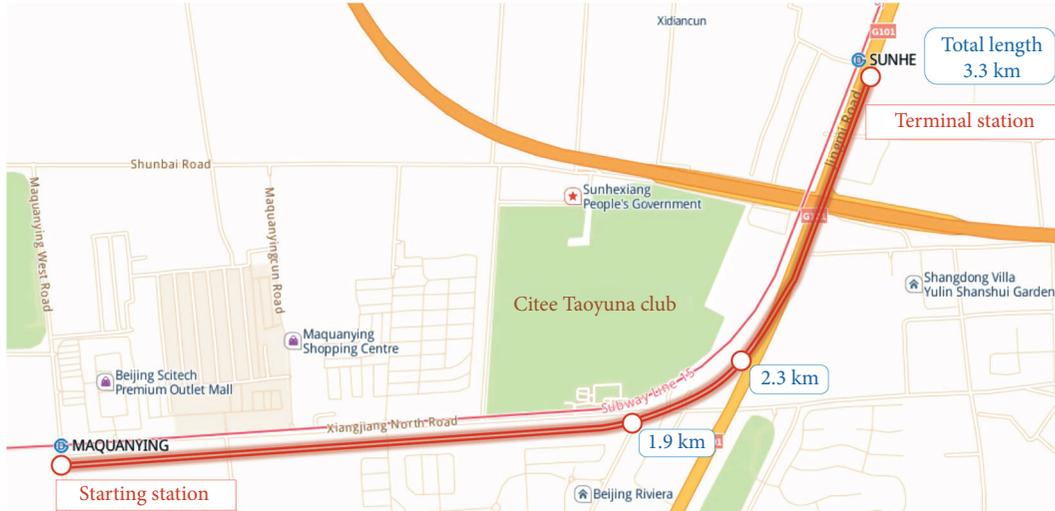
FIGURE 10: The test section of Beijing metro line 15.
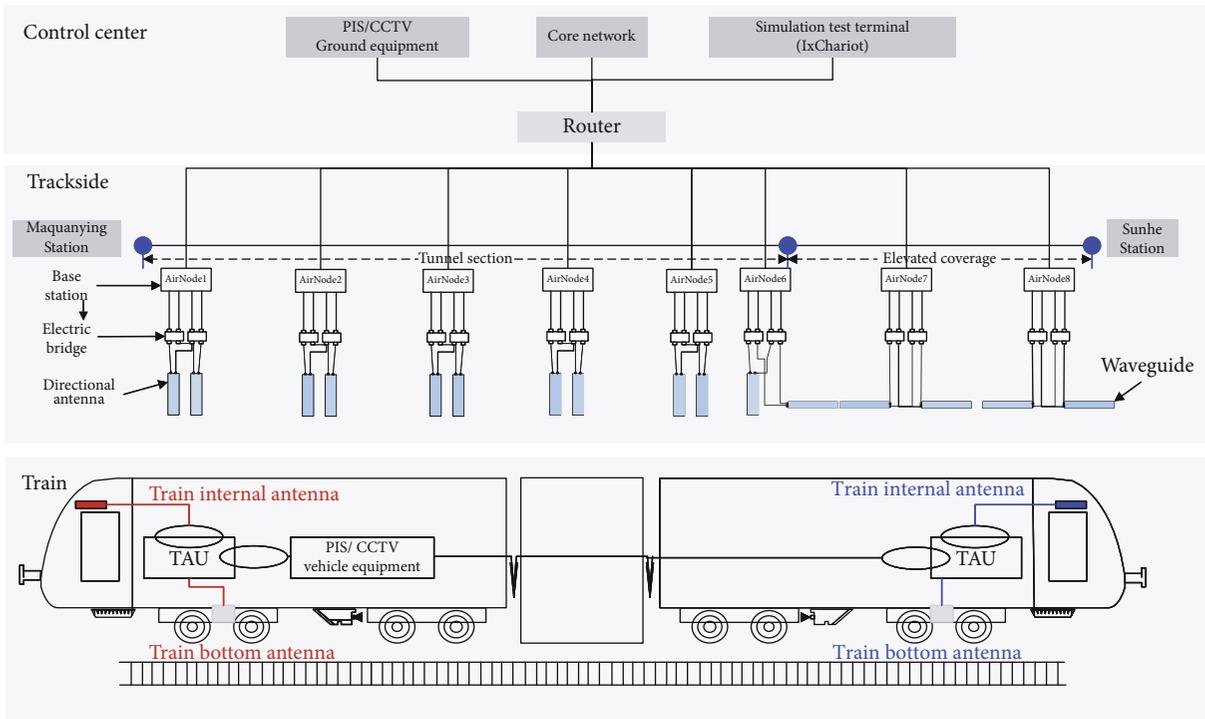
FIGURE 11: 5.8G system network structure diagram of line 15.

will affect the accuracy of the test results. Therefore, discard this set of data and perform it again to ensure the accuracy of the measured data. In the actual test, the packet loss rate test result is 0.497%. Figure 13 shows the test result of continuous packet loss.

5. The Reliability Analysis of Train Control DCS

5.1. Reliability Analysis of Communication Scenarios for Redundant Train Access Units. This paper uses the Monte Carlo simulation method, and according to the reliability definition of the communication process in 3.1, under the condition that the parameter value of the fixed transition timeout is different, the model is simulated. In practice, the parameter value of timeout represents the time limit of the signaling process performed on the train, from which the reliability under different communication time limits can be analyzed. The simulation obtains the reliability when the train is switched or reconnected separately under different timeout parameters.

According to the simulation results, as the timeout parameter value becomes more extensive, the communication scenario's reliability gradually increases. When the parameter value is 990 ms, the reliability of the communication scenario
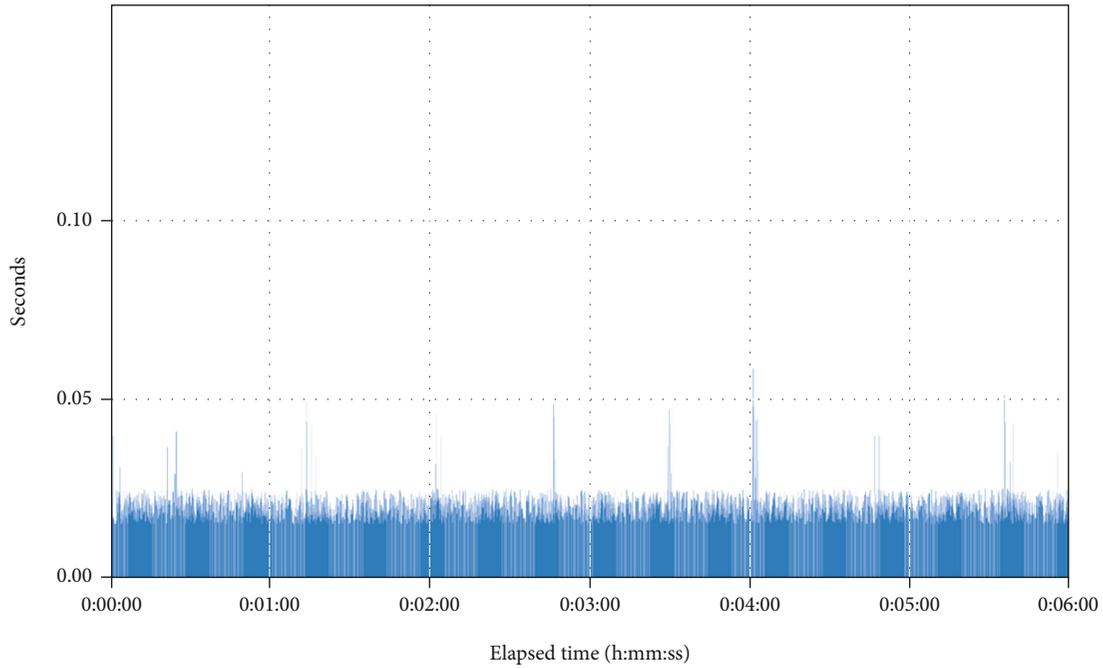
Figure 12: Test results for transmission delay in the real environment.
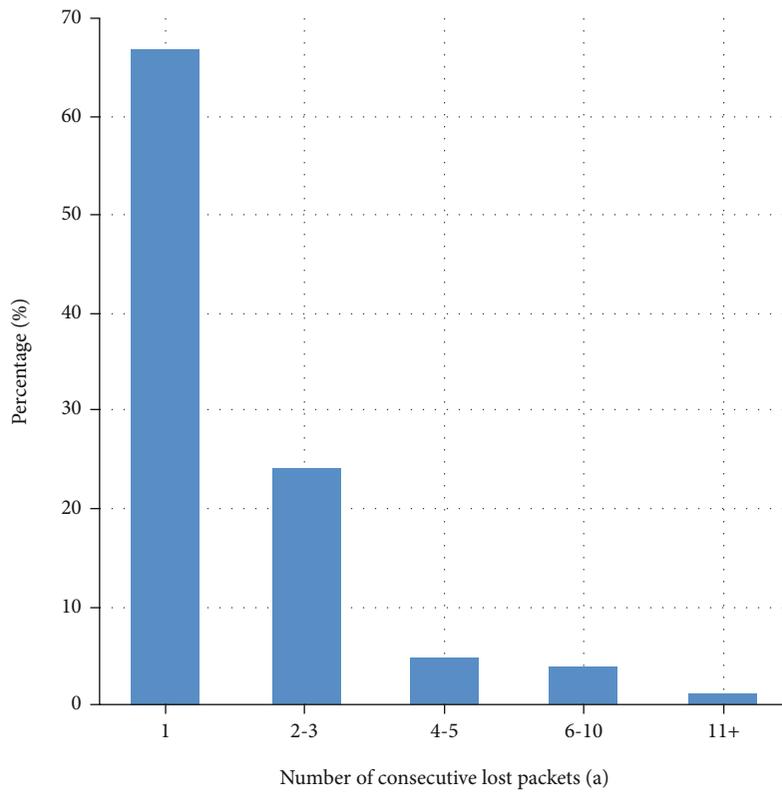
Figure 13: Test results for consecutive lost datagrams in the real environment.

for train switching reaches 99.9928%, while when the parameter value is 1050 ms, the reliability of the communication scenario for train reconnection can reach 99.9902%. Simultaneously, the results show that the reconnection process reduces train communication's reliability to a certain extent, affecting the communication between the train and other trains.

As a comparison, consider the reliability of the communication scenario when the train has only one train access unit. According to the simulation results, when the
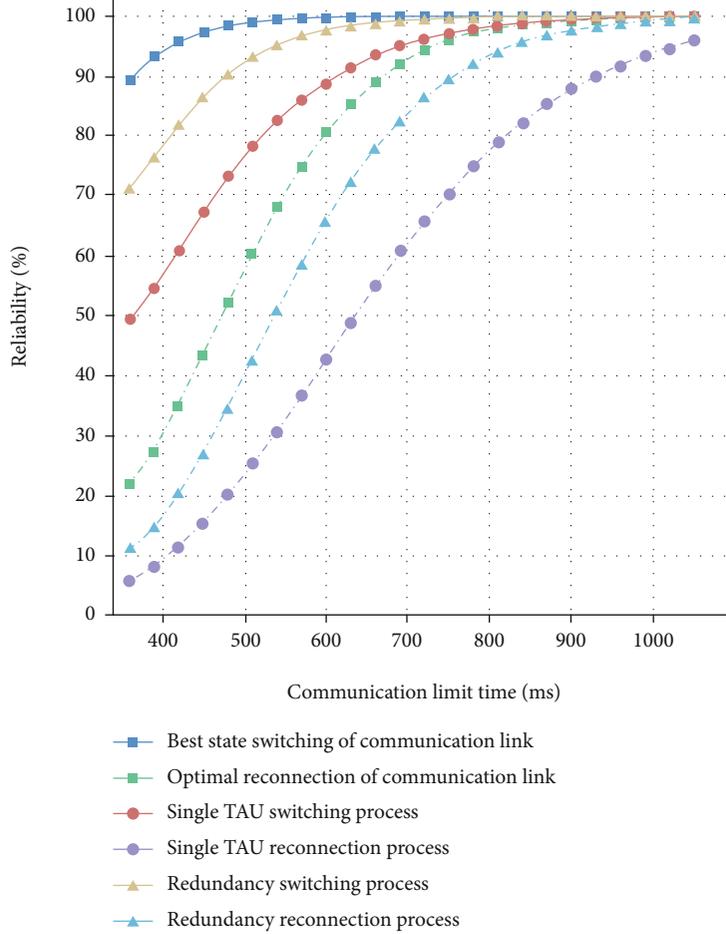
FIGURE 14: The reliability of communication scenarios of redundancy train access units.

communication limit time is short, whether it is a switching process or a reconnection process, the train access unit's reliability is greatly improved when the train access unit is redundant. As it becomes larger, the reliability improvement rate gradually decreases.

Without considering the cable transmission delay and retransmission, the time for handover and reconnection process is related to the wireless transmission delay, and its time conforms to the Gamma distribution, so it can be solved by its density distribution function to complete the wireless under different communication limit time, the probability of transmission. The density function is shown in

$$f(t) = \frac{t^{\alpha-1}\lambda^{\alpha}e^{-\lambda t}}{(\alpha-1)!}, \quad t > 0, \tag{2}$$

where α is the number of signaling that needs to be transmitted using the air interface during the signaling interaction.

In many cases, the reliability of the communication scenario is shown in Figure 14. It can be obtained that the train access unit has better communication reliability when there is redundancy, and its reliability can reach 99% when the communication limit time is 960 ms, and the reliability exceeds 99.99% when the communication limit time exceeds 1050 ms. In the case of nonredundant train access units, the communication limit time needs to be 1230 ms when the reliability reaches 99%, and when it exceeds 1710 ms, the reliability can exceed 99.99%.

5.2. Reliability Analysis of Communication Scenarios for On-Board Terminal Handoff. According to the definition of the communication process's reliability, the probability of the train being in the normal state when the parameter value of the transition timeout is different is studied. Therefore, the parameter value of the timeout transition in the top model can be set to a series of values. Under different parameter values, the Monte Carlo method simulates the model to obtain the probability that the train is in the normal state. According to the simulation results, the higher the communication process threshold, the greater the reliability of the communication. When the communication limit time is 1140 ms, the reliability exceeds 99%, and when the communication limit time is 1300 ms, the reliability of the train can reach 99.99%. Simultaneously, with the increase of the communication limit time, the improved range of reliability gradually decreases, indicating that the influence of the communication limit time on reliability gradually decreases.
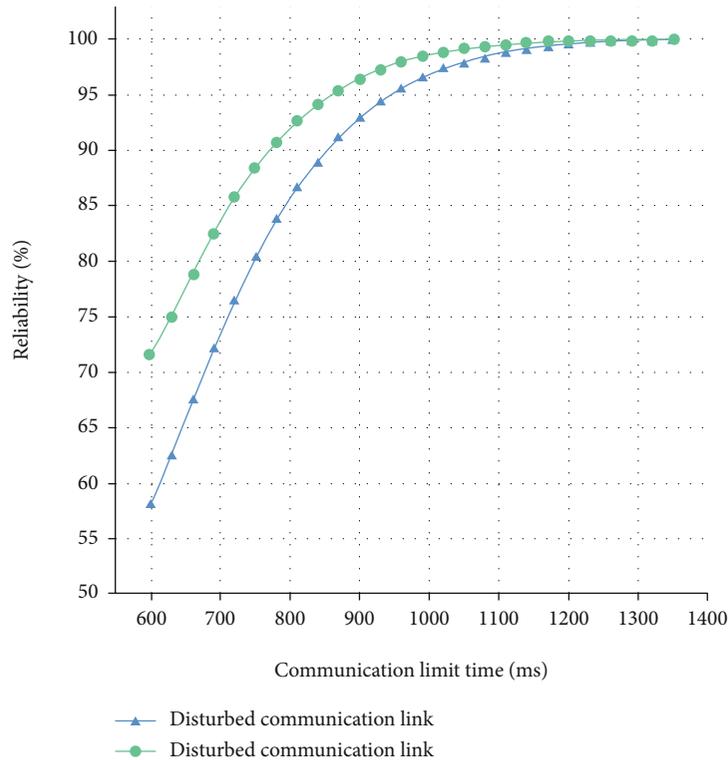
FIGURE 15: The reliability of a communication scenario for on-board terminal handoff.

Without considering the delay and retransmission of cable transmission between ground equipment, the time for the roaming switching communication process conforms to the Gamma distribution, and its density function is Equation (2). Figure 15 shows the reliability of the communication scenario. It can be obtained that when the communication link is in the best state and the communication threshold is less than 1100 ms, the reliability of the communication scenario is higher than the reliability when the link has interference. After the communication limit time gradually increases, the difference in reliability gradually decreases.

6. Conclusion

This paper combines the train control DCS based on T2T communication with the existing communication reliability analysis methods to study the reliability of communication scenarios. Based on the structure and principle of the train control system based on T2T communication, the basic structure of the train control DCS based on LTE-U communication technology is studied. The reliability of train control DCS is defined. Simultaneously, two critical communication scenarios of redundancy train access unit and on-board terminal handoff are extracted for the T2T communication scenario, and the communication process is analyzed in detail. Use the π-tool to build a DSPN model of the scene. Besides, the actual line's communication performance is tested to provide more accurate performance parameters for the model. By setting different communication limit times, the simulation obtained the reliability of crucial communication scenarios. The results show that, combined with the reliabil-

ity evaluation method proposed in this paper, DSPN can better realize the communication system models. The reliability research of the DCS in this paper provides an essential reference for the design, implementation, and optimization of the train control DCS based on T2T communication.

Data Availability

The data used to support the findings of this study are owned by a company.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] L. Zhu, F. R. Yu, B. Ning, and T. Tang, "Cross-layer handoff design in MIMO-enabled WLANs for communication-based train control (CBTC) systems," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 719–728, 2012.

[2] Y. Li, L. Zhu, H. Wang, F. R. Yu, and S. Liu, "A Cross-Layer Defense Scheme for Edge Intelligence-Enabled CBTC Systems Against MitM Attacks," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[3] J. Lin, X. Wang, D. Yao, and C. Yan, "Design and research of train collision protection system for urban rail transit," *Journal of Railway Science and Engineering*, vol. 12, no. 2, pp. 407–413, 2015.

[4] K. Guan, B. Peng, D. He et al., "Channel sounding and ray tracing for train-to-train communications at the THz

band[C]//2019 13th European Conference on Antennas and Propagation (EuCAP)," *IEEE*, pp. 1–5, 2019.

[5] L. Jian, "Research on the new generation urban rail transit signal system," *Urban Rail Transit Research*, vol. 22, no. 7, pp. 71–74, 2019.

[6] H. Wang, F. R. Yu, and H. Jiang, "Modeling of radio channels with leaky coaxial cable for LTE-M based CBTC systems," *IEEE Communications Letters*, vol. 20, no. 5, pp. 1038–1041, 2016.

[7] C. Luo, F. R. Yu, H. Ji, and V. C. M. Leung, "Cross-Layer Design for TCP Performance Improvement in Cognitive Radio Networks," *IEEE Transactions Vehicular Technology*, vol. 59, no. 5, pp. 2485–2495, 2010.

[8] Q. Lanhao, "Application of LTE-U in urban rail transit," *Urban rapid rail transit*, vol. 31, no. 5, pp. 5–10, 2018.

[9] B. Kunai, *Research on Wireless Resource and Mobility Management of LTE-U System in Urban Rail Transit*, Beijing Jiaotong University, 2019.

[10] L. Zhu, *Research on Cross-Layer Design in CBTC Train Ground Communication Systems*, Beijing Jiaotong University, 2012.

[11] L. Zhu, Y. He, F. R. Yu, B. Ning, T. Tang, and N. Zhao, "Communication-Based Train Control System Performance Optimization Using Deep Reinforcement Learning," *IEEE Transactions Vehicular Technology*, vol. 66, no. 12, 2017.

[12] J. Moreno, J. M. Riera, L. de Haro, and C. Rodriguez, "A survey on future railway radio communications services: challenges and opportunities," *IEEE Communications Magazine*, vol. 53, no. 10, pp. 62–68, 2015.

[13] L. Zhu, Y. Li, F. R. Yu, B. Ning, T. Tang, and X. Wang, "Cross-layer Defense Methods for Jamming-Resistant CBTC Systems," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[14] L. Sijia, H. Xiaoge, Z. Fan, and C. Qianbin, "Study on coexistence schemes of LTE-U and WiFi on unlicensed bands," *Journal of Chongqing University of Posts and Telecommunications (natural science edition)*, vol. 29, no. 2, pp. 182–189, 2017.

[15] Z. Luo, *LTE-U Network Resource Optimization Based on Machine Learning[D]*, Zhejiang University, Hangzhou, 2019.

[16] S. Yaqing, *DSPN Models and Verification of Vehicle Ground Communication for High Speed Maglev Control System [D]*, Beijing Jiaotong University, 2013.

[17] M. A. Marsan and G. Chiola, *On Petri Nets with Deterministic and Exponentially Distributed Firing Times[C]//European Workshop on Applications and Theory in Petri Nets*, Springer, Berlin, Heidelberg, 1986.

[18] W. Peizhong, *Design and Simulation of Petri Net Visualization Tool*, Xi'an University of Electronic Science and technology, 2012.

WILEY | Hindawi

Research Article

Measurement and Simulation for Vehicle-to-Infrastructure Communications at 3.5 GHz for 5G

Yi Zeng [1], **Haofan Yi** [2,3] **Zijie Xia** [2,3] **Shaoshi Wang** [2,3] **Bo Ai** [2,3] **Dan Fei** [2,3] **Weidan Li** [1] **and Ke Guan** [2,3]

[1]*Guangdong Communications & Networks Institute, 510700 Guangzhou, China*
[2]*State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, 100044 Beijing, China*
[3]*Beijing Engineering Research Center of High-Speed Railway Broadband Mobile Communications, Beijing, China*

Correspondence should be addressed to Haofan Yi; haofanyi@bjtu.edu.cn

Intelligent Transportation System (ITS) is more and more crucial in the modern transportation field, such as the applications of autonomous vehicles, dynamic traffic light sequences, and automatic road enforcement. As the upcoming fifth-generation mobile network (5G) is entering the deployment phase, the idea of cellular vehicle-to-everything (C-V2X) is proposed. The same 5G networks, coming to mobile phones, will also allow vehicles to communicate wirelessly with each other. Hence, 3.5 GHz, as the main sub-6 GHz band licensed in 5G, is focused in our study. In this paper, a comprehensive study on the channel characteristics for vehicle-to-infrastructure (V2I) link at 3.5 GHz frequency band is conducted through channel measurements and ray-tracing (RT) simulations. Firstly, the channel parameters of the V2I link are characterized based on the measurements, including power delay profile (PDP), path loss, root-mean-square (RMS) delay spread, and coherence bandwidth. Then, the measurement-validated RT simulator is utilized to conduct the simulations in order to supplement other channel parameters, in terms of the Ricean K-factor, angular spreads, the cross-correlations of abovementioned parameters, and the autocorrelation of each parameter itself. This work is aimed at helping the researchers understand the channel characteristics of the V2I link at 3.5 GHz and support the link-level and system level design for future vehicular communications of 5G.

1. Introduction

Intelligent Transportation System (ITS) is more and more crucial in the modern transportation field, such as the applications of autonomous vehicles, dynamic traffic light sequences, and automatic road enforcement. Since the vehicles are at high speed, the real-time and fast exchange of dynamic information should be transmitted in a short time from vehicle-to-everything (V2X) [1]. The V2X provides wireless services for vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and vehicle-to-pedestrian (V2P) communications. Such transferred information includes not only small data such as the speed, the location, and the directions of neighboring vehicles but also the big data such as video of surrounding environments and three-dimensional (3D) high-resolution maps. The transportation control system

can use these applications for congestion avoidance, control performance improvement [2], overall traffic efficiency improvement [3], and cyberattack mitigation [4].

In order to provide the reliable and low latency for the V2X services, efforts have been made in recent years to develop V2X communications using IEEE 802.11p [5] and ITS-G5 [6]. Cellular vehicle-to-everything (C-V2X) communication is now regarded as another promising and feasible solution for the fifth-generation mobile networks- (5G-) enabled vehicular communications [7]. C-V2X will allow vehicles to communicate wirelessly with each other, with traffic signs, and with other roadside infrastructures by using the same 5G networks coming to mobile phones.

In [8], the authors propose an architecture that incorporates V2V communications into C-V2X and IEEE 802.11p-based the vehicular ad hoc networks (VANETs). For the

frequency bands, 5.9 GHz band has been standardized in the IEEE 802.11p, while ITS-G5 is operating at 5 GHz frequency band. C-V2X will operate on the cellular networks using the same frequency bands as 5G new radio (NR). Frequency band for 5G NR [9] is being separated into two different ranges, including the sub-6 GHz frequency band [10] and millimeter wave (mmWave) frequency band [11]. The very high 5G frequency bands at mmWave, such as 24.25-27.5 GHz schemed in Europe, 24.25-27.5 GHz and 37-43.5 GHz schemed in China, 27.5-28.35 GHz and 37-40 GHz schemed in the United States, will allow the deployment of hotspots providing very high throughput thanks to the large available bandwidth for operators. On the other hand, the sub-6 GHz spectrum is less complicated in the development of infrastructure, deployment, and future network enhancements. Thus, it can be quickly cleared for the early deployment of 5G cellular networks across the globe. Europe has awarded the trial licenses at 3.4-3.8 GHz band [12]. China is ongoing at 2.5-2.6 GHz, 3.4-3.6 GHz, and 4.8-4.9 GHz frequency bands [13], while Korea is planning at 3.4-3.7 GHz. The United States schemes the frequency bands of 3.1-3.55 GHz and 3.7-4.2 GHz. Overall, the frequency from 2.5 GHz to 5.9 GHz is the main sub-6 GHz frequency band for V2X communications. Therefore, it is necessary to thoroughly investigate the V2X channel characteristics at this frequency band in different scenarios.

In the literature, the sub-6 GHz vehicular communication channels for 5G are characterized by academia and industry. Channel measurement is regarded as a valid approach to describe realistic propagation information. Many measurements have been taken in vehicular communications to characterize radio channel [14]. In vehicle communications, measurements are even more challenging due to high mobility scenarios and more dynamic network conditions [15]. The authors of [16] investigate channels under the conditions from Line-of-Sight (LoS) to non-Line-of-Sight (NLoS) in terms of delay spread, Doppler spread, and the RMS delay spreads at 5.9 GHz. The author of [17] present thorough characterizations for intracar propagation based on the practical measurement on-board an HSR at 2.4 and 5.8 GHz. By considering the traffic flows and the velocity of the vehicles, the wireless communications at 2.48 GHz under the NLoS condition are compared between the measurements and proposed time-variant temporal correlation function (TCF) model in [18]. The cross-polarized channel measurements are conducted in a small-cell street canyon scenario at 2.6 GHz [19]. It is found that the polarization does not influence the angles of arrival. In [20], channel measurements in an urban scenario are performed at 2.6 GHz using a virtual uniform linear array (ULA). In [21], the mobility measurements are conducted in an urban macrocellular (UMa) scenario to analyze the path loss at 2.54 GHz, 3.5 GHz, 4.9 GHz, and 5.4 GHz, respectively. Path loss, Ricean K-factor, and angular power spectrum are analyzed. The path loss of the measurements in an overpass scenario is fitted by two-ray model at 5 GHz [22]. The distributions of time-varying delay spread and Doppler spread follow a bimodal Gaussian mixture distribution in the street cross scenario at 5.5 GHz [23]. The authors of [24] survey the propa-

gation characteristics at sub-6 GHz and mmWave bands, in which the preferred bands for initial deployment are 3.3-4.2 GHz and 26/28 GHz for 5G communications. The authors of [25] summarize the advanced channel measurements for vehicular communications at sub-6 GHz frequency band.

Channel models can be classified as stochastic and deterministic ones. The geometry-based stochastic channel models (GBSMs) with wide-sense stationary (WSS) assumption for vehicular channels have been widely accepted in the past studies [26, 27]. In [28], the GBSM combined with a two-ray model is proposed for nonisotropic multiple-input multiple-output (MIMO) Ricean fading channels in regular shaped environments and is validated for V2V channel at 5.9 GHz. Moreover, several measurements show that the WSS assumption validates only in a short distance interval and time interval [29]. For the nonstationary vehicular channel modeling, most works focus on the surrounding scatterers, e.g., vehicles, pedestrians [22], and traffic signs [30]. Deterministic channel models, such as ray-tracing (RT), can provide accurate channel information in propagation scenarios. The LoS, reflected, scattered, diffracted, and penetrated multipaths can be traced as rays, which contain the amplitude, angle, and delay information. In [31], the authors integrate the radar cross sections of the small-scale structures (e.g., lampposts and traffic signs) into RT simulator and conduct extensive RT simulations for V2V channel in urban and open space environments at 5 GHz. In [32], the authors propose that RT can be utilized to detect hidden obstacles in an external environment of a vehicle with the help of a light detection and ranging (LIDAR) sensor. In [33], the authors compare the RT simulations against V2V channel measurements using a channel sounder in an urban scenario at 5.9 GHz. The received power, delay spreads, and Doppler spreads are characterized in LoS and NLoS conditions, respectively.

Nevertheless, a comprehensive characterization of V2I channel based on cellular networks at 3.5 GHz frequency band is rarely found. Due to the limits of measurement equipments, propagation scenarios, and traveling speeds of the vehicles, a tendency of vehicular channel characterization is to combine the RT simulations. Thus, in this paper, we characterize the V2I channel at 3.5 GHz based on channel measurements and simulations with the following contributions:

(i) Dynamic channel measurements are performed in a suburban scenario at 3.5 GHz frequency with the bandwidth of 100 MHz, because 3.5 GHz is the common spectrum band in Europe [12], China [13], and other countries of the world for 5G. The transmitter (Tx) is fixed, while the moving receiver (Rx) is carried by a vehicle during the measurements. Double-cone omnidirectional antennas for both Tx and Rx are utilized in the measurements. The power delay profiles (PDPs), path loss (PL), shadow fading (SF), root-mean-square (RMS) delay spread (DS), and the coherence bandwidth are analyzed based on the measured results
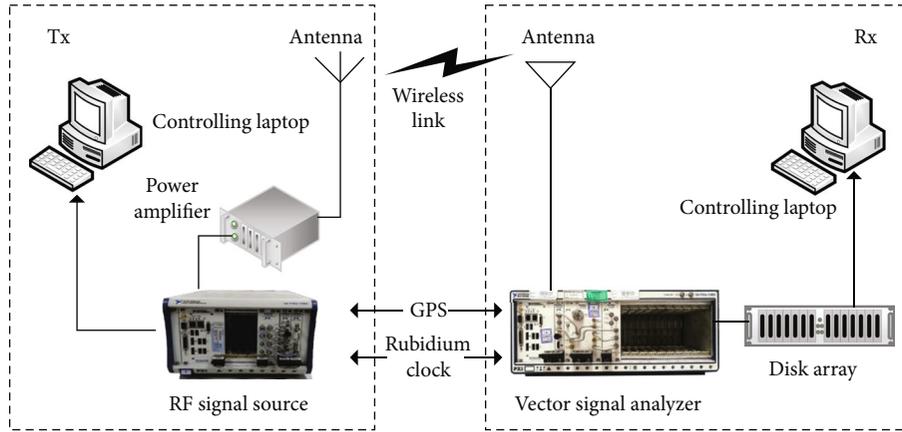
FIGURE 1: Main structure of measurement system.

FIGURE 2: Satellite image of the measurement environment and trajectories of moving Rx.

(ii) We reconstruct the suburban scenario and conduct the simulations utilizing the measurement-validated RT simulator with the same configurations as the measurement. The RT simulations breakthrough the limits of the measurement data and expand the measured two-dimensional (2D) channel into a 3D channel. The electromagnetic (EM) properties of relevant materials are calibrated. Then, the channel is characterized more comprehensively with the aid of Ricean K-factor (KF), angular spreads, and cross-polarization ratio (XPR). Combined with the channel parameters extracted from measurements, the cross-correlations of the abovementioned parameters are listed, and the autocorrelation of each parameter itself is calculated as well. The provided key parameters of this paper help to understand the V2I channel characteristics and to support the link-level and system-level design. These will enable the C-V2X communications for 5G

The remainder of this paper is organized as follows. The measurement systems and campaigns for the V2I channel are

detailed in Section 2. The analysis of the measurement results is described in Section 3. In Section 4, the RT simulations are conducted in the reconstructed measurement scenario. Finally, the conclusions are drawn in Section 5.

2. Channel Measurement Campaign

In order to measure the V2I channel, a dynamic frequency domain channel sounding method is applied in a suburban scenario. As shown in Figure 1, this measurement system is supported by a controlling laptop, a radio frequency (RF) signal source by NI PXIe-5673E module, a power amplifier, and an antenna at the Tx side. At the Rx side, it is composed of a controlling laptop, a wideband vector signal analyzer (VSA) by NI PXIe-5668 module, a disk array, and an antenna. The XHTF3311 global positioning system (GPS) and a rubidium clock are utilized to ensure the synchronization of the signal for Tx and Rx during the measurements. The carrier frequency is 3.5 GHz with a bandwidth of 100 MHz, and the number of frequency points is 821. The transmitting power is 40.77 dBm by the power amplifier. The transmit signal is a Zadoff-Chu (ZC) sequence and can provide an almost constant magnitude in the frequency domain [34].

The measurement campaigns are conducted in Haidian District, Beijing, China. We select the measurement area far from the main road of the city and only conduct the measurements in the evening so that by-passing traffics and pedestrians do not affect the measurements. We measure the LoS scenario, which exists mostly in V2I channels. As shown in Figures 2 and 3, the fixed Tx, with a height of 2.55 m, is at the intersection of the two roads. In the measurements, the Tx transmits signal sequences continuously. In order to emulate a typical vehicular environment, the two crossroads are selected in the measurements. The Rx is carried by a vehicle with a height of 1.88 m (as shown in Figure 3) forward or reversely following the track of the red arrows. Route 1 from north to south (approximately 210 m) represents the vehicle moving towards the fixed Tx, while Route 2 from west to east (approximately 340 m) represents the vehicle moving away from the Tx. Either Tx or Rx is a double-cone omnidirectional antenna with 0 dBi gain. The 3D radiation patterns and the 2D radiation patterns in the horizontal plane (H-
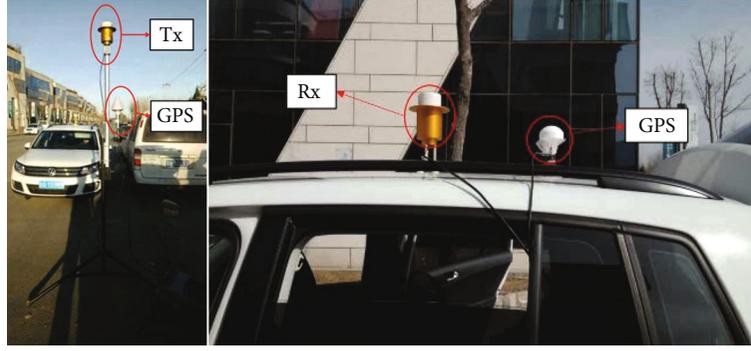
FIGURE 3: Locations of Tx and Rx.
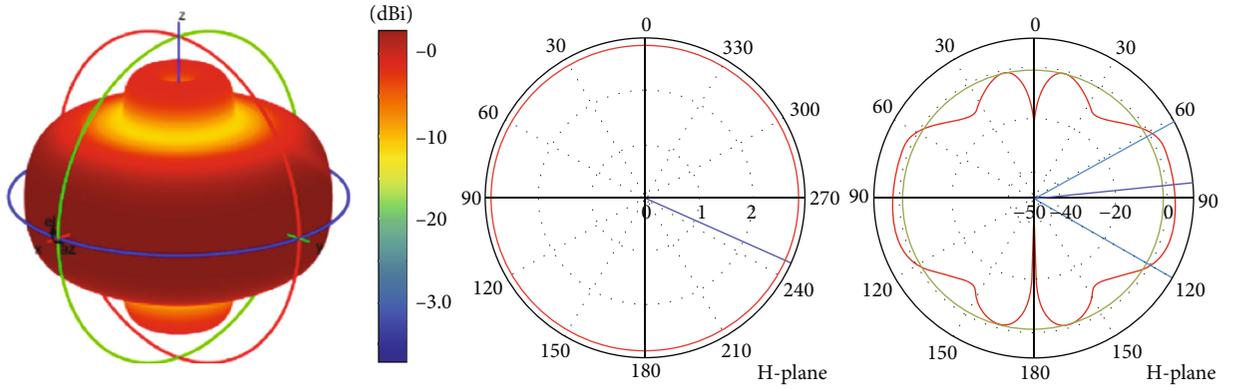
FIGURE 4: Radiation pattern of the double-cone omnidirectional antenna.
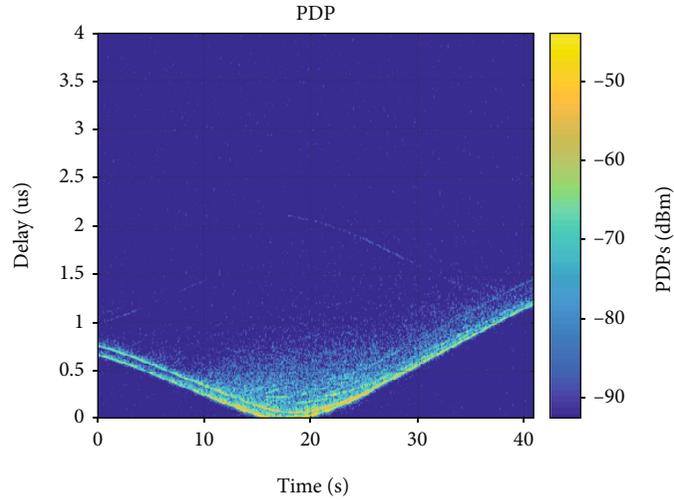
FIGURE 5: Time-varying PDP.

plane) and elevation plane (E-plane) of utilized antennas are shown in Figure 4.

3. Analysis of Measurement Results

In this section, the measured data are preprocessed in the frequency domain and time domain. The channel characteristics, including PDP, path loss, shadow fading, RMS delay spread, and coherence bandwidth, are calculated and analyzed for the V2I channel as follows.

3.1. Preprocessing. The received signal in the frequency domain can be expressed as follows [34]:

$$Y(d,f) = X(f)H_{Tx}(f)H(d,f)H_{Rx}(f), \tag{1}$$

TABLE 1: Measured path loss of V2I channel at 3.5 GHz.

	Route 1			Route 2	
$\mathrm{PL}(d_0)$	n	σ_{SF}	$\mathrm{PL}(d_0)$	n	σ_{SF}
12.15	2.60	4.91	26.11	1.66	2.19

TABLE 2: RMS delay spread and coherence bandwidth.

RMS delay spread (ns)		Coherence bandwidth (MHz)	
μ_{DS}	σ_{DS}	μ_{B_c}	σ_{B_c}
670.34	355.69	15.83	11.10

where $X(f)$ and $Y(d, f)$ are the transmitted and received signals, respectively. d is the distance between Tx and Rx, and f is the operating frequency. $H_{Tx}(f)$ and $H_{Rx}(f)$ are the transfer functions of the Tx and Rx equipment, such as antennas, VSA, and cables, while $H(d, f)$ is the transfer function of the wireless channel. In order to remove the influence of the transceivers, the reference signal measurements with the same setting as in the channel measurements are conducted in an anechoic chamber. The reference received signal can be expressed as follows:

$$Y_{\mathrm{ref}}(f) = X(f)H_{Tx}(f)H_{\mathrm{ref}}(f)H_{Rx}(f), \qquad (2)$$

where $H_{\mathrm{ref}}(f)$ is the free space transfer function.

Thus, the channel transfer function (CTF) can be calculated as follows:

$$H(d, f) = \frac{Y(d, f)}{Y_{\mathrm{ref}}(f)} H_{\mathrm{ref}}(f). \qquad (3)$$

Then, the channel impulse response (CIR) $h(\tau)$ can be calculated by the inverse Fourier transform:

$$h(\tau) = \mathrm{IFFT}\left(H(f), N_f\right), \qquad (4)$$

where τ is the time delay of the received signal and N_f is the effective number of the measured frequency points. A Hann window is used to suppress side lobes, and the sliding window is with the size of 20 wavelengths. A total of 4080 effective CIRs are extracted for the entire travel trajectory.

In addition, at the largest distance between Tx and Rx, the lowest average signal-to-noise-ratio (SNR) is about 20 dB, and at most of the measurement locations, the SNR is over 25 dB. Hence, the accurate estimation of channel parameters is possible in the following part.

3.2. Power Delay Profile. The signal power on each multipath against their respective propagation delay is defined by the PDP, which is defined as:

$$P(t, \tau) = |h(t, \tau)|^2. \qquad (5)$$

A time-varying PDP of the measurement is shown in Figure 5. The calculated optimal threshold after filtering noise is -95.69 dBm, which is 6 dB higher than the mean value of the noise power and effectively separates the signal from noise. The dynamic range is larger than 40 dB. Then, the local maximums as the "peaks" higher than the noise threshold are identified as the "paths." It can be seen clearly that in the propagation environment, there are always two energy-intensive paths. One is the LoS path, and the other is the

strongly reflected path from the ground. Besides, there are rich multipaths in the whole scenario.

3.3. Path Loss Exponent and Shadow Fading. The local wideband path gain $\mathrm{PG}(d)$ can be calculated directly from the measured CTFs in the frequency domain as follows:

$$\mathrm{PG}(d) = \frac{1}{N_f} \left(\sum_{l=1}^{N_f} |H(d, f_l)|^2 \right), \qquad (6)$$

where f_l is the sampling frequency and $N_f = 821$ is the number of the measured frequency points. As can be seen from Figure 2, the moving track of Route 1 is Rx towards the fixed Tx, while the moving track of Route 2 is Rx away from the fixed Tx. The environments and the length of these two roads are not similar. In Route 1, there are buildings along the two sides of the road, while only one side of Route 2 has buildings, and the other side is a fence with advertisements. Thus, two routes should be separated to analyze. To remove the effect of small scale fading, in this paper, we use a sliding window with the size of 20 wavelengths for averaging.

The log-distance path loss model is to fit the opposite number of measured path gain. The change of the path loss, along with the distance, is depicted by the path loss exponent. This exponent and shadow fading are extracted from the measured results by using the following expression [35]:

$$\mathrm{PL}(d) = \overline{\mathrm{PL}}(d_0) + 10n\left(\log_{10} \frac{d}{d_0}\right) + X_\sigma, \qquad (7)$$

where $\mathrm{PL}(d)$ is the path loss and d is the distance between the Tx and Rx in m. d_0 is the reference distance; we choose the measured PL at $d_0 = 10\,\mathrm{m}$ as a reference in our study. $\mathrm{PL}(d_0)$ is the media path loss at d_0, and n is the path loss exponent. By using the least-square criterion, the path loss exponent n can be obtained. X_σ is the shadow fading, which can be expressed as a Gaussian variable with zero mean value and a standard deviation of σ_{SF}. The path loss exponent and the shadow fading standard deviations based on the measurements are summarized in Table 1. The fact of higher path loss exponent n and higher σ_{SF} for Route 1 indicates a more complicated environmental condition of the Route 1. There are a large number of trees, two rows of iron fences, and concrete walls on the side of Route 1, which cause the scattering during the wireless propagation.

3.4. RMS Delay Spread and Coherence Bandwidth. The RMS delay spread is one of the key parameters for a multipath channel. It is defined in [36], namely, the square root of the
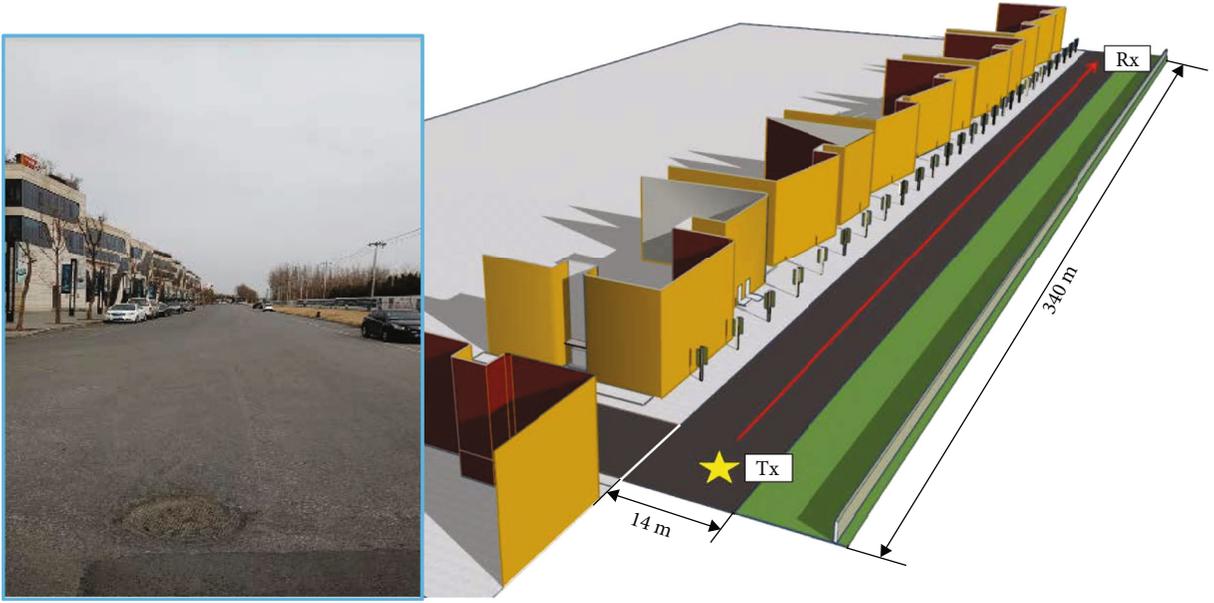
FIGURE 6: 3D model of the reconstructed suburban scenario.

TABLE 3: Simulation configurations.

Parameter	Value
Center frequency	3.5 GHz
Bandwidth	100 MHz
Frequency points	821
Tx height	2.55 m
Rx height	1.88 m
Polarization	VV, HH, VH, and HV
Tx power	40.77 dBm
Propagation mechanism	LOS + up to 3rd order of reflection + scattering + diffraction

second central moment of the PDP as follows:

$$\sigma_\tau = \sqrt{\frac{\sum_{n=1}^N \tau_n^2 \cdot P_n}{\sum_{n=1}^N P_n} - \left(\frac{\sum_{n=1}^N \tau_n \cdot P_n}{\sum_{n=1}^N P_n}\right)^2}, \qquad (8)$$

where σ_τ denotes the RMS delay spread and P_n and τ_n denote the power and the excess delay of the n-th multipaths, respectively. The distribution of RMS delay spreads for the whole trajectory is following a Gaussian distribution. The mean value (μ_{DS}) of the measurement routes is 670.34 ns with the standard deviation (σ_{DS}) 355.69 ns. It is in line with the outdoor suburban environment, where the RMS delay spread value is 0.2-2.0 μs. In [23], the RMS delay spreads around 47 ns and 75 ns are measured in urban and tunnel scenario, respectively. However, sometimes, although the LoS component exists in the channel, several significant reflected multipaths and scattered multipaths can still result in a large delay spread, such as in our study.

The inversely proportional relation between RMS delay spread σ_τ and the coherence bandwidth B_c can be expressed:

$$B_c = 1\kappa\sigma_\tau, \qquad (9)$$

where κ is a constant that depends on the environment or/and on how B_c is defined. B_c can be defined as the bandwidth at which the complex correlation function has a value of 0.5, 0.9, or even 0.95. For example, $R_H(\Delta f) = 0.5$ means the coherence bandwidth can be specifically denoted as $B_{c,0.5}$ with subscript 0.5 as the threshold [37].

The frequency correlation function of the channel $R_H(\Delta f)$ can be obtained by setting a reasonable threshold [38]:

$$R_H(\Delta f) = \int_{-\infty}^{\infty} P(\tau)e^{-j2\pi\Delta f\tau}d\tau, \qquad (10)$$

where $P(\tau)$ is the CIR at a specific time delay τ. The statistical estimation of the mean value μ_{Bc} and the standard deviation σ_{Bc} of coherence bandwidths can be obtained according to the threshold of 0.5 in our study. The values of RMS delay

spread and coherence bandwidths are summarized in Table 2. The coherence bandwidth is narrower when using the onmi-directional antennas. It is consistent with the RMS delay spread results and the inverse relation between these two parameters. As happens with the RMS delay spread, the multipath propagation causes the coherence bandwidth.

Moreover, the study on Doppler spectrum of these measurement campaigns has been published in [39]. It can be concluded that the faster the vehicle (Rx) moves, the larger the Doppler spreads. The most significant scatterers during the measurements are trees and billboards, whereas other cars parked on the roadside do not significantly contribute to the multipath propagation.

4. RT Simulation for V2I Channels at 3.5 GHZ

In order to physically interpret the measurement results, a self-developed dynamic RT simulator of Beijing Jiaotong University is used in this study. Supported by high-performance computing (HPC), this RT simulator is now furnished to an HPC cloud-based platform (CloudRT), which contains 1600 CPUs and 10 GPUs. Parallel processing can be implemented by a compute node or monitored by a management node. The first premise of accurate ray-tracing simulation is a thorough and factual description of the scenarios. The finer the scenarios modeled, the more accurate the corresponding prediction is. Also, simulation time is proportional to the accuracy. The simulation time for a single snapshot in this article is about 10 seconds, and the total simulation time depends on the number of nodes selected for parallel computing. More details of this platform can be found in [40] and the website http://raytracer.cloud.

As shown in Figure 6, the 3D model of the suburban scenario is reconstructed through SketchUp [41] and Open-Street Map [42]. The asphalt road of Route 2 is with a width of 14 m and with a length of 340 m. The height of the office buildings on the one side is about 15-20 m. In order to simulate the channel for 100 MHz bandwidth, we use the subband RT approach. The CIR for one snapshot is generated by a number of subbands at multiple center frequency points [43]. Based on the measurement configurations in the same scenario, 821 frequency points are determined. Since the relevant materials are not frequency sensitive, the constant EM properties of them are used for all subbands at 3.5 GHz. The LoS ray, reflected rays (up to third order), scattered rays, and diffracted rays are traced in the simulations for the trade-off between simulation accuracy and computational complexity. The directive scattering model [44] and the uniform theory of diffraction (UTD) model [45] are used in the simulations. In order to get the pure propagation channel without the influence of certain antenna pattern, the antenna of Tx or Rx is omnidirectional antenna, which is in line with the measurements. In addition, in order to get the fully polarimetric information of the channel, the antennas of Tx and Rx are with four cases of polarization: vertical-vertical (VV) polarization, vertical-horizontal (VH) polarization, horizontal-vertical (HV) polarization, and horizontal-
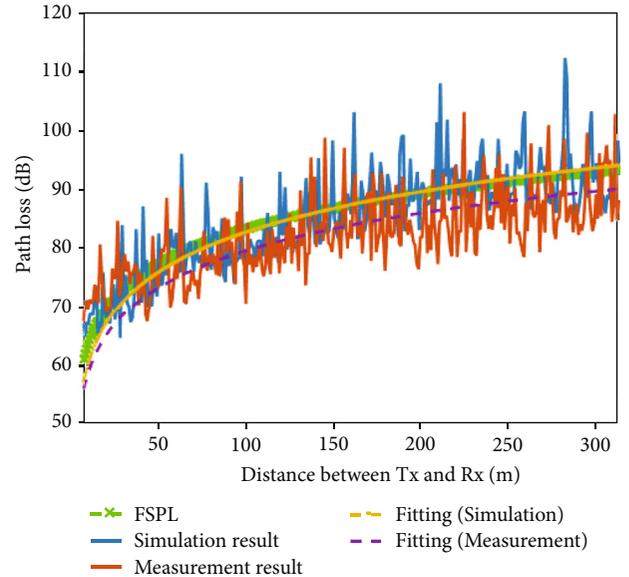
Figure 7: Path loss (measurement vs. validated and calibrated simulation).

Table 4: Material EM properties before and after calibration.

