

# Toward Reliable and Scalable Internet of Vehicles: Performance Analysis and Resource Management

*This article concerns how to ensure reliable and scalable wireless transmissions for IoV based on performance modeling and analysis.*

By YUANZHI NI, Member IEEE, LIN CAI<sup>1</sup>, Senior Member IEEE, JIANPING HE<sup>2</sup>, Member IEEE, ALEXEY VINEL<sup>3</sup>, Senior Member IEEE, YUE LI<sup>4</sup>, HAMED MOSAVAT-JAHROMI, Student Member IEEE, AND JIANPING PAN<sup>5</sup>, Senior Member IEEE

**ABSTRACT** | Reliable and scalable wireless transmissions for Internet of Vehicles (IoV) are technically challenging. Each vehicle, from driver-assisted to automated one, will generate a flood of information, up to thousands of times of that by a person. Vehicle density may change drastically over time and location. Emergency messages and real-time cooperative control messages have stringent delay constraints while infotainment applications may tolerate a certain degree of latency. On a congested road, thousands of vehicles need to exchange information badly, only to find that service is limited

due to the scarcity of wireless spectrum. Considering the service requirements of heterogeneous IoV applications, service guarantee relies on an in-depth understanding of network performance and innovations in wireless resource management leveraging the mobility of vehicles, which are addressed in this article. For single-hop transmissions, we study and compare the performance of vehicle-to-vehicle (V2V) beacon broadcasting using random access-based (IEEE 802.11p) and resource allocation-based (cellular vehicle-to-everything) protocols, and the enhancement strategies using distributed congestion control. For messages propagated in IoV using multihop V2V relay transmissions, the fundamental network connectivity property of 1-D and 2-D roads is given. To have a message delivered farther away in a sparse, disconnected V2V network, vehicles can carry and forward the message, with the help of infrastructure if possible. The optimal locations to deploy different types of roadside infrastructures, including storage-only devices and roadside units with Internet connections, are analyzed.

**KEYWORDS** | Internet of Vehicles (IoV); reliability; scalability; wireless communications.

Manuscript received June 2, 2019; revised October 1, 2019; accepted October 19, 2019. Date of publication December 23, 2019; date of current version January 22, 2020. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), Compute Canada, Chinese Scholarship Council (CSC), the 111 Project, under Grant B12018; in part by the Natural Science Foundation of China (NSFC) under Grant 61973218; in part by the Knowledge Foundation in the framework of SafeSmart "Safety of Connected Intelligent Vehicles in Smart Cities" Synergy Project from 2019 to 2023; in part by the Swedish Foundation for Strategic Research (SSF) in the framework of Strategic Mobility Program from 2019 to 2020; and in part by the Excellence Center at Linköping-Lund in Information Technology (ELLIIT) Strategic Research Network. (Corresponding author: Lin Cai.)

**Y. Ni** is with the School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China (e-mail: niyuanzhi@jiangnan.edu.cn).

**L. Cai, Y. Li, and H. Mosavat-Jahromi** are with the Department of Electrical and Computer Engineering, University of Victoria, Victoria, V8W 3P6, Canada (e-mail: cai@ece.uvic.ca; liyue331@uvic.ca; hamed.mosavat@gmail.com).

**J. He** is with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: jphe@sjtu.edu.cn).

**A. Vinel** is with the School of Information Technology, Halmstad University, Halmstad 301 18, Sweden, and also with the Department of Electrical Engineering, Western Norway University of Applied Sciences, Bergen 5020, Norway (e-mail: alexey.vinel@hh.se).

**J. Pan** is with the Department of Computer Science, University of Victoria, Victoria, BC V8W 3P6, Canada (e-mail: pan@uvic.ca).

Digital Object Identifier 10.1109/JPROC.2019.2950349

## I. INTRODUCTION

"No great civilization has been built without some well-defined system of transportation" [1]. Modern society depends heavily on modern transportation to provide a fast, safe, clean, and low-cost way to deliver materials, goods, and people. As the world population hits 7.7 billion, with the ever-growing mobility demands, today, modern transportation faces grand challenges [2]: 1) safety—car

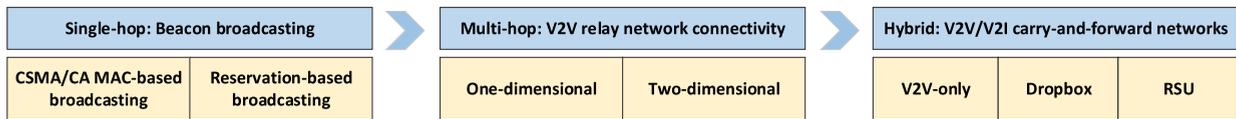


Fig. 1. Reliable and scalable IoV overview.

accidents top the cause of death for ages 8–24 [3]; 2) mobility—each commuter wastes tens to hundreds of hours per year in traffic; 3) environment—wasted fuel topped tens of billions of gallons; and 4) space—today’s cars sit unused 95% of the time, demanding a large amount of parking spaces.

Autonomous driving holds rosy promises, from car-sharing to saving hundreds of hours per year per driver, thus increasing productivity and reducing environmental impacts greatly. Current autonomous driving technologies, relying on expensive onboard sensing devices and sophisticated local data processing, were shadowed by a series of fatal accidents in 2018 and 2019 [4], [5]. Standalone vehicles, even with the most advanced artificial intelligence (AI) technologies, can hardly beat an experienced human driver with social learning skills. Alternatively, vehicles can exchange information using low-cost wireless technologies. With abundant information from other vehicles, pedestrians, roadside infrastructure, and clouds, each connected vehicle (new or even old one) can see and sense much better than any individual driver or car and make more intelligent decisions to ensure safety, improve transportation and energy efficiency, reduce emission, and provide infotainment services to people on the road [6]. There is no doubt that Internet of Vehicles (IoV) based on the vehicle-to-everything (V2X) technologies will fundamentally change our daily lives and disrupt our societies, where V2X is the umbrella term including vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), vehicle-to-cloud (V2C), vehicle-to-pedestrian (V2P), and vehicle to anything else [7], [8]. The V2X communications are mainly based on two technologies, that is, dedicated short-range communications (DSRC) and cellular networks [9], [10]. In general, IoV combines the traditional transportation infrastructure with the V2X technology to facilitate a safe, green, efficient travel and delivery system [11].

Reliable and scalable wireless transmissions for IoV are technically difficult, because: 1) each vehicle, from driver-assisted to automated one, will generate a flood of information, up to thousands of times of that by a person; 2) vehicle density may change drastically over time (off versus peak hours) and location (main versus side streets); 3) emergency and real-time cooperative control messages have stringent delay constraints (a few milliseconds), while infotainment applications may tolerate a certain degree of latency; and 4) in a congested road during rush hours or due to an accident, thousands of vehicles need

to exchange information badly, only to find that service is limited or unavailable due to the scarcity of wireless spectrum. These technical difficulties present a grand challenge to wireless communication system design [12].

Considering heterogeneous requirements of IoV applications, service guarantee relies on an in-depth understanding of network performance and major innovations in wireless resource management, which are addressed in this article. Vehicles need to exchange their state information periodically through beacon broadcasting, and reliable broadcasting requires a careful design and analysis of the medium access control (MAC) protocols. For emergency or safety messages, low-latency message flooding among connected vehicles relies on the network connectivity property, in 1-D highway or 2-D roads. For delay-tolerant messages to reach a faraway location, we can apply the carry-and-forward strategy, using vehicles as message carriers. In these cases, the delay requirement in IoV is coupled with the locations of destinations, so reliability in IoV is related to the chance that a message can be delivered to the destinations (e.g., vehicles at certain locations) before a location-dependent deadline. Given the time- and location-coupled reliability requirements, fundamental network properties and performance metrics should be quantified by rigorous modeling and analysis. Finally, given the insights from the analysis, effective wireless resource management and congestion control solutions should be applied to ensure reliability and scalability, considering the scenarios of V2V-only and hybrid V2V/V2I networks.

As shown in Fig. 1, this article focuses on the reliability and scalability issues in IoV including the beacon broadcasting, V2V relay network connectivity, and V2V/V2I carry-and-forward networks. In Section II, beacon broadcasting using random access-based MAC (IEEE 802.11p) and resource allocation-based MAC (C-V2X) is investigated and compared, and the enhancement strategies using distributed congestion control are discussed. In Section III, the 1-D and 2-D connectivity of V2V relay networks are modeled and evaluated. In Section IV, the carry-and-forward mechanism is applied in V2V networks or hybrid V2V/V2I networks. The optimal locations to deploy different types of roadside infrastructures, including storage-only devices and roadside units with Internet connections, are analyzed, followed by concluding remarks in Section V. Note that we focus on the performance analysis and resource management solutions in the MAC and above layers, assuming that the PHY layer

can ensure that any transmission within the transmission range can be successful with a high probability, if without collision. Thus, the PHY-layer details will not be discussed.