Material name	Before calibration		After calibration	
	$\varepsilon 0_r$	$\varepsilon 00_r$	$\varepsilon 0_r$	$\varepsilon 00_r$
Concrete	1.06	0.65	5.60	0.05
Tempered glass	9.99	0.43	6.27	0.33
Brick	2.00	0.25	6.09	0.08
Vegetation	29.12	0.278	10.00	0.15
Stainless steel	1.00	1E07	1.00	1E07
Granite	1.25	1.79	4.25	0.78
Canvas2.94	2.94	0.04	2.61	0.45
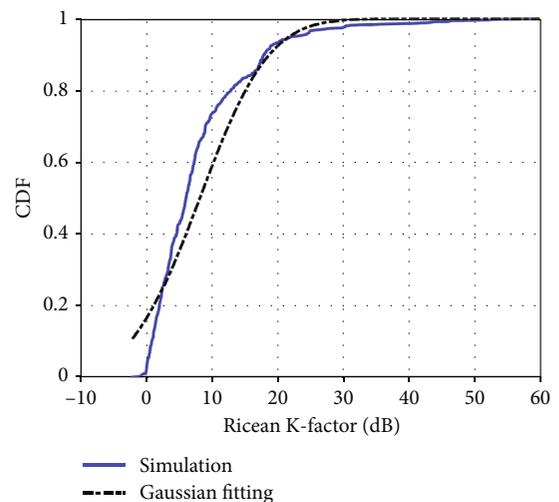
Figure 8: Ricean K-factor.

(a) ASA

(b) ESA

(c) ASD

(d) ESD
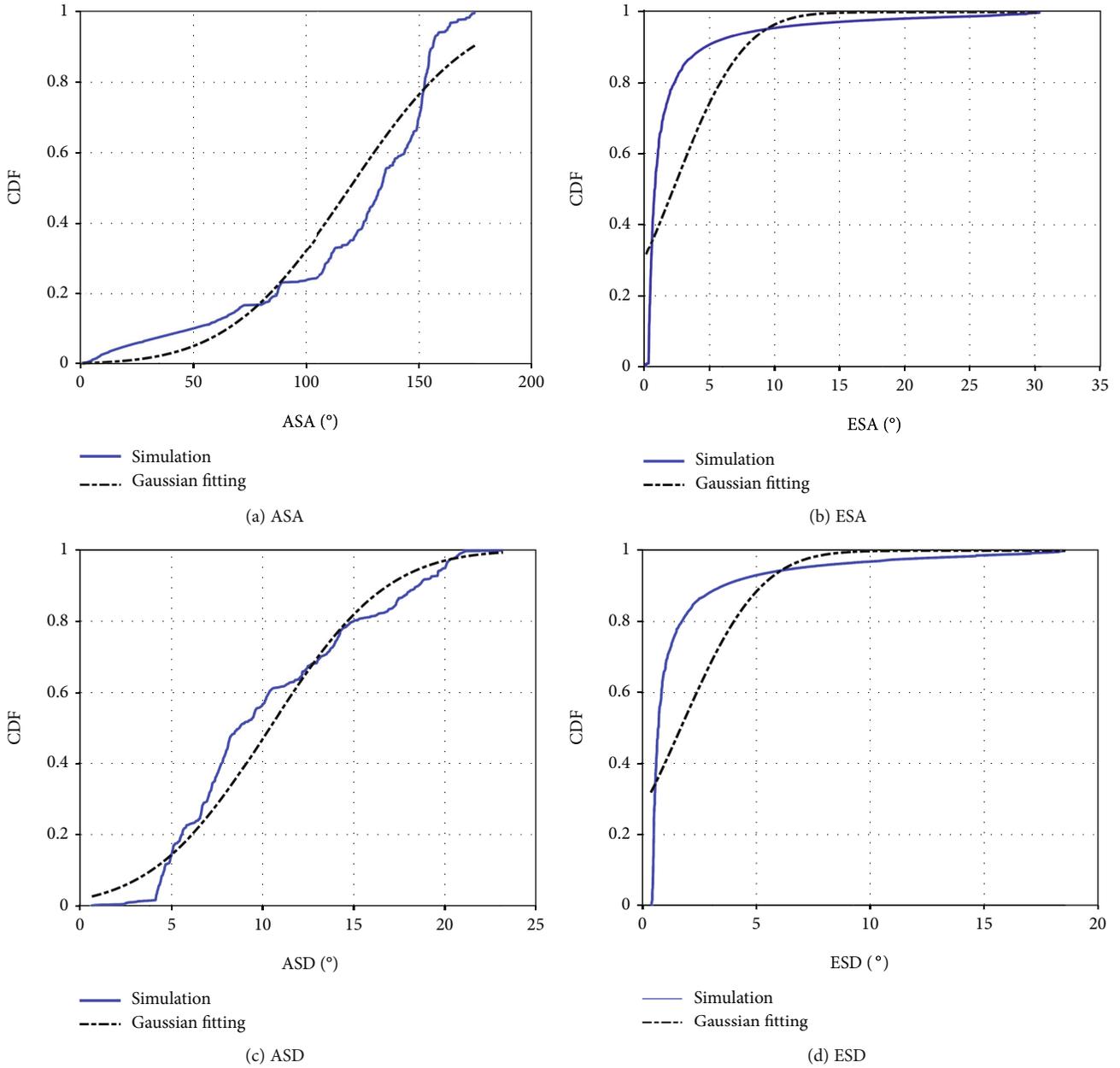
Figure 9: Angular spreads.

horizontal (HH) polarization in the extensive RT simulations. Table 3 summarizes the simulation configurations.

4.1. Validation and Calibration of RT Simulator. Several snapshots are stochastically chosen to validate the geometry and calibration of the EM properties of the suburban scenario, where the validation and calibration processes are according to our previous work [46]. Figure 7, as an example, shows the comparison of path loss between the measurement and the validated and calibrated RT simulation. The antenna influence can be neglected with the omnidirectional antennas both at Tx and Rx sides. Compared with the measurement results, the difference of path loss exponent n between simulation and measurement is 0.12, with the relative error 6.4%. The difference between the shadow fading is 0.29 dB, and the

relative error is 5.6%, which proves the accuracy of the RT simulator in channel simulation. The calibrated EM properties are summarized in Table 4, which will be used in the following extensive simulations.

4.2. Channel Characterization Based on Extensive RT Simulations. The extensive RT simulations are done for more comprehensive channel characterization. Route 2 selected as an example, the key parameters can be supplemented to fully describe for the V2I channel, in terms of the Ricean K-factor, XPR, and four angular spreads (namely, azimuth angular spread of arrival (ASA), azimuth angular spread of departure (ASD), elevation angular spread of arrival (ESA), and elevation spread of departure (ESD)). Combined with the channel parameters extracted from the measurements, the cross-
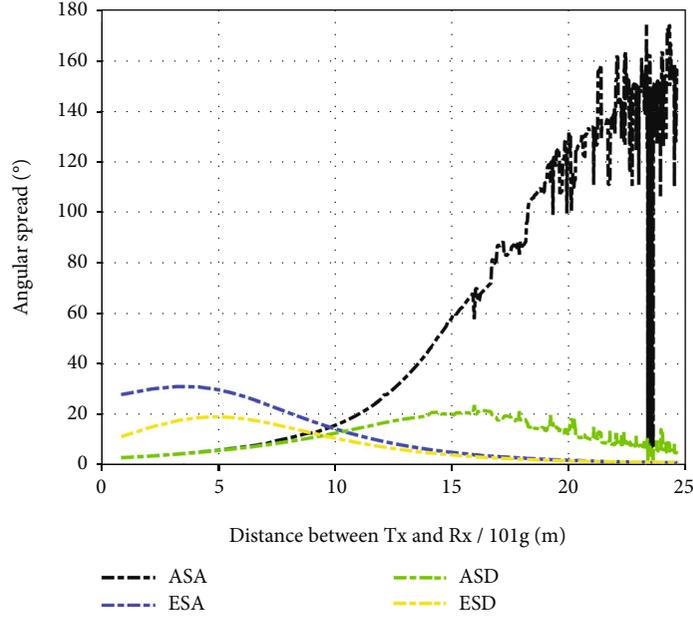
FIGURE 10: Angular domain parameters changes with distance.

correlations of the abovementioned parameters are listed, and the autocorrelation of each parameter itself is calculated as well.

4.2.1. Ricean K-Factor. The Ricean *K*-factor is defined as the ratio of the signal power of the LoS path to the signal power of other multipath components, which can be expressed as:

$$ \text{KF} = \frac{P_{LoS}}{\sum P(i) - P_{LoS}}, \tag{11} $$

where $P(i)$ is the energy of all the paths of the signal during transmission and P_{LoS} is the energy of the LoS path. As shown in Figure 8, most values of the *K*-factor are larger than 0 dB. When the probability is 0.5, the corresponding value of KF is 5.98 dB; when it reaches 0.9, the value is 18.16 dB. Such high values indicate that the LoS component plays the leading role in the propagation scenario, while the multipaths reflected from the surfaces of the buildings and the ground are relatively weak.

4.2.2. Angular Spreads. The angular spread is the distribution of the angle of arrival/departure of each multipath in 3D environments for the Tx and Rx, respectively. Each multipath is assigned an azimuth angle in the horizontal plane and an elevation angle in the vertical plane. Hence, four values of angular spreads can be calculated as follows:

$$ \sigma_{\text{AS}} = \sqrt{\frac{\sum_{n=1}^{N} (\theta_{n,\mu})^2 \cdot P_n}{\sum_{n=1}^{N} P_n}}, \tag{12} $$

where σ_{AS} denotes the angular spread (AS), P_n denotes the
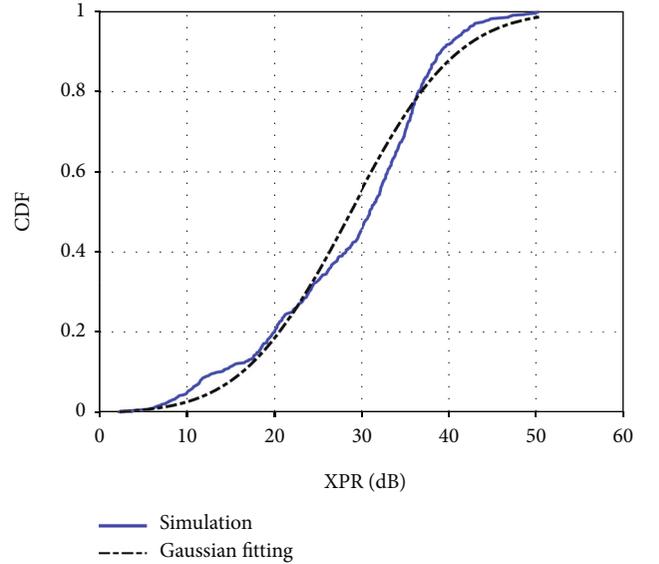
FIGURE 11: XPR.

TABLE 5: Decorrelation length of key parameters.

λ (m)	SF	KF	DS	ASD	ASA	ESD	ESA
Value	1.2	10.5	2.2	9.0	8.0	7.5	9.5

power of the *n*-th multipath, and $\theta_{n,\mu}$ is defined by:

$$ \theta_{n,\mu} = \text{mod} \left(\theta_n - \mu_\theta + \pi, 2\pi\right) - \pi, \tag{13} $$

where θ_n is the azimuth angle of arrival (AoA)/azimuth angle of departure (AoD)/elevation angle of arrival

(EoA)/elevation angle of departure (EoD) of the n-th ray. μ_θ is:

$$\mu_\theta = \frac{\sum_{n=1}^N \theta_n \cdot P_n}{\sum_{n=1}^N P_n}. \tag{14}$$

Figure 9 shows the CDFs of the four angular spreads, including ASA, ESA, ASD, and ESD, respectively. Most elevation angular spreads are below 10°, which is much smaller than azimuth angular spreads in the horizontal plane. In general, the scattering effect in the horizontal direction is stronger than it in the vertical direction, because of the multipath components reflected from the infrastructures along the roadside. We further analyze the variation of the angular spreads during the vehicle moving trajectory, as shown in Figure 10. The values of ESA and ESD are more significant than those of ASA and ASD when the distance between Tx and Rx is within 10 m. This observation implies that multipaths reflected from the ground in the vertical direction are stronger than those in the horizontal direction at the beginning. As the distance increases, the values of ASA and ASD are larger than those of ESA and ESD. Concerning the ASA, it increases sharply to 173° due to energetic scattered paths from the infrastructures far away from the Tx but close to Rx.

4.2.3. XPR. The XPR is to describe the variety of the polarization for the multipath components and is calculated by its definition as follows:

$$\text{XPR} = 10 \log_{10}\left(\frac{P_{\text{co}}}{P_{\text{cross}}}\right), \tag{15}$$

where P_{co} is copolarization defined by the received power when Tx and Rx antennas polarize in the same manner. Conversely, P_{cross} is the power received in vertical (horizontal) polarization and transmitted in horizontal (vertical) polarization. As shown in Figure 11, the high values imply little depolarization in such a suburban scenario. These values are much larger than the higher frequency band in similar scenarios, such as at 28 GHz [47] and at THz band [48]. Generally speaking, the linear polarization of the antennas is sufficient in such a scenario.

4.2.4. Autocorrelation and Decorrelation Distance. The autocorrelation function of a parameter X is to describe how fast a certain variable changes with the distance interval Δd (or time delay τ), which can be calculated as follows:

$$R(\Delta d) = \frac{E[(X_d - \mu)(X_{d+\Delta d} - \mu)]}{\sigma_X^2}, \tag{16}$$

where E is the expected value operator. Correspondingly, the decorrelation distance λ in (m) is expressed as:

$$\lambda = \underset{\Delta d > 0}{\text{argmax}}(|R(\Delta d)| < T), \tag{17}$$

where T is the threshold.

TABLE 6: Cross-correlation of the channel parameters.

Parameter	DS	KF	SF	ASD	ASA	ESD	ESA
DS	1	0.25	-0.07	0.12	-0.02	-0.09	-0.11
KF		1	-0.04	-0.61	-0.59	-0.39	-0.38
SF			1	0.01	0.02	-0.02	-0.02
ASD				1	-0.58	-0.22	0.12
ASA					1	-0.76	-0.74
ESD						1	0.99
ESA							1

TABLE 7: Key parameters of V2I channel at 3.5 GHz.

	DS (ns)	B_c (MHz)	KF (dB)	ASA (°)	ESA (°)	ASD (°)	ESD (°)	XPR (dB)
μ	670.34	15.83	6.73	117.4	2.18	10.47	1.70	28.64
σ	355.69	11.10	8.23	1.63	0.63	0.70	0.44	9.69
λ (m)	2.2	—	10.5	9.0	8.0	7.5	9.5	—

The larger the threshold is set, the stricter the correlation requirements are, and the smaller the decorrelation distance is. The threshold is generally a practical value. The common value of 0.9 is set in our study. Table 5 summarizes the decorrelation distance of each parameter. It can be preconcluded that the shadow fading is not self-correlated. The KF and DS are highly self-related. The value of the decorrelation distances of the parameters mentioned above is in line with the values provided by 3GPP 38.901 [49].

4.2.5. Cross-Correlation. We consider a wideband V2I channel, study the key parameters of the channels across 100 MHz at 3.5 GHz frequency band, and try to capture the channel correlation across different parameters. The cross-correlation coefficient of the two variables is a measure of their linear dependence. If each variable has N scalar observations, then the Pearson cross-correlation coefficient is defined as:

$$R_{A,B} = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{A_i - \mu_A}{\sigma_A}\right)\left(\frac{B_i - \mu_B}{\sigma_B}\right), \tag{18}$$

where A_i and B_i represent two different aforementioned channel parameters and μ and σ denote the mean value and the standard deviation of corresponding parameters, respectively.

Combined with the measurements and simulation results, the cross-correlations of the key parameters are summarized in Table 6. It can be found that shadow fading is not cross-correlated with other key parameters since the correlations of it with other key parameters are approximately equal to 0. Remarkably, KF is negatively correlated to the angular spreads, since with a large value of KF, a significant component comes from a single direction leading to smaller angular

spreads. In the angular domain, the parameters show relatively strong correlations between each other. If the variables of interest are highly correlated, it can be used to make accurate predictions.

4.2.6. Summary of the Channel Parameters by Measurements and Simulations. The V2I channel is comprehensively characterized based on the measurements and simulation together. The group of channel parameters can be summarized in Table 7. The validation results of path loss indicate that RT can practically reflect the target scenarios. Besides, the RT simulations can break the limits of the measurement data. Reliable channel information can be obtained with similar materials. Moreover, other important channel information, such as angles, which are not captured from the measurements easily, can be given by RT simulations.

5. Conclusion

In this paper, we comprehensively characterize the V2I channel through the combination of the measurements and simulations in a suburban scenario at 3.5 GHz band for 5G communications. To begin with, the channel measurements are conducted in a typical vehicular scenario. The track of moving vehicle is separated into two routes, where the environment of Route 1 is more complicated than Route 2. The time-varying PDP, path loss, RMS delay spread, and the coherence bandwidth are extracted from the channel measurements. From the time-variant power delay profiles, it can be clearly seen that two significant paths always exist during the measurements. It indicates that the LoS vehicular channel measurement generally follows the classic two-ray models. One is the LoS path, and the other is the reflected path from the ground. The path loss exponent n and the σ_{SF} are extracted to describe the path loss and shadow fading, respectively. Because of the more complicated environment of Route 1, the values of n and σ_{SF} are larger than these of Route 2. The coherence bandwidth is to follow the inversely proportional relation to RMS delay spreads. As the threshold is set 0.5, the coherence bandwidth then can be calculated and analyzed.

Moreover, the RT simulations are conducted in order to supply other key parameters. The high-performance self-developed RT simulator of Beijing Jiaotong University is utilized in our study. The RT simulations breakthrough the limits of the measurement data and expand the measured 2D channel into a 3D channel. After validated and calibrated by the measured PDPs and compared with the path loss, the RT simulator is utilized to do extensive simulations in a reconstructed suburban scenario. The large value of Ricean K-factor implies the LoS leading a significant role in the propagation. Four cases of the vertical/horizontal polarization of antennas for Tx and Rx are schemed to characterize the fully polarimetric information of the channel. The analysis of the XPR shows little depolarization in such a LoS condition. The angular spreads (ASA, ESA, ASD, and ESD) are analyzed. When the distance between Tx and Rx is within 10 m, the multipaths reflected from the ground are stronger than the multipaths reflected/scattered from the infrastruc-

tures, iron fences, and tress. As the distance increases, the azimuth angular spreads become larger than elevation angular spreads. Moreover, the decorrelation distance of each parameter and the cross-correlations of the abovementioned parameters are supplied. SF is neither self-correlated nor cross-correlated with other channel parameters. KF is negatively correlated to the four angular spreads, since a significant component coming from a single direction leads to a large value of KF but small angular spreads.

In summary, the group of channel parameters is thoroughly analyzed, and the V2I channel is comprehensively characterized based on the measurements and simulations. This study on channel characteristics at 3.5 GHz can be applied to the link-level and system-level design for the C-V2X communications of 5G.

Data Availability

The raw data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] Y. Ji, W. Fan, M. G. Nilsson et al., "Virtual drive testing over-the-air for vehicular communications," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 1203–1213, 2020.

[2] L. Zhu, F. R. Yu, B. Ning, and T. Tang, "Cross-layer handoff design in mimo-enabled wlans for communication-based train control (CBTC) systems," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 719–728, 2012.

[3] K. M. S. Huq, S. A. Busari, J. Rodriguez, V. Frascolla, W. Bazzi, and D. C. Sicker, "Terahertz-enabled wireless system for beyond-5G ultra-fast networks: a brief survey," *IEEE Network*, vol. 33, no. 4, pp. 89–95, 2019.

[4] L. Zhu, Y. Li, F. R. Yu, B. Ning, T. Tang, and X. Wang, "Cross-layer defense methods for jamming-resistant CBTC systems," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2020.

[5] D. Jiang and L. Delgrossi, "IEEE 802.11p: towards an international standard for wireless access in vehicular environments," in *VTC Spring 2008 - IEEE Vehicular Technology Conference*, pp. 2036–2040, Singapore, Singapore, May 2008.

[6] T. Etsi, "Intelligent Transport Systems (ITS); European profile standard on the physical and medium access layer of 5 GHz ITS," *Draft ETSI ES*, vol. 202, no. 663, p. V0, 2009.

[7] L. Zhang, Q. Ni, M. Zhai, J. Moreno, and C. Briso, "An ensemble learning scheme for indoor-outdoor classification based on KPIs of LTE network," *IEEE Access*, vol. 7, pp. 63057–63065, 2019.

[8] F. Abbas, P. Fan, and Z. Khan, "A novel low-latency V2V resource allocation scheme based on cellular V2X communications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2185–2197, 2019.

[9] F. Luo, "5G new radio (NR): standard and technology," *ZTE Communications*, vol. 1, 2017.

[10] "NR; user equipment (UE) radio transmission and reception; part 1: range 1 standalone," 3GPP TS 38.101-1, Tech. Rep., 2020.

[11] "NR; user equipment (UE) radio transmission and reception; part 2: range 2 standalone," 3GPP TS 38.101-2, Tech. Rep., 2020.

[12] A. Sgora, "5G spectrum and regulatory policy in Europe: an overview," in *2018 Global Information Infrastructure and Networking Symposium (GIIS)*, pp. 1–5, Thessaloniki, Greece, October 2018.

[13] "The ministry of industry and information technology publicly sought advice on the use of the 3.3-3.6 GHz and 4.8-5 GHz frequency bands for 5G," 2016, http://www.soecc.org.cn/show.asp?id=8885.

[14] C. Li, J. Yu, W. Chen, K. Yang, and F. Li, "Shadowing correlation and a novel statistical model for inland river radio channel," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, pp. 1–6, Shanghai, China, May 2019.

[15] W. Fan, L. Hentila, F. Zhang, P. Kyosti, and G. F. Pedersen, "Virtual drive testing of adaptive antenna systems in dynamic propagation scenarios for vehicle communications," *IEEE Access*, vol. 6, pp. 7829–7838, 2018.

[16] J. Yu, W. Chen, F. Li et al., "Channel measurement and modeling of the small-scale fading characteristics for urban inland river environment," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3376–3389, 2020.

[17] L. Zhang, J. Moreno, and C. Briso, "Experimental characterisation and modelling of intra-car communications inside high-speed trains," *IET Microwaves, Antennas Propagation*, vol. 13, no. 8, pp. 1060–1064, 2019.

[18] Q. Zhu, W. Li, C.-X. Wang et al., "Temporal correlations for a non-stationary vehicle-to-vehicle channel model allowing velocity variations," *IEEE Communications Letters*, vol. 23, no. 7, pp. 1280–1284, 2019.

[19] R. Zhang, L. Cai, Z. Zhong, J. Zhao, and J. Zhou, "Cross-polarized three-dimensional channel measurement and modeling for small-cell street canyon scenario," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 7969–7983, 2018.

[20] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3899–3911, 2015.

[21] W. Zhang, L. Tian, T. Jiang, A. Huang, J. Zhang, and Y. Zheng, "Analysis on frequency dependence of large scale fading in urban environment," in *2019 13th European Conference on Antennas and Propagation (EuCAP)*, pp. 1–5, Krakow, Poland, March 2019.

[22] P. Liu, B. Ai, D. W. Matolak, R. Sun, and Y. Li, "5-GHz vehicle-to-vehicle channel characterization for example overpass channels," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 5862–5873, 2016.

[23] L. Bernado, T. Zemen, F. Tufvesson, A. F. Molisch, and C. F. Mecklenbrauker, "Delay and Doppler spreads of nonstationary vehicular channels for safety-relevant scenarios," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 1, pp. 82–93, 2014.

[24] M. Shafi, J. Zhang, H. Tataria et al., "Microwave vs. millimeter-wave propagation channels: key differences and impact on 5G cellular systems," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 14–20, 2018.

[25] X. Cheng, R. Zhang, L. Yang, X. Cheng, R. Zhang, and L. Yang, "Vehicular channel characteristics and modeling," in *5G-Enabled Vehicular Communications and Networking*, pp. 11–40, Springer, 2019.

[26] X. C. Z. W. Huang and N. Zhang, "An improved non-geometrical stochastic model for non-WSSUS vehicle-to-vehicle channels," *ZTE Communications*, vol. 17, no. 4, 2019.

[27] L. Zhu, Y. He, F. R. Yu, B. Ning, T. Tang, and N. Zhao, "Communication-based train control system performance optimization using deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10705–10717, 2017.

[28] X. Cheng, C.-X. Wang, D. I. Laurenson, S. Salous, and A. V. Vasilakos, "An adaptive geometry-based stochastic model for non-isotropic MIMO mobile-to-mobile channels," *IEEE Transactions on Wireless Communications*, vol. 8, no. 9, pp. 4824–4835, 2009.

[29] W. Li, Q. Zhu, C.-X. Wang, F. Bai, X. Chen, and D. Xu, "A practical non-stationary channel model for vehicle-to-vehicle MIMO communications," 2020, https://arxiv.org/abs/2002.00166.

[30] K. Guan, B. Ai, M. L. Nicolas et al., "On the influence of scattering from traffic signs in vehicle- to-x communications," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 5835–5849, 2016.

[31] K. Guan, D. He, B. Ai et al., "5-GHz obstructed vehicle-to-vehicle channel characterization for internet of intelligent vehicles," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 100–110, 2019.

[32] X. Mei, K. Sakai, and N. Kamata, "Ray tracing for hidden obstacle detection," 2018, uS Patent 10,078,335.

[33] T. Abbas, J. Nuckelt, T. Kurner, T. Zemen, C. F. Mecklenbrauker, and F. Tufvesson, "Simulation and measurement-based vehicle-to-vehicle channel characterization: accuracy and constraint analysis," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 7, pp. 3208–3218, 2015.

[34] R. He, A. F. Molisch, F. Tufvesson, Z. Zhong, B. Ai, and T. Zhang, "Vehicle-to-vehicle propagation models with large vehicle obstructions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2237–2248, 2014.

[35] K. Guan, B. Ai, Z. Zhong et al., "Measurements and analysis of large-scale fading characteristics in curved subway tunnels at 920 MHz, 2400 MHz, and 5705 MHz," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2393–2405, 2015.

[36] T. Rappaport, *Wireless Communications: Principles and Practice*, Prentice-Hall, Upper Saddle River, NJ, USA, 22nd edition, 2002.

[37] X. Chen, P. Kildal, C. Orlenius, and J. Carlsson, "Channel sounding of loaded reverberation chamber for over-the-air testing of wireless devices: coherence bandwidth versus average mode bandwidth and delay spread," *IEEE Antennas and Wireless Propagation Letters*, vol. 8, pp. 678–681, 2009.

[38] M. S. Varela and M. G. Sanchez, "RMS delay and coherence bandwidth measurements in indoor radio channels in the UHF band," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 2, pp. 515–525, 2001.

[39] S. Wang, K. Guan, D. He et al., "Doppler shift and coherence time of 5G vehicular channels at 3.5 GHz," in *2018 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*, pp. 2005-2006, Boston, MA, USA, July 2018.

[40] D. He, B. Ai, K. Guan, L. Wang, Z. Zhong, and T. Kurner, "The design and applications of high-performance ray-tracing simulation platform for 5G and beyond wireless communications: a tutorial," *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 10–27, 2019.

[41] https://www.sketchup.com/.

[42] https://www.openstreetmap.org/.

[43] S. Priebe, *Towards THz Communications: Propagation Studies, Indoor Channel Modelling and Interference Investigations*, Ph.D. Dissertation, Technische Universitat Braunschweig, 2013.

[44] V. Degli-Esposti, F. Fuschini, E. M. Vitucci, and G. Falciasecca, "Measurement and modelling of scattering from buildings," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 1, pp. 143–153, 2007.

[45] P. Pathak, W. Burnside, and R. Marhefka, "A uniform GTD analysis of the diffraction of electromagnetic waves by a smooth convex surface," *IEEE Transactions on Antennas and Propagation*, vol. 28, no. 5, pp. 631–642, 1980.

[46] D. He, K. Guan, A. Fricke et al., "Stochastic channel modeling for kiosk applications in the terahertz band," *IEEE Transactions on Terahertz Science and Technology*, vol. 7, no. 5, pp. 502–513, 2017.

[47] L. Wang, B. Ai, D. He et al., "Vehicle-to-infrastructure channel characterization in urban environment at 28 GHz," *China Communications*, vol. 16, no. 2, pp. 36–48, 2019.

[48] H. Yi, K. Guan, D. He, B. Ai, J. Dou, and J. Kim, "Characterization for the vehicle-to-infrastructure channel in urban and highway scenarios at the terahertz band," *IEEE Access*, vol. 7, pp. 166984–166996, 2019.

[49] G R A N W Group et al., "Study on channel model for frequencies from 0.5 to 100 GHz (release 15)," 3GPP TR 38.901, Tech. Rep., 2018.

Research Article

Network Performance Test and Analysis of LTE-V2X in Industrial Park Scenario

Yuanyuan Fan [ID],[1] **Liu Liu** [ID],[1] **Shuoshuo Dong,**[1] **Lingfan Zhuang,**[1] **Jiahui Qiu** [ID],[2] **Chao Cai,**[2] **and Meng Song**[2]

[1]*School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China*
[2]*China United Network Communications Group Co., Ltd., Beijing 100033, China*

Correspondence should be addressed to Liu Liu; liuliu@bjtu.edu.cn

As one of the mainstream technologies of vehicle-to-everything (V2X) communication, Cellular-V2X (C-V2X) provides high reliability and low latency V2X communications. And with the development of mobile cellular systems, C-V2X is evolving from long-term evolution-V2X (LTE-V2X) to new radio-V2X (NR-V2X). However, C-V2X test specification has not been completely set in the industry. In order to promote the formulation of relevant standards and accelerate the implementation of industrialization, the field test and analysis based on LTE-V2X in the industrial park scenario is conducted in this paper. Firstly, key technologies of LTE-V2X are introduced. Then, the specific methods and contents of this test are proposed, which consists of functional and network performance tests to comprehensively evaluate the communication property of LTE-V2X. Static and dynamic tests are required in both line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios to evaluate network performance. Next, the test results verify that all functions are normal, and the performance evaluation indexes are appraised and analyzed. Finally, it summarizes the whole paper and puts forward the future work.

1. Introduction

Internet of Vehicles (IoV) refers to the realization of a comprehensive network connection of vehicle-to-everything (V2X) with the help of a new generation of information and communication technology. It can improve the intelligent level and autonomous driving ability of vehicles, thus improving traffic efficiency, building new formats of transportation services, and providing intelligent, comfortable, and efficient comprehensive services for users [1]. V2X communication technology is used to realize information sharing between vehicles and the outside world and promote the evolution of IoV to the direction of intelligence and cloud [2]. In the future autonomous driving, V2X communication technology is one of the important technologies to realize environmental perception. It can complement the advantages of traditional vehicle-mounted laser radar, camera, and other vehicle-mounted equipment, so as to provide vehicles with beyond-line-of-sight and complex environment

awareness that cannot be realized by radar. In this way, the vehicle's perception range of traffic and surroundings can be expanded from the dimension of time and space, so that the vehicle has the ability to make multi-information fusion decisions [3].

At present, the main technologies for the V2X communication in the world include dedicated short-range communication (DSRC) technology based on the IEEE 802.11P standard and V2X technology based upon the cellular mobile communication system (C-V2X) [4, 5]. The United States completed the formulation of the DSRC standard in 1999, and a lot of testing work also verified the effectiveness of DSRC, but it has obvious disadvantages like poor reliability, hidden nodes, high delay, and intermittent V2I connectivity [6]. From an industry perspective, widespread deployment of DSRC requires significant investment in the network infrastructure. In order to solve the deficiency of DSRC in testing and industrial application, 3GPP designed C-V2X and completed the formulation of the first stage LTE-based
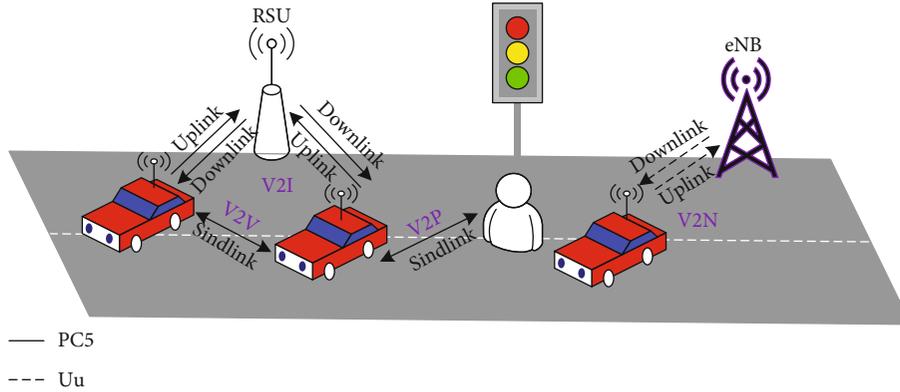
Figure 1: LTE-V2X transmission scenario.
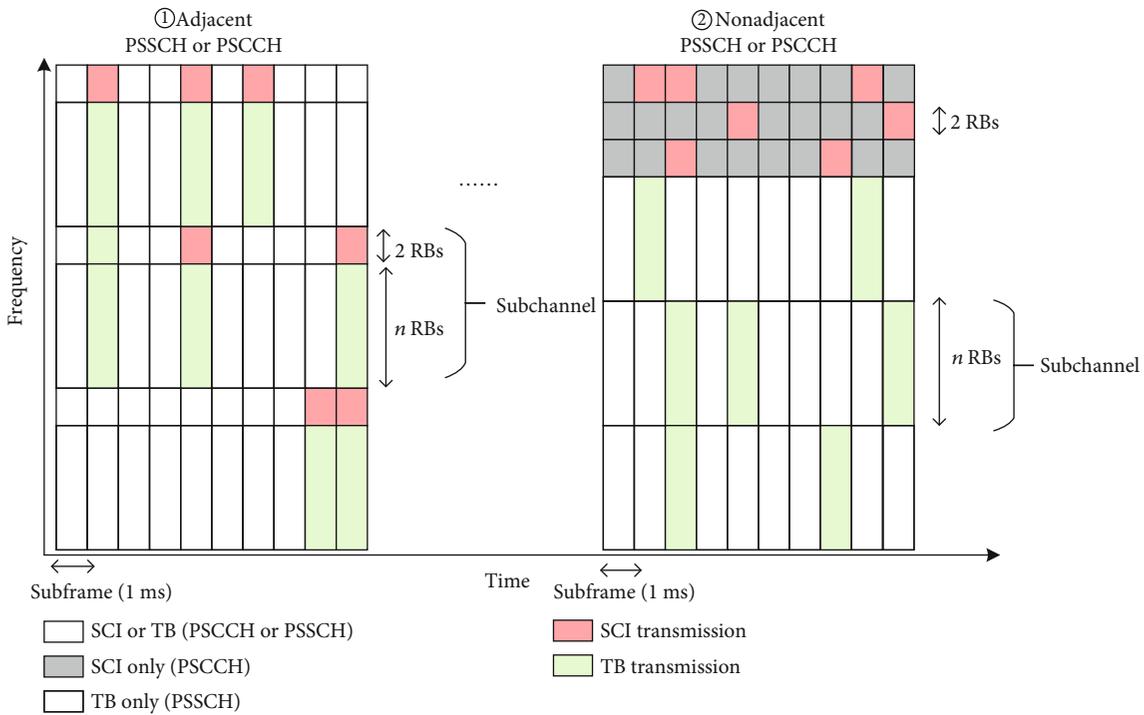
Figure 2: LTE-V2X resource configuration diagram.

standard (3GPP Release 14) in 2016. C-V2X technology on account of the cellular network can reuse cellular network infrastructure, with lower deployment cost and wider network coverage. It can realize the scenario where vehicles can travel relatively fast. In dense circumstance, C-V2X supports longer communication distances, greater capacity, better non-line-of-sight communication performance, and congestion control. In addition, C-V2X can improve communication efficiency after node synchronization through GPS, which is also not available in DSRC system, but C-V2X still has problems in roadside unit (RSU) information interaction, security certificate management, and long-term dynamic maintenance in commercial applications. From the perspective of vertical industry, to eliminate the concerns of related industries that C-V2X has not yet been tested on a

large scale, while improving the standard as soon as possible and clearing the commercial technical barriers, it is also time to prepare for testing work that complies with the C-V2X standard to verify the performance of the C-V2X communication system [7–9].

So far, manufacturers and institutions in some countries and regions have actively carried out technical research and test verification for the C-V2X communication. European countries have launched Drive C2X, C-ITS corridorl, simTD, and other projects to test and verify applications like road safety, traffic management, and environmental protection [10, 11]. In February 2012, Japan released the ARIB STD-T109 specification for 10 MHz in the 700 MHz band for V2V collision safety applications. And Japan began large-scale field testing in Hiroshima and Tokyo, respectively [1].
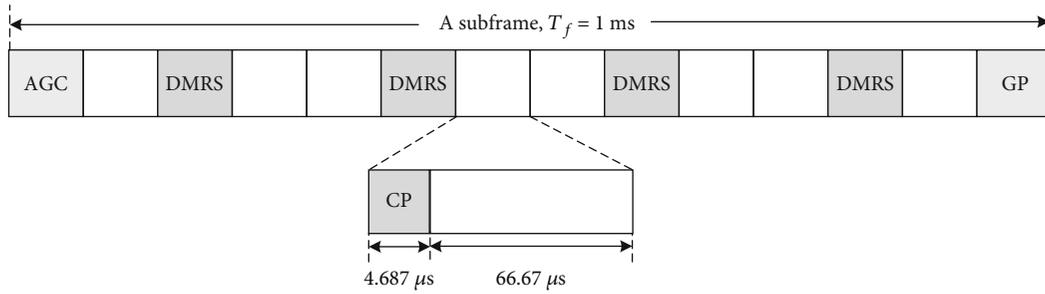
FIGURE 3: Direct link subframe structure based on the PC5 interface.
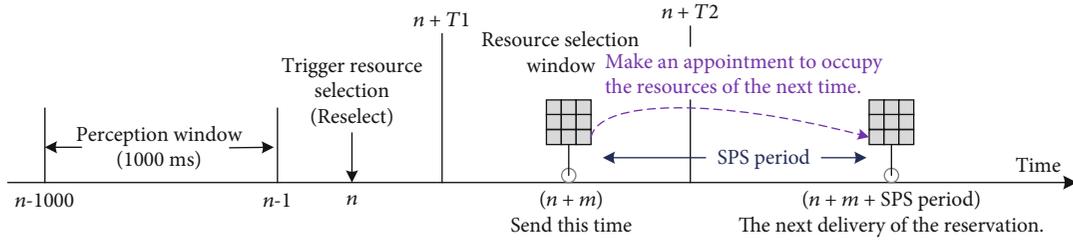
FIGURE 4: Distributed scheduling based on the PC5 interface.

Although the research of the V2X communication in the United States mainly focuses on DSRC, its domestic SAE (Society of Automotive Engineers) also established the C-V2X working group in June 2017 to implement the research on enhanced applications and direct communication [12]. In 2010, China Datang telecom technology industry group took the lead in the research of the IoV technology for intelligent transportation applications and proposed the LTE-V standard in 2013, which has now become the standard of 3GPP's LTE-V2X [3, 11]. Since 2015, China's C-V2X industry has developed rapidly, the standard system has been initially established, and the industry chain has taken shape. At the same time, enterprises related to the IoV already have a high technical strength as well as conditions for large-scale deployment and industrialization. Therefore, the CATRC (China Automotive Technology Research Center), together with CAICT (China Academy of Information and Communications Technology) [7] and other research institutes and equipment manufacturers, has actively performed laboratory and field test work in Wuxi, Shanghai, and other places. Test locations cover parks, open roads, highways, etc.

C-V2X technology is a communication technology on account of 3GPP global unified standards, including LTE-V2X for assisted driving and 5G NR-V2X for autonomous driving [12, 13]. In order to accelerate the large-scale implementation of the IoV industry, technologies and standards have been continuously improved from multiple levels. As the current C-V2X communication network technology, LTE-V2X can meet diversified IoV application scenarios and demands. In addition, it is assisted by TD-LTE [1, 3], which can make the best use of resources such as LTE deployed network and terminal chip platform, so as to save network investment and reduce chip cost. Therefore, in order to promote the industrialization of LTE-V2X and accelerate

its application as a comprehensive communication solution for vehicle-road collaboration, it is necessary to implement large-scale field tests of LTE-V2X before the formal commercial use.

Although testing based on LTE-V2X has been conducted in many parts of the world, the testing methods are not unified, and most research institutes are reluctant to publish the final test data and results. On the other hand, as the key to technology maturity and commercial use, the discussion and research of evaluation methods play a vital role in promoting technology maturity and commercial use [14, 15]. Its own scientificity and implementability are also the key factors that determine whether a certain technology or a certain product can be certified in the end. Based upon the realistic field measurement data, this paper is aimed at functional verification and network performance evaluation for typical business scenarios in LTE-V2X networking in Chongqing, China. The contribution of this paper can be summarized as follows:

(1) Aiming at the application scenario of direct vehicle communication unique to LTE-V2X, this paper introduces the key technologies applied in LTE-V2X in detail from the three aspects of physical layer, resource scheduling, and synchronization mechanism

(2) Considering the inconsistency of current IoV testing methods, this paper provides the testing scheme of IoV in the outfield. The test content consists of function and performance tests to comprehensively evaluate the communication property of LTE-V2X, where the performance test includes the static and dynamic tests. This paper also shows in detail the process of testing and the deployment of field equipment in typical scenarios

Table 1: Specific content of functional test.

Category	Mode	Test contents
Traffic safety	V2V	Forward collision warning
	V2V	Emergency brake collision warning
	V2V	Intersection collision warning
	V2V	Abnormal vehicle reminder
	V2V	Emergency vehicle reminder
Traffic efficiency	V2I	Speed guidance at intersections
Information service	V2I	Vehicle-mounted video player service

(3) In the design of the intelligent transportation system, it is required that LTE-V2X can be applied to improve road safety and traffic efficiency. Therefore, this paper shows the actual application results of devices supporting LTE-V2X, so as to verify the performance of LTE-V2X in the application layer. What is more, this paper verifies and appraise the network performance of LTE-V2X from two evaluation indexes of delay and packet loss rate (PLR) and exposes parts of the test data

The rest of this paper is organized as follows. Section 2 introduces the key technologies of LTE-V2X. Section 3 describes the methods, environment, and equipment for this test. Evaluation results and analysis are presented in Section 4. Section 5 draws a conclusion and offers ideas for future work.

2. The Key Technology

2.1. Physical Layer. LTE-V2X is an advanced information and communication technology applied in road transportation systems, with the objective of enabling information exchange between vehicle, human, infrastructure, and network. Hence, as shown in Figure 1, LTE-V2X consists of V2V (vehicle-to-vehicle), V2I (vehicle-to-infrastructure), V2P (vehicle-to-person), and V2N (vehicle-to-network) communication [15]. The vehicle establishes communication with the vehicle, RSU, or the base station in LTE (evolved Node B, eNB) through the on-board unit (OBU). In accordance with the difference of transmission modes, LTE-V2X can be divided into two communication methods: LTE-V2X-Direct and LTE-V2X-Cellular. Through the PC5 interface, V2X-Direct can not only use the dedicated frequency band of IoV (such as 5.9 GHz) to realize V2V, V2I, and V2P but also share cellular spectrum resources with cellular users. In this mode, the delay is low, and the moving speed of the vehicle is high, but good resource allocation and congestion control algorithm are needed. V2X-Cellular transmits information through the Uu interface of the cellular network and adopts the frequency band of the cellular network (such as 1.8 GHz) to make the V2X communication range wider and more stable.

The physical channel of LTE-V2X can be separated into subframes, resource blocks (RBs), and subchannels [16]. In LTE-V2X, the subframe is the most basic time series of the system. A resource block refers to a "physical resource unit" that occupies a bandwidth of 180 kHz (twelve 15 kHz subcarriers) in the frequency domain and has a duration of 1 ms in the time domain. All control signaling and data information of LTE are based on RBs. LTE-V2X supports variable bandwidth of 10-20 MHz by flexibly allocating RBs at the physical layer. Subchannel refers to a combination of RBs with the same subframe, and each subchannel may have a various number of RBs. The subchannel is used to transmit data information and control information. Data information is transmitted in the transport block (TB) of the physical sidelink shared channel (PSSCH), and sidelink control information (SCI) is transmitted in the physical sidelink control channel (PSCCH). A TB contains complete data packets to be transmitted, like beacon beacons and information transfer protocols. A node that wants to transmit TB must also transmit its associated SCI, which includes information such as the modulation and encoding scheme used to transmit TB and the RBs occupied. Since TB and its associated SCI must be transmitted in the same subframe, LTE-V2X can use either frequency division multiplexing or time division multiplexing for resources reuse. Consequently, the LTE-V2X supports HARQ (hybrid automatic repeat request transmission) which allows the same transmission to be repeated at time offset either on same frequency resources or different resources to convey the same data as needed.

Clearly by inspection of Figure 2, two resource pool configuration methods are defined in LTE-V2X. The first is the adjacent PSSCH and PSCCH, and the second is the nonadjacent PSSCH and PSCCH. The SCI of the two resource pool configurations both occupy 2 RBs to improve reliability, while the TB can occupy multiple RBs. In the first configuration mode, the SCI first occupies the first 2 RBs, and the TB occupies multiple RBs afterwards, so that one subchannel can be formed. Of course, TB can also occupy multiple subsequent subchannels (depending on its size). In the second configuration mode, the resource block is separated into multiple pools, one of which is dedicated to SCI transmission. The remaining pool is used to transmit TB and is divided into service subchannels [15, 16].

LTE-V2X uses single-carrier frequency division multiple access (SC-FDMA) technology to reduce the impact of the excessive peak-to-average power ratio (PAPR), so that it can have a larger transmission under the same power amplifier power. In order to improve the spectrum utilization ratio under the condition of high mobility, LTE-V2X transmits an
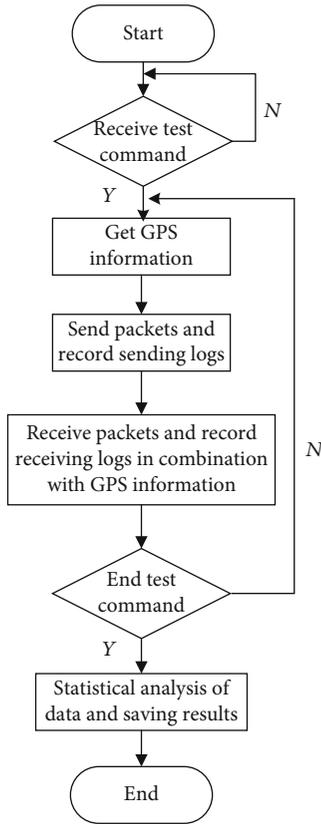
Figure 5: LTE-V2X performance test process.

OFDM waveform with a conventional cyclic prefix (CP) and sets the subcarrier spacing to 15 kHz. The structure of the direct link subframe based upon the PC5 interface is shown in Figure 3. The length of each subframe T_f is 1 ms, and the length of each symbol in the subframe is $71.357 \mu s$, so each subframe of direct link can contain 14 OFDM symbols. In a subframe, the first and last symbols are used for automatic gain control (AGC) and guard period (GP), respectively. What is more, in order to reduce the impact of the Doppler effect, the design of the demodulation reference signal (DMRS) column structure in LTE-D2D (device-to-device) is used in LTE-V2X. The DMRS in each subframe is increased from 2 columns to 4 columns, which increases the pilot density in the time domain, so that the channel detection, estimation, and compensation of high frequency in typical high-speed scenes can be effectively processed. The remaining 8 symbols are used to transmit data information. Moreover, LTE-V2X uses turbo codes, which can achieve higher reliability at the same transmission distance. LTE-V2X with turbo codes is designed to facilitate decoding capability even at lower signal-to-noise ratio (SNR), whereas for DSRC with convolutional codes requires higher SNR for successful decode [17, 18].

2.2. Scheduling of Resources. LTE-V2X supports both centralized scheduling (mode 3) and distributed scheduling (mode 4). Mode 3 implements centralized scheduling based on the Uu interface. The selection and coding method of the subchannel of the communication link are directly controlled by the eNB in the cellular network. The eNB provides dynamic scheduling or activated semipersistent scheduling (SPS) according to the service type of the terminal. In order to reduce the delay caused by signaling interaction, mode 4 provides distributed resource scheduling for vehicles. This scheduling scheme uses a "sensing + reservation SPS" approach, as shown in Figure 4. User equipment (UE) selects subchannel access by itself and then perceives resource occupancy by measuring received signal strength indication (RSSI) energy in the resource pool. Resource selection measures the RSSI energy on resources available and sorts them in descending order per energy levels. It then chooses the lowest 20% energy resources and randomly picks resources from these for transmission [18]. After selecting appropriate resources, the UE will periodically send these resources a certain number of times or until resource reselection is triggered. This method can be conducted without the support of any cellular base station, taking advantage of the periodic characteristics of V2X services. This distributed scheduling scheme based on the PC5 interface can not only carry the periodic V2X services waiting to be sent but also make full use of the sensing results to avoid resource conflicts, which helps to improve resource utilization and transmission reliability.

2.3. Synchronization Mechanism. In LTE-V2X, there are three synchronous sources: eNB, Global Navigation Satellite System (GNSS), and UE. When the eNB is used as the synchronization source, the nodes in the cellular coverage are synchronized with the eNB. Some uncovered nodes can receive the synchronization signal forwarded by the nodes in cellular coverage, so the partially covered nodes forward the synchronization information of the nodes in cellular coverage to the nodes outside cellular coverage. In the LTE-V2X system, communication nodes support GNSS module, which has high timing and frequency accuracy. Therefore, nodes that can directly obtain reliable GNSS signals are able to directly serve as synchronization sources to provide synchronization information to surrounding nodes. When LTE-V2X shares carriers with cellular systems like LTE, the transmission signals of LTE-V2X through communication may interfere with the uplink of cellular networks. In this case, eNB is still considered as the main synchronization source, and then eNB can broadcast the time deviation between eNB and GNSS to UE for adjustment compensation. In general, the synchronization source and mode are configured by the eNB in cellular coverage, and the synchronization source is determined by the preconfiguration mode outside cellular coverage, so as to achieve unified synchronization timing of the whole network [3, 12].

In accordance with the traditional LTE-D2D mechanism, the enhanced synchronization source priority can be supported by establishing a new connection to the sidelink synchronization signal (SLSS) and the physical sidelink broadcast channel (PSBCH) [19]. Considering the protection of the LTE-Uu uplink transmission and ensuring the accuracy of the timing and frequency of the synchronization source, the rules of the synchronization source priority should be implemented according to the eNB or GNSS
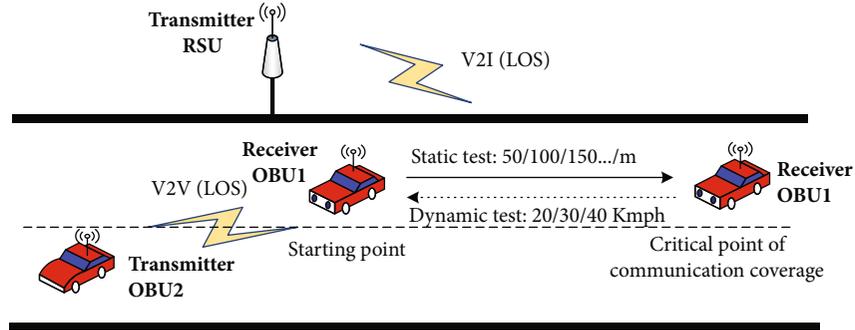
FIGURE 6: V2V/V2I performance test of the LOS scenario.
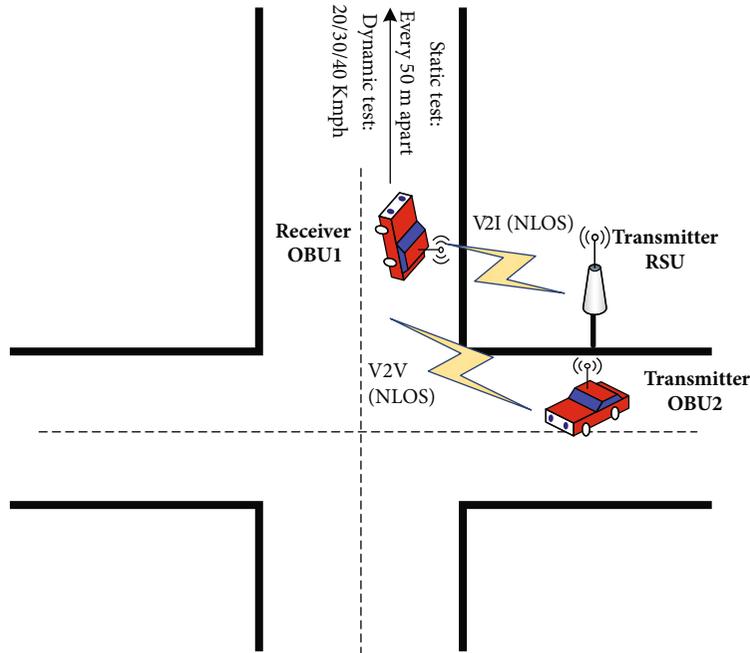
FIGURE 7: V2V/V2I performance test of the NLOS scenario.

synchronization configuration. In other words, eNB can configure to prioritize either GNSS or UE. Moreover, GNSS has higher priority when UE does not detect any cell in any carrier, and UE does not detect any SLSS transmissions that are directly synchronized to eNB.