## II. SINGLE-HOP BEACON BROADCASTING

Broadcasting of messages to immediate neighbors is a basic mode in V2V communications. For example, vehicles that are located in their direct communication range can inform surroundings about their current position, velocity, and acceleration by sending beacon messages. Beacon broadcasting can use a one-hop communication mode with omnidirectional antennas without a need for prior connection establishment, which enables low delay. Moreover, the simplicity of the approach makes it attractive for practical implementations with beacons being a universal facility for V2V communications in diverse cooperative automated driving scenarios.

### A. Carrier Sense Multiple Access With Collision Avoidance (CSMA/CA) MAC-Based Beacon Broadcasting

CSMA/CA is commonly considered for V2V beacon broadcasting. Both the DSRC standard and the European Telecommunications Standards Institute (ETSI) standard use the IEEE 802.11p MAC protocol based on CSMA/CA. In these types of random multiple access methods, a vehicle decides on the current occupancy of the channel via carrier sensing. The channel is regarded to be busy or not after an observation of its signal strength during a short time slot with duration  $\sigma$ . Note that  $\sigma \ll \tau$ , where  $\tau$  is the channel occupancy time for a beacon transmission. The access to the channel is governed by the *backoff counter* computed independently by each vehicle for every broadcast transmission. Its value is initiated via a uniform random choice from the interval  $[0, W - 1]$ , where  $W$  is a constant used by all the vehicles and referred to as a *contention window*. A vehicle decrements its backoff counter either after an idle channel period of duration  $\sigma$  or busy channel of duration  $\tau$ . If the channel is busy, either the transmission of a vehicle occurs or there are simultaneous interfering transmissions of multiple vehicles which normally assumes reception impossible (*collision*). Broadcast messages are neither acknowledged nor retransmitted.

1) *Age-of-Information-Based Broadcasting Reliability*: Vehicles will broadcast beacons periodically, so a lost beacon will result in an “aged” beacon information at the receiver side. Then, the reliability can be defined based on *age-of-information*, which measures the time elapsed since the most recently received beacon was generated. The state-of-the-art metric for CSMA/CA-based beacon broadcasting reliability analysis is the cumulative probability distribution function  $Q$  of age-of-information, which links communication reliability with the automatic networked control systems [13].

*Definition 1*: For any vehicle  $I$  in a communication range of vehicle  $J$ , the age-of-information  $D(J \rightarrow I)$  about  $J$  from the perspective of  $I$  is a random variable that represents the time elapsed since the last successfully received beacon of vehicle  $J$  by vehicle  $I$ .

The behavior of such a CSMA network with  $N$  vehicles, where each one generates  $f$  beacons per second and all vehicles are within each other’s transmission range, can be evaluated via the discrete-time embedded Markov chain  $\{n(t), s(t)\}$ , where  $n(t) \in [0, N]$  is the number of vehicles with beacons to transmit at moment  $t$ , and  $s(t) \in [0, n(t)]$  is the number of vehicles transmitting beacons at  $t$  using the same channel. The state transitions occur either after empty slots of duration  $\sigma$  or busy ones of duration  $\tau$ . Let  $\rho_{n,s}^{(f,N,\tau,\sigma,W)}$  be a stationary distribution of the above Markov chain obtained for the system parameters  $(f, N, \sigma, \tau, W)$ . For the transition probabilities of the chain, refer to [14].

*Theorem 1*: In a CSMA/CA network with  $N$  vehicles and all within each other’s transmission range, for a pair of vehicles  $J \rightarrow I$ , each having intensity of beacon arrivals  $f$ , the probability that the age-of-information does not exceed  $T_{\max}$  is

$$Q = \Pr\{D(J \rightarrow I) \leq T_{\max}\} = (1 - p_c) \sum_{i=1}^{\lfloor T_{\max} \cdot f \rfloor} \frac{i}{f} \cdot p_c^{i-1} \quad (1)$$

where

$$p_c = \frac{\sum_{n=2}^N \sum_{s=2}^n \rho_{n,s}^{(f,N,\tau,\sigma,W)}}{1 - \sum_{n=0}^N \rho_{n,0}^{(f,N,\tau,\sigma,W)}}$$

is the beacon collision probability.

The above analysis assumes that all vehicles share the same channel and are within each other’s transmission and sensing range. A more sophisticated derivation of collision probability considering hidden terminals can be found in [15]. To reduce the probability of beacon collisions, one can couple the generation frequency  $f$  with kinematics of a vehicle [16] or apply decentralized congestion control (DCC) described below.

2) *DCC*: During traffic jam, a congested highway with multiple lanes can easily reach a vehicle density of up to thousands of vehicles per  $\text{km}^2$ . When all of these vehicles need to exchange beacons with their neighbors, severe collisions in congested wireless networks may occur and lead to high packet loss rate and intolerably long age-of-information, defeating the purpose of many safety applications. In addition, congestion may occur when more beacons are needed due to vehicle and driver emergency. Essentially, we need more beacons to ensure reliability in such scenarios, but more beacons cause a higher contention overhead, thus further reducing the network throughput.

The current V2X communication systems have been identified as nonscalable due to the lack of effective

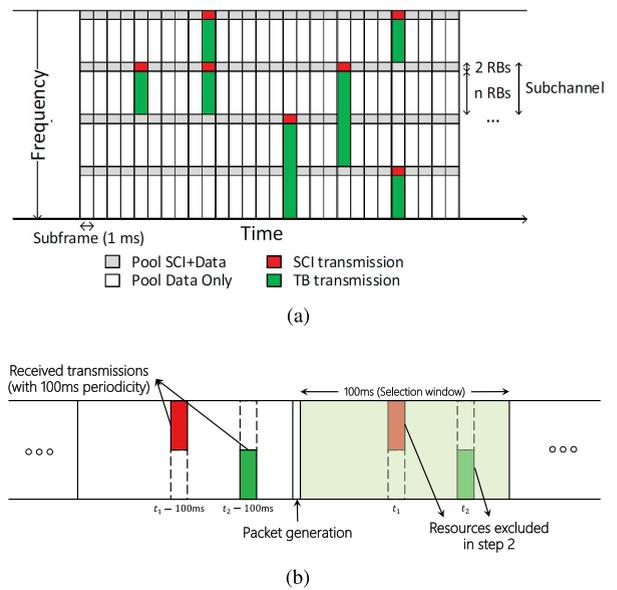
congestion control. Congestion may become severe when the V2X technologies reach a high penetration rate. It is well known that random access-based MAC protocols may suffer high loss rate due to collisions when the traffic intensity is high.

To address the scalability, we should ensure that the aggregated V2X traffic does not exceed the network capacity, requiring an effective approach to control the beacon rate. The asynchronous DCC has been adopted in the ETSI standard [17]. DCC uses the measurement of the load on the wireless channel, i.e., the channel busy ratio (CBR), to adjust the repetition rate of periodic safety messages. In [18], an adaptive congestion control algorithm, LLinear MMessage Rate Integrated Control (LIMERIC) algorithm, was proposed. LIMERIC takes advantage of full-precision control inputs available on the wireless channel, aiming to converge to a fair and efficient channel state quickly. Simulation results [19], [20] showed that, even with these congestion control schemes, given the nature of high collision loss rate, the random access-based MAC protocols fail to maintain high efficiency, high reliability, and low latency in high-density vehicular networks. For instance, it was reported that the total number of successfully received packets with DCC could be even lower than the plain IEEE 802.11p protocol under certain conditions [21].

## B. Reservation-Based Beacon Broadcasting

Another promising solution for IoV is the cellular-V2X (C-V2X) approach, which can use the existing cellular infrastructure to achieve a centralized or semicentralized control. In C-V2X, the access and transmission resource can be preconfigured per user. Compared with CSMA/CA, C-V2X reduces collisions for better QoS provisioning.