3. Test Scheme

3.1. Method and Content. In recent years, with the gradual improvement of the LTE-V2X standard, it is particularly important to accelerate the implementation of standard. Therefore, there is an urgent need to test and verify LTE-V2X-related products, which is a necessary stage for the popularization and improvement of each standard and technology [20, 21]. Considerable laboratory evaluations and field tests have been conducted in many places [22]. The test object of laboratory evaluation is module, which mainly investigates communication protocol consistency and interoperability. Conformance testing includes radio frequency consistency (signal transmission, reception, and demodulation performance), radio resource management consistency,

and communication protocol consistency. The radio frequency conformance test mainly examines whether the reception, transmission, and demodulation performance of the LTE-V2X radio frequency meet the national radio management and LTE-V2X communication requirements. The conformance test of the communication protocol includes the conformance test for the underlying layer protocol and the upper layer protocol, which ensures that both parties of the communication have a unified and unambiguous understanding of the protocol and the corresponding implementation. The object of the outfield test is vehicle, mainly from the following aspects. On the one hand, it is necessary to verify whether the designed function meets expectations. The function of the application layer of the LTE-V2X system was tested in the open road of Chongqing Automotive Research Institute in this paper. Specific test contents are exhibited in Table 1, which can be divided into traffic safety, traffic efficiency, and information service.

On the other hand, the performance of the communication system under diverse environments, road conditions, and vehicle speeds needs to be examined. The verified

(a) BBU

(b) Base station

(c) MEC equipment

FIGURE 8: Part of the equipment used in the test.

performance indicators include packet reception success probability, communication delay, and coverage. Network measurement methods include active measurement and passive measurement. Active measurement is to send probe data packets to the network and measure the network performance by analyzing the changes that are affected by the data packets. Passive measurement is to capture and analyze data packets by arranging measurement devices in the network to measure network performance. Passive measurement does not send measurement packets and will not affect the normal flow of the network, but its implementation is more complicated and requires higher performance of the measurement device. Therefore, passive measurement is more suitable for network traffic measurement, and active measurement is suitable for network performance measurement [22]. The method of the performance testing is to send probe packets to the network in this paper, and the network performance is measured by analyzing the changes that occur when the packets are affected during transmission, as shown in

Figure 5. Firstly, the GPS information of transmitter and receiver should be read. Then, the OBU/RSU at the transmitter actively sends the detection packets and records the sending logs. The time of each data communication is 1 second. The OBU at the receiving end receives the data packets and records the receiving logs in combination with GPS information, and then the statistical analysis model is used to infer the PLR and average delay of the internal link.

Measurement of one-way end-to-end delay requires clock synchronization, which is more difficult to achieve in actual measurement. Therefore, the measurement of network delay usually requires the use of round-trip time (RTT), which is the time interval required for a packet to travel from the source node to the destination node, so as to avoid the problem of clock synchronization. The specific approach is to calculate the RTT by adding a time stamp to each packet. Before sending a message, the OBU at the transmitting end adds a time stamp to each message and records it as TS1. When receiving the message, the OBU at the receiving end

FIGURE 9: The LTE-V2X test environment.

TABLE 2: Base station/OBU/RSU parameter configuration.

Equipment	Base station				
Parameter	Frequency	Bandwidth	Number of antennas	Uplink power control/HARQ	Rated transmit power
Configuration	2555~2565 MHz	10 MHz	2T×2R	Enable	2×10 W
Equipment	OBU/RSU				
Parameter	Frequency	Bandwidth	Transmit power	Message frequency	Packet size
Configuration	5855~5925 MHz	10 MHz	23 dBm	One packet/100 ms	78 bytes

adds the second timestamp and records it as TS2; then, it replies with an ACK message and marks the third timestamp as TS3. The OBU/RSU at the transmitter receives the ACK message with the fourth timestamp, which is recorded as TS4. Therefore, the RTT is calculated as follows:

$$RTT = (TS4 - TS1) - (TS3 - TS2). \tag{1}$$

In order to test the accuracy of the results, it is necessary to calculate a round-trip time (TS4 − TS1) from the sending end to the receiving end to the sending end and subtract the message processing waiting time (TS3 − TS2) at the receiving end. Hence, the end-to-end delay is half of RTT.

PLR is defined as the ratio of lost packets to all packets in the transmission, which is mainly related to network traffic, and packet loss will be caused by network congestion.

$$L(D, y) = (1 - y)^a y^b, \tag{2}$$

where a is the number of data packets received in a test time window, b is the number of unreceived data packets, $D = (a, b)$ is the set of lost data packets at one time, y is the PLR, and $L(D, y)$ is the maximum likelihood function of the packet loss rate y. Next, we take the logarithm of both

sides of equation (2),

$$lnL(D, y) = a(1 - y) + by. \tag{3}$$

Then, we need to take the derivative of both sides of equation (3) with respect to y and set the value of the derivative to be 0.

$$\frac{-a}{1 - y} + \frac{b}{y} = 0. \tag{4}$$

Referring to the above equation, the maximum likelihood estimation $\hat{y}$ of PLR can be obtained as follows:

$$\hat{y} = \frac{b}{a + b}. \tag{5}$$

The performance test of the outfield can be divided into static test and dynamic test, where the static test and dynamic test of the line-of-sight (LOS) scenario are demonstrated in Figure 6. For the static test of LOS, a fixed-point test is implemented. To put it simply, let the two cars conduct a V2V performance test every 50 meters (straight line distance) apart and record the data at this distance. Similarly, a V2I performance test is performed every 50 meters apart. When the test vehicle is unable to establish communication during movement, it is

(a) forward/emergency brake collision warning

(b) intersection collision warning

后方有车辆快速接近中请注意

Be aware that there is a vehicle approaching quickly in the rear.

(c) abnormal vehicle reminder

后方有特殊车辆请注意

Please note that there are special vehicles behind.

(d) emergency vehicle reminder

MAX-7.20km/h MAX-90.00km/h

建议加速到7公里每小时以便顺利通过红绿灯

礼嘉立交

It is recommended to accelerate to 7 km/h in order to pass traffic lights smoothly.

(e) speed guidance at intersections

FIGURE 10: Results of functional tests.

TABLE 3: Static test results of the LOS scenario.

Distance/m	Average delay/ms		Maximum delay/ms		Packet loss rate	
	V2V	V2I	V2V	V2I	V2V	V2I
50	7	7.5	7	7.5	0.66%	0%
100	7.5	7	7.5	9	0%	0%
150	7	7.5	7	7.5	0%	0%
200	7	7.5	8	7.5	0%	0%
250	7	7	7.5	7	0%	0.15%
285	7	7	8	8	1.43%	1.10%

considered to have exceeded the communication coverage of V2V/V2I. For the dynamic test of the LOS, in the initial state, one vehicle (at transmitter) keeps stationary, and the other vehicle (at receiver) moves away from the stationary vehicle at 20/30/40kmph, respectively. When the relative distance between the two vehicles exceeds their communication range, the vehicle at the transmitter moves towards the vehicle at the receiver. The above process needs to be repeated several times, and the complete test logs are recorded in order to test the network performance of LTE-V2X at different speeds.

The test of non-line-of-sight (NLOS) is similar to the it of LOS, except that the two test vehicles are located perpendicular to each other at the intersection. In Figure 7, for the static test of V2V under NLOS, the fixed-point test is also implemented, which means that two vehicles are tested for the network performance of V2V every 50 meters in a straight-line

TABLE 4: Static test results of the NLOS.

Distance/m	Average delay/ms		Maximum delay/ms		PLR	
	V2V	V2I	V2V	V2I	V2V	V2I
50	8	7	8.5	7.5	0.91%	0%
75	8	7.5	8	7.5	6.97%	0%

TABLE 5: Average and maximum delay of V2V/V2I dynamic test.

Speed km/h	Average delay /ms				Maximum delay /ms			
	LOS		NLOS		LOS		NLOS	
	V2V	V2I	V2V	V2I	V2V	V2I	V2V	V2I
20	7.25	7.5	7.5	7.17	8	8.5	9	7.5
30	8.125	8	7.83	7.17	9	9	9	8.5
40	7.5	7.5	8	7.17	8.5	8.5	9	7.5

distance. Dynamic testing, in which one vehicle stays at rest while the other moves away from the stationary vehicle at 20/30/40kmph in the communication range (in this case, the vehicle do not need to return), also requires repeated recording of delays and PLR. The V2I test method is resemblance to that of V2V, which also requires the movement of the vehicle at the receiving end. In addition, it is necessary to evaluate the interoperability between the same equipment of different manufacturers and different equipment of the same manufacturer under various working conditions and environments.

3.2. Environment and Equipment. This test took place in the open road of industrial park in China. The OBU and its display equipment were installed inside the test vehicles, and two antennas were installed on the roof of each vehicle. In order to better perceive the environment, the RSU and the camera were deployed alongside the traffic lights at the intersection. The equipment based on multiaccess edge computing (MEC) was deployed in the indoor baseband processing unit (BBU) room as seen in Figure 8(c). The actual test environment is shown in Figure 9.

The whole test system was divided into three parts: "terminal," "edge equipment," and "V2X service platform." The terminal contained OBU, person, and vehicle. The edge equipment included RSU and MEC-based device. The RSU received messages from the V2X platform or edge device and multicast to the terminal device in the area through the PC5 interface [23–25]. Moreover, RSU could collect the message from OSU to the upper layer through the PC5 or Uu interface. The equipment based upon MEC received message from RSU and distributed the processed message to RSU. It is possible to reduce the end-to-end network delay in the Uu mode by decreasing routing nodes for data transmission through the construction of an LTE network architecture based on MEC. The V2X cloud service platform was used to process information from terminals and MEC-based device, implement comprehensive scheduling and optimization, and improve driving safety and traffic efficiency. So far, the construction of seven LTE base stations based upon

the 2600 MHz frequency band, one set of MEC equipment, and a set of evolved packet core (EPC) network equipment have been completed inside the industrial park [26].

The test equipment can monitor the communication status of V2X in real time and has the recording function of sending logs or receiving logs. During the test, the antenna was installed vertically at the center of the roof of the vehicle. Single antenna transmission and dual antenna reception modes were used throughout the communication. This test network and V2X equipment were provided by Datang [3, 11]. The parameter configuration of base station and OBU/RSU are demonstrated in Table 2.

4. Evaluation Results and Analysis

The security class business, efficient traffic, and video playback business-based MEC in Table 1 were tested. Figure 10 shows the test vehicle and display device during test. It is verified that all business functions are normal. Figure 10(a) indicates that when the host vehicle (HV) and the vehicle in front of the same lane are in danger of rear-end collision or the vehicle in front is in emergency braking, the on-board device will send an early warning message to remind the rear host vehicle to avoid collision. When the HV is driving to the intersection and there is a danger of collision with a faraway vehicle traveling sideways, the driver is alerted by warning to avoid the collision, as shown in Figure 10(b). When the far vehicle rapidly approaches the host vehicle from the rear, the driver of the HV is alerted, as exhibited in Figure 10(c). While the HV is running, the emergency vehicle in the rear issues a reminder to alert the host vehicle in front, as demonstrated in Figure 10(d). As shown in Figure 10(e), the RSU multicasts real-time information to the vehicle and then the OBU reminds the driver to perform operations such as accelerating through or decelerating through the current vehicle speed, position, and remaining time of signal phase.

In the performance test, the static test results of V2V and V2I at LOS are demonstrated in Table 3 and Table 4. Table 3 indicates that different distances and LOS/NLOS have little effect on the change of delay, and the delay is about 7 ~ 8 ms. In the performance test of LOS, phenomenon of packet loss in V2V and V2I is rare when the distance is less than 250 meters, but obvious packet loss occurs when the distance is higher than 250 meters. (the site is limited to 290 meters, so the PLR at a longer distance cannot be tested.)

During the test, it is found that the NLOS has a greater impact on the communication distance, seen as Table 4. When the distance between V2I and V2V is larger than 70 and 75 in meters, respectively, communication cannot be established, but there is no phenomenon of communication failure when reaching the end of the road in the LOS. In other words, the communication distance between V2I and V2V is 70 m and 75 m, respectively, in the NLOS. Besides that, it is not difficult to find that the PLR of V2V in NLOS has increased significantly.

After undergoing multiple dynamic tests, the test results demonstrated in Table 5 can be obtained, which contains the average delay and maximum delay of V2V and V2I in the LOS and NLOS. As can be seen from Table 5, different

(a) Average packet loss rate of the V2V/V2I dynamic test

(b) Maximum packet loss rate of the V2V/V2I dynamic test
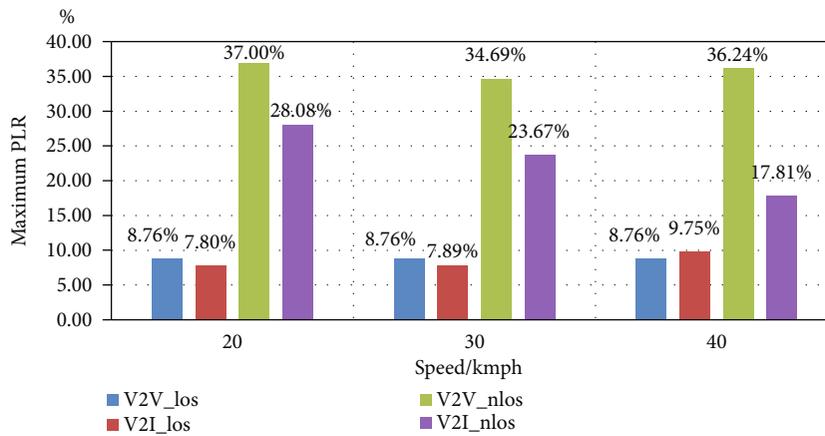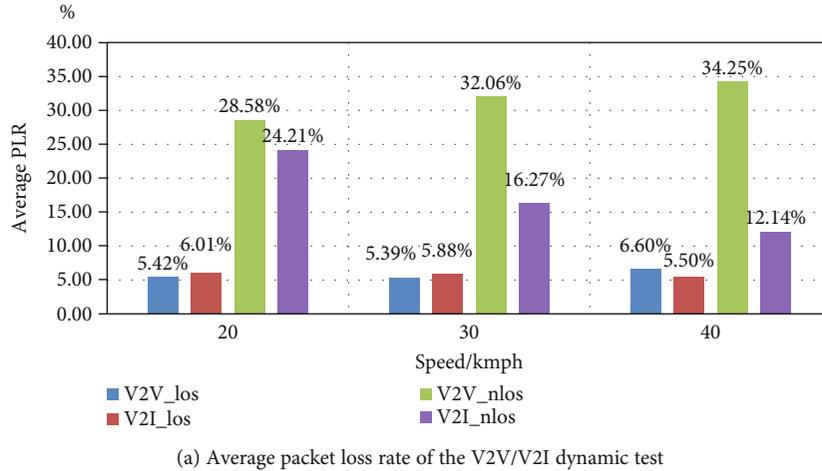
FIGURE 11: Packet loss rate of the V2V/V2I dynamic test.

scenarios and speeds have a relatively small impact on the change of delay. The average delay is maintained at 7 ~ 8 ms, and the maximum delay is 9 ms.

The average and maximum PLR of dynamic tests are exhibited in Figure 11. It can be seen from Figure 11 that the maximum PLR of V2V and V2I in the LOS is not more than 10%, and the average PLR remains around 5%~6%. The maximum PLR of V2I and V2V in the NLOS was 28.08% and 37.00%, respectively. The average PLR of V2I was around 17.5%, and the average PLR of V2V was above 30%.

As can be seen from Table 5 and Figure 11, the performance of V2I in NLOS is significantly better than V2V, the reason of which may be the difference between RSU equipment and OBU equipment, which means the former has higher antenna gain. Furthermore, by comparing dynamic test and static test results, the PLR of the dynamic test has an obvious increase compared with the static test, while the change of delay is not obvious.

5. Conclusions and Future Work

On account of the LTE-V2X networking solution, this paper conducted field testing and verification on the open road of Chongqing Automobile Research Institute in China. Firstly, the key technologies applied in LTE-V2X are introduced. In the physical layer, four DMRS signals are introduced in each subframe to counter the Doppler effect caused by high-speed movement, and two methods of resource reuse (FDM and TDM) and resource allocation (nonadjacent PSSCH and PSCCH) are adopted. In terms of resource scheduling, LTE-V2X proposes a distributed scheduling method based on the PC5 interface. In terms of the synchronization mechanism, three synchronization sources including base station, GNSS, and UE autonomy are selected. Then, it introduces the framework of IoV testing and provides the specific methods and contents of the test from the functions and network performance that IoV needs to have. Specifically, the function of the system is tested from three aspects of traffic safety, traffic efficiency and information service, and end-to-end delay and packet loss rate that are used as evaluation indexes of the performance test. The results verify the effectiveness and reliability of the application layer communication performance. We hope the overall test scheme and test results can lay a foundation for future research.

Due to space and equipment constraints, we only conducted tests on open road sections and just selected delay and packet loss rate as indicators for evaluating network performance. Therefore, in future work, we hope to carry out

tests on a variety of classic scenarios, such as viaducts, tunnels, multivehicles, and mines, and design appropriate test methods according to specific scenarios. What is more, evaluation indicators can also be expanded from multiple dimensions, such as signal-to-noise rate, signal receiving power, and data transmission rate. Certainly, it is also necessary to implement reasonable equipment deployment on the basis of particular scenarios. How to deploy RSU to make the network performance better requires the combination of testing in typical scenarios, semiphysical simulation in the laboratory, and rigorous theoretical analysis.

Data Availability

The data used to support the findings of this study are currently under embargo, so cannot be made freely available.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] S. Chen, J. Hu, Y. Shi, and L. Zhao, "LTE-V: a TD-LTE-based V2X solution for future vehicular network," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 997–1005, 2016.

[2] K. Li, Y. Dai, S. Li, and M. Bian, "State-of-the-art and technical trends of intelligent and connected vehicles," *Journal of Automotive Safety and Energy*, vol. 8, no. 1, pp. 1–14, 2017.

[3] S. Chen, J. Hu, Y. Shi, and L. Zhao, "Technologies, standards and applications of LTE-V2X for vehicular networks," *Telecommunications Science*, vol. 34, no. 4, pp. 1–11, 2018.

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

[5] "IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks–Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," in *IEEE Std 802.11-2016 (Revision of IEEE Std 802.11-2012)*, pp. 1–2793, March 2012.

[6] Y. Lv, Y. Wang, X. Liu, R. Xu, J. Fang, and X. Peng, "Research on performance testing for urban scenario based on terminal equipment of LTE-V2X vehicle network," in *2018 14th IEEE International Conference on Signal Processing (ICSP)*, pp. 993–996, Beijing, China, August 2018.

[7] CAICT, "White paper of internet of vehicles," 2017, http://www.caict.ac.cn/kxyj/qwfb/bps/202001/t20200102_273007.htm.

[8] "IMT-2020 promotion group: C-V2X task force, C-V2X white paper," June 2018, http://www.caict.ac.cn/kxyj/qwfb/bps/201806/t20180621_174522.htm.

[9] L. Zhu, Y. Li, F. R. Yu, B. Ning, T. Tang, and X. Wang, "Cross-layer defense methods for jamming-resistant CBTC systems," in *IEEE Transactions on Intelligent Transportation Systems* IEEE.

[10] H. Wang, F. R. Yu, L. Zhu, T. Tang, and B. Ning, "Finite-state Markov modeling for wireless channels in tunnel communication-based train control systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 3, pp. 1083–1090, 2014.

[11] "simTD: Shaping the future of road safety and mobility via car-to-x communication," 2018, https://www.consilium.europa.eu/.

[12] S. Chen, J. Hu, Y. Shi, L. Zhao, and W. Li, "A vision of C-V2X: technologies, field testing, and challenges with chinese development," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 3872–3881, 2020.

[13] J. Qiu, Y. Chen, and Q. Liu, "IoV standardization and its evolution strategy," *Mobile Communications*, vol. 42, no. 4, pp. 41–47, 2018.

[14] L. Liu, C. Tao, J. Qiu et al., "Position-based modeling for wireless channel on high-speed railway under a viaduct at 2.35 GHz," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 834–845, 2012.

[15] H. Zhou, W. Xu, J. Chen, and W. Wang, "Evolutionary V2X technologies toward the Internet of Vehicles: challenges and opportunities," *Proceedings of the IEEE*, vol. 108, no. 2, pp. 308–323, 2020.

[16] R. Molina-Masegosa and 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, 2017.

[17] F. Mei, S. Liu, J. Wang, Y. Ge, and T. Feng, "Negotiation-free encryption for securing vehicular unicasting communication," *Applied Sciences*, vol. 9, no. 6, p. 1121, 2019.

[18] "QualcoMM, Cellular-V2X Technology Overview," 2019, https://www.qualcomm.com/videos/cellular-v2x-overview.

[19] "Huawei Technology Co., Ltd., LTE-V2X Concept Slides," https://www.huawei.com/cn.

[20] Z. Peng, *Research and Optimization of Resource Scheduling Mechanism for 5G V2X*, Beijing University of Posts and Telecommunications, Beijing, 2019.

[21] X. Cheng, R. Zhang, and L. Yang, "Wireless Toward the Era of Intelligent Vehicles," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 188–202, 2019.

[22] K. S. Kumar, H. Nguyen, Y. H. Lee, and Y. L. Guan, "A V2X Communication System Test on Sea," in *2019 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*, pp. 929-930, Atlanta, GA, USA, July 2019.

[23] J. Zhang, Z. Li, and C. Wang, "Performance test and analysis of LTE network for internet of vehicles," *Computer Engineering*, vol. 44, no. 7, pp. 303–307, 2018.

[24] A. Martínez, E. Cañibano, and J. Romo, "Analysis of low cost communication technologies for V2I applications," *Applied Sciences*, vol. 10, no. 4, p. 1249, 2020.

[25] J. Feng, F. R. Yu, Q. Pei, J. Du, and L. Zhu, "Joint optimization of radio and computational resources allocation in blockchain-enabled mobile edge computing systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 4321–4334, 2020.

[26] L. Du, Q. Ma, J. Zhang, C. Zhang, Y. Wang, and Q. Gu, "Scenario test system for V2X applications based on cooperative system," in *18th International Symposium on Distributed Computing and Applications for Business Engineering and Science*, Wuhan, China, China, November 2019.

WILEY | Hindawi

Research Article

Joint V2V-Assisted Clustering, Caching, and Multicast Beamforming in Vehicular Edge Networks

Kan Wang [iD],[1] **Ruijie Wang** [iD],[1] **Junhuai Li** [iD],[1] **and Meng Li** [iD][2]

[1]*School of Computer and Science Engineering, Xi'an University of Technology, Xi'an 710048, China*
[2]*Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China*

Correspondence should be addressed to Meng Li; limeng720@bjut.edu.cn

As an emerging type of Internet of Things (IoT), Internet of Vehicles (IoV) denotes the vehicle network capable of supporting diverse types of intelligent services and has attracted great attention in the 5G era. In this study, we consider the multimedia content caching with multicast beamforming in IoV-based vehicular edge networks. First, we formulate a joint vehicle-to-vehicle- (V2V-) assisted clustering, caching, and multicasting optimization problem, to minimize the weighted sum of flow cost and power cost, subject to the quality-of-service (QoS) constraints for each multicast group. Then, with the two-timescale setup, the intractable and stochastic original problem is decoupled at separate timescales. More precisely, at the large timescale, we leverage the sample average approximation (SAA) technique to solve the joint V2V-assisted clustering and caching problem and then demonstrate the equivalence of optimal solutions between the original problem and its relaxed linear programming (LP) counterpart; and at the small timescale, we leverage the successive convex approximation (SCA) method to solve the nonconvex multicast beamforming problem, whereby a series of convex subproblems can be acquired, with the convergence also assured. Finally, simulations are conducted with different system parameters to show the effectiveness of the proposed algorithm, revealing that the network performance can benefit from not only the power saving from wireless multicast beamforming in vehicular networks but also the content caching among vehicles.

1. Introduction

Smart devices equipped with the capability to interact with the physical environment and the function to communicate with each other are prompting the Internet towards the so-called Internet of Things (IoT) [1]. IoT holds the promise to improve our lives and the way we interact with devices, such as actuators, sensors, cell phones, and home automation. With the ever-increasing proliferation of IoT devices, the worldwide multimedia traffic is anticipated to experience a rapid growth in the 5G era [2]. Different from the legacy IoT with information exchange at the byte level, image, audio, video, and other traffic in the 5G era are typically with a large volume of information, thereby bringing forth the new terminology, namely, multimedia IoT (MIoT). As an emerging type of IoT, MIoT denotes the IoT with multimedia traffic as outputs and inputs and has been extensively used in healthcare, smart homes,

communication-based train control system (CBTC), and Internet of Vehicles (IoV) [3–6].

The introduction of IoV-based vehicular edge networks and the increasing demand for IoV services have imposed more stringent constraints on the quality-of-service (QoS) requirements of image and video, particularly for high-quality real-time multimedia services in fast-moving vehicles. These constraints indeed pose challenges in exploiting the full potential of vehicle-to-vehicle (V2V) communications, e.g., resource allocation, multiple-input-multiple-output-(MIMO-) based beamforming, network architecture, multicast routing, and dynamic controls [7–9].

To better cope with these constrains imposed on multimedia content retrieval in IoV-based vehicular edge networks, wireless content caching begins to emerge as promising solutions [10]. By bringing multimedia contents closer to vehicles via prefetching them from the Internet or core networks, the traffic loads and network costs can be

significantly reduced. In the meantime, the QoS for IoV applications could also be improved, since lower access delay and higher data rate are enabled by caching in edge nodes (e.g., base stations (BSs) and roadside units (RSUs)). Therefore, wireless content caching contributes to the agile multimedia content distribution with a higher data rate and enhanced broadband connectivity. On the other hand, as the extension of unicast transmission and by exploiting the broadcast nature of wireless channels, the point-to-multipoint multicast transmission provides a more efficient capacity-offloading approach, to deliver the identical content to multiple vehicles on the same frequency band. Thus, the integration of caching with multicasting becomes the key enabler in IoV-based edge networks, which could not only reduce the flow (traffic) cost in core or transport networks but also improve the spectrum efficiency in vehicular edge networks [11].

Besides, there have been some works related to caching-based multicast beamforming design [12–15]. Nevertheless, the integration of caching with multicasting in vehicular edge networks is not smooth and still faces several challenges, due to the following observations. First, the caching decision needs to operate at the large timescale (e.g., several minutes), and the multicast beamforming has to be tailored at the small timescale (e.g., several seconds) to exploit the mobility. That is, it is a prerequisite to decouple the joint caching and multicast beamforming to different timescales. Second, when the edge network is confronted with IoV in the 5G era, numerous vehicles emerge to assist the V2V communications between each other. Thus, although with smaller storage size compared to the BS, caching profits in vehicles are nonnegligible, which yet has not been studied thoroughly, especially in the caching-based multicast beamforming. Third, the cooperative multicasting necessitates the clustering of vehicles (namely, deciding on which vehicle should be picked into the cluster to serve the same receiver), which needs to be optimized at the large timescale as well.

In this paper, we study the joint V2V-assisted clustering, caching, and wireless multicast beamforming in vehicular edge networks. The distinctive features of this work are as follows:

The distinctive features of this work are as follows:

(1) We present a two timescale-based V2V-assisted clustering, caching, and multicast beamforming problem, with the objective of minimizing the total network cost (involving both flow and power costs), subject to the QoS constraint for each multicast group, and the caching memory limitation for each vehicle. With the two-timescale setup, different types of variables are decoupled at separate timescales

(2) At the large timescale, we leverage the sample average approximation (SAA) technique to resolve the joint clustering and caching problem. We first reformulate and relax the stochastic original problem as a deterministic integer linear programming (ILP) one and then exhibit the equivalence of optimal solutions

between the original one and its relaxed linear programming (LP) counterpart, thereby simplifying the computation substantially

(3) We leverage the successive convex approximation (SCA) method to solve the multicast beamforming problem at the small timescale. As such, the original nonconvex problem can be transformed into a series of convex subproblems, with the convergence assured

(4) Simulations are executed with different system parameters to show the effectiveness of the proposed algorithm, as well as the convergence of the proposed two-timescale setup. Simulation results reveal that the system performance could benefit from both the power saving from multicasting and the content caching among vehicles

The remainder of this work is organized as follows. In Section 2, we list works related to caching-based multicasting and V2V-assisted caching, respectively. In Section 3, we present the system model and formulate a joint clustering, caching, and multicast beamforming problem, to minimize the total network cost subject to QoS constraints. In Section 4, we leverage the SAA and SCA methods to solve the large- and small-timescale problem, respectively. In Section 5, we present and discuss the simulation results. We conclude this paper in Section 6.

2. Related Work

Some recent studies related to caching-based multicast beamforming and V2V-assisted caching in edge networks are presented. We believe these works can motivate more achievements in the academia and industrial domain.

2.1. Caching-Based Multicasting. In the literature, multicast beamforming has emerged as an effective approach to mitigate the interference in cellular systems, e.g., in smart-grid powered cellular networks [16] or in full-duplex cellular networks [17]. Furthermore, the integration of caching and multicasting could multicast the identical content to multiple receivers in the same frequency band, and thus attracts significant interests.

In particular, in the cache-enabled cloud radio access network (RAN), Tao et al. [12] investigated the joint design of content-centric BS clustering and multicast beamforming for wireless content delivery. In particular, the problem is formulated as one mixed-integer nonlinear programming (MINLP) problem, and a sparse multicast beamforming algorithm is proposed based on the L_0 norm approximation. In the two-tier heterogeneous networks, Cui et al. [13] considered a random caching and multicasting mechanism with caching distributions as design parameters, to enable the efficient content dissemination. First, the successful transmission probabilities in the general and high signal-to-noise ratio region are derived via stochastic geometry; then, a nonconvex joint caching and

multicasting problem is formulated to maximize the successful transmission probability in the asymptotic region.

Also, in cache-enabled ultradense cellular networks, Nguyen et al. [14] proposed a cooperative multicast beamforming approach to improve the cost-efficiency. Besides, Zhou et al. [15] minimized the energy cost in a multicell multigroup setup, involving caching, computing, and communication resources.

2.2. V2V-Assisted Caching. As an analogous terminology to V2V, device-to-device (D2D) caching refers to the caching of popular contents in mobile devices. D2D could directly deliver the content to adjacent devices, thereby offloading the BS's traffic. Malak et al. [18] leveraged the stochastic geometry theory to derive the outage probability in the presence of interference and noise, designed the distributed caching strategy, and maximized the density of successful receptions. Jiang et al. [19] modelled the local caching as a backpack problem and characterized the D2D pairing as the maximum weight matching problem in the bipartite graph, to maximize the BS traffic offloading. In addition, Cacciapuoti et al. [20] claimed that in D2D networks, caching performance is closely related to devices' interests and preferences in multimedia contents. In particular, content popularity focuses on community interests, while device preference reflects the difference of request probability between different individuals. Therefore, Zhang et al. [21] designed a unified D2D-caching utility function, not only considering the D2D communication distance but also incorporating the similarity of preferences among adjacent devices.

With recent advents in the IoV paradigm, Su et al. [7] developed an edge caching mechanism in RSUs. Specifically, involving the content access pattern, vehicle velocity, and traffic density, the features of content requests are analysed, followed by a model to decide whether and where to cache the content. By proposing an analytical model, Tan et al. [22] quantified the effects of velocity, traffic density, and service rate on the content hit ratio. Also focusing on the edge caching in RSUs (termed as edge nodes therein), Zhao et al. [23] proposed one cost-efficient method to minimize the time-averaged service response delay, by jointly studied caching, request routing, and wireless resource allocation over the long run in an online manner. To solve the dynamics in IoV, Lyapunov optimization is leveraged to make near-optimal decisions, with a stable system assured. In addition, Xiao et al. [24] proposed one adaptive and dynamic user-centric virtual cell scheme to facilitate the multicasting of vehicular-to-everything (V2X) message, followed by a max-min fair problem formulation. Furthermore, focusing on 802.11p protocols rather than 3GPP standards, Wu et al. [25] proposed one semi-Markov decision process- (SMDP-) based formulation to maximize the long-term reward in vehicular edge networks, by jointly considering the transmission delay, computing delay, and task diversity.

2.3. Discussions. Edge computing and edge caching have attracted lots of attention [26–28]. The aforementioned works typically solve the clustering and multicast beamforming optimization at the same timescale. Nevertheless, it is often the case that the clustering and caching execute at a much larger timescale, while the multicast beamforming is tailored per small timescale to exploit the fast fading of wireless channels. Although Qiao et al. [28] decoupled content placement and content delivery at different timescales in vehicular edge networks, the multicast transmission was ignored, which possibly incurs low-spectrum efficiency. That is, it is a prerequisite to propose a joint clustering, caching, and multicast beamforming algorithm in vehicular edge networks on the basis of a two-timescale setup.

3. System Model and Problem Formulation

Vehicular edge network with one BS and multiple IoV devices is considered. The BS is equipped with L antennas, and each vehicle has one single antenna. As shown in Figure 1, in the context of V2V communications, the content requesting vehicle (CRV) could not only attach to the BS via cellular links but also associate with content caching vehicles (CCVs) directly. The BS is indexed by $j = 0$, while CCVs are denoted by $\mathcal{J} = \{1, \cdots, j, \cdots, J\}$. Furthermore, $\mathcal{J}_0 = \mathcal{J} \cup \{0\}$ denotes as the set of all transmitters. Meanwhile, the CRVs with the identical content request are categorized into one multicast group, and all multicast groups are denoted by $\mathcal{S} = \{1, \cdots, s, \cdots, S\}$. For each multicast group $s \in \mathcal{S}$, its all associated CRVs are denoted by $\mathcal{I}_s$, and $\cup_{s \in \mathcal{S}} \mathcal{I}_s = \mathcal{I}$ holds, where $\mathcal{I}$ is the set of all CRVs. In addition, we follow the equal content size for all contents [12], namely, each of which is chunked and normalized to the size of one.

3.1. Two-Timescale Setup. Mixed-timescale setup in [29, 30] is incorporated in this study. Each large timescale accommodates multiple successive small timescale slots, e.g., a total of T slots, denoted by $\mathcal{T} = \{1, \cdots, T\}$. In addition, in each slot, the beamformer is designed for each multicast group, to exploit the fast fading and mobility experienced by CRVs. As revealed in Figure 2, at the end of each large timescale, the network operator has to decide the CCV clustering (i.e., the CCVs serving for the same multicast group) and the content caching for transmitters. Then, following this decision unchanged throughout the next large timescale, the picked content is cached in the local memory or storage in both BS and clustered CCVs.

In this work, we assume that all multicast groups can be admitted and satisfied with their content requests. The case also occurs that insufficient resources cannot accommodate all multicast groups (e.g., in [31]), which is yet beyond the scope of this work.

Afterward, following the two-timescale setup, two types of variables can be specified as:

(i) Large-timescale variables: $x_{s,j} \in \{0, 1\}$ and $z_{s,j} \in \{0, 1\}$, indicating whether or not that transmitter j belongs to the serving cluster of group s, and whether or not that transmitter j caches the content requested by group s, respectively. That is, $x_{s,j} \in \{0, 1\}$ indicates a clustering variable: if $x_{s,j} = 1$, then transmitter j is within the cluster of s; and $x_{s,j} = 0$ vice versa
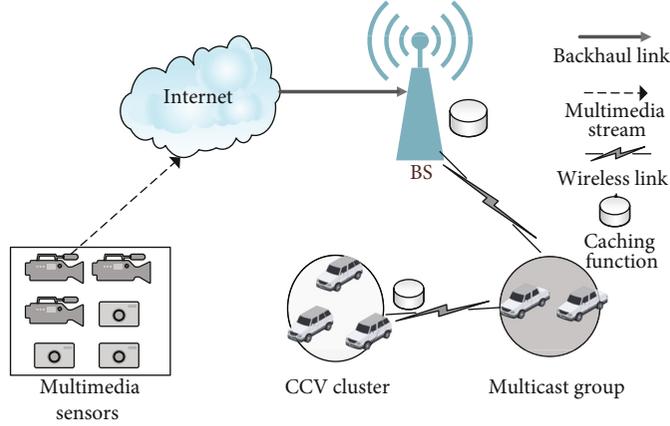
FIGURE 1: System model of vehicular edge networks.
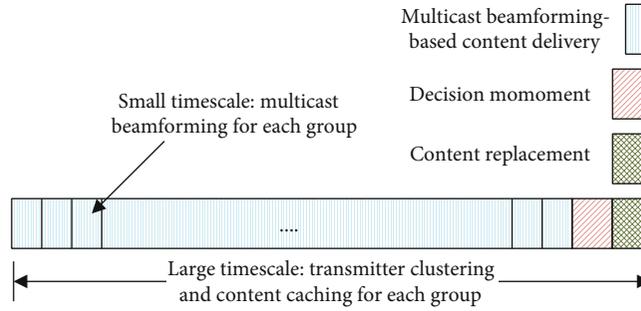
FIGURE 2: Two-timescale setup.

(ii) Small-timescale variables: $\mathbf{w}_{s,0}(t) \in \mathbb{C}^{L\times 1}$ (or $\mathbf{w}_{s,j}(t) \in \mathbb{C}, j \in \mathcal{J}$), denoting the per-slot multicast beamforming vector of the BS (or CCV j) for group s

The two-timescale setup along with its two types of variables necessitates a network cost involving both the flow and power costs, which will be introduced in the following, respectively.

3.2. Caching Model. At the end of large timescale, both the clustering variable $\{x_{s,j}\}$ and caching variable $\{z_{s,j}\}$ have to be designed for the next large timescale. It is straightforward that if the content is not cached in the network, then it must be retrieved via the backhaul link from the Internet, as shown in Figure 1, thus incurring flow cost in the backhaul link or transport networks. The data rate of fetching content from core networks should be as least as the transmission rate of its associated multicast group. Following the fixed transmission rate R_s in [12] for each multicast group s, the total flow passing through the backhaul link on the BS can be represented as $\sum_{s\in\mathcal{S}}(1-z_{s,0})R_s$. In particular, when $z_{s,0} = 1$, the immediate transmission from the BS to CRV i does not incur any flow cost on the backhaul link, and $z_{s,0} = 0$ vice versa. Besides, different from the BS, CCVs could not directly retrieve the content through the backhaul link from the Internet or core networks. Without caching, CCVs have to resort to the BS for the content retrieval, which would claim a dedicated frequency band to avert the interference to CRVs, thereby contributing to the flow cost as well. As such, the flow

cost should involve both the backhaul cost and frequency bandwidth cost, and the latter one could be equivalently translated to the former one in terms of fixed transmission rate R_s. Till now, the total flow cost (on the backhaul link and dedicated frequency band) can be represented as

$$C_F = \sum_{s\in\mathcal{S}}\sum_{j\in\mathcal{J}_0}\left(1-z_{s,j}\right)R_s. \tag{1}$$

Then, for the clustering variable $x_{s,j}$, it is irrational to cache the requested content in transmitter j when it is not picked to join the cluster of multicast group s. That is, $z_{s,j}$ could only take value zero when $x_{s,j} = 0$ holds, which can be formulated as

$$z_{s,j} \leq x_{s,j}, \forall s, j. \tag{2}$$

Next, also revealed in Equation (2), $z_{s,j}$ tends to take value one to save the flow cost, given $x_{s,j} = 1$. Yet, finite storage size in CCVs constrains the cached content number, namely, only parts of contents can be accommodated in each transmitter j, written as

$$\sum_{s\in\mathcal{S}}z_{s,j} \leq C_j, \forall j, \tag{3}$$

which indicates the caching capacity constraint C_j with a normalized content size of one.

Finally, although the caching function does also incur costs for both BS and CCVs, we only focus on the trade-off between backhaul cost and power cost, and the study on caching cost is beyond the scope of this work.

3.3. Multicast Model. The multicast beamforming vector is tailored per slot. For simplicity, let $\mathbf{w}_s(t) = [\mathbf{w}_{s,0}^H(t), \mathbf{w}_{s,1}^H(t), \cdots, \mathbf{w}_{s,J}^H(t)]^H$ be the network-wide beamforming vector for group s from all transmitters [32]. Furthermore, for each CRV $i \in \mathcal{I}_s$, $\mathbf{h}_i(t) \in \mathbb{C}^{(L+J)\times 1}$ denotes as the aggregate channel vector from all transmitters to CRV i. As such, the received signal-to-noise-ratio (SINR) for CRV i at slot t can be written as

$$\text{SINR}_i(t) = \frac{\left|\mathbf{h}_i^H(t)\mathbf{w}_s(t)\right|^2}{\sum_s{'}\neq s\left|\mathbf{h}_i^H(t)\mathbf{w}_s{'}(t)\right|^2 + \sigma_i^2}, \forall i \in \mathcal{I}_s, s, \quad (4)$$

where σ_i^2 is the additive white Gaussian noise power per CRV and $\sum_s{'}\neq s|\mathbf{h}_i^H(t)\mathbf{w}_s{'}(t)|^2$ is the intergroup interference from any other multicast group $s' \neq s$. For each group s, it is ensured that the achievable data rate of any CRV i is no smaller than its group's fixed transmission rate R_s, namely,

$$B \log_2(1 + \text{SINR}_i(t)) \geq R_s, \forall i \in \mathcal{I}_s, \quad (5)$$

which can be further recast as

$$\text{SINR}_i(t) \geq 2^{R_s/B} - 1, \forall i \in \mathcal{I}_s, s, \quad (6)$$

where B denotes the frequency bandwidth.

Furthermore, considering any pair of transmitter j and multicast group s, if j does not belong to the serving clustering of s, then it is irrational to design a nonzero beamforming vector for s. That is to say, if $x_{s,j} = 0$ holds, then we have $\mathbf{w}_{s,0}(t) = \mathbf{0}$ (or $\mathbf{w}_{s,j}(t) = 0$) at any time slot t, which can be represented as

$$(1 - x_{s,0})\mathbf{w}_{s,0}(t) = 0, \forall s, t$$
$$(1 - x_{s,j})\mathbf{w}_{s,j}(t) = 0, \forall s, j \in \mathcal{J}, t. \quad (7)$$

Moreover, since each transmitter is with the maximum transmit power threshold, the peak power budget E_j per transmitter should be imposed as

$$\sum_{s \in \mathcal{S}}\left|\left|\mathbf{w}_{s,0}(t)\right|\right|_2^2 \leq E_0, \forall t,$$
$$\sum_{s \in \mathcal{S}}\left|\left|\mathbf{w}_{s,j}(t)\right|\right|_2^2 \leq E_j, \forall j \in \mathcal{J}, t. \quad (8)$$

In addition, summing over the transmit power of all network-wide beamforming vectors, the total power cost can be computed as

$$C_P = \frac{1}{T}\sum_{t \in \mathcal{T}}\sum_{s \in \mathcal{S}}||\mathbf{w}_s(t)||_2^2, \quad (9)$$

which is averaged over all time slots in one large timescale. In particular, C_P is exactly defined as the time average of small-timescale transmit power cost $\sum_{s \in \mathcal{S}}||\mathbf{w}_s(t)||_2^2$.

3.4. Problem Formulation. Finally, involving both the flow cost and power cost, the overall cost minimization problem can be formulated as

$$\begin{aligned}
\mathscr{P}_0 : \min \quad & C_F + \eta C_P \\
\text{s.t.} \quad & (2), (3), (6), (7), (8) \\
& \text{var } x_{s,j}, z_{s,j} \in \{0, 1\}, \forall s, j, \\
& \text{var } \mathbf{w}_s(t), \forall s, t,
\end{aligned} \quad (10)$$

where η is a coefficient to balance the trade-off between flow and power costs and could be regulated artificially in line with the price of backhaul link and transmit power.

4. Joint Clustering, Caching, and Multicast Beamforming Algorithm

In this section, we first present the challenges to solve problem $\mathscr{P}_0$ and then leverage the SAA technique to decouple the two-timescale problem into one large-timescale problem and a series of independent small-timescale subproblems.

4.1. Algorithm Design Challenges. We desire to solve $\mathscr{P}_0$ at the end of each large timescale for the next one. Nevertheless, it poses challenges to solve $\mathscr{P}_0$ due to the following observations:

  (i) The channel vectors $\{\mathbf{h}_i(t)\}_{\forall t \in \mathcal{T}}$ are unknown for the operator, since all channel vectors are future ones in the next large timescale

  (ii) Both $\mathbf{x} = \{x_{s,j}\}$ and $\mathbf{z} = \{z_{s,j}\}$ are binary variables, rendering $\mathscr{P}_0$ an MINLP problem

  (iii) Even though $\mathbf{x}$ and $\mathbf{z}$ are relaxed, both Equations (6) and (7) are still nonconvex constraints

Thus, to make $\mathscr{P}_0$ tractable, an approximation approach is a prerequisite to solve it.

4.2. SAA-Based Cost Minimization. The channel vectors $\{\mathbf{h}_i(t)\}_{\forall t \in \mathcal{T}}$ are exactly stochastic at each decision moment, since they are scattered in the next large timescale and thus unpredictable. As in [29, 30], the SAA technique is leveraged to approximate the random variables $\{\mathbf{h}_i(t)\}_{\forall t \in \mathcal{T}}$, with the basic principle assuming that $\{\mathbf{h}_i(t)\}_{\forall t \in \mathcal{T}}$ is drawn from a certain distribution. As such, Equation (9) can be deemed as the time average of T random variables, and its expectation can be approximately computed as

$$E_{\widehat{\mathbf{h}}}\left(\sum_{s \in \mathcal{S}}||\widehat{\mathbf{w}}_s||_2^2\right) \approx \frac{1}{T}\sum_{t \in \mathcal{T}}\sum_{s \in \mathcal{S}}||\mathbf{w}_s(t)||_2^2, \quad (11)$$

where $\widehat{\mathbf{h}}$ is the stochastic channel vector space, and $\widehat{\mathbf{w}}_s$ denotes the $\widehat{\mathbf{h}}$-based beamforming vector. Then, by leverage the SAA, a series of samples are generated, and the expectation (11) can be approximated by its sample average. To distinguish the sample from real channel vectors per time slot, v is utilized to denote the sample index, and a total of samples are produced. Therefore, by substituting t with v at the decision moment of each large timescale, an approximate problem of $\mathscr{P}_0$ can be recast as

$$\mathscr{P}_1 : \min \ C_F + \frac{\eta}{V} \sum_{v \in \mathscr{V}} \sum_{s \in \mathscr{S}} ||\mathbf{w}_s(v)||_2^2$$

$$\text{s.t. } (2), (3), (6), (7), (8)$$

$$\text{var } x_{s,j}, z_{s,j} \in \{0, 1\}, \forall s, j, \tag{12}$$

$$\text{var } \mathbf{w}_s(v), \forall s, v,$$

where $\mathbf{h}_i(v)$ is the v-th sample from a certain distribution, and $\mathbf{w}_s(v)$ is the $\mathbf{h}_i(v)$-based beamforming vector.

From Equation (12), it follows that $\mathscr{P}_1$ turns out to be a deterministic problem rather than a stochastic one. Nevertheless, solving $\mathscr{P}_1$ still poses challenges, since both Equations (2) and (6) are nonconvex constraints. Furthermore, small-timescale variable $\mathbf{w}_s(v)$ and large-timescale variable $x_{s,j}$ are still tied in Equation (7). For this problem with two types of coupled variables, an intuitive approach is to decouple them firstly and then optimize them separately. Thus, we would propose an iterative algorithm, with the outline as follows:

(i) First, given any feasible $\mathbf{w}$, search for the optimal $\{\mathbf{x}^*, \mathbf{z}^*\}$ in $\mathscr{P}_1$

(ii) Then, given $\{\mathbf{x}^*, \mathbf{z}^*\}$, search for the optimal $\mathbf{w}^*$ with the sample $\{\mathbf{h}_i(v)\}_{\forall i, v}$

(iii) Repeat aforementioned procedures until convergence

(iv) With the acquired $\{\mathbf{x}^*, \mathbf{z}^*\}$, the optimal $\{\mathbf{w}_s(t)\}_{\forall s, t}$ for problem $\mathscr{P}_0$ can be obtained by solving the following multicast beamforming problem per slot as follows:

$$\mathscr{P}_2 : \ \min \ \frac{1}{T} \sum_{t \in \mathscr{T}} \sum_{s \in \mathscr{S}} ||\mathbf{w}_s(t)||_2^2$$

$$\text{s.t. } (6), (7), (8) \tag{13}$$

$$\text{var } \mathbf{w}_s(t), \forall s, t.$$

By comparing $\mathscr{P}_1$ and $\mathscr{P}_2$, it reaches the conclusion that $\mathscr{P}_1$ is a large-timescale problem with sample channel vectors, while $\mathscr{P}_2$ reduces to a small-timescale one with actual channel vectors. In the next, we would separately introduce approaches to solve them.

4.3. Joint Clustering and Caching Algorithm. Given any feasible $\mathbf{w}$, any group-transmitter pair with nonzero beamforming vector can be determined. Define the set $\mathscr{K}_1 = \{(s, j) \mid \mathbf{w}_{s,0}(v) \neq \mathbf{0} \text{ or } \mathbf{w}_{s,j}(v) \neq 0, \exists v \in \mathscr{V}\}$. Following Equation (7), if $\mathbf{w}_{s,0}(v) \neq \mathbf{0}$ or $\mathbf{w}_{s,j}(v) \neq 0, \exists v \in \mathscr{V}$, then $x_{s,j} = 1$ must hold. On the contrary, although $\mathbf{w}_{s,0}(v) = \mathbf{0}$ or $\mathbf{w}_{s,j}(v) = 0, \forall v$ cannot straightforwardly reach the conclusion with $x_{s,j} = 0$ from Equation (7), we can claim that $x_{s,j} = 0$ holds, namely, j does not need to belong to the serving cluster of s when $\mathbf{w}_{s,0}(v) = \mathbf{0}$ or $\mathbf{w}_{s,j}(v) = 0, \forall v$. Likewise, define the set $\mathscr{K}_2 = \{(s, j) \mid \mathbf{w}_{s,0}(v) = \mathbf{0} \text{ or } \mathbf{w}_{s,j}(v) = 0, \forall v \in \mathscr{V}\}$.

As such, the clustering variable $\{x_{s,j}\}$ is fixed given any feasible $\mathbf{w}$, and $\mathscr{P}_1$ reduces to the following large-timescale problem as

$$P_{1-I} : \min \ C_F$$

$$\text{s.t.} (2), (3),$$

$$x_{s,j} = 1, \forall (s, j) \in K_1, \tag{14}$$

$$x_{s,j} = 1, \forall (s, j) \in K_2,$$

$$\text{var } x_{s,j}, z_{s,j} \in \{0, 1\} \forall s, j,$$

which exactly involves only the caching variable $\{z_{s,j}\}$, since $\{x_{s,j}\}$ is determined.

It follows that $\mathscr{P}_{1-I}$ is a typical 0-1 ILP problem. On one hand, the celebrated cutting-plane or branch-and-bound methods have been extensively utilized to solve it, yet with the computational complexity scaled with S and J. Thus, solving it is prohibitively complicated in a larger-sized network. On the other hand, there are also extensive works (e.g., in [2]) to leverage heuristic algorithms to solve the ILP problem. Nevertheless, the optimality cannot be ensured in this case, and it poses challenges to analyse the gap between the heuristic and optimality as well.