1) *C-V2X Standardization*: In the third generation partnership project (3GPP), the technical specification group on radio access network (TSG-RAN) has completed specification work supporting basic vehicular services based on LTE sidelink in release 14 (Rel-14). Two scheduling configurations are included, i.e., the eNB-assisted scheduling (mode-3) and the distributed scheduling (mode-4), where eNB is similar to base station in traditional cellular networks. Mode-3 is based on the centralized eNB-scheduled resource allocation and interference management. Both V2I and V2V communications are supported in this mode because every transmission is scheduled by eNB's grants, including the modulation, coding, resource allocation, and power control, regardless of whether the transmission is between vehicles or to/from the eNB. We focus on mode-4, which handles the out-of-coverage scenario, where the wireless links between the eNB and vehicles are not available, and thus only the V2V communications are supported. For mode-4, compared with LTE sidelink, a more concise and compact resource allocation scheme is introduced for physical sidelink control channel (PSCCH) and physical sidelink shared channel (PSSCH) [22], as shown in Fig. 2(a). This design can



**Fig. 2. Resource allocation in C-V2X. (a) Resource arrangement [25]. (b) Sensing and SPS.**

enhance the system performance in high density and consider the latency requirements.

In C-V2X, the spectrum band is divided into subchannels, and each subchannel consists of a sidelink control information (SCI) part and a transport block (TB) transmission part. The number of resource blocks (RBs) included in one subchannel may vary while the SCI part occupies two RBs. The TB part contains the full packet which needs to be sent. The RBs used for the TB transmission and the corresponding modulation and coding scheme (MCS) are notified by the SCI. The SCI also includes a reselection counter telling others when this vehicle is going to release the occupied resource and perform the reselection procedure.

The semipersistent scheduling (SPS) and a sensing-oriented access scheme have been introduced for mode-4 C-V2X users [23], [24], specifically, to further reduce collisions. The vehicular traffic for beaconing is periodic in nature, which is utilized to estimate the congestion on a specific resource. In C-V2X, a vehicle shall always scan and sense the whole spectrum band. When a packet is generated, the vehicle searches all the available resources in a selection window and excludes those occupied by others, as shown in Fig. 2(b). Then, the vehicle randomly selects one from the rest of the available resources and starts to transmit its own SCI and packet. The length of the selection window is determined by the maximum delay. Several criteria are used to exclude resources in a selection window, such as when the average reference signal received power (RSRP) over a specific resource is higher than a threshold, or a resource is explicitly excluded by the received reselection counter (carried by SCI).

Although the 3GPP made great efforts in collision avoidance for mode-4 C-V2X users, the collision cannot be resolved when multiple vehicles choose the same resource in the selection window. Furthermore, the hidden terminal problem cannot be solved by this sense-then-transmit approach. In autonomous driving, the V2V transmission range may need to be increased to a more comprehensive surrounding situation, which naturally intensifies the collisions on air interface. In the current C-V2X system, not only are collisions inevitable as analyzed above, but also there is no hybrid automatic repeat request (HARQ) feedback for broadcast [23]. Although blind HARQ retransmissions are used to improve the reliability, without HARQ feedback, collisions are undetectable and cannot be stopped timely, and transmission failures are undetectable either. So, the reliability of V2V transmissions is not guaranteed in the current C-V2X system.

In 3GPP, latency reduction for available resource selection continually attracts attention [10], [26], where a higher requirement of 10-ms latency has been targeted by Rel-15. The most intuitive way is to reduce the size of the selection window, but it affects the number of available resources and thus intensifies collisions. Another promising way to mitigate collisions is transmission power control, by which the reuse distance of one RB should be maintained on a reasonable level and thus the number of potential competitors on the same RB is controllable. However, this design conflicts with the requirement of transmission range extension.

Given the MAC protocol in the current 3GPP standards, the advantages and shortages of the C-V2X are summarized and highlighted as follows.

*Strength:* 1) It reuses the existing cellular infrastructure. 2) Resources can be allocated in a centralized or semicentralized way to reduce collisions compared with the random access MAC protocol.

*Weakness:* 1) Collisions are still inevitable and difficult to detect. 2) The reliability of V2V communications (mode-4) cannot be guaranteed. 3) The resource reservation latency is hard to be reduced given the current selection mechanism.