Motivated by our previous works in [33, 34], we resort to the LP counterpart of $\mathscr{P}_{1-I}$. If the equivalence of optimal solutions between $\mathscr{P}_{1-I}$ and its LP counterpart can be established, then we claim that the optimal solution to LP counterpart is also integer-valued and optimal to $\mathscr{P}_{1-I}$ as well.

First, with relaxed $z_{s,j}$, $\mathscr{P}_{1-I}$ can be recast as

$$P_{1-R} : \min \ C_F$$

$$\text{s.t.} (2), (3),$$

$$x_{s,j} = 1, \forall (s, j) \in K_1, \tag{15}$$

$$x_{s,j} = 0, \forall (s, j) \in K_2,$$

$$\text{var } 0 \leq z_{s,j} \leq 1, \forall s, j,$$

where the clustering variable $\{x_{s,j}\}$ is eliminated on the basis of $\mathscr{K}_1$ and $\mathscr{K}_2$ for conciseness.

It is widely known that the optimal solution of LP must locate on the vertex of the polyhedron for a feasible set. Thus, what we only need to do next is to establish that any vertex of the polyhedron of $\mathscr{P}_{1-R}$ is integer-valued. As such, the

optimal solution to $\mathcal{P}_{1-R}$ must be integer-valued and is also optimal to $\mathcal{P}_{1-I}$. Henceforth, we further provide a sufficient condition, under which the relaxed $\mathcal{P}_{1-R}$ has the integer-valued optimal solution.

Theorem 1. *If each transmitter's storage capacity is the integer multiplier of content size, i.e., $C_j \in \mathbb{Z}, \forall j \in \mathcal{J}_0$, then the optimal solution to $\mathcal{P}_{1-R}$ is integer-valued.*

Proof. From [35], it follows that if the constraint matrix is totally unimodular, then the optimal solution is integer-valued and locates at the vertex of the polyhedron. Therefore, we prove this theorem by establishing the total unimodularity of the constraint matrix of $\mathcal{P}_{1-R}$. By introducing a caching variable vector $\mathbf{Y} = \{y_{1,0}, \cdots, y_{S,0}, y_{1,1}, \cdots, y_{S,1}, \cdots, \cdots, y_{1,J}, \cdots, y_{S,J}\}^T$ and a storage capacity vector $\mathbf{C} = \{C_0, C_1, \cdots, C_j\}^T$, Equation (3) can be recast as $\mathbf{WY} \leq \mathbf{C}$, and $\mathbf{W}$ turns out to be as

$$
\begin{aligned}
\mathbf{W} &= \begin{bmatrix} \mathbf{W}_0 \ \mathbf{W}_1 \ \mathbf{W}_2 \ \cdots \ \mathbf{W}_j \end{bmatrix} \\
&= \begin{pmatrix}
1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 & \cdots & 0 & 0 & \cdots & 0 \\
* & * & \vdots & * & * & * & \vdots & * & \ddots & * & * & \vdots & * \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & 1 & \cdots & 1
\end{pmatrix}
\end{aligned}
\tag{16}
$$

where each $\mathbf{W}_j$ is with the dimension of $(J+1) \times S$. In particular, the $(j+1)$-th row of $\mathbf{W}_j$ is all one and all zero for the other rows. Since each column of $\mathbf{W}_j$ contains only one nonzero entry belonging to $\{0,+1,-1\}$, we claim that the constraint matrix $\mathbf{W}$ meets the four subconditions of total unimodularity in [35]. As such, $\mathbf{W}$ is totally unimodular, and thus the optimal solution of $\mathcal{P}_{1-R}$ is integer-valued.

The proof is completed.

With Theorem 1, we can effortlessly solve $\mathcal{P}_{1-I}$ via classical simplex method or ellipsoid algorithm, due to the equivalence of optimal solutions between $\mathcal{P}_{1-R}$ and $\mathcal{P}_{1-I}$, thus significantly reducing the computational complexity. Although the classical cutting-plane or branch-and-bound algorithm can also work for the ILP, they are only applicable to small- and moderate-sized vehicular edge networks. When it comes to a large-sized network, Theorem 1 exhibits more effectiveness than legacy ones.

4.4. SCA-Based Multicast Beamforming. Once the aforementioned $\{\mathbf{x}^*, \mathbf{z}^*\}$ is given, $\mathcal{P}_1$ reduces to the following one as

$$
\begin{aligned}
\mathcal{P}_{1-S} : \ & \min \ \frac{1}{V} \sum_{v \in \mathcal{V}} \sum_{s \in \mathcal{S}} \|\mathbf{w}_s(v)\|_2^2 \\
& \text{s.t. } (6), (7), (8) \\
& \text{var } \mathbf{w}_s(v), \forall s, v,
\end{aligned}
\tag{17}
$$

which is exactly the same as $\mathcal{P}_2$ with the only difference in substituting the index t in $\mathcal{P}_2$ with v in $\mathcal{P}_{1-S}$. Notice that, $\mathcal{P}_2$ is the small-timescale multicast beamforming problem with actual channel vectors $\{\mathbf{h}_i(t)\}_{\forall t \in \mathcal{T}}$, while $\mathcal{P}_{1-S}$ is solved at the large timescale with channel samples $\{\mathbf{h}_i(v)\}_{\forall v \in \mathcal{V}}$, just to proceed with the algorithm iteration in solving $\mathcal{P}_{1-S}$. Thus, in the following, by putting $\mathcal{P}_2$ and $\mathcal{P}_{1-S}$ together, we take $\mathcal{P}_2$ as a goal to find its approximation algorithm, and do not distinguish them strictly.

Although the objective as well as constraints 7 and 8 turn into convex ones, the nonconvex constraint 6 renders $\mathcal{P}_2$ a nonconvex one. Thus, by resorting to the SCA technique in [12], we first recast Equation (6) as

$$
\begin{aligned}
& \left(2^{R_s/B} - 1\right) \left(\sum_{s' \neq s} \left|\mathbf{h}_i^H(t)\mathbf{w}_{s}{'}(t)\right|^2 + \sigma_i^2 \right) - \left|\mathbf{h}_i^H(t)\mathbf{w}_s(t)\right|^2 \\
& \leq 0, \forall i \in \mathcal{I}_s, s, t.
\end{aligned}
\tag{18}
$$

It is evident that the left-hand side of Equation (18) is the difference of two convex functions. Next, by substituting the second term with its first-order Taylor expansion, at the $(o+1)$-th iteration, we have

$$
\begin{aligned}
& 2Rs/B - 1 \sum_{s' \neq s} \left|\mathbf{h}_i^H(t) \mathbf{w}_{s}{'}(t)\right|^2 + \sigma_i^2 \\
& - 2\mathcal{R}\left\{ (\mathbf{w}_s^o(t))^H \mathbf{h}_i(t) \mathbf{h}_i^H(t) \mathbf{w}_s(t) \right\} - \left|\mathbf{h}_i^H(t) \mathbf{w}_s^o(t)\right|^2 \\
& \leq 0, \forall i \in \mathcal{I}s, s, t,
\end{aligned}
\tag{19}
$$

which becomes a convex constraint, with $\mathbf{w}_s^o(t)$ acting as the initial feasible solution obtained from the o-th iteration.

Till now, at the $(o+1)$-th iteration, $\mathcal{P}_2$ can be transformed to the following one as

$$
\begin{aligned}
\mathcal{P}_{2-C} : \ & \min \ \frac{1}{T} \sum_{t \in \mathcal{T}} \sum_{s \in \mathcal{S}} \|\mathbf{w}_s(t)\|_2^2 \\
& \text{s.t. } (7), (8), (19) \\
& \text{ar } \mathbf{w}_s(t), \forall s, t,
\end{aligned}
\tag{20}
$$

which is a convex quadratically constrained quadratic programming (QCQP) problem and can be solved via many mature convex algorithms effortlessly, e.g., interior-point method [36]. At each iteration, a convex QCQP problem needs to be solved until convergence. From Equation (19), it follows that the obtained $\mathbf{w}_s^o(t)$ from the o-th iteration is always feasible to the subproblem at the $(o+1)$-th iteration. Besides, since the objective of $\mathcal{P}_{2-C}$ is the minimization of power cost, $\mathbf{w}_s^{o+1}(t)$ must be a better solution than $\mathbf{w}_s^o(t)$, revealing the monotonicity of objective.

4.5. Algorithm Outline and Computational Complexity. At the end of a large timescale, the operator has to iteratively resolve problems $\mathcal{P}_{1-S}$ and $\mathcal{P}_{1-I}$, until the termination

---

Initialization: $x_{s,j} = 1, \forall s \in \mathcal{S}, \forall j \in \mathcal{J}_0$ (Full cooperative multicasting is as the initial solution).
Repeat
      Solve problem $\mathcal{P}_{1-S}$ via SCA to acquire $\mathbf{w}^*$;
      Obtain $\mathcal{K}_1$ and $\mathcal{K}_2$ from Equation (14);
      Solve problem $\mathcal{P}_{1-I}$ via Theorem 1 to acquire $\{\mathbf{x}^*, \mathbf{z}^*\}$.
Until the error tolerance is met or maximum iterations are reached.
for $t = 1, 2 \cdots, T$
      Solve problem $\mathcal{P}_2$ via SCA with the available $\{\mathbf{x}^*, \mathbf{z}^*\}$.
end for

ALGORITHM 1: Joint V2V-assisted clustering, caching, and multicasting algorithm.

criteria are satisfied, namely, one convergent solution is acquired or the maximum iteration number is reached, as described in Algorithm 1. Then, with the available $\{\mathbf{x}^*, \mathbf{z}^*\}$ from large timescale, the operator proceeds to tailor the multicast beamforming per slot, to solve $\mathcal{P}_2$ via SCA for the optimal $\mathbf{w}_s^o(t)$.

From Algorithm 1, it follows that the SCA is a prerequisite at both timescales. Assume the maximum iteration numbers for joint V2V clustering and caching algorithm, and SCA method reach ϱ_{max} and o_{max}, respectively. From [37], it follows that the interior point method to solve each per-iteration subproblem for $\mathcal{P}_{1-S}$ or $\mathcal{P}_2$ is with the computational complexity of $\Phi = \mathcal{O}((J + L)^3 S^3 V^3)$ or $\Phi = \mathcal{O}((J + L)^3 S^3 T^3)$. Thus, the overall computational complexity reaches $\mathcal{O}((\varrho_{max} + 1)o_{max}\Phi) = \mathcal{O}(\varrho_{max} o_{max} \Phi)$.

5. Simulation Results and Discussions

In this section, we demonstrate the performance of proposed joint V2V-assisted clustering, caching, and multicast beamforming algorithm in vehicular edge networks via simulation results. In particular, the impact of the following two parameters is studied: (1) the number of CCVs; and (2) the average data rate requirement per CRV. Meanwhile, two costs are leveraged as performance metrics: (1) total flow cost; and (2) total transmit power cost. In addition, for performance comparison, three benchmark schemes are also evaluated, listed as follows:

(i) Multicast without (w.o.) V2V caching refers to the multicast beamforming scheme proposed in [11], which only involves the BS caching but neglecting the V2V caching. For fair comparison, the two-timescale setup is also used in this scheme

(ii) Unicast with (w.) V2V caching refers to the unicast transmission scheme proposed in [23], where both V2V and BS caching are involved. In this scheme, each CRV is assigned an individual beamforming vector, regardless of its requested content

(iii) Unicast w.o. V2V caching. In this scheme, each CRV has to access the BS to fetch its requested content

5.1. Simulation Parameters.
We investigate a time-slotted wireless network consisting of one BS and 20 CRVs, with a radius of 500 m for the cell coverage. The BS is equipped with 20 antennas, and each CRV and CCV is identically equipped with one antenna. Overall, 20 types of multimedia contents exist in the system. The BS has a storage size of 6 contents, while each CCV could accommodate at most 1 content. For simplicity, we do not employ the well-known Zipf distribution in [2]; instead, each CRV equally requests any content with a probability of 5%.

The system bandwidth is 5 MHz, and the additive white Gaussian noise power spectral is -174 dBm/Hz. The transmit power budgets for the BS and CCV are 46 dBm and 24 dBm, respectively. The path loss model is $35.3 + 37.6 \log_2(d(m))$, the log-normal shadowing parameter is 8 dB, and the multipath channel model with Rayleigh fading is assumed [20]. Overall, 200 samples are produced to simulate stochastic channels vectors. Besides, the Monte Carlo approach is utilized, both CRVs and CCVs are uniformly distributed and dropped in the cell coverage, and all simulation results are averaged over 1000 droppings. In addition, the vehicle speed follows a truncated normal distribution ranging from 15 m/s and 31 m/s.

5.2. Convergence Performance.
Figure 3 illustrates the convergence performance of joint V2V clustering and caching algorithms. In this setup, the number of CCVs is 20, and the data rate requirement per group is 2 Mbps. To overcome the impact of η with too large or small values on the objective of $\mathcal{P}_0$ and for fair comparison, only the system power cost is evaluated. First, it can be observed from Figure 3 that the joint V2V clustering and caching algorithm tends to saturate within 30 iterations. In spite of moderate convergence, the result is acceptable since it is operated at a large timescale and does not need to work per small timescale slot. Second, with the increase of η, the total power cost gradually reduces. This is because the objective of $\mathcal{P}_0$ is the minimization of system cost; a larger η will increase the weight of power cost, thus reducing the power proportion in the optimal value.

Figure 4 reveals the convergence of the SCA-based multicast beamforming algorithm. Likewise, the number of CCVs is 20, and the data rate requirement per group is 2 Mbps. We leverage the L_2 norm on the difference of successive optimal solutions, i.e., $||\mathbf{w}_s^{o+1}(t) - \mathbf{w}_s^o(t)||_2$ as the metric. As shown in Figure 4, all settings under different values of η have good convergence, saturating within 10 steps.

5.3. Performance Comparison.
Figure 5 shows the power-flow cost tradeoff curve of the proposed algorithm under different
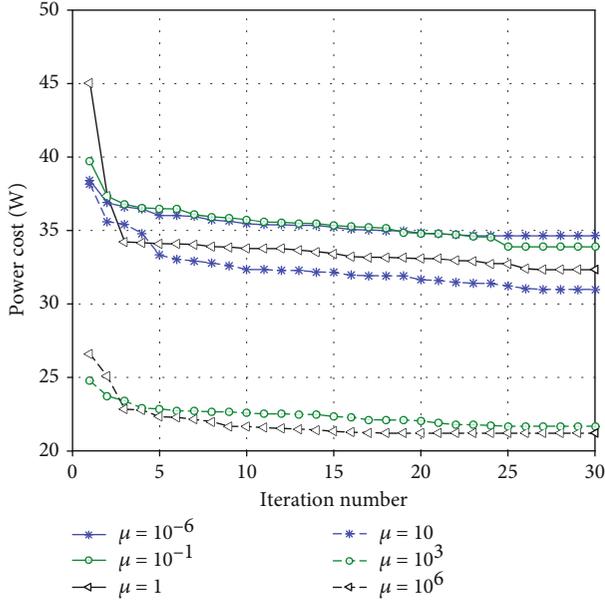
Figure 3: The convergence of the proposed algorithm.
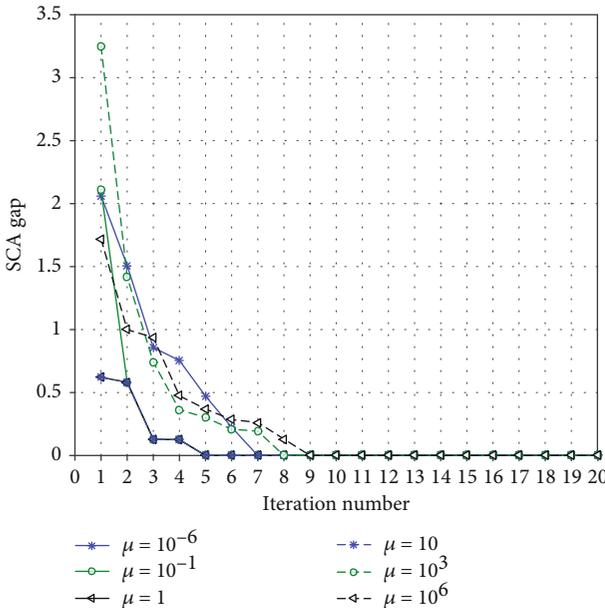

Figure 5: Flow-power cost tradeoff.

Figure 4: The convergence of the SCA method.

values of η. Similar to the result in Figure 3, when η approaches infinity, the proportion of power cost is getting larger, and thus the power cost decreases in the minimization of the weighted sum. In particular, the flow cost reduces to 21 W while the flow cost reaches 34 Mbps, given $\eta = 10^6$. On the contrary, when η approaches zeros, the proportion of power cost gets smaller, and more weights are imposed on the flow cost. In particular, the flow cost increases to 35 W while the flow cost reduces to 18 Mbps, given $\eta = 10^{-6}$.

Figure 6 shows the impact of CCV number on the flow cost. In this scenario, the average data rate requirement is 2 Mbps, $\eta = 10^3$, and the CRV number ranges from 15 to 30. As shown in Figure 6, the curve is broken line-shaped,

since the flow cost must be the integer multiplier of the data rate requirement. The flow cost reduces with the increase of CCV number. This is because the wireless network incorporating more CCVs would introduce multiuser diversity gain and a more flexible clustering combination. In the unicast w.o. V2V caching scheme, the caching functions are only available in the BS, and each CCV provided a requested content individually, thus resulting in the largest flow cost. In addition, unicast w. V2V caching outperforms multicast w.o. V2V caching in terms of flow cost, indicating that the V2V caching gains overwhelm the multicasting gains in this setting, especially when the CCV number is relatively large.

Figure 7 shows the impact of CCV number on the power cost, with identical settings as in Figure 6. Likewise, unicast w.o. V2V caching still gets the largest power cost compared to the other three schemes. Meanwhile, unicast w. V2V caching is also with worse performance compared to multicast w.o. V2V caching. This is because multicast w.o. V2V caching tends to save more transmit power than unicast w. V2V caching, while the latter one must allocate an individual beamformers to each CCV. The performance gap also indicates the multicast gains over unicast. Furthermore, the proposed algorithm gets close performance with multicast w.o. V2V caching, only with a slightly small gap, since in this setting, multicast gains overwhelm V2V caching gains.

In Figure 8, we compare the performance with different data rate requirements R_s, ranging from 2 Mbps to 3.8 Mbps. In this scenario, there are 20 CCVs, and $\eta = 10^3$. Figure 8 shows the flow cost is almost proportional to the data rate requirement. In the meantime, there exists a significant gap between the proposed algorithm and the other three schemes (approximately 30 Mbps on average compared to multicast w.o. V2V caching), which can be explained as follows: on one hand, the unicast transmission would result in more flow costs, since each CRV is equipped with a CCV cluster and a content flow; on the other hand, the multicast w.o. V2V
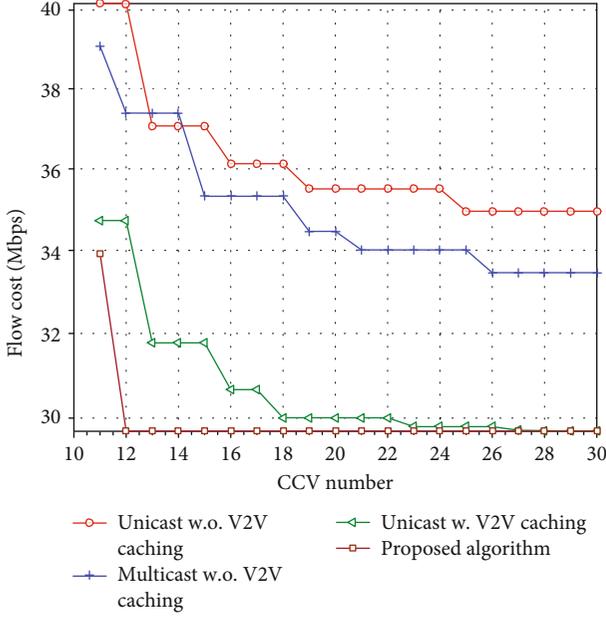
FIGURE 6: Flow cost versus the CCV number.
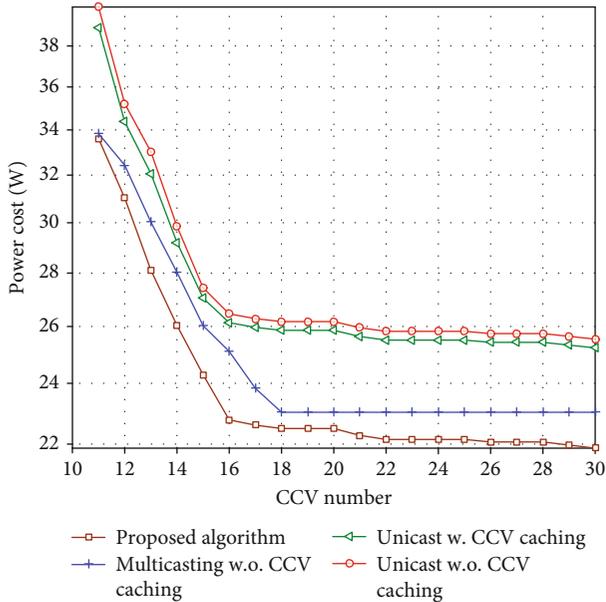
FIGURE 7: Power cost versus the CCV number.

FIGURE 8: Flow cost versus the average data rate requirement.
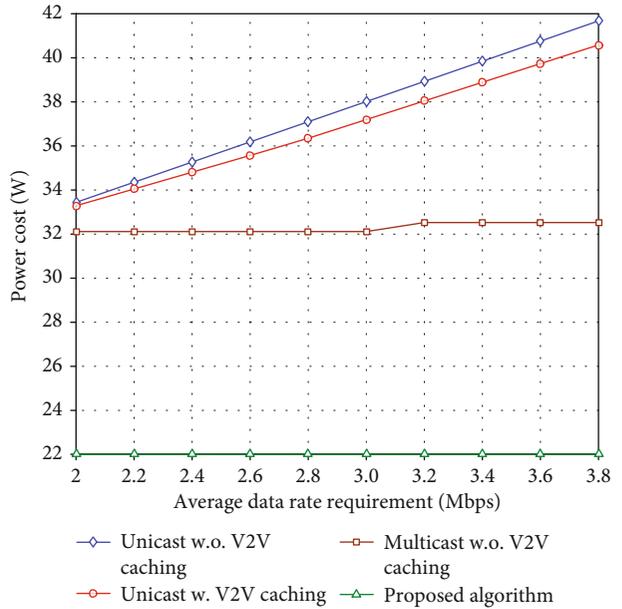
FIGURE 9: Power cost versus the average data rate requirement.

caching neglects the caching functions in CCVs, thus incurring the traffic congestion.

Figure 9 shows the impact of data rate requirement on power cost, with identical settings as in Figure 8. Different from the results in Figure 8, both the proposed algorithm and multicast w.o. V2V almost remain unchanged with the increase of data rate requirements. This is because in this setting, sufficient power is available to support a larger data rate, and thus multicast transmissions could save more power than unicast. Meanwhile, the performance gap between the proposed algorithm and multicast w.o. V2V caching reveals that, the lack of V2V caching would incur that all CRVs tend

to access the BS, the power resource of CCVs is wasted, and thus a worse feasible solution is produced. In addition, unicast would allocate an individual beamformer to each CRV, thus getting the worst performance.

6. Conclusions

In this study, wireless content caching along with multicast beamforming was studied in vehicular edge networks. First, a two-timescale setup was proposed to decouple the joint clustering, caching, and multicast beamforming problem at separate timescales. Next, at the large timescale, with the

SAA technique, a theorem was verified to reveal the equivalence of optimal solutions between the original ILP problem and its relaxed LP counterpart. Then, with the SCA method, the multicast beamforming-based power minimization problem was solved per slot. Simulation results revealed that the network performance could benefit from not only the power saving from wireless multicast beamforming but also the content caching and sharing among vehicles.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] S. A. Alvi, B. Afzal, G. A. Shah, L. Atzori, and W. Mahmood, "Internet of multimedia things: vision and challenges," *Ad Hoc Networks*, vol. 33, pp. 87–111, 2015.

[2] C. Liang, Y. He, F. R. Yu, and N. Zhao, "Enhancing video rate adaptation with mobile edge computing and caching in software-defined mobile networks,," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 7013–7026, 2018.

[3] Y. Mei, F. Li, L. He, and L. Wang, "Joint source and channel rate allocation over noisy channels in a vehicle tracking multimedia internet of things system," *Sensors*, vol. 18, no. 9, article 2858, 2018.

[4] J. Zheng and Q. Wu, "Performance modelling and analysis of the IEEE 802.11p EDCA mechanism for VANET," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2673–2687, 2016.

[5] L. Zhu, Y. Li, F. R. Yu, B. Ning, T. Tang, and X. Wang, "Cross-layer defense methods for jamming-resistant CBTC systems," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2020.

[6] L. Zhu, Y. He, F. R. Yu, B. Ning, T. Tang, and N. Zhao, "Communication-based train control system performance optimization using deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10705–10717, 2017.

[7] Z. Su, Y. Hui, Q. Xu, T. Yang, J. Liu, and Y. Jia, "An edge caching scheme to distribute content in vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 5346–5356, 2018.

[8] M. Fallgren, T. Abbas, S. Allio et al., "Multicast and broadcast enablers for high-performing cellular V2X systems," *IEEE Transactions on Broadcasting*, vol. 65, no. 2, pp. 454–463, 2019.

[9] L. Zhu, F. R. Yu, B. Ning, and T. Tang, "Cross-layer handoff design in MIMO-enabled WLANs for communication-based train control (CBTC) systems," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 719–728, 2012.

[10] G. Xylomenos, C. N. Ververidis, V. A. Siris et al., "A survey of information-centric networking research," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 1024–1049, 2014.

[11] X. Duan, Y. Liu, and X. Wang, "SDN enabled 5G-VANET: adaptive vehicle clustering and beamformed transmission for aggregated traffic," *IEEE Communications Magazine*, vol. 55, no. 7, pp. 120–127, 2017.

[12] M. Tao, E. Chen, H. Zhou, and W. Yu, "Content-centric sparse multicast Beamforming for cache-enabled cloud RAN," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6118–6131, 2016.

[13] Y. Cui, Z. Wang, Y. Yang, F. Yang, L. Ding, and L. Qian, "Joint and competitive caching designs in large-scale multi-tier wireless multicasting networks," *IEEE Transactions on Communications*, vol. 66, no. 7, pp. 3108–3121, 2018.

[14] H. T. Nguyen, H. D. Tuan, T. Q. Duong, H. V. Poor, and W. J. Hwang, "Collaborative multicast beamforming for content delivery by cache-enabled ultra dense networks," *IEEE Transactions on Communications*, vol. 67, no. 5, pp. 3396–3406, 2019.

[15] Y. Zhou, F. R. Yu, J. Chen, and Y. Kuo, "Cache-aware multicast beamforming design for multicell multigroup multicast," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 11681–11693, 2018.

[16] Y. Dong, M. J. Hossain, J. Cheng, and V. C. M. Leung, "Cross-layer scheduling and beamforming in smart-grid powered cellular networks with heterogeneous energy coordination," *IEEE Transactions on Communications*, vol. 68, no. 5, pp. 2711–2725, 2020.

[17] Y. Dong, A. el Shafie, M. J. Hossain, J. Cheng, N. al-Dhahir, and V. C. M. Leung, "Secure beamforming in full-duplex MISO-SWIPT systems with multiple eavesdroppers," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6559–6574, 2018.

[18] D. Malak, M. Al-Shalash, and J. G. Andrews, "Optimizing content caching to maximize the density of successful receptions in device-to-device networking," *IEEE Transactions on Communications*, vol. 64, no. 10, pp. 4365–4380, 2016.

[19] W. Jiang, G. Feng, and S. Qin, "Optimal cooperative content caching and delivery policy for heterogeneous cellular networks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 5, pp. 1382–1393, 2017.

[20] A. S. Cacciapuoti, M. Caleffi, M. Ji, J. Llorca, and A. M. Tulino, "Speeding up future video distribution via channel-aware caching-aided coded multicast,," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 8, pp. 2207–2218, 2016.

[21] T. Zhang, H. Fan, J. Loo, and D. Liu, "User preference aware caching deployment for device-to-device caching networks," *IEEE Systems Journal*, vol. 13, no. 1, pp. 226–237, 2019.

[22] W. L. Tan, W. C. Lau, O. Yue, and T. H. 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, 2011.

[23] J. Zhao, X. Sun, Q. Li, and X. Ma, "Edge caching and computation management for real-time internet of vehicles: an online and distributed approach," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–15, 2020.

[24] H. Xiao, X. Zhang, A. T. Chronopoulos, Z. Zhang, H. Liu, and S. Ouyang, "Resource management for multi-user-centric V2X communication in dynamic Virtual-Cell-Based ultra-dense networks," *IEEE Transactions on Communications*, vol. 68, no. 10, pp. 6346–6358, 2020.

[25] Q. Wu, H. Liu, R. Wang, P. Fan, Q. Fan, and Z. Li, "Delay-sensitive task offloading in the 802.11p-based vehicular fog computing systems," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 773–785, 2020.

[26] Y. Dong, M. Z. Hassan, J. Cheng, M. J. Hossain, and V. C. M. Leung, "An edge computing empowered radio access network with UAV-mounted FSO fronthaul and backhaul: key challenges and approaches," *IEEE Wireless Communications*, vol. 25, no. 3, pp. 154–160, 2018.

[27] H. Zhang, Y. Dong, J. Cheng, M. J. Hossain, and V. C. M. Leung, "Fronthauling for 5G LTE-U ultra dense cloud small cell networks," *IEEE Wireless Communications*, vol. 23, no. 6, pp. 48–53, 2016.

[28] G. Qiao, S. Leng, S. Maharjan, Y. Zhang, and N. Ansari, "Deep reinforcement learning for cooperative content caching in vehicular edge computing and networks," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 247–257, 2020.

[29] J. Tang, B. Shim, and T. Q. S. Quek, "Service multiplexing and revenue maximization in sliced C-RAN incorporated with URLLC and multicast eMBB," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 4, pp. 881–895, 2019.

[30] J. Tang, T. Q. S. Quek, T. Chang, and B. Shim, "Systematic resource allocation in cloud RAN with caching as a service under two timescales," *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 7755–7770, 2019.

[31] Y. L. Lee, J. Loo, T. C. Chuah, and L. C. Wang, "Dynamic network slicing for multitenant heterogeneous cloud radio access networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2146–2161, 2018.

[32] S. Gong, S. X. Wu, A. M. So, and X. Huang, "Distributionally robust collaborative beamforming in D2D relay networks with interference constraints," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5048–5060, 2017.

[33] K. Wang, F. R. Yu, L. Wang et al., "Interference alignment with adaptive power allocation in full-duplex-enabled small cell networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 3010–3015, 2019.

[34] K. Wang, W. Ji, J. Li, H. Wang, and T. Cao, "Wireless content caching in sliced cellular networks with multicast beamforming," in *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–6, Xi'an, China, 2019.

[35] L. Pu, L. Jiao, X. Chen, L. Wang, Q. Xie, and J. Xu, "Online resource allocation, content placement and request routing for cost-efficient edge caching in cloud radio access networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 8, pp. 1751–1767, 2018.

[36] S. Boyd and L. Vandenberghe, "Interior-point methods," in *Convex Optimization*, pp. 561–623, Cambridge University Press, Cambridge, UK, 2004.

[37] M. El-Absi, M. Shaat, F. Bader, and T. Kaiser, "Interference alignment with frequency-clustering for efficient resource allocation in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 7070–7082, 2015.

WILEY | Hindawi

Research Article

Machine Learning-Based Resource Allocation Strategy for Network Slicing in Vehicular Networks

Yaping Cui [1,2,3,4] **Xinyun Huang,**[1,3,4] **Dapeng Wu,**[1,3,4] **and Hao Zheng**[1,3,4]

[1]*School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*
[2]*School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu 611731, China*
[3]*Chongqing Key Laboratory of Optical Communication and Networks, Chongqing 400065, China*
[4]*Chongqing Key Laboratory of Ubiquitous Sensing and Networking, Chongqing 400065, China*

Correspondence should be addressed to Yaping Cui; cuiyp@cqupt.edu.cn

The diversified service requirements in vehicular networks have stimulated the investigation to develop suitable technologies to satisfy the demands of vehicles. In this context, network slicing has been considered as one of the most promising architectural techniques to cater to the various strict service requirements. However, the unpredictability of the service traffic of each slice caused by the complex communication environments leads to a weak utilization of the allocated slicing resources. Thus, in this paper, we use Long Short-Term Memory- (LSTM-) based resource allocation to reduce the total system delay. Specially, we first formulated the radio resource allocation problem as a convex optimization problem to minimize system delay. Secondly, to further reduce delay, we design a Convolutional LSTM- (ConvLSTM-) based traffic prediction to predict traffic of complex slice services in vehicular networks, which is used in the resource allocation processing. And three types of traffic are considered, that is, SMS, phone, and web traffic. Finally, based on the predicted results, i.e., the traffic of each slice and user load distribution, we exploit the primal-dual interior-point method to explore the optimal slice weight of resources. Numerical results show that the average error rates of predicted SMS, phone, and web traffic are 25.0%, 12.4%, and 12.2%, respectively, and the total delay is significantly reduced, which verifies the accuracy of the traffic prediction and the effectiveness of the proposed strategy.

1. Introduction

Autonomous driving is one of the key scenarios in 5G networks. In order to achieve road safety of intelligent transportation systems (ITS), the ultrareliable and low-latency communications (URLLC) must be guaranteed in vehicular networks. Moreover, for vehicle-to-everything (V2X) communications, the large data transmission of diversified service requirements poses the challenges to improve the transportation efficiency of ITS [1]. Therefore, it is urgent to tailor the vehicular networks to cater to these different requirements that come from different services. Along this line of thought, multiple virtual networks are created via network slicing as a feasible way to meet its diverse needs [2].

The main function of network slicing is creating multiple logical separate networks based on public shared physic

infrastructure [3]. Specially, these logical networks are independent with each other. Through network slicing, mobile network operator (MNO) can allocate network resources dynamically and flexibly to each logical network slice on demand [4, 5] so as to support extensive use cases with various performance and service needs. Network slicing can be realized by software-defined network (SDN) and network function virtualization (NFV) technologies [6], and the future network will evolve into the flexible and programmable network architecture gradually [7]. One important issue in network slicing is the scheduling policy which allocates limited resources dynamically to vehicles with various quality of service (QoS) requirements according to the traffic change and network state. Most of the existing researches about network slicing focused on core networks, and the requirements (i.e., QoS) for RAN slicing are usually assumed to be

guaranteed perfectly. However, it may be unreasonable when the radio access network (RAN) is not considered due to the resource scheduling processing [8].

Through network slicing operation, multiple network slices will share the resources of the basic network according to their specific needs [9, 10]. The traffic characteristics of each network slice were analyzed and predicted in [11, 12], and the admission control decisions for network slice requests were then studied. In [13], the problem of QoS-aware joint admission control and network slicing was studied, and a heuristic algorithm was proposed to solve this problem. A model for orchestrating network slices based on service requirements and available resources was proposed in [14]. In addition, a Markov decision process framework was proposed to formulate and determine the optimal strategy of resource allocation for the 5G networks. In [15], a novel radio resource slicing framework for 5G networks was proposed; then, radio resources were allocated to different slices based on reinforcement learning. In [16], the slicing resource allocation problem was modeled as an online winner determination problem to maximize the social welfare of auction participants. Similarly, in [17], a novel auction mechanism-based network slicing strategy is presented in which resources and revenue were jointly optimized. In [18], the authors studied a simple dynamic resource sharing policy and indicate that the slices are able to maximize their carried loads subject to performance requirements by admission control to manage users' performance.

However, due to the characteristics of the vehicular networks, such as stringent delay constraints, and complicated communication environments [19], the general network slice resource allocation strategy may be insufficient in such scenario. Fortunately, under the complex and changeable vehicular network environment, machine learning would be an effective solution to this problem. For the dynamic nature of 5G vehicular scenarios, reference [20] proposed an online learning algorithm, namely fast machine learning, to solve the problem of beam selection to achieve higher context-awareness and adaptability in millimeter-wave vehicular communications. Furthermore, in [21], the authors discussed the prospect of managing vehicular network resources by reinforcement learning, and some open issues are also highlighted. A novel resource allocation scheme based on deep learning was proposed in [22]; the proposed scheme can optimize resource allocations according to the changing demands and network dynamics in SDN-based vehicular networks.

With the accumulation of cellular traffic data and the development of machine learning and artificial intelligence [23], the idea of traffic prediction based on machine learning is becoming more and more popular in the field of communications [24]. In [25], the authors designed a hybrid deep learning model for spatiotemporal prediction, where the spatial dependence was modeled by autoencoder and the temporal dependence was captured by LSTM [26]. In [27], a deep transfer learning-based prediction architecture has been proposed to predict different service traffic more conveniently. Moreover, in order to allocate resources appropriately, literatures [28, 29] used the LSTM to predict future traffic in RAN network slicing.

Motivated by the above analysis, we use machine learning to predict the services traffic of each slice, so as to allocate radio resources to each slice to reduce the delay. In more detail, firstly, we propose a new radio resource management, namely, shared proportion fairness (SPF), to keep resource management in accordance with slicing vehicle activity, and then, we use it for resource allocation representation. Moreover, we formulated the system delay minimization problem of resource allocation as a convex optimization problem. Secondly, we use ConvLSTM, which combines CNN and LSTM, to model the temporal-spatial dependency of the slice service traffic in the vehicular communication networks. Using the ConvLSTM for traffic prediction, we can predict different service traffic to obtain the user load distribution. Finally, according to the predicted results, a primal-dual interior-point-based resource allocation strategy is used to explore the optimal slice weight.

The rest of this paper is organized as follows. Section 2 describes the system model and assumptions. Using the above model, the resource allocation problem is formulated as a convex optimization problem to obtain optimal slice weight allocation in Section 3; then, an LSTM-based resource allocation is presented to minimize the system delay. In Section 4, we propose a primal-dual interior-point-based resource allocation strategy to solve the optimal slice weight problem. Simulation results are provided in Section 5 to evaluate the performance of the proposed traffic prediction and resource allocation strategy, followed by concluding remarks in Section 6.

2. System Model

We consider the cellular network consists of B Road Side Units (RSU) and V network slices. And the sets of RSUs and slices are denoted by $\mathscr{B}$ and $\mathscr{V}$, respectively. As shown in Figure 1, RSU is virtualized into three layers, namely RSU interface layer, RSU virtualization layer, and RSU virtual resource layer. Among them, the RSU interface layer provides related interface for each slice. The RSU virtualization layer implements the functions of slice management, SDN control, and slice coordination. Besides, the RSU virtual resource layer provides the virtual resources required by each slice, which are obtained from the resource sharing layer of the base station. Also, the base station controls the service traffic prediction of the entire system and schedules the resources among slices.

System states $\mathscr{U}_b^v$, $\mathscr{U}_b$, and $\mathscr{U}^v$ represent the sets of vehicles that communicate with RSU b on slice v, communicate with RSU b, and on slice v, respectively. Specially, we use n_b^v and n^v denote the cardinalities of these sets, i.e., $|\mathscr{U}_b^v| = n_b^v$, and $|\mathscr{U}^v| = n^v$. Further, we assume that each vehicle communicates with only one RSU and connects to one slice.

In our model, the vehicle communicates with the RSU that provides it with the strongest SINR; thus, the downlink SINR can be expressed as

$$\text{SINR}_{ub} = \frac{P_b G_{ub}}{\sum_{k \in \mathscr{B}} P_k G_{uk} + \sigma^2}, \tag{1}$$
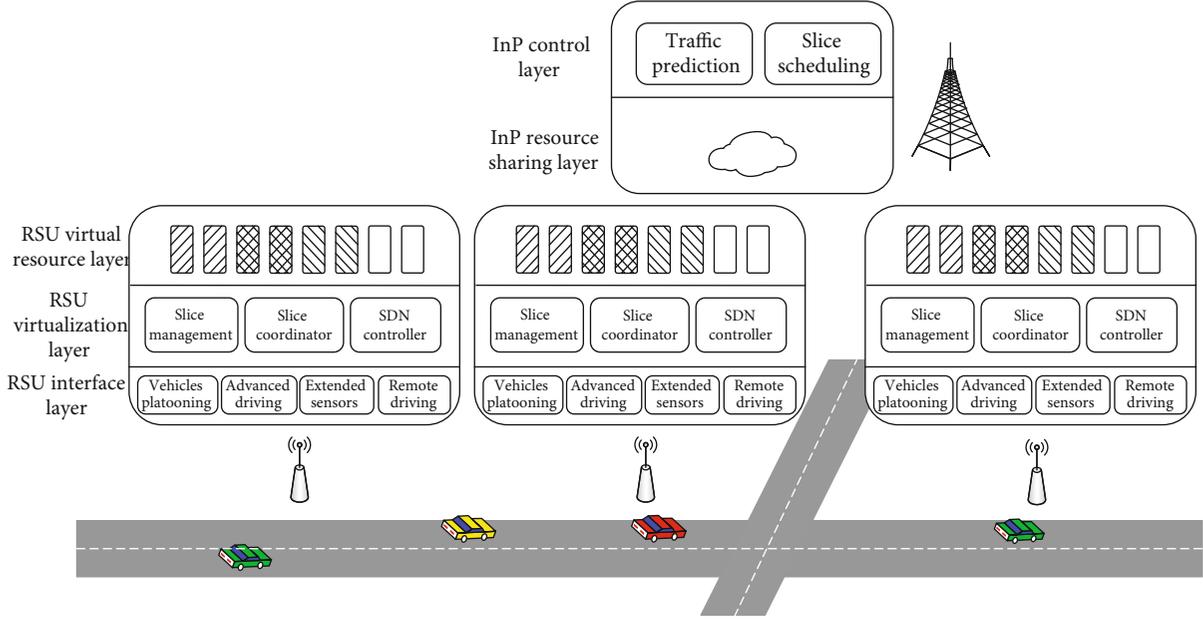
Figure 1: Network slicing framework in vehicular networks.

where the spectrum of the RSU is set to 20 MHz, and the transmit power P_b is set to 44 dBm [30]. The channel gain G_{ub} is related to path loss, shadow fading, and fast fading. The path loss is equals to $39 \log_{10}(d_{ub}) + 25 + 20 \log_{10}(f_c)$ [31], where d_{ub} and f_c represent the distance between the vehicle and the RSU and carrier frequency, respectively, and we set $f_c = 4$ GHz. Shadow fading follows logarithmic normal distribution with mean square deviation 4 dB. Fast fading is Rayleigh distribution depending on vehicles speed. The noise σ^2 depends on the noise spectral density $\eta = -174$ dBm/Hz and the noise figure $\gamma = 9$ dB.

According to the Shannon capacity formula, we get the spectrum efficiency of vehicle u at RSU b, which can be expressed as

$$e_{ub} = \log_2(1 + \mathrm{SINR}_{ub}). \tag{2}$$

The limited resources of each RSU are shared by all connected vehicles, so vehicle $u \in \mathcal{U}_b$ can be allocated a fraction of resources from RSU b. The allocated resources may be some resource blocks or time slots. For simplicity, we use $f_u \in [0, 1]$ that represents the allocated resources to vehicle u, which implies a proportion of total resources of RSU with $\sum_{u \in \mathcal{U}_b} f_u = 1$. Therefore, the transmission rate of RSU b to vehicle u is formulated as $r_u = f_u c_u$, where $c_u = B e_{ub}$ denotes transmission rate when all resources of RSU b are allocated to vehicle v, and B is bandwidth.

3. LSTM-Based Resource Allocation

In this section, we will describe the LSTM-based resource allocation. The shared proportional fairness resource allocation is first introduced, which can keep resource management in accordance with slicing vehicle activity. Then, the system delay minimization problem is formulated as a convex optimization problem. Finally, the ConvLSTM-based cellular traffic prediction is proposed at the end of this section.

3.1. Shared Proportional Fairness. For each slice in vehicular networks, we assume each slice is allocated a certain percentage of the radio resources, which is denoted by s^v, $v \in \mathcal{V}$, so we have $s^v > 0$, $\forall v \in \mathcal{V}$, and $\sum_{v \in \mathcal{V}} s^v = 1$. Next, the vehicle gets a subweight from the serving slice that depends on the number of active vehicles, i.e., for a vehicle $u \in \mathcal{U}^v$, $\forall v \in \mathcal{V}$, where $\omega_u = s^v / n^v$ means the subweight of vehicle u. Finally, The RSU allocates its resource to vehicles in proportion to their weights. Consequently, the transmission rate from RSU b to vehicle u can be obtained and written as

$$r_u = \frac{\omega_u}{\sum_{u' \in \mathcal{U}_b} \omega_{u'}} c_u = \frac{s^v / n^v}{\sum_{v' \in \mathcal{V}} \left(n_b^{v'} s^{v'} / n^{v'} \right)} c_u. \tag{3}$$

Considering there are many vehicles on slice v at RSU b so based on some further notations that are introduced in Table 1, the average transmission rate provided by RSU b to slice v is expressed as

$$r_b^v = \frac{s^v / \rho^v}{\tilde{g}_b} c_u. \tag{4}$$

According to (4), the average bit transmission delay (BTD) of the vehicle on slice v can be given by

$$\mathrm{BTD}^v = \sum_{b \in \mathcal{B}} \tilde{\rho}_b^v \mathrm{BTD}_b^v = \frac{\rho^v \langle \tilde{\rho}^v, \tilde{g} \rangle_{\Delta_v}}{s^v}, \tag{5}$$

TABLE 1: Key notations.

Notation	Definition	Interpretation
ρ^v	n^v	Overall load of slice v
$\boldsymbol{\rho}^v$	$(\rho_b^v \triangleq n_b^v : b \in \mathscr{B})$	Load distribution of slice v
$\tilde{\boldsymbol{\rho}}^v$	$(\tilde{\rho}_b^v \triangleq \rho_b^v/\rho^v : b \in \mathscr{B})$	Relative load distribution of slice v
$\tilde{\boldsymbol{g}}$	$\left(\tilde{g}_b \triangleq \sum_{v \in V} s^v \tilde{\rho}_b^v : b \in \mathscr{B}\right)$	Overall weight relative load distribution
$\boldsymbol{\delta}^v$	$(\delta_b^v \triangleq \mathbb{E}[1/c_b^v] : b \in \mathscr{B})$	Mean reciprocal capacity of slice v
Δ_v	diag $(\boldsymbol{\delta}^v)$	Diagonal matrix of mean reciprocal capacity of slice v
θ^v	$(\theta_b^v \triangleq \mathbb{E}[1/(1-\sigma_b^v)] : b \in \mathscr{B})$	Waiting parameter of slice of slice v
$\boldsymbol{\theta}^v$	diag (θ^v)	Diagonal matrix of waiting parameter of slice v

where BTD^v represents the average BTD of the vehicle at RSB b on slice v. Besides, we use $\langle x_1, x_2 \rangle_M \triangleq x_1^T M x_2$ and $\|x\|_M \triangleq \sqrt{x^T M x}$ that denote the weighted inner product of the vectors and the weighted norm of a vector, respectively. M denotes a diagonal matrix.

We assume the message e handling process follows the $GI/M/1/\infty$ queue model [32]. Among which, the random variable of message arrival interval obeys general distribution, $F(t), t \ge 0$, while $F(t)$ in the different time slots is independent and identically distributed. Its expectation $1/\lambda = \int_0^\infty t dF(t), \lambda > 0$, where λ is the arrival rate. The message service time is exponentially distributed, i.e., $G(t) = 1 - e^{-\mu t}, t \ge 0$, and the mean value of the message service times depends on the number of allocated resource blocks. To facilitate the analysis of the system delay, $1/\mu_0$ is used to denote the average service time when only one resource block for the message processing. Similarly, we denote the average service time as $1/RB_b^v \mu_0$ when RB_b^v resource blocks for the message processing.

The average waiting delay of the vehicle $u_b^v, u_b^v \in \mathscr{U}_b^v$ is given by

$$\mathrm{WD}_b^v = \frac{1}{\mathrm{RB}_b^v}(1 - \sigma_b^v), \tag{6}$$

where $\mathrm{RB}_b^v = (s^v/n^v/\sum_{v' \in \mathscr{V}}(n_b^{v'} s^{v'}/n^{v'}))c_u$, and σ_b^v can be obtained by solving the following equation:

$$\int_0^\infty e^{-\mathrm{RB}_b^v \mu^v(1-\sigma_b^v)t} \frac{r_b^v \rho_b^v e^{-\rho_b^v + \rho_b^v e^{-\rho_b^v t} - r_b^v t}}{1 - e^{-\rho_b^v}} dt = \sigma_b^v. \tag{7}$$

As a result, the average waiting delay of the vehicle on slice v is

$$\mathrm{WD}^v = \sum_{b \in B} \tilde{\rho}_b^v \mathrm{WD}_b^v = \frac{\rho^v \langle \tilde{\boldsymbol{\rho}}^v, \tilde{\boldsymbol{g}} \rangle_{\theta^v}}{s^v \mu^v}. \tag{8}$$

According to formula $\langle x_1, x_2 \rangle_{M_1 + M_2} \triangleq \langle x_1, x_2 \rangle_{M_1} +$

$\langle x_1, x_2 \rangle_{M_2}$ and (5) (8), the total average delay of a vehicle on slice v can be obtained by

$$D_{\mathrm{Total}}^v = \mathrm{BTD}^v + \mathrm{WD}^v = \frac{\rho^v \langle \tilde{\boldsymbol{\rho}}^v, \tilde{\boldsymbol{g}} \rangle_{\mu^v \Delta_v + \theta^v}}{s^v \mu^v}. \tag{9}$$

3.2. Problem Formulation. In the real implementation of network slicing, each slice would provide a guarantee of service to the vehicles, that is, the total delay on slice v does not exceed a deadline d_v. In this subsection, we will explore how to obtain the optimal solution of minimizing system delay by allocating weight to each slice.

Considering the network with just one slice v, so we have $s^v = 1, \tilde{\boldsymbol{g}} = \tilde{\boldsymbol{\rho}}^v$. To satisfy the deadline d_v, from (9), we can obtain that

$$\rho^v \le l(d_v, \tilde{\boldsymbol{\rho}}^v) \triangleq \frac{\mu^v d_v}{\|\tilde{\boldsymbol{\rho}}^v\|_{\mu^v \Delta_v + \theta^v}^2}, \tag{10}$$

where $l(d_v, \tilde{\boldsymbol{\rho}}^v)$ is the acceptable maximum load of slice v.

Next, considering multislice networks, each slice has its self-requirement. According to (9) (10), each slice would satisfy the following constraint to meet their requirements: $\forall v \in \mathscr{V}$

$$s^v \ge \frac{\rho^v}{l(d_v, \tilde{\boldsymbol{\rho}}^v) - \rho^v} \sum_{u \ne v} s^u \frac{\langle \tilde{\boldsymbol{\rho}}^v, \tilde{\boldsymbol{\rho}}^u \rangle_{\mu^v \Delta_v + \theta^v}}{\|\tilde{\boldsymbol{\rho}}^v\|_{\mu^v \Delta_v + \theta^v}^2}. \tag{11}$$

Equation (11) can be written in a simplified form, i.e.,

$$\sum_{v \in V} s^v \boldsymbol{h}^v \ge \boldsymbol{0}, \tag{12}$$

where $\boldsymbol{h}^v = (h_u^v : u \in \mathscr{V})$ is share coupling vector of slice v and can be expressed by

$$h_u^v = \begin{cases} 1, & v = u \\ -\dfrac{\rho^u}{l(d_u, \tilde{\boldsymbol{\rho}}^u) - \rho^u} \dfrac{\langle \tilde{\boldsymbol{\rho}}^u, \tilde{\boldsymbol{\rho}}^v \rangle_{\mu^u \Delta_v + \theta^u}}{\|\tilde{\boldsymbol{\rho}}^u\|_{\mu^u \Delta_v + \theta^u}^2}, & v \ne u \end{cases} \tag{13}$$

Our objective is to satisfy the requirements of vehicles in each slice and minimize the system overall delay, so the objective function is the summing average delay of all slices. Consequently, the optimization problem can be formulated as

$$\min_{s^v} \sum_{v \in V} \frac{\rho^v \langle \tilde{\boldsymbol{\rho}}^v, \tilde{\boldsymbol{g}} \rangle_{\mu^v \Delta_v + \theta^v}}{s^v \mu^v}$$
$$\text{s.t.} C1 : \sum_{v \in V} s^v h_i^v \geq 0, i = 1, 2, \cdots, V$$
$$C2 : \sum_{i=1}^{V} s^i = 1 \qquad (14)$$
$$C3 : s^i \geq 0, i = 1, 2, \cdots, V$$

where s^v is the optimization variable. Constraint $C1$ ensures the requirements of slices can be satisfied. Constraints $C2$ and $C3$ state the weight of each slice is nonnegative and is constrained by total resources. We can see that the problem (14) is an inequality constrained convex optimization problem, the method to solve it is described in Section 4.

3.3. Cellular Traffic Prediction. In practice, searching the optimal weights of each slice through minimizing the system total delay according to the current load distribution will cause some delay; we define that as arranging delay D_{ARR}. To reduce the arranging delay, traffic prediction is a feasible solution. Through traffic prediction, the system can acquire complicated traffic information in advance to calculate the optimal weighs for each slice. Hence, the arranging delay can be largely reduced.

ConvLSTM neural networks are adapted to predict cellular traffic. ConvLSTM networks can not only model the sequences information of cellular traffic accurately as same as Long Short-Term Memory (LSTM) networks [12] but also the local feature as same as convolutional neural networks (CNN). In short, it can easy to capture the spatial-temporal dependencies. Consequently, ConvLSTM networks are appropriate for predicting slice traffic in complex vehicular networks. This memory cell consists of cell states and three neural network units, i.e., input gate, forget gate, and output gate. For this specific framework, it is able to effectively store information chronically from long-term sequences.

As shown in Figure 2, the forget gate outputs a value $f_t \in [0, 1]$ to the cell according to the current input x_t and past output H_{t-1}, which determines what information would be abandon in past cell status C_{t-1} now. The calculation formula of forget gate is given by

$$f_t = \sigma \left(W_f^x * x_t + W_f^h H_{t-1} + b_f \right). \qquad (15)$$

The input gate decides update when a new input comes to the ConvLSTM unit through a sigmoid function, which can
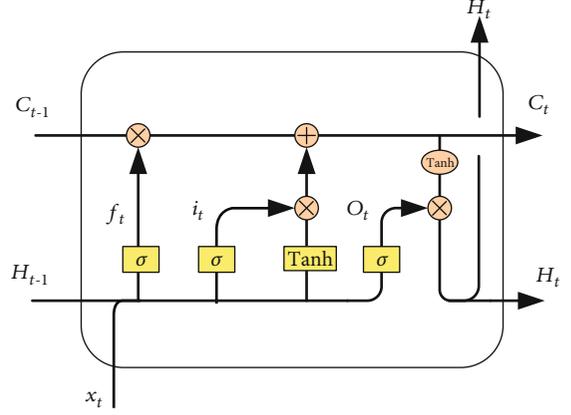
FIGURE 2: LSTM cell structure.

be further effects present states C_t, and expressed as

$$i_t = \sigma \left(W_i^x * x_t + W_i^h H_{t-1} + b_i \right),$$
$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh \left(W_c^x * x_t + W_c^h H_{t-1} + b_c \right). \qquad (16)$$

The output gate decides the output of this cell through a sigmoid function. After that, the output H_t is given by

$$o_t = \sigma \left(W_o^x * x_t + W_o^h H_{t-1} + b_o \right),$$
$$H_t = o_t \circ \tanh \left(C_t \right), \qquad (17)$$

where f_t, i_t, o_t, C_t, x_t, and H_t denote the output of forget gate, the output of input gate, cell status, cell input, and cell output, respectively, and $W_i^x, W_f^x, W_c^x, W_o^x, W_i^h, W_f^h, W_c^h, W_o^h, b_f, b_i, b_c$, and b_o are the parameters of the LSTM network. The functions $\sigma(\bullet)$ and $\tanh(\cdot)$ are sigmoid function and hyperbolic tangent function, respectively. In the above equations, the notation $*$ denotes the convolution operation, and the notation $\circ$ denotes the Hadamard product. Different from the common LSTM networks, the inputs or outputs in the ConvLSTM unit are all three-dimensional tensors. More specifically, the citywide service traffic data can be treated as a matrix or picture. Then, previous multiple service data are fed into the ConvLSTM networks to obtain future results. So, the multiply operation of common LSTM networks is replaced by convolution operation in ConvLSTM. The neural network can be accomplished via updating various parameters in each iteration, e.g., $W_i^x, W_f^x, W_c^x, W_o^x, W_i^h, W_f^h, W_c^h, W_o^h, b_f, b_i, b_c$, and b_o, so that the networks can minimize the error between forecasted values and ground trues.

4. A Primal-Dual Interior-Point Method-Based Resource Allocation Algorithm

In Section 3, the LSTM-based resource allocation is presented to minimize the system delay. Then, we will explore how to obtain the optimal slice weight of resources in

Input: Initial x_0, λ_0, scale factor k, residual error ϵ_{feas}, duality gap error ϵ, J times observed load distribution $\hat{\rho}_{t-J+1}$, $\hat{\rho}_{t-J+2}$, $\cdots$, $\hat{\rho}_t$
Output: Optimal solution x^*_{t+1}, $\cdots$, x^*_{t+K}
Phase 1: Predict service traffic
Training the ConvLSTM networks to obtain parameters W^x_i, W^x_f, W^x_c, W^x_o, W^h_i, W^h_f, W^h_c, W^h_o, b_f, b_i, b_c, b_o
According to J times observed load distribution predict K sequences in the future $\tilde{\rho}_{t+1}$, $\cdots$, $\tilde{\rho}_{t+K} = \underset{\rho_{t+1},\cdots,\rho_{t+K}}{\mathrm{argmax}}\ p(\rho_{t+1}, \cdots, \rho_{t+K} \mid \tilde{\rho}_{t-J+1}$,

$\tilde{\rho}_{t-J+2}$, $\cdots$, $\tilde{\rho}_t)$
Phase 2: To obtain optimal slice weight
while True **do**
    Calculate initial value of the surrogate gap $\eta \leftarrow f(x)^T \lambda$
    if$\{\|\gamma_{\mathrm{pri}} < \epsilon_{feas}\|\}$&&$\{\|\gamma_{\mathrm{dual}} < \epsilon_{feas}\|\}$&&$\{\|\hat{\eta}\| < \epsilon\}$**then**
        break
    end if
    Determine $t \leftarrow 2kV/\eta$
    Compute primal-dual search direction Δy_{pd}
    Determine initial step length $s_0 = \min\ \{0.99, \min\ \{-\lambda_i/\Delta\lambda_i \mid \Delta\lambda_i < 0\}\}$
    while $\min\ \{f_i(x + s\Delta x)| = 1,\cdots,2V\} > 0$**do**
        Ensure satisfy the constraint condition $s \leftarrow \beta s$
    end while
    while$\|\gamma_t(x + s\Delta x, \lambda + s\Delta\lambda, v + s\Delta v) < \epsilon_{feas}\|_2 > (1 - \alpha s)\|\gamma_t(x, \lambda, v)\|_2$
do
        Determine backtracking search step length $s \leftarrow \beta s$
    end while
    Update search direction $y \leftarrow y + \Delta y_{pd}$
end while

ALGORITHM 1: Resource allocation algorithm based on the primal-dual interior-point method.

problem (14) and apply the LSTM in resource allocation processing.

The resource allocation strategy includes two phases: the first phase is service traffic prediction by using machine learning described in the preceding section, and the second phase is the optimizing procedure that based on the primal-dual interior-point method. Based on the predicted results in the first phase, the slice weight distribution process can be performed in advance so that the delay is decreasing obviously.

Considering there are inequality constraints in the convex optimization problem (14), so it can be solved by the primal-dual interior-point method. We will transform the inequality constrained optimization problem as an equality constrained optimization problem so that the central path of this problem can be found. Therefore, we rewrite the problem (14) as

$$\min \sum_{v\in\mathcal{V}} \frac{\rho^v \langle \tilde{\rho}^v, \tilde{g}\rangle_{\mu^v\Delta_v+\theta^v}}{s^v\mu^v} + \sum_{i=1}^{V}\left(-\frac{\log\ (\sum_{v\in\mathcal{V}} s^v h^v_i)}{t}\right) + \sum_{i=1}^{V} -\frac{\log\ (s^i)}{t}$$
$$\text{s.t. } \sum_{i=1}^{V} s^i = 1\ i = 1, 2, \cdots, V.$$
$$(18)$$

For simplicity, let $f_0(x) = \sum_{v\in\mathcal{V}}(\rho^v \langle \tilde{\rho}^v, \tilde{g}\rangle_{\mu^v\Delta_{vv}+\theta^v}/s^v\mu^v)$,

$\phi(x) = \sum_{i=1}^{V}(-\log\ (\sum_{v\in\mathcal{V}} s^v h^v_i)) + \sum_{i=1}^{V} -\log\ (s^i)$, where $x = [s^1, s^2, \cdots, s^V, -f_0(x)]$ is the optimization variable. Considering the equivalent problem

$$\min tf_0(x) + \phi(x)$$
$$\text{s.t. } Ax = 1$$
$$(19)$$

Due to the number of slices is three in the simulation setting, so $x = [s^1, s^2, s^3, -f_0(x)]$ and $A = [1, 1, 1, 0]$. According to reference [33], the Newton step Δ_y is given by the modified KKT equations

$$\begin{bmatrix} \nabla f_0(x) + \sum_{i=1}^{m}\lambda_i\nabla^2 f_i(x) & Df(x)^T & A^T \\ -\mathbf{diag}\ (\lambda)Df(x) & -\mathbf{diag}\ (f(x)) & 0 \\ A & 0 & 0 \end{bmatrix}$$
$$\begin{bmatrix} \Delta x \\ \Delta\lambda \\ \Delta v \end{bmatrix} = \begin{bmatrix} \nabla f_0(x) + Df(x)^T\lambda + A^Tv \\ -\mathbf{diag}\ f(x) - (1/t)\mathbf{1} \\ Ax - \mathbf{1} \end{bmatrix} = \begin{bmatrix} \gamma_{\mathrm{dual}} \\ \gamma_{\mathrm{cent}} \\ \gamma_{\mathrm{pri}} \end{bmatrix},$$
$$(20)$$

(a)

(b)

With prediction
Without prediction

(c)
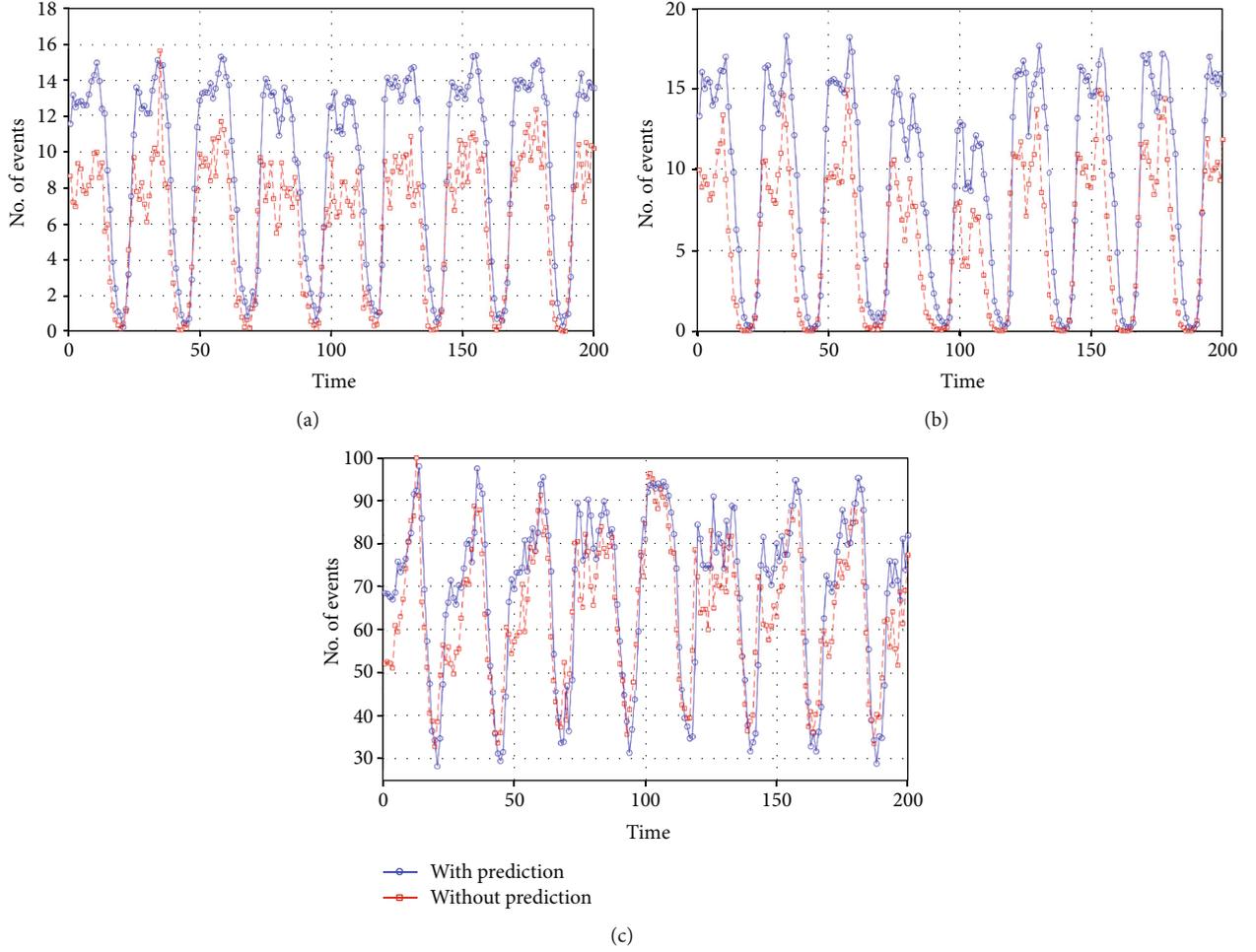
FIGURE 3: The comparison of traffic with and without prediction for three types of service. (a) SMS. (b) Phone. (c) Web.

- - - With prediction
——— Without prediction
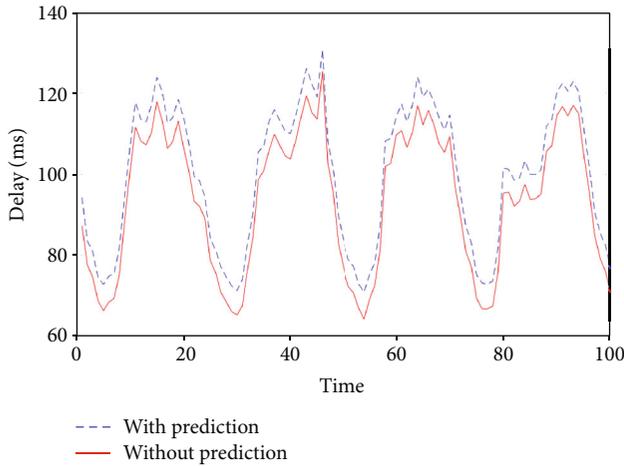
FIGURE 4: Network slicing delay with and without prediction.

where $\lambda_i = -(1/tf_i(x))\boxtimes i = 1, \cdots, 2,$

$$(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_{2V}(x) \end{bmatrix} \quad (21)$$

is the inequality constrained function of the original problem,

$$Df(x) = \begin{bmatrix} \nabla f_1(x)^T \\ \vdots \\ \nabla f_2(x)^T \end{bmatrix} \quad (22)$$

is its derivative matrix. γ_{dual} and γ_{pri} are dual residual and primal residual, respectively, and these residuals are used for the termination condition in the primal-dual interior-point method. The solution of (20) is the current primal-dual search direction $\Delta y_{\text{pd}} = (\Delta x_{\text{pd}}, \Delta \lambda_{\text{pd}}, \Delta v_{\text{pd}})$.

The step length can be obtained by backtracking line search which is based on the norm of the residual. Through continuous iteration, the best solution will be returned when

(a)

(b)

(c)
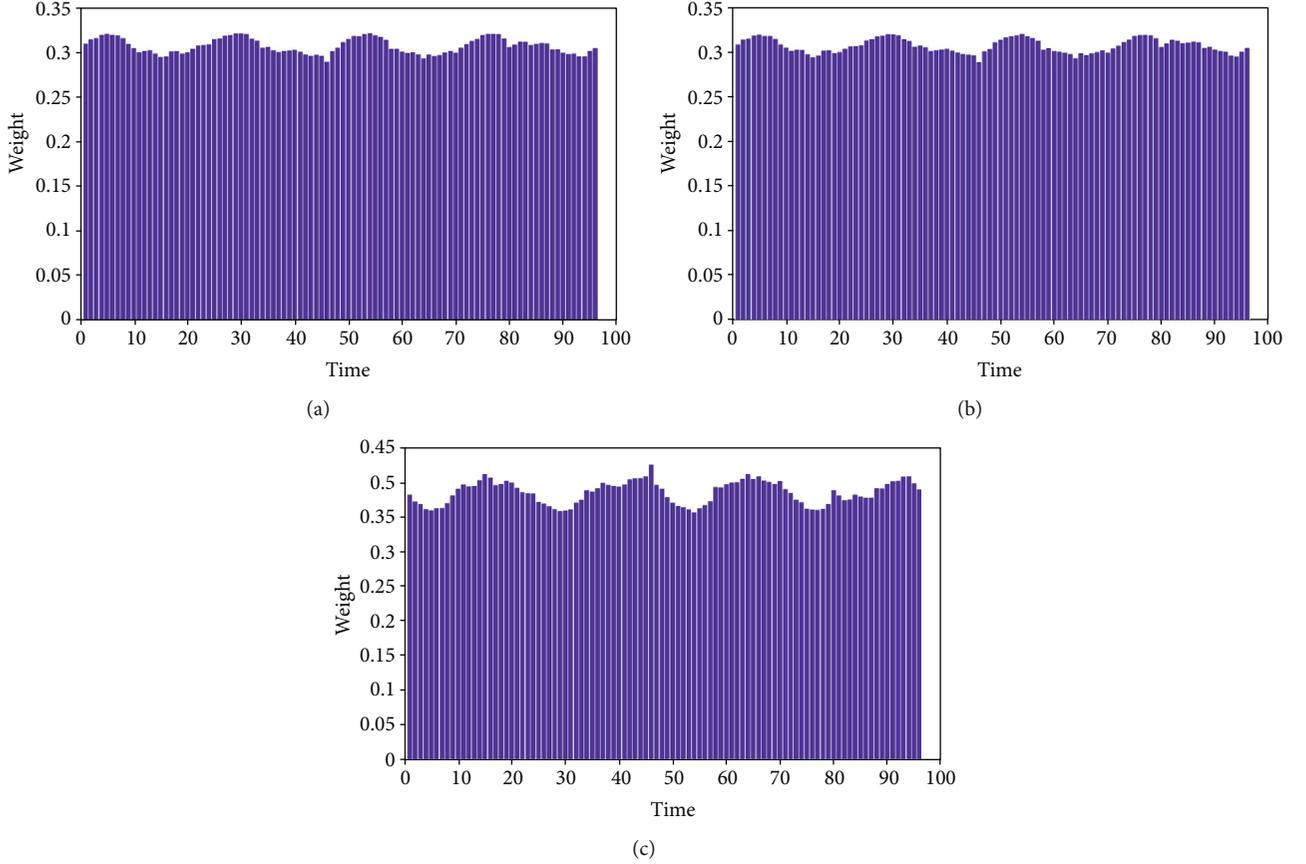
FIGURE 5: Optimal slice weight allocation at different times of three types of slice. (a) SMS. (b) Phone. (c) Web.

the present state satisfies the termination condition; the details are elaborated in Algorithm 1.

5. Performance Evaluation

In this section, we present numerical experiments to illustrate the performance of the proposed traffic prediction and resource allocation strategy. We use the cellular traffic dataset [34] of Milan, Italy, to train our neural networks; this dataset contains three categories of service traffic, i.e., SMS, phone, and web traffic, and we can deem it as three slices. The area of Milan is divided into a grid overlay of 100×100 squares; the dataset records 1000 samples of each square with a temporal interval of one hour. We choose the 800 samples from this dataset as the training set, the rest of the samples as the testing set. In the simulation setting of ConvLSTM, the neural network is set to three layers, and each layer has three ConvLSTM cells. In the training stage, we set the learning rate as 0.01, batch size as 32. After 100 times iteration, the trained model is used to predict the three types of service traffic. Then, the primal-dual interior-point resource allocation strategy is used to solve the optimal slice weight allocation according to the predication. In simulation parameters, we set scale factor $k = 2$, residual error $\epsilon_{feas} = 10^{-6}$, and duality gap error $\epsilon = 10^{-8}$.

Figure 3 shows the difference between predicted traffic and real traffic in a certain cell. The vertical axis of Figure 3

is the number of user access for service, and the horizontal axis of Figure 3 is the temporal dimension. It can be seen from Figure 3 that the temporal activities of three services show a strong daily pattern and follow weekly-periodic properties. Besides, the traffic volume expresses a slight difference among the three services. Specifically, the number of web access is always larger than that of the other two services. It can be seen that the service traffic shows a strong daily pattern. The average error rates of SMS, phone, and web traffic are 25.0%, 12.4%, and 12.2%, respectively.

Figure 4 shows the network slicing delay with and without prediction when slice weight is fixed, and the default of arranging delay D_{ARR} was set to 20 ms. The system delay also exhibits daily characteristics. Further, we can get that the more the number of user access, the greater the total average delay of the system. It can be seen that the periodicity of service traffic is mapped dynamically to system delay completely. Using the service traffic prediction, the slice weight can be distributed on average 15.33 ms in advance.

The predicted results express system load distribution in the future; the optimal slice weight can be solved accordingly. In Figure 5, we present the optimal slice weight under various times. Especially, at the 12th, 36th, 60th, and 84th time slots, the optimal slice weights of SMS are 0.3011, 0.3028, 0.3001, and 0.3076, respectively; the slice weights of phone and web traffic are 0.3032, 0.3059, 0.3018, and 0.3015; and 0.3957, 0.3913, 0.3982, and 0.3819. The results clearly show that the

optimal resource allocation problem is resolved under different times; the optimal slice weight is dynamic and varies with the service traffic pattern.

From the above results, it can be seen that the traffic prediction can be used to predict the service traffic in the future, so the user load distribution can be obtained. Therefore, the slice weight can be distributed in advance to save arranging delay. Furthermore, the primal-dual interior-point-based resource allocation strategy can calculate the current optimal slice weight distribution.

6. Conclusions

For slicing resource allocation problem in vehicular networks, this paper proposes an LSTM-based resource allocation, which contains two phases, i.e., traffic prediction phase and resource allocation phase. Moreover, we use ConvLSTM to capture the spatial-temporal dependencies of service traffic of each slice. Therefore, the user load distribution is obtained through traffic prediction. Then, based on the predicted user load distribution, we propose a primal-dual interior-point-based resource allocation strategy to solve the optimal slice weight problem for minimizing the system total delay. The resource allocation strategy proposed in this paper can make the slice weight be allocated in advance, thus greatly saving the delay. Moreover, with the development of AI, the research field of wireless communication will be impacted deeply. In the future, we will investigate the implementation of resource allocation with machine learning methods in vehicular network slicing.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

References

[1] S. Seng, C. Luo, X. Li, H. Zhang, and H. Ji, "User matching on blockchain for computation offloading in ultra-dense wireless networks," *IEEE Transactions on Network Science and Engineering*, 2020.

[2] S. Abdelwahab, B. Hamdaoui, M. Guizani, and T. Znati, "Network function virtualization in 5G," *IEEE Communications Magazine*, vol. 54, no. 4, pp. 84–91, 2016.

[3] NGMN, *Description of Network Slicing Concept Version 1.0*, NGMN, Frankfurt, Germany, 2016.

[4] J. Tang, B. Shim, and T. Q. S. Quek, "Service multiplexing and revenue maximization in sliced C-RAN incorporated with URLLC and multicast eMBB," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 4, pp. 881–895, 2019.

[5] H. Halabian, "Distributed resource allocation optimization in 5G virtualized networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 3, pp. 627–642, 2019.

[6] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, and V. C. M. Leung, "Network slicing based 5G and future mobile networks: mobility, resource management, and challenges," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 138–145, 2017.

[7] R. Li, Z. Zhao, X. Zhou et al., "Intelligent 5G: when cellular networks meet artificial intelligence," *IEEE Wireless Communications*, vol. 24, no. 5, pp. 175–183, 2017.

[8] P. Caballero, A. Banchs, G. De Veciana, and X. Costa-Perez, "Network slicing games: enabling customization in multi-tenant networks," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pp. 1–9, Atlanta, GA, USA, 2017.

[9] *System Architecture for the 5G System; Stage 2 (Release 15), document 3GPP TS 23.501 v1.0.0*, 2018.

[10] W. Paper, "5G Radio Access Capabilities and Technologies, Ericsson, White Paper Uen 284 23-3204 Rev C," 2016, April 2019, https://www.ericsson.com/assets/local/publications/white-papers/wp5g.pdf.

[11] V. Sciancalepore, K. Samdanis, X. Costa-Perez et al., "Mobile traffic forecasting for maximizing 5G network slicing resource utilization," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pp. 1–9, Atlanta, GA, USA, 2017.

[12] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5G wireless communications: a deep learning approach," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 227–236, 2020.

[13] H. M. Soliman and A. Leon-Garcia, "QoS-aware frequency-space network slicing and admission control for virtual wireless networks," in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Washington, DC, USA, 2016.

[14] D. T. Hoang, D. Niyato, P. Wang, A. De Domenico, and E. C. Strinati, "Optimal cross slice orchestration for 5G mobile services," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, pp. 1–5, Chicago, IL, USA, 2018.

[15] A. Aijaz, "Hap-SliceR: a radio resource slicing framework for 5G networks with haptic communications," *IEEE Systems Journal*, vol. 12, no. 3, pp. 2285–2296, 2018.

[16] L. Liang, Y. Wu, G. Feng, X. Jian, and Y. Jia, "Online auction-based resource allocation for service-oriented network slicing," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 8063–8074, 2019.

[17] M. Jiang, M. Condoluci, and T. Mahmoodi, "Network slicing in 5G: an auction-based model," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, Paris, France, May 2017.

[18] J. Zheng, P. Caballero, G. de Veciana, S. J. Baek, and A. Banchs, "Statistical multiplexing and traffic shaping games for network slicing," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2528–2541, 2018.

[19] H. Peng, L. Liang, X. Shen, and G. Y. Li, "Vehicular communications: a network layer perspective," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1064–1078, 2019.

[20] G. H. Sim, S. Klos, A. Asadi, A. Klein, and M. Hollick, "An online context-aware machine learning algorithm for 5G mmWave vehicular communications," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2487–2500, 2018.

[21] L. Liang, H. Ye, and G. Y. Li, "Toward intelligent vehicular networks: a machine learning framework," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 124–135, 2019.

[22] S. Khan, H. A. Khattak, A. Almogren et al., "5G vehicular network resource management for improving radio access through machine learning," *IEEE Access*, vol. 8, pp. 6792–6800, 2020.

[23] J. Wang, B. Gong, H. Liu, and S. Li, "Multidisciplinary approaches to artificial swarm intelligence for heterogeneous computing and cloud scheduling," *Applied Intelligence*, vol. 43, no. 3, pp. 662–675, 2015.

[24] L. Nie, D. Jiang, S. Yu, and H. Song, "Network traffic prediction based on deep belief network in wireless mesh backbone networks," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–5, San Francisco, CA, USA, 2017.

[25] J. Wang, J. Tang, Z. Xu et al., "Spatiotemporal modeling and prediction in cellular networks: a big data enabled deep learning approach," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pp. 1–9, Atlanta, GA, USA, 2017.

[26] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: continual prediction with LSTM," *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, 2000.

[27] C. Zhang, H. Zhang, J. Qiao, D. Yuan, and M. Zhang, "Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1389–1401, 2019.

[28] M. Yan, G. Feng, J. Zhou, Y. Sun, and Y. C. Liang, "Intelligent resource scheduling for 5G radio access network slicing," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7691–7703, 2019.

[29] R. Li, C. Wang, Z. Zhao, R. Guo, and H. Zhang, "The LSTM-based advantage actor-critic learning for resource management in network slicing with user mobility," *IEEE Communications Letters*, vol. 24, no. 9, pp. 2005–2009, 2020.

[30] 3GPP, "TR 38.802 v14.2.0. Technical Specification Group Radio Access Network; Study on new radio access technology physical layer aspects," in 3GPP, 2017.

[31] Y. J. Bultitude and T. Rautiainen, "IST-4-027756 WINNER II D1. 1.2 V1. 2 WINNER II Channel Models," EBITG, TUI, UOULU, CU/CRC, NOKIA, Tech. Rep, 2007.

[32] L. Kleinrock, *Queueing Systems. Volume I: Theory*, Wiley, New York, NY, USA, 1972.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, Cambridge, UK, 2004.

[34] G. Barlacchi, M. de Nadai, R. Larcher et al., "A multi-source dataset of urban life in the city of Milan and the Province of Trentino," *Scientific Data*, vol. 2, no. 1, article 150055, 2015.

Research Article

Fog-Based Distributed Networked Control for Connected Autonomous Vehicles

Zhuwei Wang ⓘ**, Yuehui Guo** ⓘ**, Yu Gao** ⓘ**, Chao Fang** ⓘ**, Meng Li** ⓘ**, and Yang Sun** ⓘ

Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

Correspondence should be addressed to Chao Fang; fangchao.bupt@gmail.com

With the rapid developments of wireless communication and increasing number of connected vehicles, Vehicular Ad Hoc Networks (VANETs) enable cyberinteractions in the physical transportation system. Future networks require real-time control capability to support delay-sensitive application such as connected autonomous vehicles. In recent years, fog computing becomes an emerging technology to deal with the insufficiency in traditional cloud computing. In this paper, a fog-based distributed network control design is proposed toward connected and automated vehicle application. The proposed architecture combines VANETs with the new fog paradigm to enhance the connectivity and collaboration among distributed vehicles. A case study of connected cruise control (CCC) is introduced to demonstrate the efficiency of the proposed architecture and control design. Finally, we discuss some future research directions and open issues to be addressed.

1. Introduction

Nowadays, networking technology has become the key element in various application areas with the increasing number of connected devices. Networked control systems (NCSs), which consist of spatially distributed sensors, actuators, and controllers in the close-loop communication network, are the powerful enablers to meet the demands for both information dissemination and data analysis [1, 2]. This characteristic of distributed architecture design makes NCSs easier to implement, maintain, and share information resources, meeting the requirements of future connected system [3]. With the development of wireless communication technology, the transmission rate and reliability of information exchange through wireless link gradually meet the needs of future NCSs. In particular, the use of wireless communication eliminates unnecessary wiring among system components and can be upgraded easily [4]. Therefore, in wireless networked control systems (WNCS), such as wireless sensor networks and wireless multihop networks, more flexible system architecture design, more easily resource utilization, and increased safety can be achieved [5].

However, the usage of wireless networks in the data transmission introduces message delay and packet loss which cannot meet the emerging requirements for real-time control applications [6]. For example, higher requirements are emerged such as sufficiently reliability and low latency. Previous cloud-based architecture has a high response delay and is not suitable for delay-sensitive applications. Delays may cause the performance degradation and even more serious safety issues. In addition, for real-time control applications, the requirements of transmission reliability and stability introduce further research challenges. The complex network structures, large network scales, flexible communication topologies, and variable communication environments can also destabilize a WNCS and degrade system performance. In addition, traditional single-controller-based closed-loop control architecture is difficult to meet the needs of real-time control applications [7–10]. In order to improve overall system performance and efficiency, multicontroller design is adopted with the advantages in modularity, scalability, and robustness. With the growing number of connected vehicles require more communication, computing, and storage resources in the network.

Considering the current situation of WNCS, traditional cloud-based system architectures have pronounced limitations with extracommunication cost and waste on networking resources. Therefore, it is necessary to introduce edge computing and use cooperative features of wireless networks to optimize system performance and stability [11, 12]. In this case, all underlying distributed controllers are able to share information within a certain neighboring area and cooperate to make decisions alleviating computation burden and design complexity [13]. Fog computing expands the cloud computing paradigm, which is seen as a promising technique to spread the computing resources from the cloud computing servers to the edge in order to balance the load [14]. The integration of cloud, fog, and edge nodes at the lowest perception layer takes the advantages of low latency and flexible network topology [15]. Among all the service types, the WNCS based on fog computing provides a new solution for the development of real-time control applications. Fog computing makes it available for the central data processing and storage on-line at the cloud and distributed caching and computing resources at the edge [16].

In recent years, with the increasing number of vehicles on the road, intelligentization and networking of vehicles have been widely concerned by automobile manufacturers and research institutions [17, 18]. With the new round of technological revolution from the Internet to the Internet of Things (IoT) [19], the construction of Intelligent Transportation Systems (ITSs) has become a tendency to deal with the serious problems of traffic safety and road congestion [20–22]. In order to achieve safe, energy-saving, and efficient control, the ability of flexibly information sharing, environmental awareness, intelligent decision-making, coordinated control, and execution is needed among vehicles [23]. Applying WNCS technologies to the intelligent control of vehicles can strengthen the information exchange, make full use of system resources, and promote the realization of ITS. Since fog platform can provide computing, storing, and networking services, it is considered as a promising solution in the connected vehicles scenario [24–26].

In the safety aspect, the collision avoidance (CA) system plays a vital role in preventing the driver-caused accidents [27, 28]. Especially in the cruise control problem, applying wireless communication among vehicles, rather than completely relying on sensors equipped on each vehicle, has the potential to provide better system performance [29, 30]. In order to deal with above challenges, more flexible design for the vehicular platoon becomes an emerging trend. Connected cruise control (CCC), as a potential solution, has been proposed to hold a smooth traffic flow with flexible connectivity structure and communication topology [31, 32].

Recently, advanced driver assistance systems (ADAS) have attracted widespread attention in order to improve the flow rate and safety of the traffic flow [17, 33]. In the cruise control problem, the CA technology plays a vital role in preventing traffic accidents, and the use of wireless communication technology to replace sensors directly equipped between traditional vehicles can further provide better system performance [28, 30]. Therefore, the design of more flexible vehicle platoon system based on wireless communication technology

has become as an emerging trend [27, 29]. The CCC system, guarantee a stable traffic flow through flexible communication topologies and connection structures, has become a potential solution for future development [31, 32]. In this paper, we investigate the CCC system as a case study by implementing the proposed architecture. The main contributions of this paper are listed as follows.

(i) We propose a novel fog-based distributed network architecture toward real-time control applications. The critical system variables involved within the communication and control system that may influence system performances are analyzed

(ii) We formulate the CCC problem as a case study of the proposed fog-based distributed network architecture, and the control scheme is analyzed thoroughly with the cloud, fog, and connected vehicles. Then, a linear quadratic optimization problem is formulated to regulate vehicle's longitudinal motion when the time delay is considered

(iii) Based on the fog control nodes, the whole vehicle system can successfully split into multiple distributed vehicle platoons that each has an automated vehicle at the tail. Then, a two-step control scheme is proposed to solve the optimal CCC problem. (i) At the offline step, a backward recursion approach is used to derive the optimal control gain. (ii) At the online step, the optimal control strategy can be real-time calculated based on the current system states and local cached information

(iv) In a distributed control scheme, the optimal CCC strategy can be derived as a linear function of current platoon state information and the last control strategy

The remainder of this article is organized as follows. We propose a novel fog-based distributed network architecture toward real-time application in the following section. Next, a practical case study is provided to validate the high efficiency of the proposed architecture. Afterward, some discussions are presented in future research directions and open issues. Finally, the conclusion is made.

2. Proposed Architecture

With the recent revolution in wireless networking applications, such as IoT, Tactile Internet [34], and Vehicular Ad Hoc Networks (VANETs) [35], fog computing becomes a promising enabler to carry out a massive amount of communication, computation, storage, and networking services between edge devices and traditional center cloud servers [36]. More specifically, fog computing is more suitable for real-time control applications with the advantages of low latency, location awareness, wider geographical distribution, wider range of mobility, suitable for more nodes, and support for network heterogeneity [37].
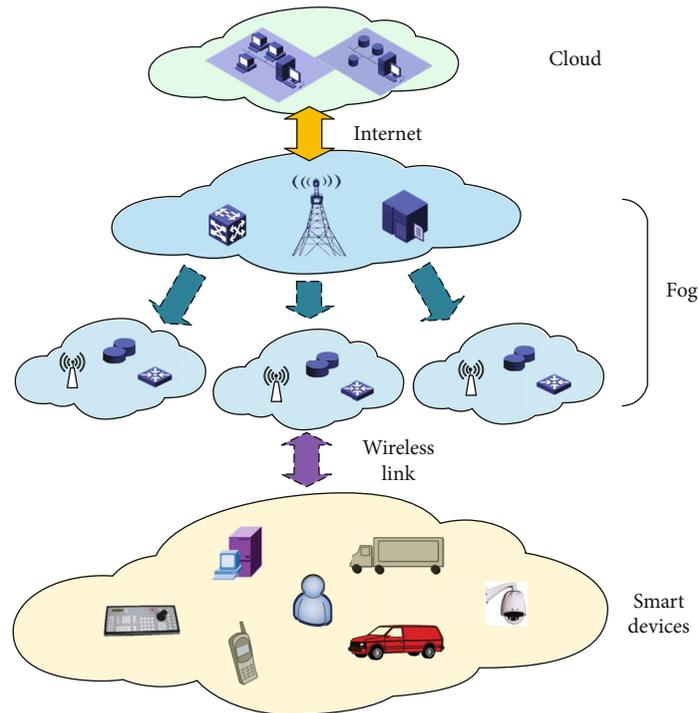
FIGURE 1: Overall system model of fog-based WNCS.

2.1. Overall Architecture Design. In this paper, we propose a fog-based distributed network architecture toward real-time control applications. Figure 1 depicts the overall system model consisting of three logical layers, the cloud, distributed fog, and edge smart devices. The cloud can serve multiple fog nodes, and a fog node can serve several dozens of edge nodes. The edge nodes can be sensors, embedded smart devices, mobile terminals, and vehicles. Each fog node can not only act as an independent service provider but also work together with other fog nodes to achieve collaborative control and services to edge nodes. The fog layer can be further classified into two categories, determined by their capability. Normal fog nodes are governed by the more powerful nodes. In the proposed architecture, the fog nodes have the ability of communicating, processing, and caching to complete two kinds of functions. On the one hand, the fog nodes realize the control of the edge local network, such as resource allocation and transmission path planning [38]. On the other hand, the fog nodes act as remote controllers, generate control strategies in real time according to the state information of the controlled platform and network characteristics to implement close-loop feedback control. Therefore, the fog nodes can both govern the wireless network as well as the real-time control of the controlled platform, which can effectively avoid entirely relying on the cloud. This scheme relieves the problem of high latency, and the usage of distributed cooperative control alleviates the transmission instability caused by wireless channels in traditional single-controller systems. The above fog-based system architectures can be applied to delay-sensitive systems in smart grid, vehicular network, wireless sensor and actuator network, and smart medical systems.

Figure 2(a) shows the typical WNCS model with the controlled plant, sensors, actuators, controllers, and relay nodes. Compared with the architecture in Figure 1, the plant along with sensors and actuators is seen as the edge node; the routers and controllers are seen as the fog nodes. The plant states are sampled at periodic intervals by the sensors capable of wireless communication. In order to realize the closed-loop control, each packet consisting of the plant states is transmitted to the controller over the wireless link; the controller computes the control signal based on the sensor measurements, then sent it to the actuator to ensure desirable dynamic and steady-state response.

2.2. Critical System Variables. In the proposed system, there are several critical variables involved with the combination of control system and wireless communication as shown in Figure 2(b). The practical physical systems are continuous-time systems where typical WNCS considers a discrete-time plant model with digital controllers. In the WNCS, periodic sampling is widely used, and shorter sampling period increases the network traffic causing higher message delay and message loss probability.

In particular, the network-induced delays mainly consist of the communication delays among sensors, controllers, and actuators, varying with communication channel quality. With different applications using various communication networks, the delays may possess diverse characteristics such as random or deterministic, constant, or time-varying. The uncertainty of delays significantly degrades the system performance. In addition, inevitably channel fading or shadowing in the dynamic wireless communication increases the packet loss probability, and retransmission is unfitted in
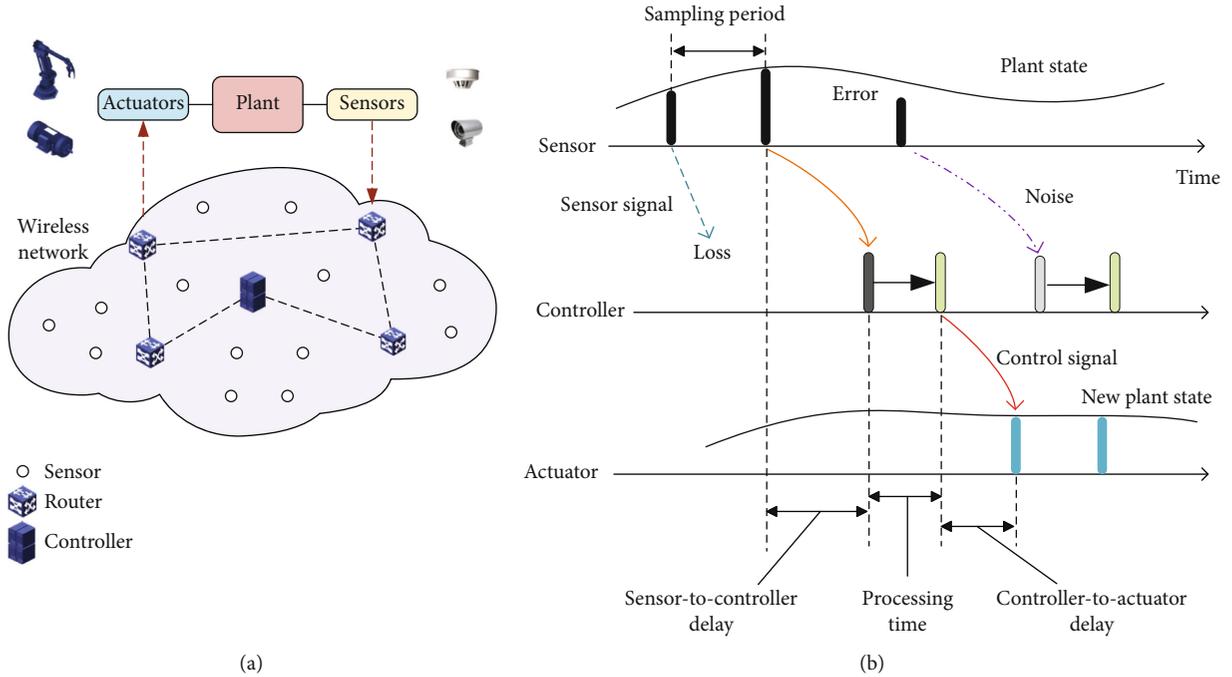
(a)

(b)

FIGURE 2: (a) WNCS model. (b) Timing diagram for the control loop over a wireless communication network.

real-time applications. In this case, seamless channel and bandwidth managements are required to optimize the network performance. In WNCS, the communication channel is often seen as an uncertain linear system with additive Gaussian noise. In order to design control strategies with partial state observation, estimation of the unknown variables based on measurements observed over time can be used.

In the proposed architecture, novel design is needed to optimize the system performance of control and communication. Since the raw sensor data in local networks are converged into the fog nodes with unavoidable invalid data, the fog nodes are responsible of filtering and temporarily storing the data. Fog nodes are also capable of the management of the local network with resource allocation, transmission path planning, and congestion control. In addition, the dynamically changes of nodes result in the time-varying network topology. In this case, it is difficult to maintain the system stability with single centralized controller [8]. In the modern control system, multiple distributed controller collaboration can enhance the control performance and provide better system stability.

3. Case Study: Connected Cruise Control

Today's automakers, Internet and leading technology companies are striving to develop innovative technologies to provide a fully integrated and highly intelligent vehicle experience. Electric and self-driving cars, the advanced state of artificial intelligence (AI) technology, and connectivity through wireless communication are revolutionizing the future of mobility [39, 40]. Connected vehicles are the emerging tendency in automobile manufacturing industries providing advanced quality of service (QoS) and attractive performance enhancements. The VANET has been widely

perceived as a promising concept for the realization of ITS with both safety and efficiency promotion. Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications are implemented in VANETs to support road safety, navigation, entertainments, and other application services.

In this section, we further investigate the system architecture toward real-time autonomous driving applications. In particular, to fulfill low latency requirements, traditional cloud-based network architecture may be insufficient with the growing number of vehicles. Based on the fog-based WNCS architecture shown in Figure 1, Figure 3 presents its application in the vehicular network scenario. In this architecture, vehicles are seen as edge devices equipped with sensors and On Board Units (OBUs). The perception of real-time traffic conditions is fulfilled by roadside sensors and vehicle's onboard sensors. Vehicles are capable of wireless communication broadcasting their kinematic data and transportation information in the local network. Each vehicle plays the role of packet sender, receiver, or even router within their reach via wireless medium. Rather than moving at random, vehicles tend to move in an organized fashion with the slowly varying network topology. RSUs and RSU controllers (RSUCs) are seen as fog nodes to provide more powerful packet transmission and computing functions as well as storage capabilities in the network in a timely, highly efficient, and coordinated way. RSUs are managed by the RSUC which has more resources for computing, storage, and communication through Internet to the cloud. The traditional core cloud facilitates the large-scale data processing and storage on-line and provides global view with big data and AI technology.

3.1. Problem Formulation and Control Design. As one of the widely used applications of Automatic Driving Assistance Systems, cruise control systems can effectively improve road
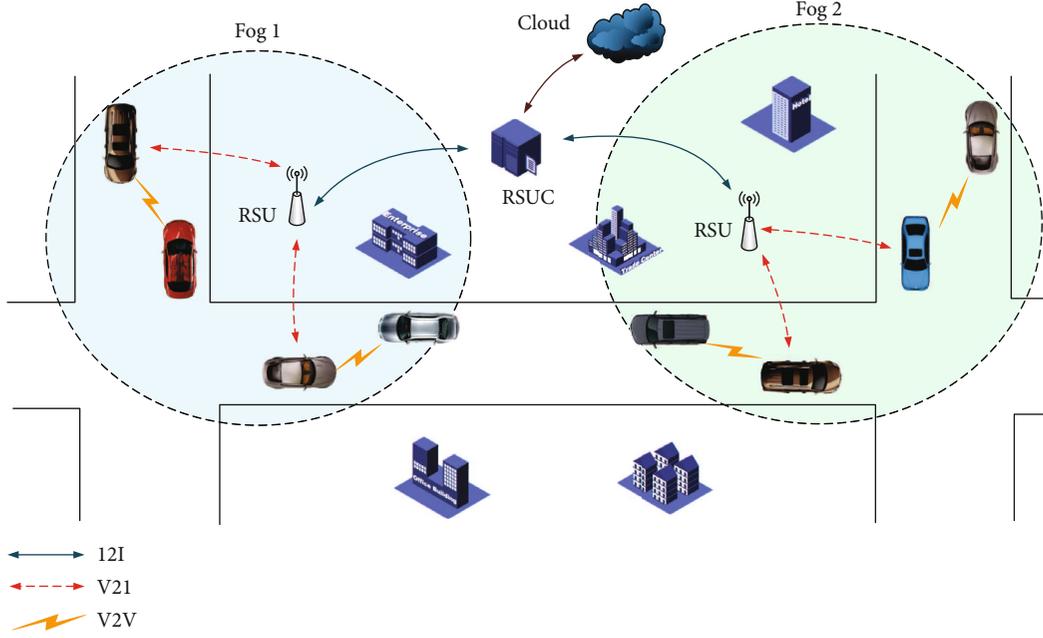
FIGURE 3: Fog-based vehicular network scenario.

traffic efficiency, reduce traffic accidents, improve vehicle fuel economy, and promote the realization of the ITS [41]. The CCC system, as an integration of the sensor-based cruise control system and wireless communication, can make full use of system resources to effectively avoid collision along with smooth drive [31]. Figure 4 depicts the considered scenario where vehicles generally move at a relatively constant speed, such as following a paved highway, and can be controlled as a whole platoon. The vehicular platoon consists of two kinds of vehicles: the human-driven vehicles and the connected and automated vehicles (CAVs). The CAVs are able to communicate with each other within a designated area through V2V communication and communicate with RSUs through V2I communication [20, 42]. The system is a location-based distributed control system, and the connectivity structure is basically stable or changing slowly. The local wireless network includes a RSU and vehicles governed by the RSU, and the transmission path through V2V and V2I is planned dynamically in RSUs to collaborate the resources and balance the load. The close-loop control of the CAV is realized with the collaboration of the cloud, fog, and vehicles. The cloud can provide big data processing and storage capability which is connected through the Internet. The fog, namely, the RSUs and RSUCs, brings powerful communicating and computing capabilities to the connected vehicles via V2I communication. The RSUs are responsible for local information processing and filtering, thereby effectively relieving the workload on the cloud, making it more suitable for real-time control applications. The large amount of moving vehicles is capable of sensing the real-time traffic information. In particular, at each sampling interval, the state information of all vehicles in the platoon is sent to the RSU. Then, the vehicle platoon model is formulated, and the optimal control strategy for the CAV is calculated by the RSU. The cooperative control is achieved by the information shar-

ing among the RSUs of the control strategies. In addition, the transmission path through V2V and V2I is planned dynamically in RSUs to allocate the resources and balance the load. Finally, the control signal is downloaded to the CAV to regulate the longitudinal motion based on its real-time states, and the close-loop control is realized.

In this subsection, the optimal control design is investigated for the CCC system with communication delays in a heterogeneous platoon including CCC and human-driven vehicles. First, the mathematical model of the vehicle dynamics for the connected vehicles is formulated. Here, we investigate the longitudinal control of CAVs where the velocity and distance between two consecutive vehicles are considered. The CAV's dynamics is formulated as

$$\dot{h}_j(t) = v_{j+1}(t) - v_j(t),$$

$$\dot{v}_j(t) = u_j(t - \tau), \tag{1}$$

where $u_j(t)$, namely, the acceleration, is the control signal of j-th CAV, τ is the time delay introduced by the V2V communication, and $v_j(t)$ and $h_j(t)$ are the velocity and headway, respectively.

The behavior of the human-driven vehicle can be formulated as [31, 43].

$$\dot{h}_i(t) = v_{i+1}(t) - v_i(t), \tag{2}$$

$$\dot{v}_i(t) = \alpha_i(V(h_i(t)) - v_i(t)) + \beta_i(v_{i+1}(t) - v_i(t)), \tag{3}$$

where $V(h)$ denotes the range policy, β_i and α_i are the system parameters determined by the human driver, and $v_i(t)$ and $h_i(t)$ are the velocity and headway of i-th human-driven vehicle.

FIGURE 4: A platoon of connected vehicles.

Considering n vehicles in the whole system as shown in Figure 4, we assume there are m CAVs, and then, the whole system can be split into m vehicular platoons with single CAV in each platoon. The j-th vehicular platoon has n_j vehicles that $\sum_j n_j = n$. We define the state vector as $x = [\tilde{h}_1, \tilde{v}_1, \cdots, \tilde{h}_n, \tilde{v}_n]^T$, where $\tilde{h}$ and $\tilde{v}$ denote the state errors to the desired states h^* and v^* Then, base on (1) and (3), the mathematical model for the whole system can be formulated as

$$\dot{x}(t) = Ax(t) + Bu(t - \tau), \tag{4}$$

where the control signal u is an m-dimensional vector corresponding to m CAVs, and the coefficient matrices can be expressed as

$$A \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_m \end{bmatrix}, = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_m \end{bmatrix}, \tag{5}$$

and the block matrices are

$$A_j = \begin{bmatrix} 0 & -1 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \Gamma_2 & \Phi_2 & 0 & \beta_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 & -1 \\ 0 & 0 & \cdots & \cdots & 0 & 0 & \cdots & \Gamma_{n_j} & \Phi_{n_j} \end{bmatrix}, B_j = [0, 1, 0, \cdots, 0]^T, \tag{6}$$

where $\Gamma_i = \alpha_i f^*$, $\Phi_i = -\alpha_i - \beta_i$, and $i = 2, 3, \cdots, n_j$.