2) *C-V2X Performance Analysis:* The performance of the current C-V2X system has been analyzed in [27]. Given a dense scenario where the traffic density is 0.3 vehicles/m and packet transmission frequency is 10 Hz, the packet delivery ratio (PDR) is about 85% for vehicles 200 m away, 70% for those 300 m away, and the farther the lower. If the packet transmission frequency is increased to 20 Hz, or the number of subchannels decreases from 4 to 2 in each subframe, the PDR on 300 m is reduced to less than 50%.

PDR can be analyzed as follows [27]:

$$\text{PDR}(d_{t,r}) = (1 - \delta_{\text{HD}}) \cdot (1 - \delta_{\text{SEN}}(d_{t,r})) \cdot (1 - \delta_{\text{PRO}}(d_{t,r})) \cdot (1 - \delta_{\text{COL}}(d_{t,r})) \quad (2)$$

**Table 1** Comparison of C-V2X and IEEE 802.11p [29]

|                       | C-V2X                                    | 802.11p       |
|-----------------------|--|---------------|
| Synchronization       | Synchronous                              | Asynchronous  |
| Resource multiplexing | FDM and TDM Possible                     | TDM Only      |
| Channel Coding        | Turbo                                    | Convolutional |
| Waveform              | SC-FDM                                   | OFDM          |
| Retransmission        | HARQ (repetition)                        | No HARQ       |
| Resource Access       | SPS with relative energy-based selection | CSMA-CA       |

where  $d_{t,r}$  is the distance between the transmitter and receiver, and  $\delta_{\text{HD}}, \delta_{\text{SEN}}(d_{t,r}), \delta_{\text{PRO}}(d_{t,r}), \delta_{\text{COL}}(d_{t,r})$  are the probabilities of four types of errors defined below.

a)  $\delta_{\text{HD}}$  is due to half-duplex (HD) transmissions, when a packet to be received overlaps the receiver's transmitting subframe,  $\delta_{\text{HD}} = \Pr(e = \text{HD})$ ; b)  $\delta_{\text{SEN}}(d_{t,r})$  is due to the received signal power below the sensing power threshold (SEN),  $\delta_{\text{SEN}}(d_{t,r}) = \Pr(e = \text{SEN} | e \neq \text{HD})$ ; c)  $\delta_{\text{PRO}}(d_{t,r})$  is due to propagation effect (PRO), i.e., the received SNR cannot support a successful decoding,  $\delta_{\text{PRO}}(d_{t,r}) = \Pr(e = \text{PRO} | e \neq \text{HD}, e \neq \text{SEN})$ ; and d)  $\delta_{\text{COL}}(d_{t,r})$  is due to packet collisions (COL), i.e., more than one vehicles transmit on the same resource,  $\delta_{\text{COL}}(d_{t,r}) = \Pr(e = \text{COL} | e \neq \text{HD}, e \neq \text{SEN}, e \neq \text{PRO})$ . These four types of errors are mutually exclusive and modeled in [27]. The analytical model matches the simulation closely. The mean absolute deviation of the analytical model is generally below 2.5% compared with the simulation.

Molina-Masegosa and Gozalvez [28] compared the performance of C-V2X and 802.11p. In a given scenario, one outperforms the other depending on the configurations. For instance, 802.11p outperforms C-V2X when we set the channel load to 50 packets per second (pps) per vehicle and increase the data rate for 802.11p to 18 Mb/s (the default is 6 Mb/s) while keeping the MCS for C-V2X unchanged. However, when C-V2X chooses 16-QAM with a coding rate of 0.5, it outperforms 802.11p in the same setting. Generally, C-V2X achieves a higher PDR. Molina-Masegosa and Gozalvez [28] also investigated how the redundant transmissions affect reliability. Comparing the single transmission (1 SCI + TB/packet) and 2-time repeating (2 SCI + TB/packet) schemes in simulation, the results showed that repeated transmissions improve the PDR when the channel load is low (10 pps). However, they intensify the collisions and thus reduce the PDR if the channel load is high (20–50 pps).

The differences between C-V2X and IEEE 802.11p are summarized in Table 1.

In addition, the synchronization mechanism of C-V2X has a lower channel access overhead, so it can achieve an overall higher spectral efficiency than 802.11p [30]. Furthermore, the techniques including FDM, turbo code, and SC-FDM adopted by C-V2X can extend the communication range and achieve a higher reliability than 802.11p. The soft combining enabled by HARQ further increases the reliability. The ideal selection of resource should avoid the collision whenever possible. The MAC protocol of C-V2X

chooses those resources by avoiding contention overhead as much as possible, by excluding the RBs for all potential periodical transmissions in the selection window. On the contrary, the CSMA-CA can only select the first “good enough” resource (sensed idle) without considering the nature of periodical traffic in the vehicular network, thus leading to high contention cost [29].

In [31], the performance evaluation of C-V2X communications was studied in terms of scheduling and resource allocation. For the scheduling, the SPS is deployed and the resource allocation is done based on mode-3 and -4 of sidelink communications. Since the transmission collisions among different UEs are usual in the standardized SPS, a cooperative solution is proposed in order to prevent a concurrent resource reselection. The main idea is to provide future reselection information in the transmitted packets. The authors examined their proposed approach through simulations which clearly show a lower collision probability compared to the standard. After 1000 s of simulation, the collision probabilities of the proposed solution and the standard are 0.34% and 1.34%, respectively. They also showed that the collision probability is still below 1% by deploying a retransmission scheme which guarantees the reliability of 99%.

Though C-V2X outperforms 802.11p in some scenarios, the current C-V2X standards (especially mode-4) cannot fully meet the scalability and reliability requirements. Given the weaknesses of standardized C-V2X summarized before, reliability and latency tradeoff remains open. A vehicle may adopt a longer transmission interval with a low data rate to enhance the reliability, which, conversely, increases the latency. Therefore, an advanced MAC protocol is beckoning for further study to jointly consider collision avoidance, reliability enhancement, scalability, and latency reduction.

### III. MULTIHOP V2V RELAY NETWORK CONNECTIVITY

Messages may need to be delivered beyond the transmission range. For instance, emergency electronic brake message can alert the following vehicles, even though their views have been blocked by others; vehicles in congested areas can alert the incoming vehicles to detour well in advance if possible; parking lots, gas/charging stations, restaurants, and hotels can announce their services and availability to vehicles several blocks away. In these scenarios, to propagate the messages beyond the one-hop transmission range, a multihop relay system can be built to connect vehicles farther away.

A motivating example of a 1-D highway vehicle network is shown in Fig. 3. Vehicle  $a$  experiences an accident at time 0 and sends an emergency message out. The following vehicles  $b$  and  $c$  are within the transmission range of  $a$ , so they can receive the broadcast emergency message. Following a Geocast routing protocol, a vehicle will relay (retransmit) the message it received if it did not hear the retransmission of the same message by other vehicles

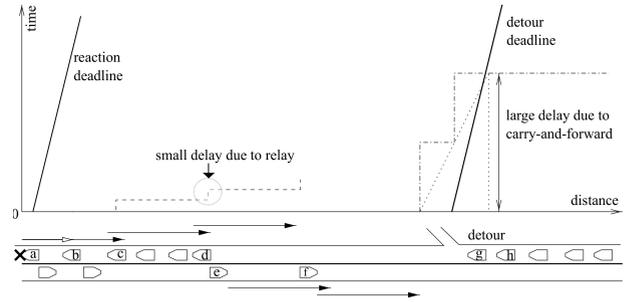


Fig. 3. Time and location coupled data propagation in V2V networks.

farther away from the source, like  $b$  or  $c$  did after a small delay. In the example,  $d$  also receives the message and broadcasts it out. Although there is no vehicle within the transmission range of  $d$  following it, the vehicles in the opposite direction,  $e$  and  $f$ , can further relay the message to the right. In this example, since the distance between  $f$  and  $g$  is larger than the transmission range,  $f$  can carry and forward the message and rebroadcast it periodically till the message reaches  $g$ . However, given the limited vehicle speed compared to the wireless signal propagation speed, such carry-and-forward transmissions involve much longer delay. In the top of Fig. 3, the  $x$ -axis represents the distance to vehicle  $a$ , the point-of-interest (POI), and the  $y$ -axis shows the time deadline at different locations. The reaction deadline is for vehicles following  $a$ : the closer to the POI, the less time a vehicle has to avoid a chain-reaction crash. The detour deadline is for vehicles receiving the message to take the detour exit.

From the above example, for emergency messages, relaying in a connected group of vehicles in the same or opposite directions is effective and critical to ensure that the messages can be delivered to others promptly before the location-dependent deadline. However, given the mobility and dynamic topology of vehicles, it is difficult to quantify the connectivity of V2V networks. In this section, we first study the 1-D V2V network connectivity, followed by the 2-D V2V network connectivity. The carry-and-forward network performance will be studied in Section IV.