Then, the corresponding discrete-time system model can be formulated as

$$x_{k+1} = A_k x_k + B_k^1 u_k + B_k^2 u_{k-1}, \tag{7}$$

where

$$x_k = x(kT), u_k = u(kT), A_k = e^{AT},$$
$$B_k^1 = \int_0^{T-\tau} e^{At} dt B, B_k^2 = \int_{T-\tau}^T e^{At} dt B, \tag{8}$$

and T is the sampling period, and τ is the network-induced delay. Here, the short-delay case is considered that τ is smaller than the sampling period.

The objective is to design an optimal control strategy to reduce energy consumption and improve traffic capacity. In the uniform traffic flow, each vehicle is tracking the desired velocity and headway. Thus, in order to minimize the deviations of CAV's headway, velocity, and acceleration, the optimization problem of the control system is typically designed as

$$\min J_N = x_N^T Q x_N + \sum_{k=0}^{N-1} \left(x_k^T Q x_k + u_k^T R u_k \right),$$
$$s.t. x_{k+1} = A_k x_k + B_k^1 u_k + B_k^2 u_{k-1}, \tag{9}$$

where N is the finite time horizon and Q and R are the symmetric positive definite weight matrices. Solving the optimization problem, we can obtain the control signals for the CAVs to regulate their longitudinal motion.

In the proposed architecture, RSUs and RSUCs can provide powerful communicating and computing capabilities in a coordinated way, enabling connected vehicles in the local

wireless network to interact with each other. Here, vehicles tend to move in an organized fashion in order to maximize the throughput of the traffic flow. Therefore, a uniform desired velocity is to be realized for the whole system. Since the RSUs can share the local information with each other, the desired states for the traffic flow can be determined by the RSUCs with global view. Then, with multiple RSUs implemented in the system to provide seamless experience for all vehicles, we can split the whole system into several parts with a single RSU serving one corresponding vehicular platoon. In this case, when each platoon reaches the desired states, the total cost can be minimized, and the traffic capacity is maximized. Thus, the whole system can split into m vehicular platoons as $x = [x_1, x_2, \cdots, x_m]^T$. In particular, each vehicle platoon has multiple vehicles with a CAV at the tail, and other vehicles are human-driven. For the given j-th vehicular platoon, there are total n_j vehicles that we have $x_j = [\tilde{h}_1, \tilde{v}_1, \cdots, \tilde{h}_{n_j}, \tilde{v}_{n_j}]^T$.

In this case, for the decentralised control system, the optimization problem in (9) can be equivalent to a noncooperative control game for each vehicle platoon as

$$\min J_{j,N} = x_{j,N}^T Q_j x_{j,N} + \sum_{k=0}^{N-1} \left(x_{j,k}^T Q_j x_{j,k} + u_{j,k}^T R_j u_{j,k} \right),$$

$$s.t. x_{j,k+1} = A_{j,k} x_{j,k} + B_{j,k}^1 u_{j,k} + B_{j,k}^2 u_{j,k-1}, \tag{10}$$

where for CAV j, $j = 1, 2, \cdots, m$; $A_{j,k}$, $B_{j,k}^1$, and $B_{j,k}^2$ are the $2n_j \times 2n_j$, $2n_j \times 1$, and $2n_j \times 1$ coefficient matrices as parts of the original $2n \times 2n$ matrix A_k, $2n \times 1$ matrix B_k^1, and $2n \times 1$ matrix B_k^2 in (7), respectively; the control signal $u_j(t)$ is for the j-th CAV; and Q_j and R_j are the corresponding symmetric positive definite weight matrices.

In order to solve the optimal control strategy $u_{j,k}$, we define a new state vector $z_{j,k} = [x_{j,k}, u_{j,k-1}]^T$, and we have

$$z_{j,k+1} = F_{j,k} z_{j,k} + G_{j,k} u_{j,k}, \tag{11}$$

where the coefficient matrices are

$$F_{j,k} = \begin{bmatrix} A_{j,k} & B_{j,k}^2 \\ 0 & 0 \end{bmatrix},$$

$$G_{j,k} = \begin{bmatrix} B_{j,k}^1 \\ I \end{bmatrix}. \tag{12}$$

Then, by using z_k, we can rewrite the optimization problem in (10) as

$$\min J_{j,N} = z_{j,N}^T \bar{Q}_j z_{j,N} + \sum_{k=0}^{N-1} \left(z_{j,k}^T \bar{Q}_j z_{j,k} + u_{j,k}^T R_j u_{j,k} \right),$$

$$s.t. z_{j,k+1} = F_{j,k} z_{j,k} + G_{j,k} u_{j,k}, \tag{13}$$

where

$$\bar{Q}_j = \begin{bmatrix} Q_j & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{14}$$

Then, as a part of cost function $J_{j,N}$, the residual cost function is defined as

$$V_{j,k} = z_{j,N}^T \bar{Q}_j z_{j,N} + \sum_{i=k}^{N-1} \begin{bmatrix} z_{j,i} \\ u_{j,i} \end{bmatrix}^T \begin{bmatrix} \bar{Q}_j & 0 \\ 0 & R_j \end{bmatrix} \begin{bmatrix} z_{j,i} \\ u_{j,i} \end{bmatrix}. \tag{15}$$

Based on our previous works [44], the residual cost function can be derived as a quadratic function of platoon states that

$$V_{j,k} = z_{j,k}^T H_{j,k} z_{j,k}. \tag{16}$$

Then, the optimal control strategy is derived as

$$u_{j,k} = -W_{j,k} z_{j,k}, \tag{17}$$

where

$$W_{j,k} = \left[G_{j,k}^T H_{j,k+1} G_{j,k} + R_j \right]^{-1} G_{j,k}^T H_{j,k+1} F_{j,k},$$

$$H_{j,k} = F_{j,k}^T H_{j,k+1} F_{j,k} + \bar{Q}_j - W_{j,k}^T G_{j,k}^T H_{j,k+1} F_{j,k}, H_{j,N} = \bar{Q}_j. \tag{18}$$

The optimal controller design can be summarized as in Algorithm 1. The message flows for the control system are shown in Figure 5. In particular, numerous RSUs are seen as multiple remote controllers in the WNCS model. CAVs equipped with sensors and actuators connect to RSUs through wireless link. This distributed and collaborate design promotes the system reliability and stability. Fog nodes deal with real-time responsiveness while the cloud server focuses on delay-tolerate data processing and storage.

The control scheme is designed in the following steps. First, at each sampling interval, the state information of the whole platoon is captured by vehicles and uploaded to the RSUs via V2I communication. The RSUs are responsible for local information processing and filtering. To enhance the capacity and efficiency of the system, the selected traffic flow information is transmitted to the cloud server. Based

Off-line:

1:   Initialize $H_{j,N} = \bar{Q}_j, j \in [1, m]$

2:   **for** $k = N - 1 : -1 : 0$ do.

3:       Calculate $W_{j,k}$ using

$W_{j,k} = [G_{j,k}^T H_{j,k+1} G_{j,k} + R_j]^{-1} G_{j,k}^T H_{j,k+1} F_{j,k}$

4:       Calculate $S_{j,k}$ using

$H_{j,k} = F_{j,k}^T H_{j,k+1} F_{j,k} + \bar{Q}_j - W_{j,k}^T G_{j,k}^T H_{j,k+1} F_{j,k}.$

5:   **end for**

On-line:

1:   Determine the desired states h^*, v^*

2:   Initialize $x_j(0), u_{j,k} = 0, k \le 0$

3:   **for** $=k = 0 : 1 : N - 1$ **do**

4:       Update $z_{j,k} = [x_{j,k}, u_{j,k-1}]^T$

5:       Calculate $u_{j,k}$ using $u_{j,k} = -W_{j,k} z_{j,k}$

6:   **end for**

ALGORITHM 1. Optimal CCC strategy.

on the historical and real-time data with global insight, the cloud employs a powerful AI to learn the generic trends of traffic flow in the near future. In the cloud, the desired speed of vehicles can be determined to realize dynamic traffic flow management. The computing results in the cloud are then downloaded to the RSUs. Then, the vehicle platoon model and the optimal control problem are formulated in the RSUs. The off-line part of the control strategy proposed in Algorithm 1, namely, the coefficient $W_{j,k}$, can be calculated by the RSU which is then transmitted to the CAV. Finally, the real-time control signal is computed by the on-board controller based on the coefficient $W_{j,k}$ and the real-time measurements of vehicle's states, namely, the on-line part of Algorithm 1, to regulate vehicle's longitudinal motion. In the next time period, the updated vehicle states are fed back to the RSU to realize the close-loop control.

3.2. Simulations and Results. In this subsection, we provide simulations of the connected vehicle system to study the performance of the proposed control scheme. Here, the typical scenario of the three-vehicle platoon is considered [31, 43], which can be easily extended to more vehicle scenarios. In particular, each platoon consists of 2 normal human-driven vehicles and a single CAV at the tail. In addition, in order to validate the performance of the distributed control, 3 distributed platoons with different initial states are considered in the simulation, which can also be easily extended to more complex distributed control system. For the first platoon, the initial states for the CAV are set to be $h(0) = 12$ [m] and $v(0) = 5$ [m/s]. For the second platoon, the initial states for the CAV are set to be $h(0) = 15$ [m] and $v(0) = 8$ [m/s]. For the last platoon, the initial states for the CAV are set to be $h(0) = 25$ [m] and $v(0) = 20$ [m/s]. The global desired velocity for the traffic flow is set as $v^* = 15$ [m/s], and the desired intervehicle distance is set to be $h^* = 20$ [m]. The sampling periods and communication delays are different in each platoon, which are set as $T = 0.2s, \tau = 0.1s$; $T = 0.2s, \tau = 0.3s$;

and $T = 0.4s, \tau = 0.3s$ to verify the effectiveness of the proposed algorithm when the communication topology changes.

Figures 6 and 7 show the headway h and velocity trajectories v of the CAVs in 3 different vehicle platoons, respectively. It can be seen that the CAVs' velocity and headway are tracking the desired states in the presence of various time delays, and the remained state errors are close to almost zero. Figure 8 shows the acceleration u of the CAVs, all of which go down to near zero within about 5 seconds. The results depict that all 3 CAVs gradually reach the same desired states with the proposed control strategy. Therefore, all of the 3 platoons can simultaneously track the desired velocity no matter what the initial states are. This indicates that the proposed network architecture and control algorithm can regulate the uniform traffic flow with multiple distributed controlled vehicle platoons.

4. Future Research Challenges and Open Issues

Current research has achieved great success in the design of low latency and high efficiency network system, but many of the technologies are not fully resolved in practical applications. In the following, we conclude the most talked key research directions and challenging issues to be addressed.

4.1. Tradeoff between Communication and Control. Based on the analysis of the critical system variables involved with wireless communication, it is obvious that the condition of the communication network directly influences the control performance. In the existing research of WNCS, due to the complexity of the control system itself, researchers often separate the design of communication network and control strategies independently. However, in practical applications, the network characteristic tends to change dynamically, and the corresponding control strategies also need to be changed. If the network resource management and control strategy optimization can be jointly designed, the performance and the system stability can be further improved. In the independent design from the control aspect, control strategies are normally carried out with certain communication requirements, which results in the loss of serviceability in other environments. Even the communication system can satisfy the control requirements, predefined communication variables may result in the waste of wireless resources with low system efficiency. Therefore, it is essential to jointly design more integrated control algorithms that consider both communication and control so that the whole system can be compatible with various delay-sensitive and safe-critical applications. However, in order to realize the integration of control systems with communication, multiple interaction and coordination among system components should be reevaluated. For example, higher information rate brings better control performance, which at the same time leads to traffic congestion and increases computational burden. In the view of resource allocation of communication networks, since the wireless resource consumption observably increases with growing QoS requirements, the joint design approach may use various communication techniques to satisfy different QoS requirements. In the case study, we consider the
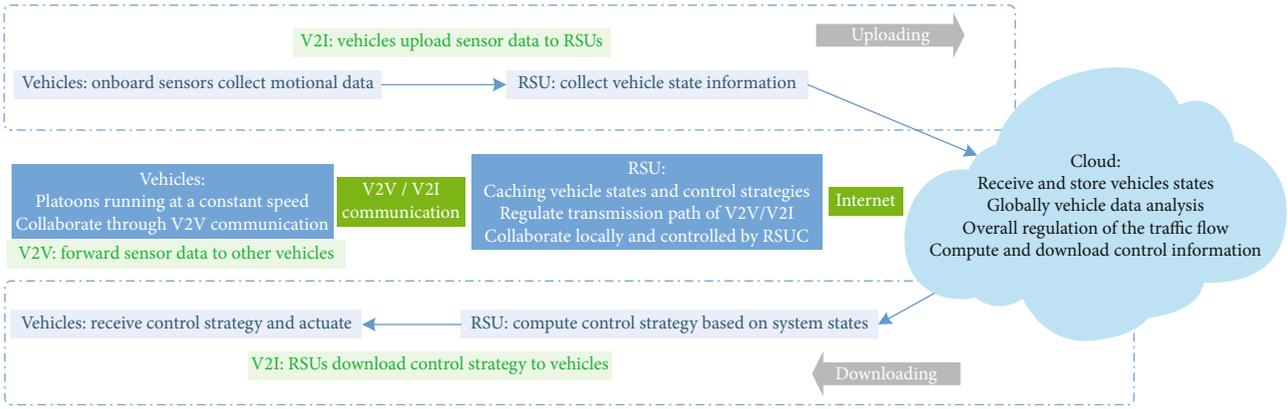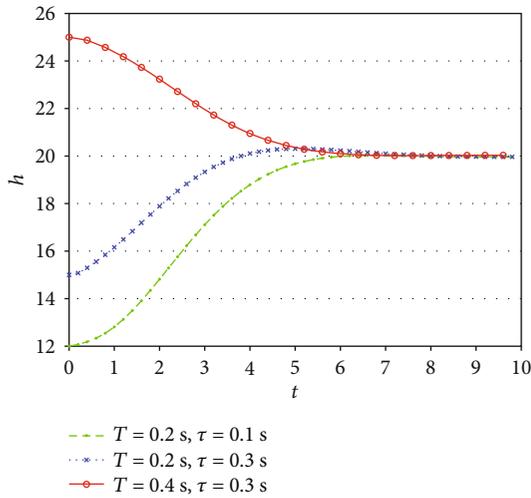
Figure 5: The procedure of CCC.

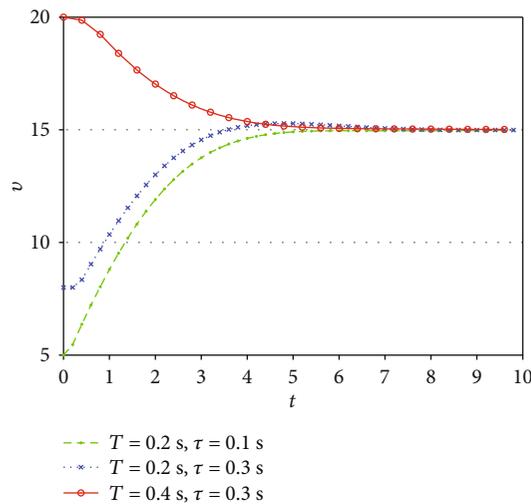Figure 6: Headway of the CAVs in different platoons.

Figure 7: Velocity of the CAVs in different platoons.

network-induced delay and packet dropout in the design of the control strategy. Further research is needed in the controller design with other consideration such as transmission path planning, energy consumption, and resource allocation. In addition, the dynamic nature of the vehicles results in the time-varying network topology. In this case, it is difficult to maintain the system stability, and the controller design remains challenging.

4.2. Security and Privacy Problems. Security issues should be carefully considered in the whole system, including radio access, data transmission, and controller design. However, current solutions based on classic secrecy and authentication methods do not meet the needs of real-time control applications. Existing centralized security protection protocols might not be applicable in the distributed network system and may bring significant delay in delay-sensitive applications. Therefore, it is a great challenge to achieve both high security and control performance under limited resources. In the connected vehicle scenario, the sensor-based and V2X-enabled automated vehicles rely heavily on electronic equipments with sensitive personal information needed to be kept secure, such as the position data, social network information, and multimedia records. In this case, some specific procedures need to be taken into consideration to avoid potential cybersecurity attacks [45]. Therefore, information security and privacy problems need further investigation.

4.3. Application Considerations. In the whole network, fog not only acts as middle-ware between the cloud and edge but also provides quick and efficient control services. Depending on the type of service, the fog servers take the responsibility of deciding the kind of tasks to be processed locally and that to be offloaded to the cloud [12]. In particular, fog has relatively weaker storage and computation resources than the cloud, so whether a certain application's data are stored locally or transmitted to the cloud becomes a problem. In addition, what kind of network structure is suitable for what type of applications, what kind of communication technology will be used between fogs, and how cooperation will be realized between fogs are major research issues that need further investigation. Different use cases
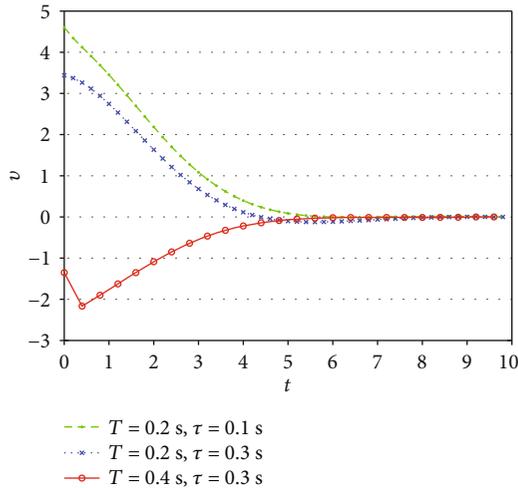
FIGURE 8: Acceleration of the CAVs in different platoons.

come with diverse requirements. How to customize the network to provide corresponding network resources with specific control applications remains challenging.

With the removal of wiring in WNCS, permanently power supply is often removed, so additional limitations on the energy consumption of the system components arise. The tradeoff between control performance and network lifetime should be considered thoroughly. More fog nodes mean more energy consumption, less computing resource result in longer time delay. How to address the tradeoff in energy consumption and computation in order to improve the performance of green communications also needs further research.

5. Conclusion

In this paper, we propose a novel fog-based distributed network architecture toward real-time control applications. Three logical layers, the cloud, distributed fog, and edge smart devices, are considered. We discuss the critical system variables of communication and control systems, including sampling period, message delay, message dropout, and noisy channel. Fog enriches the network resources making real-time control possible. A case study of CCC is introduced to demonstrate the feasibility of the proposed architecture in the field of vehicular networks. The optimal control design for the CAV is proposed to realize uniform traffic flow. Simulations indicate that the performances are improved with the low latency and real-time control capabilities. Finally, we discuss some future research directions and open issues to be addressed.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

References

[1] O. Bello and S. Zeadally, "Intelligent device-to-device communication in the internet of things," *IEEE Systems Journal*, vol. 10, no. 3, pp. 1172–1182, 2016.

[2] L. Zhu, F. R. Yu, B. Ning, and T. Tang, "Cross-layer handoff design in mimo-enabled wlans for communication-based train control (cbtc) systems," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 719–728, 2012.

[3] X. Ge, F. Yang, and Q.-L. Han, "Distributed networked control systems: a brief overview," *Information Sciences*, vol. 380, pp. 117–131, 2017.

[4] A. W. Al-Dabbagh and T. Chen, "Design considerations for wireless networked control systems," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 9, pp. 5547–5557, 2016.

[5] P. Park, S. Coleri Ergen, C. Fischione, C. Lu, and K. H. Johansson, "Wireless network design for control systems: a survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 978–1013, 2018.

[6] G. Zhao, M. A. Imran, Z. Pang, Z. Chen, and L. Li, "Toward real-time control in future wireless networks: communication-control co-design," *IEEE Communications Magazine*, vol. 57, no. 2, pp. 138–144, 2019.

[7] J. Wang, J. Liu, and N. Kato, "Networking and communications in autonomous driving: a survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1243–1274, 2019.

[8] Z. Wang, X. Wang, L. Liu, M. Huang, and Y. Zhang, "Decentralized feedback control for wireless sensor and actuator networks with multiple controllers," *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 5, pp. 1471–1483, 2017.

[9] Z. Wang, L. Liu, and Z. Wang, "Stochastic optimal linear control of wireless networked control systems with delays and packet losses," *IET Control Theory & Applications*, vol. 10, no. 7, pp. 742–751, 2016.

[10] L. Zhu, Y. Li, F. R. Yu, B. Ning, T. Tang, and X. Wang, "Cross-layer defense methods for jamming-resistant cbtc systems," *IEEE Transactions on Intelligent Transportation Systems (Early Access)*, pp. 1–13, 2020.

[11] T. X. Tran, A. Hajisami, P. Pandey, and D. Pompili, "Collaborative mobile edge computing in 5G networks: new paradigms, scenarios, and challenges," 2016, https://arxiv.org/abs/1612.03184.

[12] M. Aazam, S. Zeadally, and K. A. Harras, "Fog computing architecture, evaluation, and future research directions," *IEEE Communications Magazine*, vol. 56, no. 5, pp. 46–52, 2018.

[13] Y. Liu, J. E. Fieldsend, and G. Min, "A framework of fog computing: architecture, challenges, and optimization," *IEEE Access*, vol. 5, pp. 25445–25454, 2017.

[14] M. Li, P. Si, and Y. Zhang, "Delay-tolerant data traffic to software-defined vehicular networks with mobile edge computing in smart city," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 9073–9086, 2018.

[15] S.-C. Hung, H. Hsu, S.-Y. Lien, and K.-C. Chen, "Architecture harmonization between cloud radio access networks and fog networks," *IEEE Access*, vol. 3, pp. 3019–3034, 2015.

[16] L. Zhou, L. Yu, S. Du, H. Zhu, and C. Chen, "Achieving differentially private location privacy in edge-assistant connected vehicles," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4472–4481, 2019.

[17] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, "Three decades of driver assistance systems: review and future perspectives," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 4, pp. 6–22, 2014.

[18] L. Zhu, Y. He, F. R. Yu, B. Ning, T. Tang, and N. Zhao, "Communication-based train control system performance optimization using deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10705–10717, 2017.

[19] H. F. Atlam, R. J. Walters, and G. B. Wills, "Fog computing and the Internet of Things: a review," *Big Data and Cognitive Computing*, vol. 2, no. 2, pp. 10–18, 2018.

[20] J. Liu, J. Wan, D. Jia et al., "High-efficiency urban traffic management in context-aware computing and 5g communication," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 34–40, 2017.

[21] 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*, vol. 19, no. 8, pp. 2391–2406, 2018.

[22] Z. Wang, Y. Gao, C. Fang, L. Liu, S. Guo, and P. Li, "Optimal connected cruise control with arbitrary communication delays," *IEEE Systems Journal*, vol. 14, no. 2, pp. 2913–2924, 2020.

[23] X. Wang, Z. Ning, X. Hu et al., "A city-wide real-time traffic management system: enabling crowdsensing in social internet of vehicles," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 19–25, 2018.

[24] E. Baccarelli, P. G. V. Naranjo, M. Scarpiniti, M. Shojafar, and J. H. Abawajy, "Fog of everything: energy-efficient networked computing architectures, research challenges, and a case study," *IEEE Access*, vol. 5, pp. 9882–9910, 2017.

[25] T. S. Darwish and K. Abu Bakar, "Fog based intelligent transportation big data analytics in the internet of vehicles environment: motivations, architecture, challenges, and critical issues," *IEEE Access*, vol. 6, pp. 15679–15701, 2018.

[26] W. Zhang, Z. Zhang, and H.-C. Chao, "Cooperative fog computing for dealing with big data in the internet of vehicles: architecture and hierarchical resource management," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 60–67, 2017.

[27] C. Bila, F. Sivrikaya, M. A. Khan, and S. Albayrak, "Vehicles of the future: a survey of research on safety issues," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 5, pp. 1046–1065, 2017.

[28] M. Mazzola, G. Schaaf, A. Stamm, and T. Kurner, "Safety-critical driver assistance over LTE: toward centralized ACC," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 9471–9478, 2016.

[29] T. Zeng, O. Semiari, W. Saad, and M. Bennis, "Integrated communications and control co-design for wireless vehicular platoon systems," in *2018 IEEE International Conference on Communications (ICC)*, pp. 1–6, Kansas City, MO, USA, 2018.

[30] L. Zhang and G. Orosz, "Beyond-line-of-sight identification by using vehicle-to-vehicle communication," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 6, pp. 1962–1972, 2018.

[31] G. Orosz, "Connected cruise control: modelling, delay effects, and nonlinear behaviour," *Vehicle System Dynamics*, vol. 54, no. 8, pp. 1147–1176, 2016.

[32] T. Zhang, Y. Zou, X. Zhang, N. Guo, and W. Wang, "A cruise control method for connected vehicle systems considering side vehicles merging behavior," *IEEE Access*, vol. 7, pp. 6922–6936, 2019.

[33] D. Jia, K. Lu, J. Wang, X. Zhang, and X. Shen, "A survey on platoon-based vehicular cyber-physical systems," *IEEE Communication Surveys and Tutorials*, vol. 18, no. 1, pp. 263–284, 2016.

[34] G. P. Fettweis, "The tactile internet: applications and challenges," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 64–70, 2014.

[35] J. Wan, D. Zhang, S. Zhao, L. T. Yang, and J. Lloret, "Context-aware vehicular cyber-physical systems with cloud support: architecture, challenges, and solutions," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 106–113, 2014.

[36] P. Patil, A. Hakiri, and A. Gokhale, "Cyber foraging and off-loading framework for internet of things," in *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, vol. 1, pp. 359–368, Atlanta, GA, USA, 2016.

[37] C. Yu, B. Lin, P. Guo, W. Zhang, S. Li, and R. He, "Deployment and dimensioning of fog computing-based internet of vehicle infrastructure for autonomous driving," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 149–160, 2019.

[38] Q. Fan and N. Ansari, "Towards workload balancing in fog computing empowered IoT," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 253–262, 2020.

[39] S. Zhang, J. Chen, F. Lyu, N. Cheng, W. Shi, and X. Shen, "Vehicular communication networks in the automated driving era," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 26–32, 2018.

[40] X. Ge, H. Cheng, G. Mao, Y. Yang, and S. Tu, "Vehicular communications for 5G cooperative small-cell networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 7882–7894, 2016.

[41] A. Vinel, L. Lan, and N. Lyamin, "Vehicle-to-vehicle communication in C-ACC/platooning scenarios," *IEEE Communications Magazine*, vol. 53, no. 8, pp. 192–197, 2015.

[42] J. Guanetti, Y. Kim, and F. Borrelli, "Control of connected and automated vehicles: state of the art and future challenges," *Annual reviews in control*, vol. 45, pp. 18–40, 2018.

[43] M. Huang, W. Gao, and Z.-P. Jiang, "Connected cruise control with delayed feedback and disturbance: an adaptive dynamic programming approach," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 2, pp. 356–370, 2019.

[44] M. Huang, L. Liu, X. Wang, and Z. Wang, "Optimal state feedback control for wireless networked control systems with decentralised controllers," *IET Control Theory & Applications*, vol. 9, no. 6, pp. 852–862, 2015.

[45] M. Amoozadeh, A. Raghuramu, C. N. Chuah et al., "Security vulnerabilities of connected vehicle streams and their impact on cooperative driving," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 126–132, 2015.

WILEY | Hindawi

Research Article

Joint Task Offloading and Resource Allocation Strategy for DiffServ in Vehicular Cloud System

Ya Kang [iD], Zhanjun Liu, Qianbin Chen, and Yingdi Dai

School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China

Correspondence should be addressed to Ya Kang; 417184514@qq.com

A vehicular cloud (VC) can reduce latency and improve resource utilization of the Internet of vehicles by effectively using the underutilized computing resources of nearby vehicles. Although the task offloading of the VC enhances road safety and traffic management on the Internet of vehicles and meets the low-latency requirements for driving safety services on the Internet of vehicles business, there are still some key challenges such as the resource allocation mechanism of differentiated services (DiffServ) and task offloading mechanism of improving user experience. To address these issues, we study the task offloading and resource allocation strategy of the VC system where tasks generated by vehicles can be offloaded and executed cooperatively by vehicles in VC. Specifically, the computing task is further divided into independent subtasks and executed in different vehicles in VC to maximize the offloading utility. Considering the mobility of vehicles, the deadline of tasks, and the limited computing resources, we propose the optimization problem of task offloading in the VC system in the cause of improved user experience. To characterize the difference in service requirements resulting from the diversity of tasks, a DiffServ model focusing on the pricing of a task is utilized. The initial pricing of a task is tailored by the characteristics of the task and the uniqueness of the network status. In this model, tasks are sorted and processed in order according to task pricing, so as to optimize resource allocation. Numerical results show that the proposed scheme can effectively increase the resource utilization and task completion ratio.

1. Introduction

In a vehicle network [1], the cloud composed of vehicles with strong computing power is called VC [2]. The task offloading of the VC [3] has a significant impact on the performance of the Internet of vehicles [4], which enhances vehicle road safety and traffic management by effectively utilizing the underutilized computation resources of nearby vehicles and reduces traffic congestion and accidents, thus improving the network performance of the Internet of vehicles. The processing of emerging message-type safety applications in the vehicle network requires a large amount of computing resources, such as $4\,K/8\,K$ ultrahigh real-time traffic video streaming, multiview video stitching processing of autonomous vehicles, and AR-assisted applications related to safety warnings. However, the collaborative scheduling of information and data by offloading computing tasks to the vehicular cloud can effectively enhance the utilization of computing

resources due to the limited computing resources of vehicles. At the same time, due to its proximity to users and rich computing resources, the vehicular cloud can meet the low delay requirements for safe driving on the Internet of vehicles business, which is a key technology in the existing Internet of vehicles.

Owing to the fast-moving characteristics of vehicles, computing resources of vehicles are dynamic. For example, the connection of mobile vehicles can be connected or disconnected at any time, which will bring about unstable computing resource availability. Accordingly, in order to ensure the offloading of the task, the task should be offloaded within vehicle contact time [5]. In vehicular cloud computing, task offloading is inseparable from efficient resource allocation technology. Resources involved in the entire process of offloading the task of the vehicle to the vehicular cloud are mainly divided into the communication resources and the computing resources [6]. [7–9]. For the task offloading of
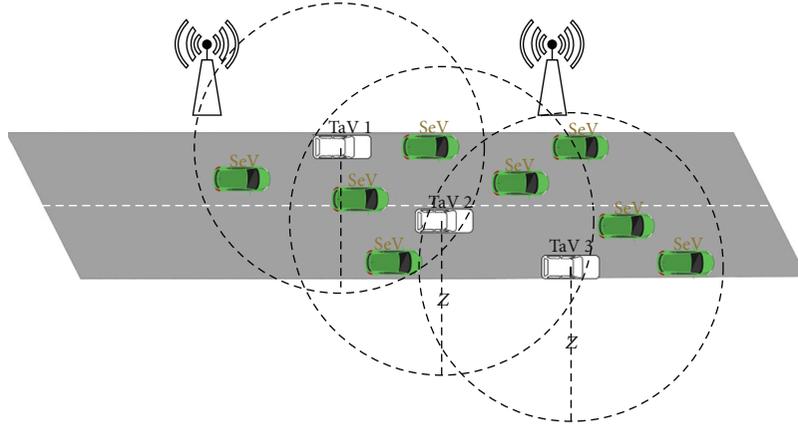
Figure 1: Vehicular cloud system model.

VC, Wu et al. assume that each vehicle task can be divided into several independent subtasks, and then, the task allocation problem is solved [10]. Under the condition of ensuring user service experience, Liu et al. study how to allocate limited computing resources to maximize the economic benefits of MEC service providers when vehicles offload tasks to MEC servers [11]. In order to solve the users' selfishness, Liu et al. have established a kind of new type of computing sharing market in the vehicle network [12]. Several other works, such as [13, 14], use the method of task replication to allow a task to be executed by multiple vehicles so as to maximize the possibility of completing the task before a given deadline and improve the reliability.

In recent years, game theory has become an incentive mechanism to analyze and allocate resources. A game model can effectively capture the interaction between users and servers [15], so game theory has become a way to analyze and allocate resources [16]. For the sake of creating greater social and economic benefits in the process of resource management, a pricing strategy as an effective way of network resource management on the basis of game theory is proposed, in [16]. The forementioned research believes that the pricing mechanism is an effective way to improve resource utilization. However, how to use the pricing mechanism to distinguish different services in the VC system in game theory has not been well explored. It also can be seen from the above literatures that most of the resource allocation process VC system does not DiffServ. That is to say, the relationship between default services is equal. Unlike the resource management [17] scheme that considers the priority of a task in the general VC computing scenario, we propose a game theory-based DiffServ resource allocation mechanism that DiffServ through a task's differentiated pricing. In other words, the higher the initial pricing of the task, the higher the service requirements of the business. The initial pricing of a task mainly involves two factors, namely, the urgency of the task vehicle and the network condition around the task. Finally, we maximize the offloading utility of tasks in the VC system and consider the mobility of vehicles and meet task delay constraints to maximize the completion of tasks. The main contributions of our paper are summarized as follows.

(i) To meet the requirements of the diversity of tasks while effectively using computing resources, we propose a resource allocation mechanism for differentiated services and offload each task according to the pricing order

(ii) To improve the completion ratio of whole tasks in the vehicular cloud system, a task offloading mechanism that maximizes the offloading utility is proposed and considers the mobility of vehicles and the deadline of tasks

(iii) To address the problem of task offloading and resource allocation, a distributed resource allocation algorithm and Lagrange dual method are used, respectively

2. System Model

We consider a dynamic traffic scenario in which the vehicle can be either a service requester or a service provider in a finite two-way straight road. Let $i \in (1, 2, \cdots, M)$ be the group of M TaVs (task vehicles). We use three parameters (d_i, c_i, T_i^{deadline}) to indicate the generated task of TaV i, by which d_i (in bits) specifies the data size of the task, where $c_i = \kappa d_i$, c_i (in CPU cycles per bit) is the computing resource required to processing the task, κ is the mapping between CPU cycles and task size, and T_i^{deadline} is the deadline for task execution. When the task processing delay exceeds the deadline, the task processing fails. f_{i0} indicates the local computing ability of TaV i. At the same time, $j \in (i, 2, \cdots, N)$ is denoted as a group of vehicles with free resources, which is called SeVs (service vehicles), where the computing resource of SeV j is defined as F_j^{max}. Figure 1 depicts a simple example of a vehicular cloud system. Here, TaV 1, TaV 2, and TaV 3 form a vehicular cloud with SeVs within their covering radius, respectively. It is assumed that the communication radius of each vehicle is Z.

2.1. Task Offloading Model. Since the size of the output task is usually much smaller than the size of the input task, we can ignore the return time in offloading [18, 19]. Each TaV first offloads the computing task to one or more surrounding

SeVs for processing and then handles the size of the remaining task by itself to ensure that the task is completed within the deadline. The total offloading delay for SeV j to process the size of a task assigned by TaV i is given as

$$T_{ij} = \frac{d_{ij}}{R_{ij}} + \frac{c_{ij}}{f_{ij}}, \tag{1}$$

where d_{ij} is the task size of the TaV i offloading to SeV j, R_{ij} is the transmission rate of TaV i to SeV j, c_{ij} is the required computation resource to complete task d_{ij}, and f_{ij} is the computing resource provided by SeV j to TaV i.

2.2. Local Processing Model. When the task is not completely offloaded, the size of the remaining task is processed locally, and the local processing delay is expressed as follows:

$$T_{i0} = \frac{c_i - \sum_{j=1}^N c_{ij}}{f_{i0}}. \tag{2}$$

2.3. Mobility Model. Assuming that the initial position of TaV i and the initial position of SeV j is, respectively, $(x_{\text{TaV}_i}, y_{TaV_i})$ and (x_{SeV_j}, y_{SeV_j}), the speed of each vehicle is defined as $v_{\text{TaV}_i}, v_{\text{SeV}_j} \in [v_{\min}, v_{\max}]$ that the vehicle maintains a uniform linear motion during task processing, so the motion angle $\theta = \{0, \pi\}$, after moving the time slot Δt, the position of the TaV i is changed to $x_{\text{TaV}_i}^{\Delta t} = x_{\text{TaV}_i} + v_{\text{TaV}_i} \Delta t \cos \theta$, $y_{\text{TaV}_i}^{\Delta t} = y_{\text{TaV}_i} + v_{\text{TaV}_i} \Delta t \sin \theta$. The position of SeV j is changed to $x_{\text{SeV}_j}^{\Delta t} = x_{\text{SeV}_j} + v_{\text{SeV}_j} \Delta t \cos \theta$, $y_{\text{SeV}_j}^{\Delta t} = y_{\text{SeV}_j} + v_{\text{SeV}_j} \Delta t \sin \theta$. According to Euclid's theorem, after TaV i and SeV j move through the time slot Δt, the distance between the two vehicles is

$$s = \sqrt{\left(x_{\text{TaV}_i}^{\Delta t} - x_{\text{SeV}_j}^{\Delta t}\right)^2 + \left(y_{\text{TaV}_i}^{\Delta t} - y_{\text{SeV}_j}^{\Delta t}\right)^2}. \tag{3}$$

Communication between each other can be achieved only when SeV j is within the communication radius Z of TaV i. The connection time between vehicles is predicted as follows:

$$Z^2 = \left(x_{\text{TaV}_i}^{T_{ij}^{\text{re}}} - x_{\text{SeV}_j}^{T_{ij}^{\text{re}}}\right)^2 + \left(y_{\text{TaV}_i}^{T_{ij}^{\text{re}}} - y_{\text{SeV}_j}^{T_{ij}^{\text{re}}}\right)^2. \tag{4}$$

2.4. Offloading Utility Function. Considering all the TaVs in the VC system, the total offloading utility obtained of offloading TaVs to SeVs can be defined as

$$U = \sum_{i=1}^M (u_i - C_i), \tag{5}$$

where u_i is denoted as the offloading revenue of TaV i and C_i is the offload cost of the TaV i.

The total offloading revenue obtained by TaV i for offloading tasks to the VC is expressed as follows [20]:

$$u_i = \alpha_i \sum_{j=1}^N \log_2 (1 + d_{ij}), \tag{6}$$

where α_i represents the task offloading revenue weight parameter of TaV j, and its value is greater than zero.

The transmission cost required by TaV i to offload tasks to the VC is expressed as follows:

$$C_i = \chi \sum_{j=1}^N d_{ij}, \tag{7}$$

where χ is the unit transmission cost required for transmiting 1 Mb data to any SeV.

With the aim of maximizing the task offloading of TaVs in the VC system, the following optimization is formulated P1:

$$\max_{d_{ij}} U$$

$$\text{s.t.} \quad \text{C1}: T_{ij} \leq T_i^{\text{deadline}}, \forall i \in M, \forall j \in N$$

$$\text{C2}: \sum_{j=1}^N d_{ij} \leq d_i, \forall i \in M, \forall j \in N \tag{8}$$

$$\text{C3}: d_{ij} \geq 0, \forall i \in M, \forall j \in N$$

$$\text{C4}: T_{ij} \leq T_{ij}^{\text{re}}, \forall i \in M, \forall j \in N.$$

C1 means that the processing delay of each subtask is limited by the deadline of the task. C2 and C3 ensure that the size of total tasks offloading is not greater than the generated task size and the size of offloading is positive. C4 indicates that TaV i and SeV j should not spend more time on task processing than the connection time between the two vehicles.

3. Computing Resource Allocation Based on Game Theory

For the sake of enabling SeVs to share their own resources to others, we propose an incentive method which is based on game theory as well as introduce a paid resource allocation strategy. The offloading delay for TaV i to offload the task to SeV j in the process of task offloading is as follows:

$$T_{ij} = \frac{d_i}{R_{ij}} + \frac{c_i}{f_{ij}}. \tag{9}$$

With the purpose of simplifying the problem, we will consider R_{ij} as a fixed value. According to (9), the delay for task offloading is affected by computing resources provided by SeV j to TaV i. It is impossible for SeV j to provide computing resources to TaV i for free because of the selfishness of vehicles. According to messages exchanged between vehicles,

we design a distributed incentive mechanism for resource allocation. The specific details are shown as follows.

3.1. Utility and Cost Functions

(1) With the aim of maximizing utility of SeV j, the following optimization problem is formulated:

$$\max_{f_{ij}} \quad V_j = \delta_j \ln\left(1 + p_i f_{ij}\right) - cc_j f_{ij}$$
$$\text{s.t.} \qquad \text{C1} : f_{ij} > 0 \tag{10}$$
$$\text{C2} : f_{ij} \leq F_j^{\max},$$

where cc_j represents the unit resource cost price of SeV j; δ_j is the willingness of SeV j, with $\delta_j \geq 1$; p_i represents the bid for TaV i; C1 indicates that the resources provided by SeV j are larger than zero; and C2 indicates that the total resources provided by SeV j are not larger than the maximum resources provided by itself.

Theorem 1. *For any price given by each TaV, there must be a unique Nash equilibrium solution in the noncooperative game between SeVs.*

Proof. The first-order derivative of V_j with respect to f_{ij} is expressed as follows:

$$\frac{\partial V_j}{\partial f_{ij}} = \frac{\delta_j p_i}{1 + p_i f_{ij}} - cc_j. \tag{11}$$

The second-order derivative of V_j with respect to f_{ij} is expressed as follows:

$$\frac{\partial^2 V_j}{\partial f_{ij}^2} = -\frac{\delta_j p_i}{\left(1 + p_i f_{ij}\right)^2}. \tag{12}$$

As the values of all parameters in (12) are positive, the second-order derivative of utility function is lower than zero, namely, $\partial^2 V_j / \partial f_{ij}^2 < 0$. Therefore, the utility function of SeV j is a concave function with maximum value. The corresponding optimal solution $f_{ij} = \delta_j / cc_j - 1/p_i$ can be obtained by $\partial V_j / \partial f_{ij} = 0$, and there is a unique Nash equilibrium in the noncooperative game between SeVs given the price of the TaV.

(2) In order to maximize economic benefits, we should minimize offloading cost. With the aim of minimizing cost function of TaV i, the following optimization problem is formulated:

$$\min_{f_{ij}} \quad p_i \sum_{j=1}^{N} f_{ij}$$
$$\text{s.t.} \qquad \text{C1} : f_{ij} \geq 0 \tag{13}$$
$$\text{C2} : f_i^{\min} < \sum_{j=1}^{N} f_{ij} \leq f_i^{\max}$$
$$\text{C3} : P_{\min} \leq p_i \leq P_{\max},$$

where C2 means that the total computing resources provided by SeVs are larger than the minimum required resources $f_i^{\min}$, but it is smaller than the maximum required resources $f_i^{\max}$. C3 indicates that the bids of TaVs are larger than the lowest price $P_{\min}$, but it is smaller than the highest price $P_{\max}$.

3.2. TaV Pricing Rules.

This section defines the concept of "pricing" for tasks by their service requirements, for the purpose of solving the problem of resource allocation. The pricing function is used to classify tasks requested by TaVs and further distinguish the priority of tasks. The definition of the pricing function of the TaVs task mainly involves two factors, namely, the urgency of the task and the network conditions around the TaV. Accordingly, the pricing of the TaV is obtained by the urgency of the task and the network status around the TaV. Then, the urgency of the task is related to the size of the task generated by the TaV and the deadline of the task. The network condition around the TaV is related to the number of requests received and the number of answers received in the previous slot.

3.2.1. Distribution of the Number of Vehicles within the TaV Coverage.

Owing to the stability of the network, the network condition of the previous and the next moment is unchanged. Suppose that TaV i and SeV j are randomly distributed on the road. Then, it is known that the TaV $k(k = i)$ receives R_k requests and S_k responses at the previous time. According to the conditional probability, the distribution function of the number of TaVs m_k and the distribution function of the number of SeVs n_k in the coverage area is obtained.

(a) Distribution function of the number of TaVs m_k within the coverage of TaV k is expressed as follows:

$$P\left(\frac{m_k}{R_k}\right) = \frac{P(m_k, R_k)}{P(R_k)} = \frac{P(m_k, R_k)}{\sum_{m_k=R_k}^{\infty} P(m_k, R_k)}$$
$$= \frac{C_{m_k}^{R_k} P_1^{R_k} (1 - P_1)^{m_k - R_k}}{\sum_{m_k=R_k}^{\infty} C_{m_k}^{R_k} P_1^{R_k} (1 - P_1)^{m_k - R_k}}, \tag{14}$$

where (m_k, R_k) indicates that there are totally m_k TaVs within the coverage of TaV k, among which TaV k receives R_k requests, and P_1 indicates the probability of any TaV sending out requests.

(b) Distribution function of the number of SeVs n_k within the coverage of TaV k is expressed as follows:

$$
\begin{aligned}
P\left(\frac{n_k}{S_k}\right) &= \frac{P(n_k, S_k)}{P(S_k)} = \frac{P(n_k, S_k)}{\sum_{n_k=S_k}^{\infty} P(n_k, S_k)} \\
&= \frac{C_{n_k}^{S_k} P_2^{S_k} (1 - P_2)^{n_k - S_k}}{\sum_{n_k=S_k}^{\infty} C_{n_k}^{S_k} P_2^{S_k} (1 - P_2)^{n_k - S_k}},
\end{aligned}
\tag{15}
$$

where (n_k, S_k) means there are a total of n_k SeVs within the coverage of TaV k, among which TaV k receives S_k responses, and P_2 means the probability that any SeV will respond to a TaV after receiving it.

3.2.2. Average Number of Vehicles within the Coverage of TaV k

(a) The average number of TaVs within the coverage of TaV k is expressed as follows:

$$
E(m_k) = \sum_{m_k=R_k}^{\infty} m_k P\left(\frac{m_k}{R_k}\right)
\tag{16}
$$

(b) The average number of SeVs within the coverage of TaV k is expressed as follows:

$$
E(n_k) = \sum_{n_k=S_k}^{\infty} n_k P\left(\frac{n_k}{S_k}\right)
\tag{17}
$$

3.2.3. Remaining Resources within the Coverage of TaV k. Suppose that the resource requirements of each TaV obey the normal distribution of $Q_1 \sim (\mu_1, \delta_1^2)$ and the resources provided by each SeV obey the normal distribution of $Q_2 \sim (\mu_2, \delta_2^2)$, the remaining resources around TaV k obtained by (16) and (17) are expressed as follows:

$$
Q_k = E(n_k)Q_2 - E(m_k)Q_1.
\tag{18}
$$

On account of the independence of Q_1 and Q_2, we can obtain the mean and variance value of Q_k:

$$
E(Q_k) = E(n_k)\mu_2 - E(m_k)\mu_1,
\tag{19}
$$

$$
D(Q_k) = E(n_k)^2 \delta_2^2 - E(m_k)^2 \delta_1^2.
\tag{20}
$$

According to (19) and (20), we can get the distribution of the remaining resources Q_k. In other words, a new normal distribution is expressed as follows:

$$
F(Q_k) = \frac{1}{\sqrt{2\pi D(Q_k)}} \int_{-\infty}^{Q_k} e^{-\left((t - E(Q_k))^2 / 2D(Q_k)\right)}.
\tag{21}
$$

3.2.4. Initial Pricing of TaV k. According to (21), the remaining resources covered by TaV k are expressed as follows:

$$
E(Q_k) = E(n_k)\mu_2 - E(m_k)\mu_1.
\tag{22}
$$

The initial pricing of TaV k [21] is

$$
P_k = \begin{cases} \lambda_k(P_{\mathrm{mid}} - P_{\mathrm{min}}) + P_{\mathrm{min}}, & E(Q_k) \geq 0, \\ -\lambda_k(P_{\mathrm{max}} - P_{\mathrm{mid}}) + P_{\mathrm{mid}}, & E(Q_k) < 0, \end{cases}
\tag{23}
$$

$$
\lambda_k = \frac{2}{\pi} \arctan\left(\frac{f_k^{\mathrm{min}}}{E(Q_k)}\right),
\tag{24}
$$

where f_k^{min} means the minimum resources obtained by TaV k, $f_k^{\mathrm{min}} = d_k / T_k^{\mathrm{deadline}}$. From the pricing function, we can see that when the surrounding remaining resources are greater than zero, the price will be raised by the lowest price; otherwise, the price will be raised by the middle price.

After getting the best initial pricing, we develop a distributed resource allocation algorithm. The algorithm dynamically changes the price of the TaV to ensure the maximum utility of the SeV, and the TaV can obtain resources to meet its needs. The algorithm is introduced in detail in Algorithm 1. Firstly, the TaV sends the business request (including the price); meanwhile, all vehicles within the coverage of the TaV will receive the request message. Then, the SeV that received the request message maximizes its utility to obtain the size of computing resources that can be provided (Steps 4-14 of Algorithm 1). Next, the SeV sends a response (resources that the SeV can be provided) to the TaV; meanwhile, the TaV receives the response and adds up resources provided by SeVs. Finally, if the sum of the total resources provided is greater than the maximum resource demand of the TaV, the price will be reduced; otherwise, the price will be increased if the total resources provided are less than the minimum resource demand of the TaV. Otherwise, the price is the optimal price, and the optimal resource allocation result of the SeV is obtained (Steps 15-30 of Algorithm 1).

4. Task Allocation of Vehicular Cloud System

This section is aimed at solving the problem of task allocation of the VC system. Because the second-derivative of U for d_{ij} in equation (5) can be obtained $\partial^2 U / \partial d_{ij}^2 < 0$, the objective function of optimization problem P1 is the concave function of variable d_{ij}. Since the constraints C1, C2, C3, and C4 in P1 are all convex sets, the optimization problem P1 is a convex optimization problem. In the end, the optimal d_{ij} is obtained by Lagrange theory.

```
1:   Initialization: The cost price ccⱼ of the SeV j and the maximum resource it can provide are Fⱼᵐᵃˣ; the TaV i the required computing
     resource is fᵢᵐⁱⁿ < fᵢ ≤ fᵢᵐᵃˣ
2:   For t=0, TaV i to set the price pᵢ(t), and send it to the vehicles within the coverage
3:   for t = 1, 2, 3.···do
4:      for i = 1, 2, ···, M do
5:        for j = 1, 2, ···, N do
6:           According to ∂V/∂fᵢⱼ = 0, compute fᵢⱼ
7:           for k = 1 : i do
8:              a = a + fₖⱼ
9:           end for
10:           If a > Fⱼᵐᵃˣ then
11:              fᵢⱼ = 0
12:           end if
13:        end for
14:      end for
15:      If ∑ⱼ₌₁ᴺ fᵢⱼ ≥ fᵢᵐᵃˣ then
16:         decrease the price, pᵢ(t + 1) = pᵢ(t) − v
17:      else
18:         if ∑ⱼ₌₁ᴺ fᵢⱼ ≤ fᵢᵐⁱⁿ the
19:            increase the price, pᵢ(t + 1) = pᵢ(t) + v
20:         else
21:            pᵢ(t + 1) = pᵢ(t)
22:         end if
23:         If pᵢ(t + 1) == pᵢ(t) then
24:            break
25:         else
26:            report the new price pᵢ(t + 1)
27:            back to 3
28:         end if
29:      end if
30:   end for
```

ALGORITHM 1: Distributed resource allocation algorithm.

The Lagrangian function of optimization problem P1 is defined as follows (without considering $d_{ij} \geq 0$):

$$
L\left(d_{ij}, \varpi, \sigma, \psi\right) = U + \sum_{i=1}^{M} \sum_{j=1}^{N} \varpi_{ij} \left[T_i^{\text{deadline}} - \frac{d_{ij}}{R_{ij}} - \frac{\kappa d_{ij}}{f_{ij}} \right]
$$
$$
+ \sum_{i=1}^{M} \sigma_i \left(d_i - \sum_{j=1}^{N} d_{ij} \right)
$$
$$
+ \sum_{i=1}^{M} \sum_{j=1}^{N} \psi_{ij} \left(T_{ij}^{\text{re}} - \frac{d_{ij}}{R_{ij}} - \frac{\kappa d_{ij}}{f_{ij}} \right),
$$

(25)

where ϖ, σ, ψ are the Lagrangian multipliers.

According to the original optimization problem P1, a Lagrangian dual problem can be obtained as follows:

$$
\text{P2}: \quad \min_{\varpi_{ij}, \delta_i, \psi_{ij}} \max_{d_{ij}} L\left(d_{ij}, \varpi, \delta, \psi\right). \tag{26}
$$

For the optimization problem P2, it can be decomposed into a maximum task assignment problem and a minimum

Lagrangian factor update problem, which can then be solved through iteration. The maximized part of the optimization problem P2 is expressed as follows:

$$
\text{P21}: \max_{d_{ij}} L(\varpi, \sigma, \psi). \tag{27}
$$

Since the optimization problem P1 is a convex optimization problem, according to the KKT condition, the optimal solution of the optimization problem P21 is equal to the optimal solution of the optimization problem P1. By solving equation $\partial L(\varpi, \sigma, \psi)/\partial d_{ij} = 0$ and considering constraint $d_{ij} \geq 0$, the optimal task assignment d_{ij} is obtained as follows:

$$
d_{ij} = \begin{cases} 0, & d_{ij} \leq 0, \\ \dfrac{\alpha}{\ln 2\left(\chi + \left(1/R_{ij} + k/f_{ij}\right)\left(\varpi_{ij} + \psi_{ij}\right) + \delta_i\right)} - 1, & d_{ij} > 0. \end{cases}
$$

(28)

Require: α, d_i, χ
Ensure: $d_{ij}, \forall d$
1:     Define t for counting iteration, initialize $t = 0$, define ε as convergence threshold
2:     Initialization $\omega_{ij}, \sigma_j, \psi_{ij}$
3:     According to $\omega_{ij}, \sigma_j, \psi_{ij}$ and (28), compute d_{ij}
4:     According to (25), compute $L(d_{ij}^{*(t)}, \omega^{(t)}, \sigma^{(t)}, \psi^{(t)})$, let $L^{(t)} = L(d_{ij}^{*(t)}, \omega^{(t)}, \sigma^{(t)}, \psi^{(t)})$
5:     According to (30)(31)(32), compute $\omega_{ij}^{(t+1)}, \sigma_j^{(t+1)}, \psi_{ij}^{(t+1)}$
6:     According to (25), compute $L(d_{ij}^{*(t)}, \omega^{(t)}, \sigma^{(t)}, \psi^{(t)})$, let $L^{(t)} = L(d_{ij}^{*(t)}, \omega^{(t+1)}, \sigma^{(t+1)}, \psi^{(t+1)})$
7:     **while** $|L^{(t+1)} - L^{(t)}| > \varepsilon$ **do**
8:        According to $\omega_{ij}^{(t+1)}, \sigma_j^{(t+1)}, \psi_{ij}^{(t+1)}$, compute $d_{ij}^{(t+1)}$
9:        $L^{(t)} = L^{(t+1)}$
10:       $d_{ij}^{*(t)} = d_{ij}^{*(t+1)}$
11:       Repeat steps 5, 6
12:       $t = t + 1$
13:       **if** $\sum_{j=1}^{N} d_{ij}^{*} \sim = d_i$ **then**
14:          $d_{i0}^{*} = d_i - \sum_{j=1}^{N} d_{ij}^{*}$
15:          **if** $T_{i0} \leq T_i^{\max}$ **then**
16:             d_{i0}^{*} and d_{ij}^{*}
17:          **else**
18:             $d_{i0}^{*} = 0$ and $d_{ij}^{*} = 0$
19:          **end if**
20:       **else**
21:          $d_{i0}^{*} = 0$ and d_{ij}^{*}
22:       **end if**
23:     **end while**

ALGORITHM 2: Task assignment algorithm based on Lagrange theory.

The Lagrange multipliers updating problem of P2 is expressed as

$$\text{P22}: \quad \min_{\omega_{ij}, \sigma_i, \psi_{ij}} L(d_{ij}, \omega, \sigma, \psi). \tag{29}$$

It can be seen from the above analysis that P22 is also a convex problem, so the gradient descent method is used to update the Lagrangian multipliers $\omega_{ij}, \sigma_i, \psi_{ij}$, and the Lagrangian multipliers $\omega_{ij}, \sigma_i, \psi_{ij}$ are all nonnegative real numbers.

We denote $\omega_{ij}^{(t)}, \sigma_i^{(t)}, \psi_{ij}^{(t)}$ as the Lagrangian multipliers of the tth iteration, and then, the Lagrange multipliers $\omega_{ij}^{(t+1)}$, $\sigma_i^{(t+1)}, \psi_{ij}^{(t+1)}$ of the $(t + 1)$th iteration can be obtained as follows:

$$\omega_{ij}^{(t+1)} = \max \left\{ 0, \omega_{ij}^{(t)} - \tau^{(t)} \left(T_i^{\max} - \frac{d_{ij}}{R_{ij}} - \frac{c_{ij}}{f_{ij}} \right) \right\}, \tag{30}$$

$$\psi_{ij}^{(t+1)} = \max \left\{ 0, \psi_{ij}^{(t)} - \tau^{(t)} \left(T_{ij}^{\text{re}} - \frac{d_{ij}}{R_{ij}} - \frac{c_{ij}}{f_{ij}} \right) \right\}, \tag{31}$$

$$\sigma_i^{(t+1)} = \max \left\{ 0, \sigma_i^{(t)} - \tau^{(t)} \left(d_i - \sum_{j=1}^{N} d_{ij} \right) \right\}, \tag{32}$$

where $\tau^{(t)} = \tau^{(0)} / \sqrt{t}$ is the tth update step length. In order to ensure the convergence of the algorithm, this update step length [22] is considered.

The task allocation algorithm can be divided into the following four steps: (1) Initialize the Lagrange multipliers. (2) According to the given Lagrange multipliers, solve the optimization problem P21 through KKT conditions to obtain the optimal task allocation result; update the Lagrange multipliers with the method in formula (30)–(32). (3) Get the task allocation result by the convergence condition of Lagrange function. (4) Judge whether the task has been completely allocated; if it has been completely allocated, return the allocation result; otherwise, judge whether the remaining unallocated task can be executed locally. If possible, the rest of the tasks are executed locally and the task assignment results are returned. Otherwise, the task cannot be completed on time, and the task assignment will fail. In Algorithm 2, the task allocation algorithm will be described in detail.

5. Numerical Results

The numerical results given in this section show that the proposed algorithm improves the resource utilization and total task completion ratio in the VC system. We evaluate and compare the effects of different vehicle speed, vehicle density, task size, and number of TaVs on the performance of the proposed scheme and other baseline schemes. We will compare the proposed resource allocation scheme in this paper (PRAS) with

two benchmark schemes, including (i) LCS (Lowest Cost Scheme): the server bids by the network conditions, and the buyer always chooses the server with the lowest unit price, and (ii) HRS (Highest Revenue Scheme): the buyer bids by resource requirements, and the server always provides services to the buyer with the highest unit bid. In addition, the proposed offloading scheme in this paper (POS) is compared with the benchmark scheme by three different comparison methods: (i) LPS (Local Processing Scheme); (ii) AOS (Average Offloading Scheme): TaVs distribute tasks to the surrounding SeVs equally; and (iii) UOS (Unified Offloading Scheme): TaVs consider its own computing resources and assign the task equally to the surrounding SeVs and itself.

In this paper, the MATLAB simulation platform is used to verify the proposed algorithm. Vehicles are randomly distributed to a two-way lane with a length of 1000 m and a width of 50 m. Furthermore, the computation capacity and task size of the vehicle follow normal distribution. Relevant simulation parameters are given in Table 1.

5.1. Effect of the Number of TaVs.

Figure 2(a) shows resource utilization for different algorithms when the number of TaVs varies from 10 to 20, where the resource utilization ratio is equal to the ratio of the resources used by TaVs to the total resources of SeVs in the system. It is observed from Figure 2(a) that the resource utilization monotonically increases with the number of TaVs, and the growth rate gradually decreases; in other words, increasing the number of TaVs from 10 to 14 makes more TaVs benefit from the resources of SeVs than that from 16 to 20. The reason is that the available resources are no longer enough to respond to the requests of new TaVs with limited computation resources. As a result, the resource utilization gradually increased slowly. Furthermore, it can also be observed that PRAS can obtain higher resource utilization than HRS and LCS. When the number of TaVs is more, it can better reflect the superiority of PRAS. This is because the number of TaVs increases, and the more TaVs compete for limited computing resources, the probability of successful acquisition of resources by TaVs is reduced. For the sake of obtaining better resource utilization, compared with HRS and LCS, PRAS further optimizes the network status and DiffServ of different tasks based on resource allocation. In other words, PRAS adopts a differentiated pricing strategy. Compared with HRS which only considers network condition and LCS which only considers its own condition, PRAS considers the condition of the TaV, including whether the TaV has urgent tasks and whether the network around the TaV is busy. Therefore, the total resource utilization rate of the network is improved. In addition, when the number of TaVs is less than 12, the HRS and PRAS have the same resource utilization as the computing resources are sufficient and both two algorithms can provide the resources required for TaVs.

Figure 2(b) depicts the task completion ratio of different algorithms where the number of TaVs varies from 10 to 20, where the task completion ratio is equal to the number of tasks successfully completed within the deadline over the total number of tasks in the system. It can be seen from Figure 2(b) that POS can achieve a relatively higher task completion ratio compared with the three baseline schemes. This

TABLE 1: Simulation parameter.

Parameter	Value
v	$[80 - 120]$ km/h
Z	300 m
R	5 Mb
T_{max}	$1 \sim 3$ s
F_j^{max}	$N(4 * 10^{10}, 0.5 * 10^{10})$ (CPU cycle/s)
f_i	$N(2 * 10^{10}, 1 * 10^9)$ (CPU cycle/s)
d	$[3, 7]$ (Mb)
k [15]	18000 cycles/bit
$M + N$	$[20 - 30]$

is because UOS and AOS do not consider mobility of vehicles and task deadline constraints. However, the proposed scheme jointly optimizes mobility of vehicles and the deadline of the task, which is conducive to making full use of resources. Hence, more tasks can be offloaded and task completion is significantly increased. In addition, since LPS does not consider the choice of SeVs, the increase in the number of TaVs does not affect the task completion ratio of LPS.

Figure 3 shows the price for different task sizes of the TaV when the number of TaVs varies from 10 to 20. It is observed from Figure 3 that the price monotonically increases with the number of TaVs, and the price also monotonically increases with task size. The reason is that increasing the size of tasks and the number of TaVs will reduce the TaV surrounding idle resources. The price function of the TaV is a piecewise function of the surrounding remaining resources. When the remaining resources of the TaV are greater than zero, the price is increased based on the lowest price. When the remaining resources around the TaV are greater than zero, the price is increased based on the middle price.

5.2. Effect of Vehicle Density.

Figure 4 shows the resource utilization and task completion ratio of the VC system versus the vehicle density. Numerical results demonstrate that vehicle density is inversely related to resource utilization and positively related to the task completion ratio. This is because the number of SeVs increases with increasing vehicle density, and resource utilization decreases due to the constant number of TaVs. Moreover, with the increase in the number of SeVs, there will be more available resources around TaVs and more computation tasks can be offloaded to SeVs, which will lead to the improvement of the task completion ratio. On the contrary, when the vehicle density in the system is constant, the resource utilization increases and the task completion ratio decreases with the increase of the number of TaVs. As the number of TaVs increases, more TaVs compete for limited resources, which will lead to the increase in resource utilization. The reason is that when the number of TaVs increases to a certain extent, the resources of SeVs are fully utilized but still cannot meet resources required for the excessive growth of the number of TaVs. In the end, the size of total offloading is no longer changed and the task completion ratio is reduced.
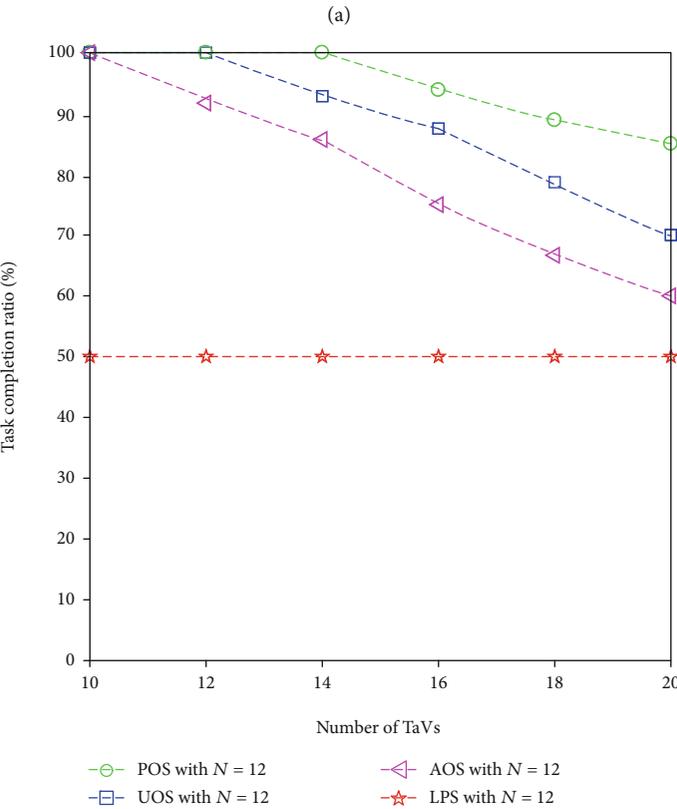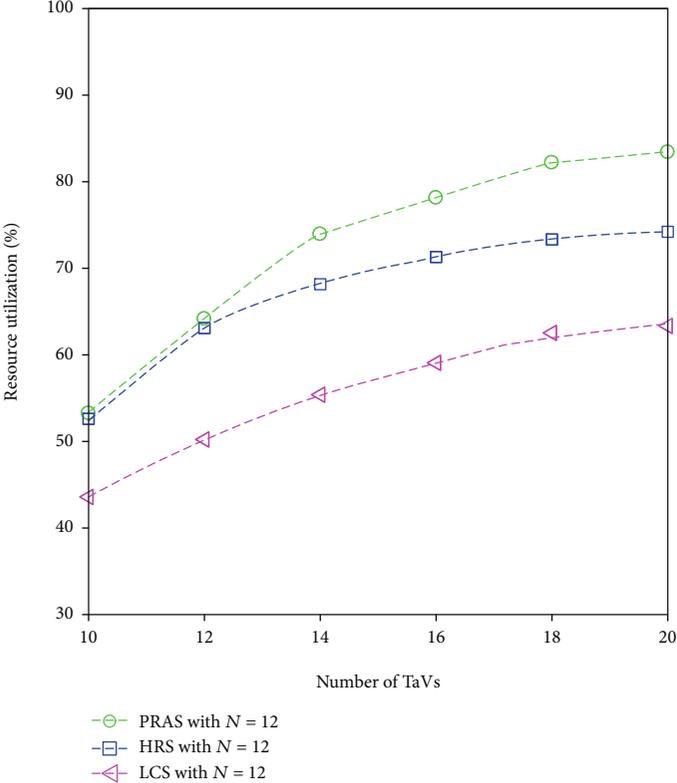
(a)

(b)

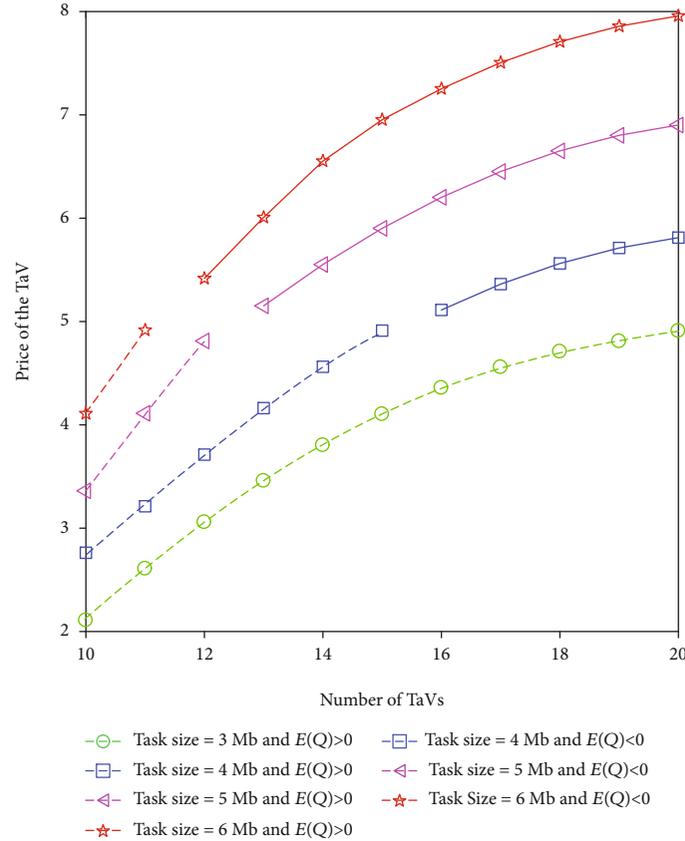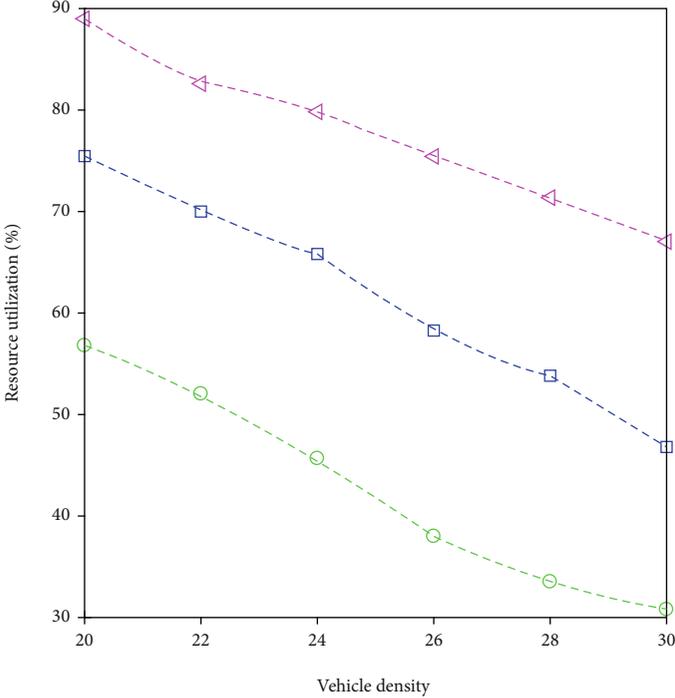Figure 2: (a) Resource allocation versus the number of TaVs. (b) Task completion ratio versus the number of TaVs.

FIGURE 3: Price versus the number of TaVs.

5.3. Effect of TaV Speed. Here, we make comparison of TaV speed performance under different schemes. Figure 5 shows offloading utility and task completion ratio for different algorithms when TaV speed varies from 80 km/h to 120 km/h. It can be concluded that the task completion ratio and offload utility of POS, UOS, AOS, and LPS all decrease with increasing speed from Figures 5(a) and 5(b). As the speed of TaVs increases, the contact time between TaVs and SeVs becomes shorter, and the size of offloading task data decreases. Therefore, the offloading utility and task completion ratio decrease. Then, by comparing POS with other schemes, it can be seen that the task completion ratio and offload utility of POS are optimal. This is because POS optimizes the mobility of vehicles, the deadlines of TaVs, local computing resources of TaVs, and the service differentiation of different types of tasks. The resource allocation is superior to other schemes, and task allocation is also more reasonable than other schemes. Hence, the task completion ratio and offloading utility of POS are superior to other solutions. In addition, with the decrease of the task size processed by SeVs, the task offloading time is shortened and the success rate of task offloading is increased accordingly. What is more, since the choice of SeVs is not considered, LPS is not sensitive to the increase of speed on the task completion ratio of LPS.

5.4. Effect of Task Size. Figure 6(a) shows offloading utility versus the task size of TaVs. Numerical results demonstrate

that task size is positively related to offloading utility. The numerical results show that when the task size is less than 4 Mb, the offloading utility of AOS is optimal and the offloading utility of POS is close to UOS. Since POS and UOS process a part of tasks locally, the size of offloading tasks is reduced, resulting in lower offloading utility than AOS. Moreover, it can be seen from Figure 6(a) that when the sizes of tasks are greater than 4 Mb, the offloading utility of POS is the highest. It can also be obtained that with the increase of task size of TaVs, the offloading utility of POS, HRS, and LCS first increases rapidly and then slowly increases. This is because the number of SeVs remains unchanged. When the task size of TaVs begins to increase, the offloading utility increases rapidly. However, with the increase of the size of the task to a certain extent, the resources of SeVs are fully utilized, which still cannot meet the resource requirements of the excessive increase in TaVs' task size, so the growth trend of offloading utility gradually slows down.

We also compare the task completion ratio in terms of the task size in Figure 6(b). It is interesting to see that the task completion ratio decreases with the increase of the task size. The reason is that as the sizes of the tasks processed by each vehicle increases, the risk of task exceeding the deadline also increases due to the constant deadline. As a result, the task completion rate is reduced. In addition, it can be seen that the UOS and AOS methods do not take the task processing time beyond the deadline into account, which results in the
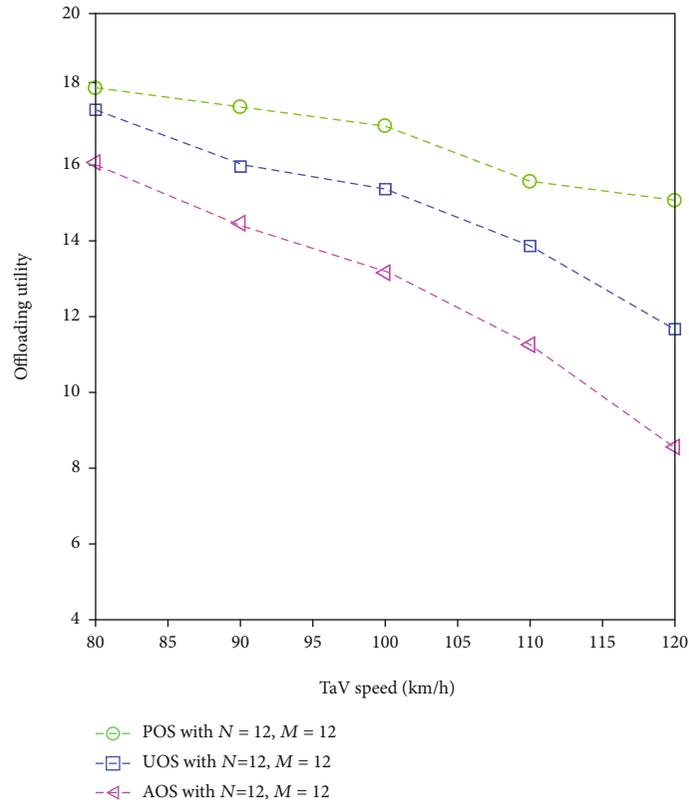
(a)
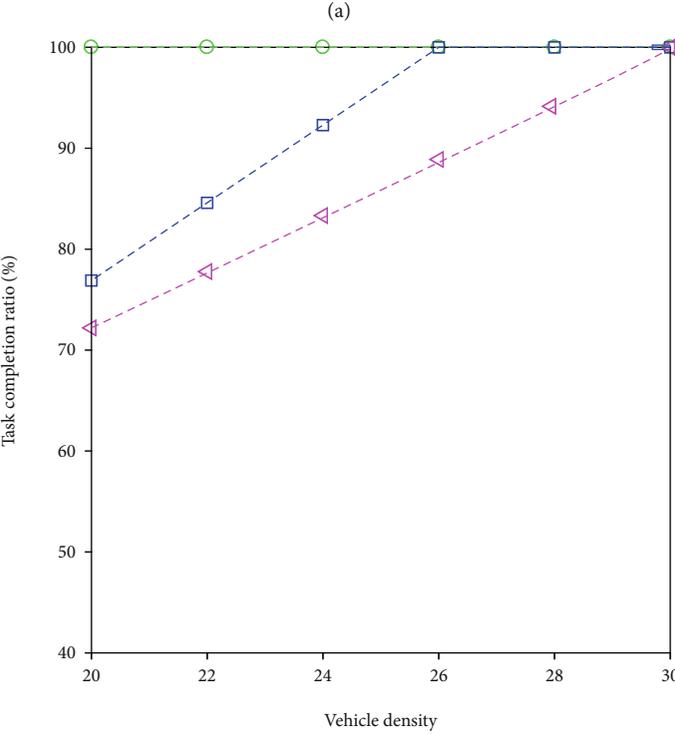
(b)

Figure 4: (a) Resource allocation versus vehicle density. (b) Task completion ratio versus vehicle density.

(a)
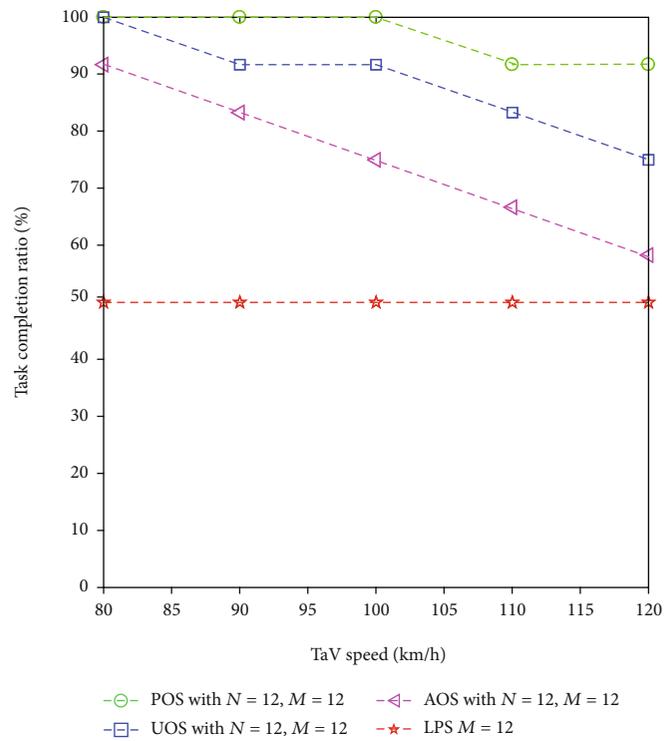
(b)

Figure 5: (a) Offloading utility versus TaV speed. (b) Task completion ratio versus TaV speed.
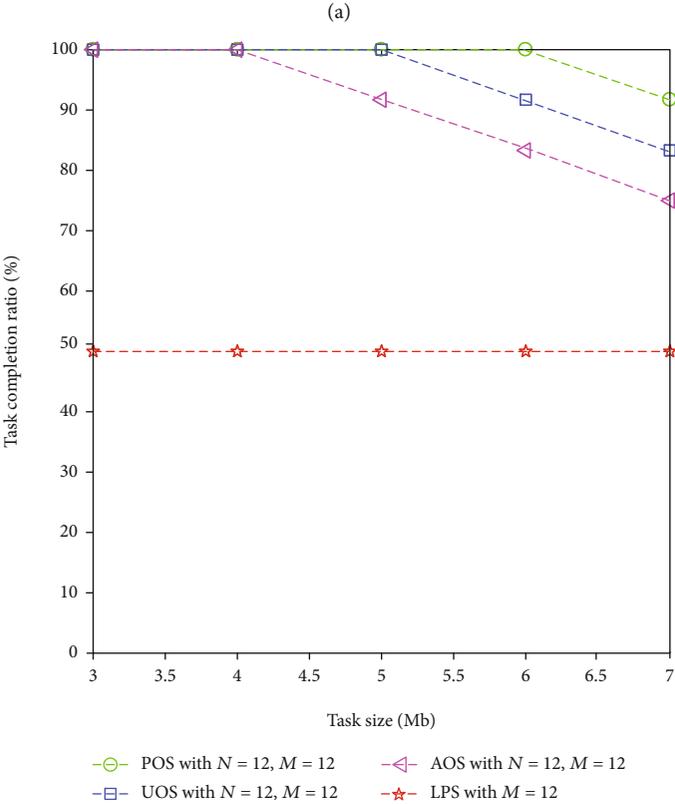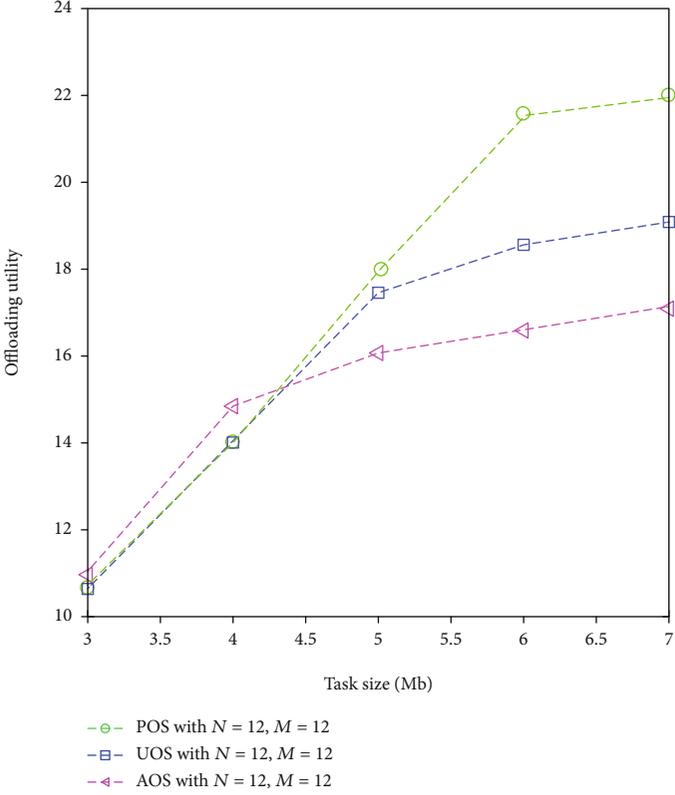
(a)

(b)

FIGURE 6: (a) Offloading utility versus task size. (b) Task completion ratio versus task size.

task offloading success rate lower than the POS, and the task completion rate is lower than the POS. What is more, LPS is executed locally without computational offloading, so it takes more time than the offloading mode.

6. Conclusion

In this paper, we investigate the joint task offloading and resource allocation mechanism in the VC system and maximize the total offloading revenue of tasks under the constraints of task completion deadlines and rapid vehicle movement characteristics. Considering the different types of services required by tasks in VC, this paper proposes a resource allocation model based on differentiated services and designs the pricing function of a task. Its pricing relies on the size of the task, the deadline for the task, and the number of requests and responses received by the vehicle that generated the task in the previous time slot. Furthermore, considering the mobility of vehicles, the deadline of tasks, and the limited computing resources, the centralized Lagrange dual method is used to solve the task offloading problem of the VC system. The simulation results show that the optimization effect of DiffServ offloading strategy is significantly improved, and the resource utilization and task completion ratio are significantly increased.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None of the authors have any conflicts of interest

Acknowledgments

References

[1] S. Basagni, M. Conti, S. Giordano, and I. Stojmenovic, *The Next Paradigm Shift: From Vehicular Networks to Vehicular Clouds*, Mobile Ad Hoc Networking, 2013.

[2] R. E. Sibai, T. Atechian, J. B. Abdo, R. Tawil, and J. Demerjian, "Connectivity-aware service provision in vehicular cloud," in *2015 International Conference on Cloud Technologies and Applications (CloudTech)*, Marrakech, Morocco, June 2015.

[3] M. Whaiduzzaman, M. Sookhak, A. Gani, and R. Buyya, "A survey on vehicular cloud computing," *Journal of Network Computer Applications*, vol. 40, pp. 325–344, 2014.

[4] F. Yang, S. Wang, J. Li, Z. Liu, and Q. Sun, "An overview of internet of vehicles," *China Communications*, vol. 11, no. 10, pp. 1–15, 2014.

[5] L. Liu, C. Chen, Q. Pei, S. Maharjan, and Y. Zhang, "Vehicular edge computing and networking: a survey," *Mobile Networks and Applications*, vol. 25, pp. 1–42, 2020.

[6] J. Sun, Q. Gu, T. Zheng, P. Dong, and Y. Qin, "Joint communication and computing resource allocation in vehicular edge computing," *International Journal of Distributed Sensor Networks*, vol. 15, no. 3, Article ID 155014771983785, 2019.

[7] H. R. Arkian, R. E. Atani, and A. Pourkhalili, "A cluster-based vehicular cloud architecture with learning-based resource management," in *2014 IEEE 6th International Conference on Cloud Computing Technology and Science*, pp. 162–167, Singapore, Singapore, December 2014.

[8] L. Hou, K. Zheng, P. Chatzimisios, and Y. Feng, "A continuous-time markov decision process-based resource allocation scheme in vehicular cloud for mobile video services," *Computer Communications*, vol. 118, pp. 140–147, 2018.

[9] R. I. Meneguette, A. Boukerche, A. H. M. Pimenta, and M. Meneguette, "A resource allocation scheme based on semi-Markov decision process for dynamic vehicular clouds," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, Paris, France, May 2017.

[10] S. Wu, W. Xia, W. Cui, Q. Chao, and L. Shen, "An efficient offloading algorithm based on support vector machine for mobile edge computing in vehicular networks," in *2018 10th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–6, Hangzhou, China, October 2018.

[11] Q. Liu, Z. Su, and Y. Hui, "Computation offloading scheme to improve qoe in vehicular networks with mobile edge computing," in *2018 10th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–5, Hangzhou, China, October 2018.

[12] M. Liwang, S. Dai, Z. Gao, Y. Tang, and H. Dai, "A truthful reverse-auction mechanism for computation offloading in cloud-enabled vehicular network," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4214–4227, 2019.

[13] Y. Sun, J. Song, S. Zhou, X. Guo, and Z. Niu, "Task replication for vehicular edge computing: a combinatorial multi-armed bandit based approach," in *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–7, Abu Dhabi, United Arab Emirates, December 2018.

[14] Z. Jiang, S. Zhou, X. Guo, and Z. Niu, "Task replication for deadline-constrained vehicular cloud computing: optimal policy, performance analysis, and implications on road traffic," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 93–107, 2018.

[15] M. LiWang, J. Wang, Z. Gao, X. Du, and M. Guizani, "Game theory based opportunistic computation offloading in cloud-enabled iov," *IEEE Access*, vol. 7, pp. 32551–32561, 2019.

[16] M. Liu and Y. Liu, "Price-based distributed offloading for mobile-edge computing with computation capacity constraints," *IEEE Wireless Communications Letters*, vol. 7, no. 3, pp. 420–423, 2018.

[17] T. Zhao, S. Zhou, X. Guo, Y. Zhao, and Z. Niu, "A cooperative scheduling scheme of local cloud and internet cloud for delay-aware mobile cloud computing," in *2015 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, San Diego, CA, USA, December 2015.

[18] J. Liu and Q. Zhang, "Offloading schemes in mobile edge computing for ultra-reliable low latency communications," *IEEE Access*, vol. 6, pp. 12825–12837, 2018.

[19] Y. Dai, D. Xu, S. Maharjan, and Y. Zhang, "Joint load balancing and offloading in vehicular edge computing and

networks," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4377–4387, 2019.

[20] C.-K. Tham and B. Cao, "Stochastic programming methods for workload assignment in an ad hoc mobile cloud," *IEEE Transactions on Mobile Computing*, vol. 17, no. 7, pp. 1709–1722, 2018.

[21] A. N. Toosi, R. N. Calheiros, R. K. Thulasiram, and R. Buyya, "Resource provisioning policies to increase iaas provider's profit in a federated cloud environment," in *2011 IEEE International Conference on High Performance Computing and Communications*, pp. 279–287, Banff, AB, Canada, September 2011.

[22] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1310–1322, 2006.

WILEY | Hindawi

Research Article

Evaluation of Localization by Extended Kalman Filter, Unscented Kalman Filter, and Particle Filter-Based Techniques

Inam Ullah,[1] Xin Su,[1] Jinxiu Zhu [ID],[1] Xuewu Zhang,[1] Dongmin Choi [ID],[2] and Zhenguo Hou[3]

[1]*College of Internet of Things (IoT) Engineering, Hohai University (HHU), Changzhou Campus, 213022, China*
[2]*Division of Undeclared Majors, Chosun University, Gwangju 61452, Republic of Korea*
[3]*China Construction Seventh Engineering Division CoRP, Ltd, 108 Chengdong Road, Jinshui District, Zhengzhou City, Henan Province, China*

Correspondence should be addressed to Dongmin Choi; jdmcc@chosun.ac.kr

Mobile robot localization has attracted substantial consideration from the scientists during the last two decades. Mobile robot localization is the basics of successful navigation in a mobile network. Localization plays a key role to attain a high accuracy in mobile robot localization and robustness in vehicular localization. For this purpose, a mobile robot localization technique is evaluated to accomplish a high accuracy. This paper provides the performance evaluation of three localization techniques named Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), and Particle Filter (PF). In this work, three localization techniques are proposed. The performance of these three localization techniques is evaluated and analyzed while considering various aspects of localization. These aspects include localization coverage, time consumption, and velocity. The abovementioned localization techniques present a good accuracy and sound performance compared to other techniques.

1. Introduction

Accurate localization is a very important aspect of several wireless sensor networks (WSNs) and Internet of Things (IoT) applications. These applications included underwater navigation, indoor positioning, bridges monitoring, industrial monitoring, health monitoring, and security systems [1–4]. For the improvement of localization accuracy, a variety of localization techniques have been investigated in the previous work. However, in this paper, the authors focused on three basic localization techniques named Extended Kalman Filter (EKF-) based localization, Unscented Kalman Filter (UKF-) based localization, and Particle Filter (PF-) based localization.

In the early period, Kalman Filter (KF) is used in an iterative manner that considers the prior information of the noise features to compensate for and to filter out the noise. But still, the issues arise during localization when attempting to model the noise that is only an approximation and does not specify the noise real distribution [5–7]. KF is only appli-

cable for the linear stochastic procedures; however, for the nonlinear procedures, the EKF can be applied. The supposition of these two methods (KF and EKF) is that noise and process measurements are self-governing and with a normal probability distribution [8].

The authors analyzed the pertinency of the KF to the mobile robot self-positioning in [9–11]. These algorithms are only appropriate for linear systems. On the other hand, for robot self-positioning, EKF provides an alternative to the Bayesian filter. Therefore, in [12], the author proposed an EKF approach for the localization of four-wheel encoders and laser range-finder nodes. EKF is basically used for the nonlinear functions, which apply the Taylor series expansion to linearize the measurement models. Thus, the first-order nonlinear functions of the Taylor series is used. In the predictable statistics of the subsequent distributions, this linearization often encourages higher error. This is particularly obvious when the systems are vastly nonlinear, and it can lead to the deviation in the filter. Besides, the UKF does not estimate the nonlinear method, and the actual nonlinear

model is applied to the observation models. UKF basically estimate the distribution of random variables [13] by applying the scaled unscented alteration method. For the autonomous robot localization or Autonomous Underwater Vehicles (AUV) localization, the EKF and UKF performance is equated in [14]. The author concludes that the UKF performance is better concerning numerous characteristics such as average localization accuracy relative to the EKF. Besides this, UKF does not involve applying the Jacobian matrix; therefore, the UKF precision is higher.

The PFs have taken further consideration by the investigators because of their advantages against the EKF and the UKF [15]. PF can solve the issue of a measurement system that is affected by the non-Gaussian noise and accomplish the localization globally in the case in which there is no prior knowledge about the target [16]. The PFs are easier in implementation as compared to the EKF, but also exist some disadvantages such as letdown due to sample penury and substantial computational load. The letdown due to the sample penury occur more commonly and resultantly causes the failure of the PF algorithm. To address this issue, various PFs have been investigated such as Markov Chain Monte Carlo (MCMC) move step and Regularized PF (RPF) [17]. The proposed approaches cannot prevent it completely but can mitigate the sample penury. If the conventional PF flops because of the sample penury, it is unable to return to the normal conditions. Furthermore, a Hybrid Particle/Finite Impulse Response (FIR) Filter (HPFF) is investigated to solve this issue [18–21]. FIR filters are more stable as compared to other filters and robust, but FIR filters characteristically drop inaccuracy to the PF in the case of nonlinear systems [22–25]. Therefore, PF is applied as an elementary filter, which is susceptible to failure and deviation. The FIR filter is applied to recuperate the PF by restarting and reorganizing.

1.1. Contributions of This Work. This work aims to propose an effective vehicular localization technique for the mobile robot. For this purpose, the authors introduced three localization techniques. The proposed techniques' performance is better in simulation concerning the abovementioned localization techniques in the literature. The contributions of this paper are divided into the following three phases.

(1) EKF-based localization

(2) UKF-based localization

(3) PF-based localization

The performance of the abovementioned techniques is evaluated and analyzed in different scenarios. While considering numerous aspects of localization, the accuracy of these techniques is higher. A variety of vehicular localization techniques are investigated in the literature to analyze the presentation of the proposed techniques. These techniques are evaluated while considering numerous aspects of localization such as accuracy, time consumption, vehicle velocity, coverage area, and localization coverage. Furthermore, the presentation of the proposed localization techniques is compared with each other's and also with other localization techniques.

1.2. Organization. The rest of the paper is sorted as follows. The related work is detailed in Section 2, and Section 3 presents the proposed techniques used for localization. The subsections of Section 3 present the proposed EKF, UKF, and PF localization techniques. Section 4 presents the discussion on the simulation results and comparison while Section 5 concludes the work.

2. Related Work

In order to expand the performance of mobile robot localization, several approaches have been introduced [26–28] to address vehicular localization [29] issues and errors of the NLoS environment. The method in [30] uses two receivers, and the Frequency Difference of Arrival (FDoA) signals to approximate the moving emitter velocity. There is a nonlinear measurement error which is occurred by the RF system noise in the geolocation environment with the Time Difference of Arrival (TDoA) [31, 32] and FDoA. For this problem, the authors proposed the iterated dual-EKF approach to recompense for the nonlinear estimation error. Using the iterated dual-EKF approach, the parameter estimation (PE) filter updates model uncertainties caused by exterior noises and has a higher convergence rate of the system parameters by the iteration method.

EKF is a traditional technique for positioning estimation [33, 34]. EKF extends models of nonlinear functions in the Taylor series close to the estimation state and shortens them to consider the linearization of the model in the first order. In the Line-of-Sight (LoS) environment, EKF based on the linear models achieves higher accuracy. But, if the channel is in the NLoS environment, the EKF shows a high error in localization because of measurement data deviation [35]. For mobile node localization, a variety of NLoS approaches have been presented [36–40]. The method of unscented transformation is used for the standard KF to produce the UKF, which attains a better estimation than the other methods. Another method is presented for the positioning of targets named Adaptive Iterated Unscented KF (AIUKF) which combines the adaptive factor and iterative approach to improve the localization performance. A PF localization approach based on the Monte Carlo technique is used for the positioning which exploits random sample group information for the approximation of the state Probability Density Function (PDF). Therefore, the performance of the PF-based localization in non-Gaussian is much better while demanding a high number of sample points. The authors collected all metrics of the range and obtained the final estimation of the state on the basis of position estimation of the fusion subgroup. But on the other hand, in the NLoS, the authors find discarding and detecting the range calculation from the beacon nodes. However, most of the above methods perform accurately only in a particular noise distribution field, which is not authentic.

Perhaps, EKF is a well-known procedure for estimating a noisy measurement of the nonlinear system state. EKF is based on the nonlinear maps of the system around the estimated route. It is also based on the idea that the noises of measurement, input noise, and initial state are Gaussian. It

is acknowledged that the EKF is likely to deviate, mostly because of bad primary estimates and higher noises, but our experience does not know of any testable convergence conditions. The estimation divergence is extra probable when measurements are lost due to communication faults [41, 42], which is a communal disorder in mobile robotics.

In recent years, UKF was developed to tackle some problems such as the need for Gaussian noises and the properties of poor approximation [43]. The basic idea of UKF is to find a transformation that allows a random vector having a length of n to approximate the covariance and mean when it is changed by the nonlinear map. It can be achieved by calculating a set of $2n + 1$ points, known as σ-points, based on the innovative vector mean and variance, transforming those points through the nonlinear map and then resembling the transformed vector mean and variance from the transformed σ-points. The authors investigated that the EKF estimate is precise to the first order, while in the case of Gaussian noises, the UKF is precise to the third order. Similarly, for the EKF, the covariance of the estimate is precise to the first order and second order for the UKF.

A procedure of tracking through PFs is presented in [44] to produce a probability distribution over the targeted area. The procedure applies the recursive Bayesian filters which are based on the sequential MC technique of sampling to estimate the location of the target posterior distribution by applying another distribution that can be prior arbitrary. PF begins with a regularly primed set of particles. Then, according to a movement model, all particle positions are modified. The authors considered the measure in the (x, y) space that follows an arbitrary walk model for the representation of the human movement. The $x(t + \tau) = x(t) + \varphi(\tau)$ shows a random path of the user, where φ denotes the random parameter which characterizes the probability of following a known direction in the tracking procedure next phase. In the measures achieved from the tags deployed in RFID, the model also takes into consideration the error model.

A statistical method named Kullback-Leibler Distance (KLD) sampling is presented to increase the efficiency of the PFs by adapting the sample sets size on-the-fly [45]. The approach is used to link the error of approximation which is introduced by the PF sample-based illustration. The method selects a smaller number of samples if the density is observant to a small part of the state space, and rising samples if the uncertainty of the state is high. The mobile robot localization problem is used to test and demonstrate the approach to the adaptive PF. The localization of robots is the problem of estimating the position of a robot relative to the map of its operation area. This issue has been identified as one of the most important mobile robotics challenges which come in diverse essences. The simplest problem with localization is position tracking, where the primary location of the robot is determined, and location pursues to exact small, gradual errors in the robot's odometry. The global localization is another stimulating problem for mobile robots, where a robot does not have prior knowledge about his position, but it has to be decided from scratch, instead.

Following the above discussion, it is reflected that to present efficient and accurate localization techniques, the issue of vehicular localization with limited features is previously important. Most of the previous localization techniques focused only on the vehicle localization accuracy while ignoring several aspects such the vehicle velocity, time consumption, and coverage. Besides this, we must take into account the positioning of the mobile robot in a precise manner. The author, therefore, considers three localization techniques named EKF, UKF, and PF in this paper. The proposed techniques are evaluated on the basis of their performances. The authors consider several aspects for localization such the vehicle velocity, time consumption, accuracy, and localization coverage.

3. Proposed Techniques of Localization

This section discusses the proposed localization techniques based on EKF, UKF, and PF. In the next pages, the proposed localization techniques are examined. Table 1 shows the notations used in this work.

3.1. Extended Kalman Filter-Based Localization. EKF is typically implemented by substitution of the KF for nonlinear systems and noise models. The models of observation and state transformation are nonlinear functions, but these can be differentiable functions. The observation and state transition models [46] are delineated as

$$x_k = f(x_{k-1}, u_k) + w_k, \tag{1}$$

$$z_k = h(x_k) + v_k. \tag{2}$$

In the above equations, w_k and v_k denote the process and observation noises and zero-mean Gaussian with the covariance Q_k and R_k, where x_k and x_{k-1} are the current and previous robot state, while u_k represents the input vector.

The state covariance is calculated as

$$P_k = \begin{bmatrix} P_{x_v} & P_{x_v l} \\ P_{lx_v} & P_{ll} \end{bmatrix} = \begin{bmatrix} P_{x_v x_v} & P_{x_v l_1} & \cdots & P_{x_v l_1} \\ P_{l_1 x_v} & P_{l_1 l_1} & \cdots & P_{l_1 l_n} \\ \cdots & \cdots & \cdots & \cdots \\ P_{l_n x_v} & \cdots & \cdots & P_{l_n l_n} \end{bmatrix}, \tag{3}$$

where P_k is the covariance of the estimated $n \times n$ matrix. Prediction and update states are the two stages of the EKF algorithm. To calculate the prediction state

$$\hat{x}_{k|k-1} = F \times \hat{x}_{k-1|k-1} + B_k \times u_k, \tag{4}$$

where F denotes the state transition and B_k denotes the input matrix. Covariance from equation (3) will become

$$P_{k|k-1} = F \times P_{k-1|k-1} \times F^T + Q, \tag{5}$$

where $P_{k|k-1}$ and $P_{k-1|k-1}$ are the current and previous states. Q represents a covariance matrix for the noise process. The

TABLE 1: Index of notation.

Notation	Description
T	Time interval
φ	Random parameter
X_k	Current state
X_{k-1}	Previous state
Z_k	Observation model
f	Prediction state function
h	Prediction measurement function
$P_{k/k-1}$	Current state covariance
$P_{k-1/k-1}$	Previous state covariance
Q	Covariance matrix of the process noise
R	Covariance matrix of the measurement noise
$h\left(\widehat{X}_k\right)$	Predictable estimation
S_k	Covariance matrix of $\tilde{y}_k$
H_k	Predictable estimation Jacobian matrix
K_k	Kalman optimal gain
$P_{k/k}$	Covariance matrix subsequent state
$F\&B$	State transition matrices
δ_x^2	Standard deviation
x, y	Denote locations
$P_\circ$	State information vector
H	Output measurement matrix
v_{k-1}	Process noise
n_k	Measurement noise
w	Weights
$x_\circ$	Sigma point
x_i	Sigma point where $i = 1, 2, \cdots, 2n$
x_{est}	Estimation state
x_{tru}	True state
n	Dimension of x
s_k	Set of particles
$p\left(2k/x_k^i\right)$	Importance factor
$f_k(x_k)$	Positive state function
T_s	Robot sampling time
$p(x_k/z_1 : k)$	Successive approximation
$\delta(.)$	Dirac function
N	Denote the samples

origination vector and covariance matrix for the EKF update state is

$$\tilde{Y}_k = Z_k - h\left(\widehat{X}_k\right),$$
$$S_k = H_k \times P_{k|k-1} \times H_k^T + R_k, \tag{6}$$

where $h(\widehat{X}_k)$ is the predictable estimations, S_k denotes the covariance matrix of $\tilde{Y}_k$, and H_k is the predictable estimations Jacobian matrix. To calculate the subsequent state vector:

$$\widehat{x}_{k|k} = \widehat{X}_{k|k-1} + K_k \times \tilde{Y}_k.$$
$$K_k = P_{k|k-1} \times H_k^T \times S_k^{-1}, \tag{7}$$
$$P_{k|k} = (I_n - K_k \times H_k) \times P_{k|k-1},$$

where K_k is the Kalman optimal gain and $P_{k|k}$ denotes the covariance matrix subsequent state. To apply the prediction, motion, and observation models, the state vector is computed as

$$X_k = [x\, y\, \text{yaw}\, v]^T. \tag{8}$$

Therefore, the state transition matrices F and B are calculated as

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$
$$B = \begin{bmatrix} dt.\cos(x(3)) & 0 \\ dt.\sin(x(3)) & 0 \\ 0 & dt \\ 1 & 0 \end{bmatrix}, \tag{9}$$

In the same way, the measurement noise R is calculated as

$$R = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}. \tag{10}$$

The Jacobian of the measuring system's motion model and covariance can be computed as

$$\text{JF} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -dt.\sin(x(3)) & dt.u(1).dt.\cos(x(3)) & 1 & 0 \\ dt.\cos(x(3)) & dt.\sin(x(3)) & 0 & 1 \end{bmatrix},$$
$$P_0 = \begin{bmatrix} \sigma_x^2 & 0 & 0 & 0 \\ 0 & \sigma_y^2 & 0 & 0 \\ 0 & 0 & \sigma_{vx}^2 & 0 \\ 0 & 0 & 0 & \sigma_{vy}^2 \end{bmatrix}, \tag{11}$$

where σ_x^2 denotes the standard deviations of the estimated x, y locations and P_0 is the state information vector. The value of

the standard deviation, in this case, depends on certain variables such as the accuracy of GPS and the distance as shown in Figure 1. Thus, the measuring system Jacobian can be written as

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{12}$$

where H is the output measurement matrix and the Jacobian matrices of the state equation. The localization results of the EKF algorithm is shown in Figure 1.

3.2. Unscented Kalman Filter-Based Localization. UKF is a nonderivative filtering approach that immediately propagates the mean measurement covariance and the state estimates. Therefore, it precisely represents the allied distributions by any nonlinear transformations that include the measurement and state dynamics. In the process model, because of the nonlinearity in the state functions of the velocity, unscented approximation to the finest filtering explanation can be derived by implementing two steps of measurement and time update. In the implementation of the two phases, an unscented transform is approved for the sigma points formation. Furthermore, for the derivation of UKF, the Jacobian matrix calculation is not required as required for the EKF technique. The technique can proceed in the following way. Provided a nonlinear system with a discrete time-system model [47, 48]:

$$\begin{aligned} x_k &= f(x_{k-1}, v_k, u_{k-1}), \\ y_k &= h(x_k, n_k, u_k), \end{aligned} \tag{13}$$

where v_{k-1} and n_k show the process noise and measurement noise. Further, to compute the sigma points, time, and measurement update, the x_k estimates can be calculated. A set of test points and its related weights w at time $k-1$ is utilized for sigma points measurement such as $x_0, k-1 = x_{k-1}$, where x_0 denotes the sigma point at state:

$$x_{\text{est}} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}^T, \tag{14}$$

where x_{est} is the estimated state. The state will be true if $x_{\text{tru}} = x_{\text{est}}$. The corresponding weights for the sigma points can be calculated as

$$\begin{aligned} x_{i,k-1} &= x'_{k-1} + \left[\sqrt{(n+\lambda) \times P_{k-1}} \right]_i, \\ x_{i,k-1} &= x'_{k-1} - \left[\sqrt{(n+\lambda) \times P_{k-1}} \right]_{i-n}, \end{aligned} \tag{15}$$

where x_i denotes the sigma points and $i = 1, 2, \cdots, 2n$. Normally, the weights w are characterized as

$$\begin{aligned} w^m &= \left[\frac{\lambda}{n+\lambda} \right], \\ w^c &= \left[\left(\frac{\lambda}{n+\lambda} \right) + \left(1 - \alpha^2 + \beta \right) \right]. \end{aligned} \tag{16}$$

Similarly, at $t = 0$, the weights w^m and w^c can be written as

$$\begin{aligned} w_0^m &= \left[\frac{\lambda}{n+\lambda} \right], \\ w_0^c &= \frac{\lambda}{(n+\lambda) + (1 - \alpha^2 + \beta)}, \\ w_i^m &= w_i^c = \frac{1}{2(n+\lambda)}, \end{aligned} \tag{17}$$

where the dimension of x represented by n and $\lambda = \alpha^2(n+k) - n$ is the parameter of scaling. Moreover, α is applied to find the sigma points spreading around the state variable x' mean, and k is the second scaling parameter. To compute the prediction of sigma points with observation and motion model, the covariance and mean will become

$$\begin{aligned} x_{i,k|k-1} &= f(x_{i,k-1}), \\ \bar{x}' &= \sum_{i=0}^{2n} w_i^m x_{i,k|k-1}, \\ p'_k &= \sum_{i=0}^{2n} w_i^c \left[x_{i,k|k-1} - \bar{x}' \right] \left[x_{i,k|k-1} - \bar{x}' \right]^T + Q_k, \\ \dot{x}_{i,k|k-1} &= \left[x_{0:2n,k|k-1} \quad x_{0,k|k-1} + v\sqrt{Q_k} \quad x_{0,k|k-1} - v\sqrt{Q_k} \right]_i. \end{aligned} \tag{18}$$

With supplementary points gained from the process noise covariance matrix square root, the sigma points were increased.

$$\begin{aligned} \dot{y}_{i,k|k-1} &= h(\dot{x}_{i,k|k-1}), \\ \bar{y}' &= \sum_{i=0}^{2n^a} \dot{w}_i^m \dot{Y}_{i,k|k-1}, \\ p_{yy,k} &= \sum_{i=0}^{2n^a} \dot{w}_i^c \left[\dot{Y}_{i,k|k-1} - \bar{y}'_k \right] \left[\dot{Y}_{i,k|k-1} - \bar{y}'_k \right]^T + R_k, \\ p_{xy,k} &= \sum_{i=0}^{2n^a} \dot{w}_i^c \left[\dot{x}_{i,k|k-1} - \bar{x}'_k \right] \left[\dot{y}_{i,k|k-1} - \bar{y}'_k \right]^T, \end{aligned} \tag{19}$$
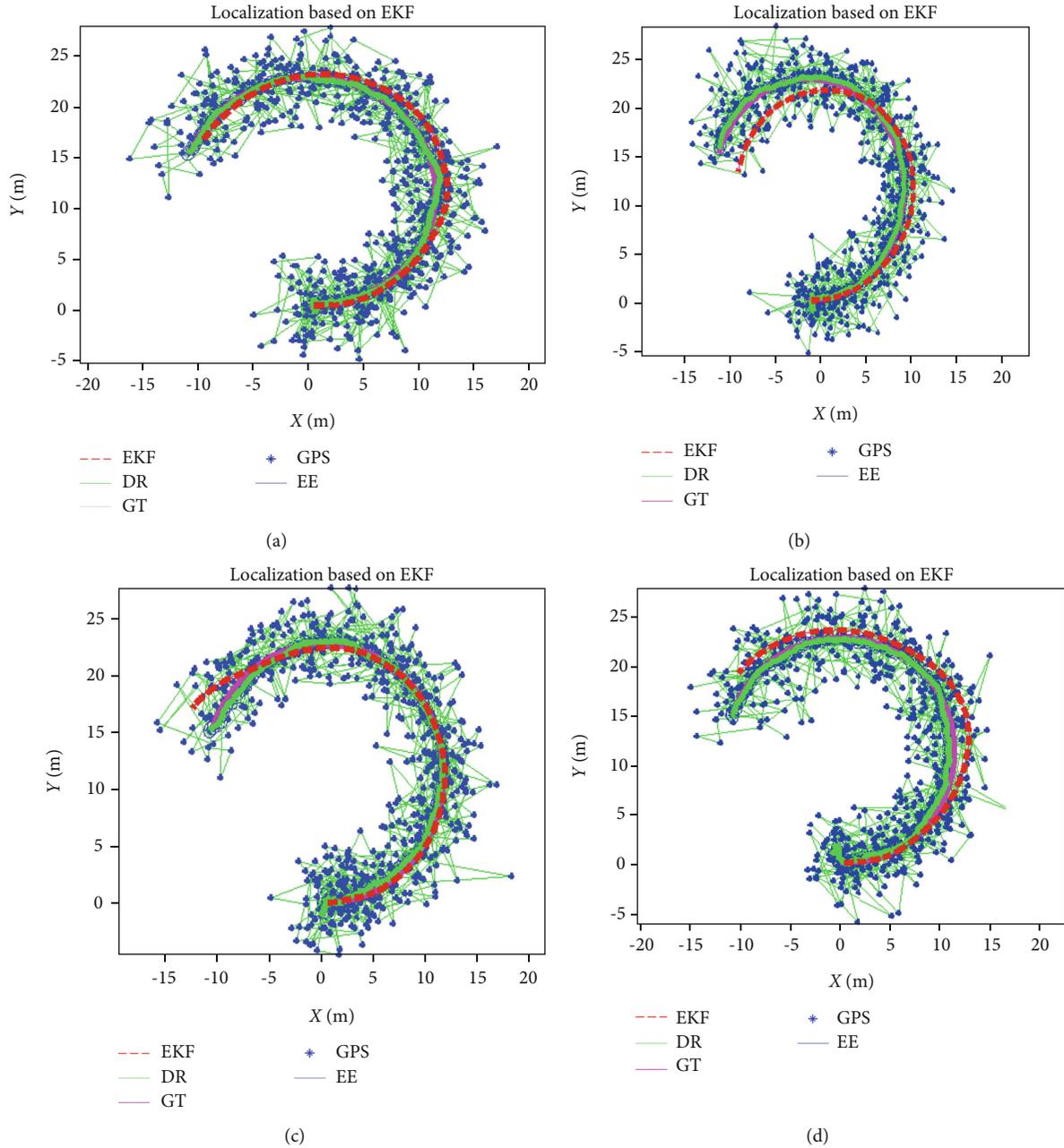
(a)

(b)

(c)

(d)

FIGURE 1: Localization comparison of EKF. In this phase, for all four iterations, the velocity is $v = 1$ m/s and time is $t = 60$ sec. In the legends, the red dashed line represents the EKF localization, where the green is the dead reckoning (DR), the pink solid line is the ground truth (GT), the blue color asterisks denote the GPS signals, and the blue line denotes the error ellipse (EE) during localization.

where R_k denote the covariance matrix. The Kalman gain can be computed to update the state and the covariance.

$$k_k = p_{xy,k} \times p'^{-1}_{yy,k},$$

$$\bar{x}_k = \bar{x}' + k_k \left(y_k - \bar{y}' \right), \tag{20}$$

$$p_k = p'_k - k_k \times p_{xy,k} \times k_k^T,$$

By considering the primary estimation state, i.e., $x'_0 = E[x_0]$ and $p_0 = E[(x_0 - x'_0) \times (x_0 - x'_0)^T]$

$$R = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}, \tag{21}$$

$$x_{k+1} = Fx_k + Bv_k,$$

where R denotes the covariance matrix for measuring noise and F and B can be calculated as

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$B = \begin{bmatrix} dt.\cos(x(3)) & 0 \\ dt.\sin(x(3)) & 0 \\ 0 & dt \\ 1 & 0 \end{bmatrix}, \tag{22}$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

In the above equations, F and B denote the state matrices of transition, while H represents the observation model output measurement matrix. The localization results of the UKF technique is presented in Figure 2.

3.3. Particle Filter-Based Localization. The authors analyzed the performance of the PF-based localization technique in this part. As similar to the tracking method, a PF is implemented to construct a probability distribution over the field of the targeted area of operation. This technique uses a recursive Bayesian filter based on the sampling of MC to calculate the target location subsequent distribution by applying another distribution which can be random a priori [40]. The PF approach is less intensive computationally in comparison to the EKF and UKF techniques. Besides this, in contrast to the KF, PF can avoid any supposition related to the intrinsic prominent attribute of the process and uncertainty related to the node information. The PF required an opti-

mum interval of average but does not include the statistics on noise. Moreover, when the robot is traveling during the process of localization, the sensor particles in the surrounding of the robot can communicate with the robot and transfer information to the robot [45]. The mobile robot can receive a range of data from RFID that its position is known. The location of a robot in a regular PF can be defined by vector $x_k = [x_k \quad y_k]^t$. At the first step, the state estimation vector is $x_{est} = [0 \quad 0 \quad 0]^t$ with an estimate $\bar{x}_k$ of x_k. As mentioned in equations (1) and (2), the prediction and measurement models are calculated. Let's suppose, to describe a group of particles at instant k time:

$$s_k = \left[\left(x_k^i, w_k^i \right) \mid i = 1, 2, 3, \cdots, n_s \right],$$

$$w_k^i = \frac{p\left(x_k^i \mid z^k, u^k \right)}{p\left(x_k^i \mid u_k, x_{k-1} \right) p\left(x_{k-1} \mid z^{k-1}, u^{t-1} \right)},$$

$$w_k^i = \frac{\eta p\left(z^k \mid x_k^i \right) p\left(x_k^i \mid u_k, x_{k-1} \right) p\left(x_{k-1} \mid z^{k-1}, u^{k-1} \right)}{p\left(x_k^i \mid u_k, x_{k-1} \right) p\left(x_{k-1} \mid z^{k-1}, u^{k-1} \right)},$$

$$w_k^i = \eta p\left(z_k \mid x_k^i \right), \tag{23}$$

where s_k is the set of particles, the numerator is the target distribution, and the denominator is the distribution proposal. Moreover, η is the constant and $p(z_k \mid x_k^i)$ denotes the importance factor. For the state function, $f_k(x_k)$ is supposed to be a positive function where the PF algorithm produces the samples from $f_k(x_k)p(x_k \mid z^k, u^k)$ where at the initial stage, the samples are at position $f_0(x_0)$.

To calculate the innovative samples, a random particle x_{k-1}^i is computed from X_{k-1} and being distributed for n in relation to $f_{k-1}(x_{k-1})p(x_{k-1} \mid z^{k-1}, u^{k-1})$. At state $x_k^i \sim p(x_k \mid u_k, x_{k-1}^i)$, the importance weights w can be written with the $f_k(x_k)$ function.

$$w_k^i = \frac{f_k\left(x_k^i \right) p\left(x_k^i \mid z^k, u^k \right)}{f_{k-1}\left(x_{k-1}^i \right) p\left(x_k^i \mid u_k, x_{k-1} \right) p\left(x_{k-1} \mid z^{k-1}, u^{t-1} \right)},$$

$$w_k^i = \frac{f_k\left(x_k^i \right) \eta p\left(z^k \mid x_k^i \right) p\left(x_k^i \mid u_k, x_{k-1} \right) p\left(x_{k-1} \mid z^{k-1}, u^{k-1} \right)}{f_{k-1}\left(x_{k-1}^i \right) p\left(x_k^i \mid u_k, x_{k-1} \right) p\left(x_{k-1} \mid z^{k-1}, u^{k-1} \right)}. \tag{24}$$

To replace the constant of proportionality, the above equation can be written as

$$w_k^i \propto p\left(z^k \mid x_k^i \right) \times \frac{f_k\left(x_k^i \right)}{f_{k-1}\left(x_{k-1}^i \right)}. \tag{25}$$

The particle motion can be anticipated at the time k by applying the particles that contain the robot position or location, such as $x_k^i = f(x_{k-1}^i) + w_k$.

$$x_k^i = x_{k-1}^i + \begin{bmatrix} v_k \times T_s \times \cos \theta_k \\ v_k \times T_s \times \sin \theta_k \end{bmatrix} + w_k. \tag{26}$$
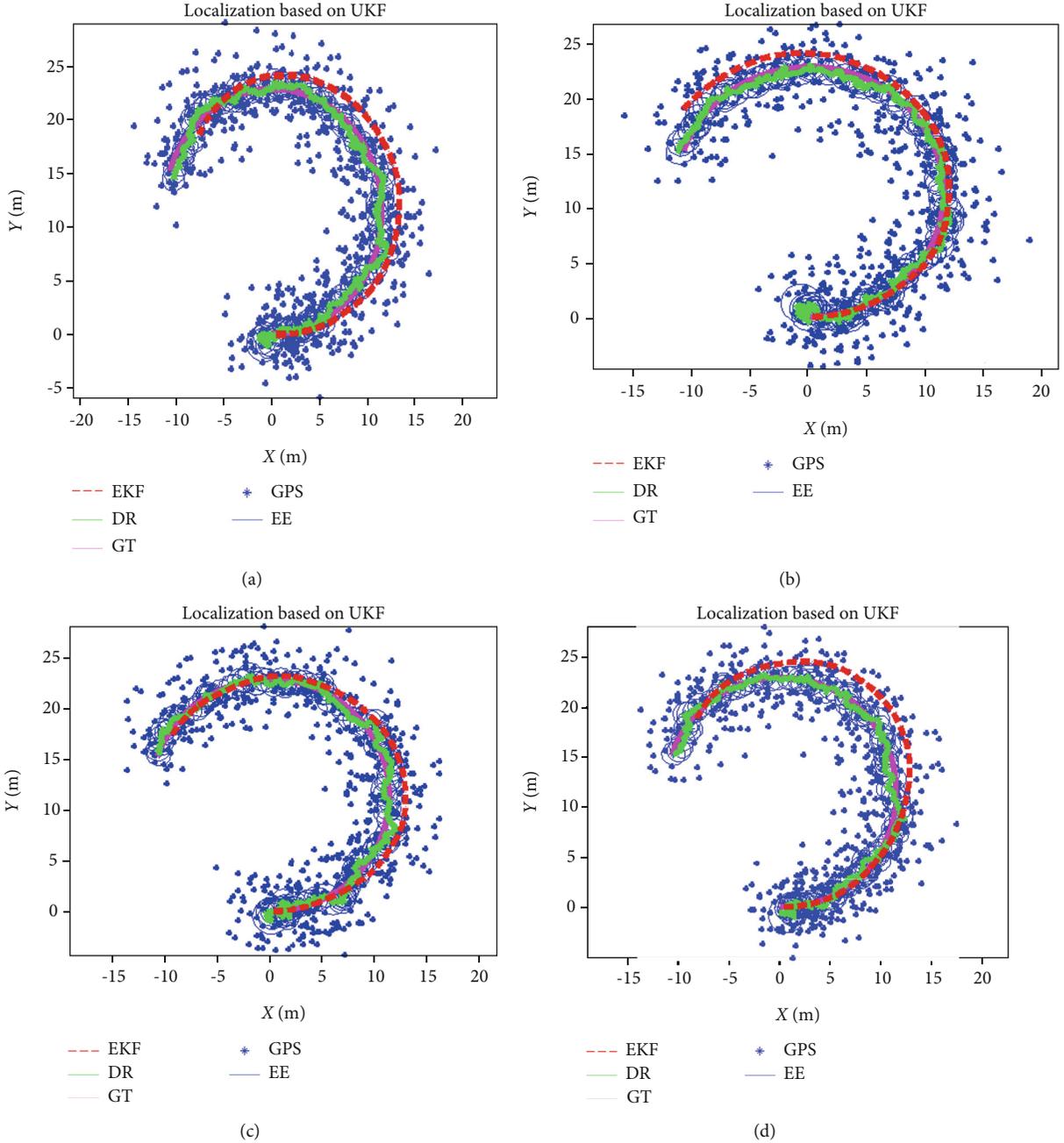
(a)

(b)

(c)

(d)

FIGURE 2: Localization comparison of UKF. In this phase, for all four iterations, the velocity is $v = 1$ m/s and time is $t = 60$ sec.

The state $\bar{x}_{k|k-1}$ estimate objectives can also be delineated by $\bar{x}_{k|k-1} = f(\bar{x}_{k-1|k-1})$.

$$p_{k|k-1} = \sum_{i=1}^{n_s} w_k^i \times \left[x_k^i - \bar{x}_{k|k-1}\right] \times \left[x_k^i - \bar{x}_{k|k-1}\right]^T + Q. \quad (27)$$

In the above equation, T_s is the robot sampling time, Q represents the covariance matrix, whereas the weight is represented by w. The successive approximation $p(x_k | z_{1:k})$ and the state $\bar{x}_{k|k}$ estimation can be calculated by:

$$p(x_k | z_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \times \delta\left(x_k - x_k^i\right), \quad (28)$$

$$\bar{x}_{k|k} = E[x_k | z_{1:k}], \quad (29)$$

$$\bar{x}_{k|k} = \int x_k \times p(x_k \mid z_{1:k}) dx_k \approx \sum_{i=1}^{N_s} w_k^i \times x_k^i, \qquad (30)$$

$$p_{k|k} = \sum_{i=1}^{N_s} w_k^i \times \left[x_k^i - \bar{x}_{k|k} \right] \times \left[x_k^i - \bar{x}_{k|k} \right]^T, \qquad (31)$$

where $\delta(.)$ is the function of Dirac delta. In the above equation, the subsequent approximation is set as a subsequent function which is approximated by N samples set [49].

$$F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad (32)$$

$$B = \begin{bmatrix} dt.\cos(x(3)) & 0 \\ dt.\sin(x(3)) & 0 \\ 0 & dt \end{bmatrix}. \qquad (33)$$

As mentioned above in equations (32) and (33), F and B denote the motion and measurement model state transition matrices.

4. Simulation Discussion and Comparison

In the above sections, the authors investigated the localization performance of the proposed EKF, UKF, and PF techniques. Each technique of localization performs well in its context; however, their efficiency differs from one another in some factors. A separate simulation is conducted for each localization technique, but some parameters are consistent through different iterations in these approaches. These parameters include such as those of the first case; the velocity is chosen as $v = 1$ m/s and the time is $T = 10$ sec as presented in Figures 1, 2, and 3, respectively. Table 2 shows the parameters used in the simulations. Furthermore, the initial time is chosen to be $t = 0$ and $t = 60$ sec which represent the end time, while the global time is chosen to be $dt = 0.1$ sec. For navigation, Dead Reckoning (DR) is the method of measuring one's current location by using a previously defined location, using speed and course estimates over time. On the other hand, Ground Truth (GT) is required for real-time localization. The first case state vector is $x_{\text{Est}} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^t$ where the true state can be $x_{\text{Tru}} = x_{\text{Est}}$. At the initial stage, the observation vector is calculated in such way $z = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}^t$; also, the matrix of covariance Q for motion and covariance matrix for observation R are calculated. While calculating Q and R, the sigma points are calculated. For the PF technique, the function $f_k(x_k)$ is used to calculate the importance of weights w_i. The subsequent approximation $p(x_k \mid z_{1:k})$ and the state estimation $\bar{x}_{k|k}$ are calculated in equations (28) to (31). The Jacobian matrices F of the state equation calculated in equation (32) and Jacobian of motion model JF is calculated in equation (33). The coverage area in all three methods is almost similar, while the localization accuracy is varying. Besides this, the time consumption for

all three techniques is dissimilar as can be seen in Table 3 where the velocity and time are reserved constant during four iterations.

In the next step, the velocity v is varied to $v = 2, v = 3, v = 4, v = 5, v = 6, \cdots, v = 10$ m/s. By the variable speed, the coverage range is enlarged as shown in Figures 4 and 5. The velocity variation is considered for the EKF and UKF techniques similarly, but the effect is different in each technique. The coverage of localization in the UKF-based technique is higher (see Figure 5) than the EKF algorithm (see Figure 4). In this phase, the localization coverage is higher, but unfortunately, the localization accuracy is also affected. When the robot is traveling with a higher speed during the process of localization, the mobile robot has a lower communication with the sensors in the surrounding, but if the robot is traveling with a lower speed, the process of communication is more reliable and convenient as compared to the case of high speed. Therefore, the author concludes that the proposed techniques are more accurate and reliable for lower velocities.

Furthermore, the proposed localization techniques' performance is evaluated in different scenarios. The performance is assessed by varying the time of T in all three techniques of localization. However, the velocity is kept constant in all three approaches as can be seen in Table 4. By changing the time T, the localization performance is also varying as can be seen in Figures 6, 7, and 8. In the case of EKF-based localization, by changing the time T, the localization coverage is decreasing as shown in Figure 6. Secondly, in the case of UKF-based localization, by varying the time T, the localization coverage is decreasing as presented in Figure 7. Similarly, by changing the time of T in the PF-based localization, the coverage area is decreasing gradually. It shows that the time T is inverse proportional to the coverage area of localization in all three techniques of localization. However, the time consumption is different in all three techniques as can be seen in Table 4. Among all three localization approaches, PF is the lowest time consumer as compared to other approaches. To compare the time consumption of our proposed PF technique with other techniques, the authors in [50] presented a PF technique Wi-Fi target localization. The technique is used to solve the localization problem of radio source by applying the RSSI measurements. The technique exploits the behavior of wireless signals in free space which obtains the estimates of positions from the signal strength. The authors considered the time-varying strategy to evaluate the performance of localization. The time is enlarged to $T = 1$ sec, $T = 20$ sec, $T = 50$ sec, and $T = 100$ sec, but unfortunately, by increasing the time more and more, the performance is gradually decreasing. However, in our proposed localization techniques, by varying the time T, the performance is still much better as compared to the other techniques.

As mentioned before, every method performs well in its domain, but concerning some aspects, their performance is varying. To compare the proposed methods, the authors evaluated their performance while considering several factors such as localization coverage, time consumption, and localization accuracy. In the PF algorithm,
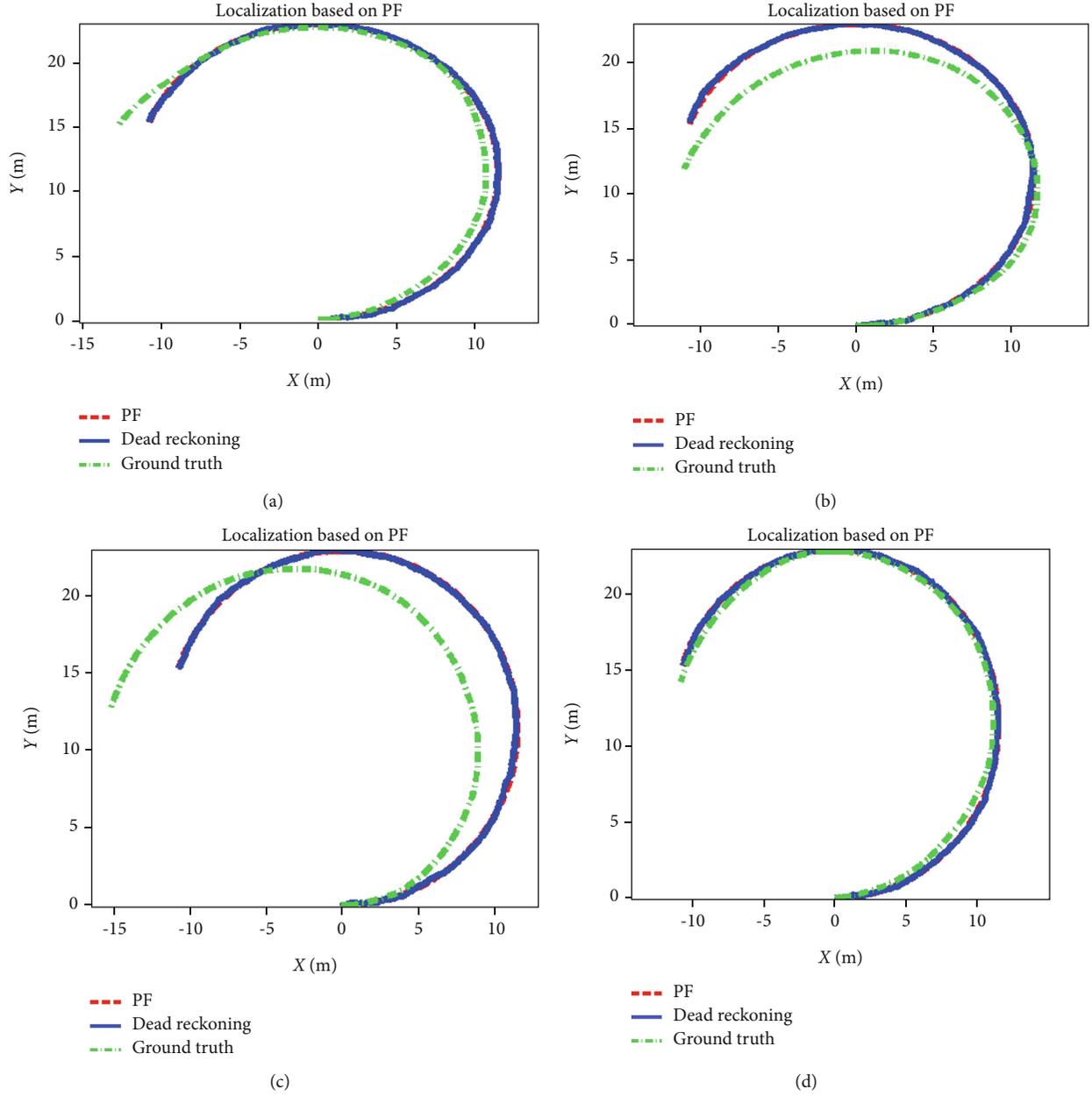
(a)

(b)

(c)

(d)

Figure 3: Localization comparison of PF technique. In this phase, for all four iterations, the velocity is $v = 1$ m/s and time is $t = 60$ sec.

the time consumption is lower than in the EKF and UKF algorithms as shown in Table 3. Moreover, UKF is less time consuming as compared to the EKF technique. However, the localization accuracy of EKF and UKF is higher than the PF localization technique. In comparison with other techniques, the performance of the proposed technique is better than the previous as investigated in the literature. A number of localization approaches are presented by the researchers [14, 43, 51, 52]. Each approach focused on the localization performance, but a limited number of aspects are considered such as most of them focused only on the accuracy of localization. However, our proposed methods consider several aspects at once such as the localization coverage, localization accuracy, and consumption of time. Therefore, to the best of

the author's knowledge, the proposed techniques are performing well in comparison with other techniques in this field.

5. Conclusion

To conclude, in this paper, the authors addressed three localization strategies based on EKF, UKF, and PF techniques. The authors evaluated the efficiency of the proposed techniques of localization by considering many factors such as the scope of localization, the accuracy of localization, and time of consumption. Basically, two steps are used to investigate the proposed localization techniques. Firstly, the authors kept the velocity of the robot constantly i.e., $v = 1$ m/s, and the process is repeated for four iterations as shown in

TABLE 2: Parameters used in simulations.

Parameters	Values
Time (T)	10 sec
Initial time (t)	0 sec
Final time (t)	60 sec
Global time (t)	0.1 sec
Initial velocity (v)	1.0 m/s
Updated velocities (v)	1, 2, 3,...,10 m/s
Degree to radian	180°
Yaw rate	5 deg/sec
Sigma points (i)	$i = 1, 2, 3, \cdots, n$
α	0.001
β	2
κ	0
Prediction covariance matrix (Q)	0.1
Observation covariance matrix (Q)	1
N_{eff}	1.0
Total range	360°
Weights (w_i)	$i = 1, 2, 3, \cdots, 2n$

TABLE 3: Comparison of the time consumption by EKF, UKF, and PF localization techniques.

Velocities	Iterations	Time$_{\text{EKF}}$	Time$_{\text{UKF}}$	Time$_{\text{PF}}$
	1	6.7431 sec	5.9950 sec	5.5588 sec
V1 = 1 m/s	2	6.7009 sec	5.9716 sec	5.5312 sec
	3	6.6051 sec	5.9627 sec	5.5436 sec
	4	6.4641 sec	5.9503 sec	5.5521 sec
V2 = 2 m/s	1	6.3744 sec	5.9005 sec	-
V3 = 3 m/s	1	6.3919 sec	5.9609 sec	-
V4 = 4 m/s	1	6.3835 sec	5.9306 sec	-
V5 = 5 m/s	1	6.1829 sec	5.9171 sec	-
V6 = 6 m/s	1	6.2834 sec	5.7584 sec	-
…	…	…	…	…
V10 = 10 m/s	1	6.2380 sec	5.8450 sec	-

Figures 1, 2, and 3. In the second case, the velocity of the robot is varied to $v = 2$ m/s, $v = 3$ m/s, and so on. As a result, the localization scope often differs by changing the velocity of the robot, as shown in Figure 4 (EKF) and Figure 5 (UKF). Furthermore, the authors evaluated the time consumption of the proposed localization techniques. Among all these techniques, the PF's time consumption is lower compared to the localization techniques of EKF and UKF as shown in Table 2. However, the localization accuracy of EKF and UKF is better than the PF-based localization. Finally, the proposed methods are compared with each other's and also with other standard approaches. Therefore, the proposed localization methods performed better as compared to the other techniques as mentioned in the above section of comparison.
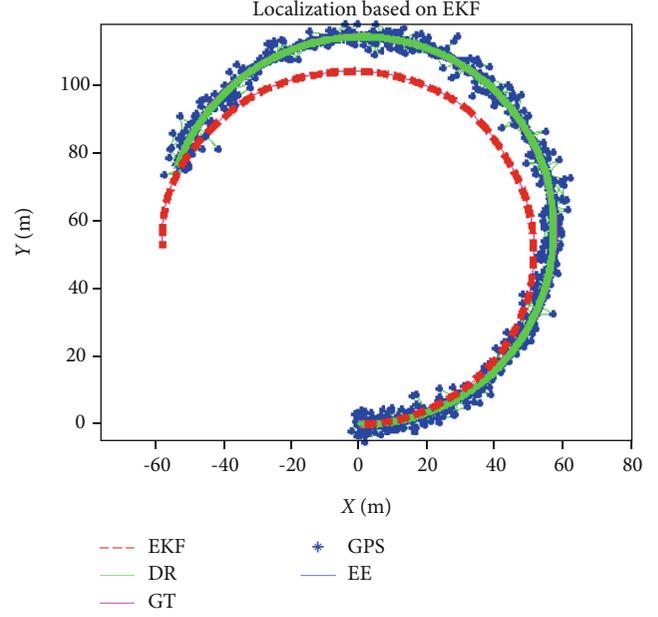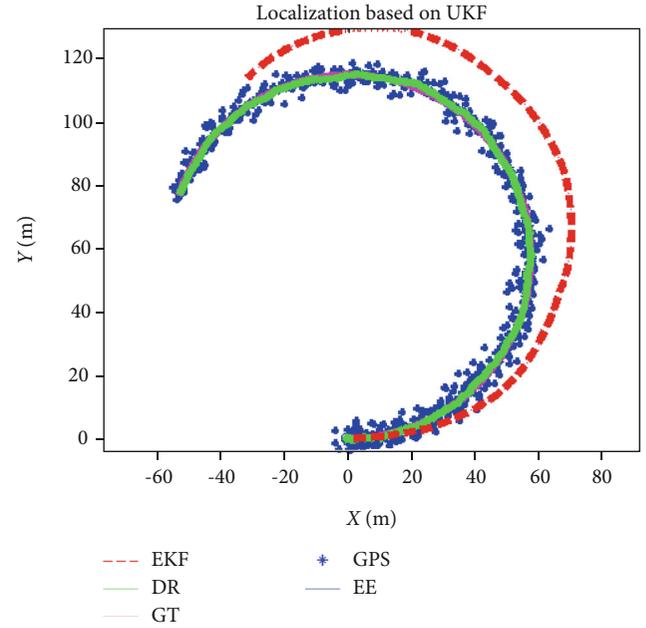

FIGURE 4: EKF localization with velocity $v = 5$ m/s.

FIGURE 5: UKF localization with velocity $v = 5$ m/s.

TABLE 4: Performance comparison of EKF, UKF, and PF localization algorithms by varying the time T.

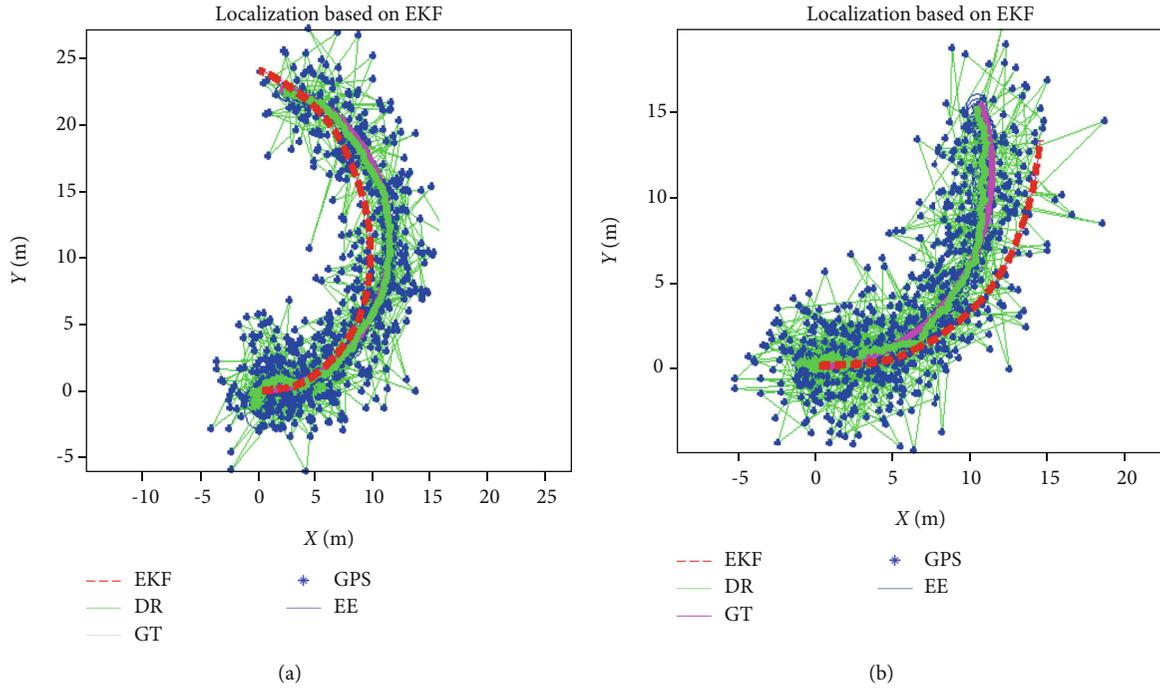| Time (T)	Velocity (V)	Time$_{\text{EKF}}$	Time$_{\text{UKF}}$	Time$_{\text{PF}}$
$T = 10$ sec	$V = 1$ m/s	6.3845 sec	5.9253 sec	5.5112 sec
$T = 30$ sec	$V = 1$ m/s	6.5245 sec	5.7664 sec	5.3805 sec
$T = 60$ sec	$V = 1$ m/s	6.6377 sec	5.7263 sec	5.7213 sec
$T = 90$ sec	$V = 1$ m/s	6.9117 sec	5.7601 sec	5.1393 sec
$T = 120$ sec	$V = 1$ m/s	6.2887 sec	5.8666 sec	5.2038 sec

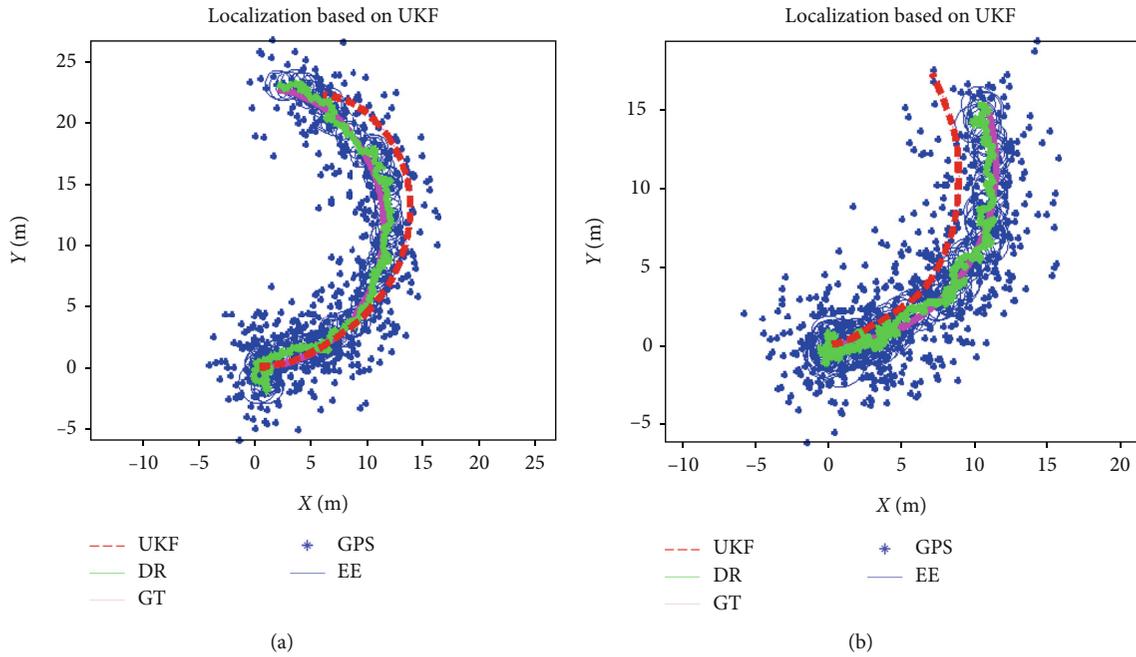FIGURE 6: Localization performance of EKF with (a) $T = 30$ sec and (b) $T = 90$ sec.

FIGURE 7: Localization performance of UKF with (a) $T = 30$ sec and (b) $T = 90$ sec.
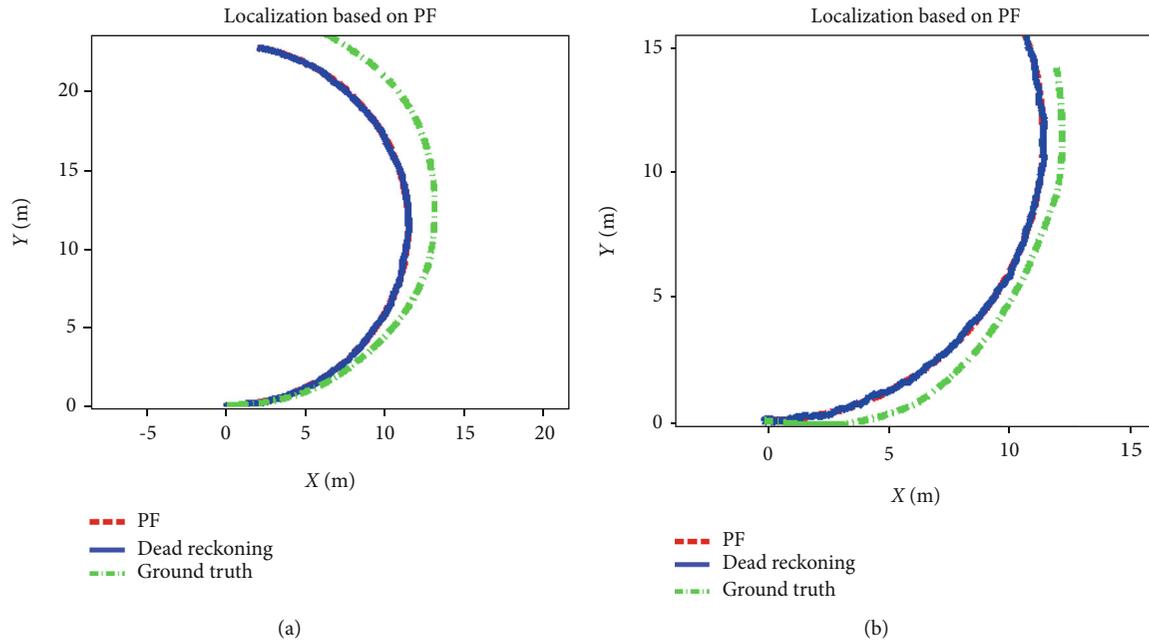
FIGURE 8: Localization performance of PF with (a) $T = 30$ sec and (b) $T = 90$ sec.

In the future, the authors will perform more experiments to study the effects and performance of these localization methods. In addition, our future study will also have a look on the performance of these localization techniques in combination with the simultaneous robotic localization and mapping.

Data Availability

Since the funding project is not closed and related patents have been evaluated, the simulation data used to support the findings of this study are currently under embargo while the research findings are commercialized. Requests for data, based on the approval of patents after project closure, will be considered by the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] R. Khan, S. U. Khan, S. Khan, and M. U. A. Khan, "Localization performance evaluation of extended Kalman filter in wireless sensors network," *Procedia Computer Science*, vol. 32, pp. 117–124, 2014.

[2] I. Ullah, J. Chen, X. Su, C. Esposito, and C. Choi, "Localization and detection of targets in underwater wireless sensor using distance and angle based algorithms," *IEEE Access*, vol. 7, pp. 45693–45704, 2019.

[3] K. Sha, T. A. Yang, W. Wei, and S. Davari, "A survey of edge computing based designs for IoT security," *Digital Communications and Networks*, vol. 6, no. 2, pp. 195–202, 2020.

[4] X. Su, H. Yu, W. Kim, C. Choi, and D. Choi, "Interference cancellation for non-orthogonal multiple access used in future wireless mobile networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, 2016.

[5] A. Shareef, Y. Zhu, V. M. Moreno, and A. Pigazo, "Localization using extended Kalman filters in wireless sensor networks," *Kalman Filter Recent Advances and Applications*, pp. 297–320, 2009.

[6] P. Thulasiraman and K. A. White, "Topology control of tactical wireless sensor networks using energy efficient zone routing," *Digital Communications and Networks*, vol. 2, no. 1, pp. 1–14, 2016.

[7] L. Zhu, Y. Li, F. R. Yu, B. Ning, T. Tang, and X. Wang, "Cross-layer defense methods for jamming-resistant CBTC systems," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[8] I. Ullah, X. Su, X. Zhang, and D. Choi, "Simultaneous localization and mapping based on Kalman filter and extended Kalman filter," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 2138643, 12 pages, 2020.

[9] W. Yu, J. Peng, X. Zhang, S. Li, and W. Liu, "An adaptive unscented particle filter algorithm through relative entropy for mobile robot self-localization," *Mathematical Problems in Engineering*, vol. 2013, Article ID 567373, 9 pages, 2013.

[10] E. Ivanjko, A. Kitanov, and I. Petrovic, "Model based Kalman filter mobile robot self-localization," *Robot Localization and Map Building*, pp. 59–89, 2010.

[11] S. Safavat, N. N. Sapavath, and D. B. Rawat, "Recent advances in mobile edge computing and content caching," *Digital Communications and Networks*, vol. 6, no. 2, pp. 189–194, 2020.

[12] L. Teslić, I. Škrjanc, and G. Klančar, "EKF-based localization of a wheeled mobile robot in structured environments," *Journal of Intelligent & Robotic Systems*, vol. 62, no. 2, pp. 187–203, 2011.

[13] R. Kandepu, B. Foss, and L. Imsland, "Applying the unscented Kalman filter for nonlinear state estimation," *Journal of Process Control*, vol. 18, no. 7-8, pp. 753–768, 2008.

[14] A. Giannitrapani, N. Ceccarelli, F. Scortecci, and A. Garulli, "Comparison of EKF and UKF for spacecraft localization via angle measurements," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 1, pp. 75–84, 2011.

[15] J. M. Pak, C. K. Ahn, P. Shi, Y. S. Shmaliy, and M. T. Lim, "Distributed hybrid particle/fir filtering for mitigating NLOS effects in toa-based localization using wireless sensor networks," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 5182–5191, 2017.

[16] I. Ullah, Y. Liu, X. Su, and P. Kim, "Efficient and accurate target localization in underwater environment," *IEEE Access*, vol. 7, pp. 101415–101426, 2019.

[17] W. R. Gilks and C. Berzuini, "Following a moving target-Monte Carlo inference for dynamic Bayesian models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 1, pp. 127–146, 2001.

[18] J. M. Pak, C. K. Ahn, Y. S. Shmaliy, and M. T. Lim, "Improving reliability of particle filter-based localization in wireless sensor networks via hybrid particle/fir filtering," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 5, pp. 1089–1098, 2015.

[19] C. K. Ahn, P. Shi, and M. V. Basin, "Deadbeat dissipative fir filtering," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 63, no. 8, pp. 1210–1221, 2016.

[20] M. Raja, "Application of cognitive radio and interference cancellation in the L-band based on future air-to-ground communication systems," *Digital Communications and Networks*, vol. 5, no. 2, pp. 111–120, 2019.

[21] C. K. Ahn, "Strictly passive FIR filtering for state-space models with external disturbance," *AEU-International Journal of Electronics and Communications*, vol. 66, no. 11, pp. 944–948, 2012.

[22] J. M. Pak, S. Y. Yoo, M. T. Lim, and M. K. Song, "Weighted average extended FIR filter bank to manage the horizon size in nonlinear fir filtering," *International Journal of Control, Automation and Systems*, vol. 13, no. 1, pp. 138–145, 2015.

[23] P. M. Pradhan and G. Panda, "S-transformation based integrated approach for spectrum estimation, storage, and sensing in cognitive radio," *Digital Communications and Networks*, vol. 5, no. 3, pp. 160–169, 2019.

[24] J. M. Pak, C. K. Ahn, Y. S. Shmaliy, P. Shi, and M. T. Lim, "Switching extensible fir filter bank for adaptive horizon state estimation with application," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 3, pp. 1052–1058, 2016.

[25] Y. Li, S. Xia, B. Cao, and Q. Liu, "Lyapunov optimization based trade-off policy for mobile cloud offloading in heterogeneous wireless networks," *IEEE Transactions on Cloud Computing*, 2019.

[26] W. Cui, B. Li, L. Zhang, and W. Meng, "Robust mobile location estimation in NLOS environment using GMM, IMM, and EKF," *IEEE Systems Journal*, vol. 13, no. 3, pp. 3490–3500, 2018.

[27] T.-J. Ho, "Robust urban wireless localization: synergy between data fusion, modeling and intelligent estimation," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 685–697, 2014.

[28] M. Lipka, E. Sippel, and M. Vossiek, "An extended Kalman filter for direct, real-time, phase-based high precision indoor localization," *IEEE Access*, vol. 7, pp. 25288–25297, 2019.

[29] L. Zhu, Y. He, F. R. Yu, B. Ning, T. Tang, and N. Zhao, "Communication-based train control system performance optimization using deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10705–10717, 2017.

[30] K. Lee, J. Oh, and K. You, "TDOA-/FDOA-based adaptive active target localization using iterated dual-EKF algorithm," *IEEE Communications Letters*, vol. 23, no. 4, pp. 752–755, 2019.

[31] X. Su, I. Ullah, X. Liu, and D. Choi, "A review of underwater localization techniques, algorithms, and challenges," *Journal of Sensors*, vol. 2020, Article ID 6403161, 24 pages, 2020.

[32] Y. Li, J. Liu, B. Cao, and C. Wang, "Joint optimization of radio and virtual machine resources with uncertain user demands in mobile cloud computing," *IEEE Transactions on Multimedia*, vol. 20, no. 9, pp. 2427–2438, 2018.

[33] Y. Wang, H. Jie, and L. Cheng, "A fusion localization method based on a robust extended Kalman filter and track-quality for wireless sensor networks," *Sensors*, vol. 19, no. 17, p. 3638, 2019.

[34] C.-D. Wann, Y.-J. Yeh, and C.-S. Hsueh, "Hybrid TDOA/AOA indoor positioning and tracking using extended Kalman filters," in *2006 IEEE 63rd Vehicular Technology Conference*, pp. 1058–1062, Melbourne, Vic., Australia, May 2006.

[35] X. Su, K. Fan, and W. Shi, "Privacy-preserving distributed data fusion based on attribute protection," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 10, pp. 5765–5777, 2019.

[36] X. Ou, X. Wu, X. He, Z. Chen, and Q.-A. Yu, "An improved node localization based on adaptive iterated unscented Kalman filter for WSN," in *2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)*, pp. 393–398, Auckland, New Zealand, June 2015.

[37] P. Del Moral, A. Doucet, and A. Jasra, "An adaptive sequential Monte Carlo method for approximate Bayesian computation," *Statistics and Computing*, vol. 22, no. 5, pp. 1009–1020, 2012.

[38] L. Mihaylova, A. Y. Carmi, F. Septier, A. Gning, S. K. Pang, and S. Godsill, "Overview of bayesian sequential Monte Carlo methods for group and extended object tracking," *Digital Signal Processing*, vol. 25, pp. 1–16, 2014.

[39] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[40] J. Hammersley, *Monte Carlo Methods*, Springer Science & Business Media, 2013.

[41] S. Kluge, K. Reif, and M. Brokate, "Stochastic stability of the extended Kalman filter with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 55, no. 2, pp. 514–518, 2010.

[42] L. Zhu, F. R. Yu, B. Ning, and T. Tang, "Cross-layer handoff design in MIMO-enabled WLANs for communication-based train control (CBTC) systems," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 719–728, 2012.

[43] L. D. Alfonso, A. Grano, P. Muraca, and P. Pugliese, "Extended and unscented Kalman filters in a cells-covering method for environment reconstruction," in *2019 IEEE 15th International Conference on Control and Automation (ICCA)*, pp. 1557–1562, Edinburgh, United Kingdom, United Kingdom, July 2019.

[44] M. Moreno-Cano, M. A. Zamora-Izquierdo, J. Santa, and A. F. Skarmeta, "An indoor localization system based on artificial

neural networks and particle filters applied to intelligent buildings," *Neurocomputing*, vol. 122, pp. 116–125, 2013.

[45] D. Fox, "Kld-sampling: adaptive particle filters," in *Advances in Neural Information Processing Systems: Natural and Synthetic*, Vancouver, British Columbia, Canada, December 3-8, 2001.

[46] C.-H. Park and J.-H. Chang, "Robust LMedS-based WLS and Tukey-based EKF algorithms under LOS/NLOS mixture conditions," *IEEE Access*, vol. 7, pp. 148198–148207, 2019.

[47] B. T. Burchett, "Unscented Kalman filters for range-only cooperative localization of swarms of munitions in three-dimensional flight," *Aerospace Science and Technology*, vol. 85, pp. 259–269, 2019.

[48] I. Ullah, Y. Shen, X. Su, C. Esposito, and C. Choi, "A localization based on unscented Kalman filter and particle filter localization algorithms," *IEEE Access*, vol. 8, pp. 2233–2246, 2019.

[49] B.-F. Wu and C.-L. Jen, "Particle-filter-based radio localization for mobile robots in the environments with low-density wlan aps," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 12, pp. 6860–6870, 2014.

[50] N. Wagle and E. Frew, "A particle filter approach to wifi target localization," in *AIAA guidance, navigation, and control conference*, p. 7750, Toronto, Ontario, Canada, August 2010.

[51] C. Yang, W. Shi, and W. Chen, "Comparison of unscented and extended Kalman filters with application in vehicle navigation," *The Journal of Navigation*, vol. 70, no. 2, pp. 411–431, 2017.

[52] K. Ahn and Y. Kang, "A particle filter localization method using 2d laser sensor measurements and road features for autonomous vehicle," *Journal of Advanced Transportation*, vol. 2019, Article ID 3680181, 11 pages, 2019.