#### A. One-Dimensional V2V Networks

We simplify the problem by assuming that vehicles within the transmission range of each other can always be connected, given reliable link-layer protocols, e.g., applying link-layer automatic repeat request (ARQ). For the 1-D scenario, note that vehicles in a “connected” cluster can reach each other by fast-speed relaying. As shown in Fig. 3, vehicles from  $a$  to  $f$  can be viewed as a cluster. We define a vehicle cluster as a group of vehicles where there exists at least one multihop relay path between any two vehicles. For 1-D cases, *cluster size*, the distance between the first and last vehicles in the same cluster can be used to measure the connectivity.

Since the road length is much longer than the width, we can use a linear topology to study the 1-D V2V network connectivity. Assuming intervehicle distances,  $X_1, X_2, \dots$ , are i.i.d. random variables in free-flow scenarios (confirmed in [32]), we first derive the average cluster size and the cluster size distribution [33].

$c_s$  denotes the cluster size. Let  $\lambda$  be the average vehicle density, and  $R$  be the transmission distance. Conditioning on  $X_1$ , for the expected distance between the first vehicle and its follower, we have

$$\mathbf{E}\{c_s\} = \Pr\{X_1 \leq R\}(\mathbf{E}\{X_1|X_1 \leq R\} + \mathbf{E}\{c_s\}).$$

This is because if  $X_1$  is no larger than  $R$ , the cluster size will be equal to  $X_1$  plus the size of a cluster starting from the second vehicle; otherwise, the cluster size is zero. Similarly, we can calculate the higher order moments of  $c_s$ , for example,

$$\mathbf{E}\{c_s^2\} = \Pr\{X_1 \leq R\}\mathbf{E}\{(X_1 + c_s)^2|X_1 \leq R\}.$$

With some algebraic manipulation and plugging in the PDF function of  $X_1$ ,  $f_X(x)$ , we have the following theorem [34].

*Theorem 2:* Suppose that the vehicle density in a path is  $\lambda$  and the intervehicle distance has the PDF denoted by  $f_X(x)$ . Then, we have the mean and second moment of  $c_s$  by

$$\mathbf{E}\{c_s\} = \frac{\int_0^R x f_X(x) dx}{\int_R^\infty f_X(x) dx} \quad (3)$$

$$\mathbf{E}\{c_s^2\} = \frac{\int_0^R (2x\mathbf{E}\{c_s\} + x^2) f_X(x) dx}{\int_R^\infty f_X(x) dx}. \quad (4)$$

The variance of  $c_s$  is  $\mathbf{Var}\{c_s\} = \mathbf{E}\{c_s^2\} - \mathbf{E}\{c_s\}^2$ , so both the mean and variance of  $c_s$  can be obtained. Then, we use the following corollary as an example to identify a general approach for estimating the distribution of  $c_s$  [35].

*Corollary 1:* Suppose that the vehicle density in a path is  $\lambda$  and the intervehicle distances are i.i.d. with an exponential distribution. Then, the PDF of  $c_s$  can be approximated as

$$f_c(x) = (x^{k-1} e^{-x/\theta}) / (\theta^k \Gamma(k)), \quad x > 0 \quad (5)$$

where  $k = (\mathbf{E}\{c_s\}^2) / (\mathbf{E}\{c_s^2\} - \mathbf{E}\{c_s\}^2)$ ,  $\theta = (\mathbf{E}\{c_s^2\} - \mathbf{E}\{c_s\}^2) / (\mathbf{E}\{c_s\})$ ,  $\Gamma(\cdot)$  is the Gamma function, and  $\mathbf{E}\{c_s\}$  and  $\mathbf{E}\{c_s^2\}$  are obtained from (3) and (4), respectively.

*Proof:* Since the intervehicle distances are i.i.d. with an exponential distribution, we have  $f_X(x) = \lambda e^{-\lambda x}$ . By substituting the expression of  $f_X(x)$  into (3) and (4), one can obtain the closed forms of  $\mathbf{E}\{c_s\}$  and  $\mathbf{E}\{c_s^2\}$ ,

respectively, for example,

$$\mathbf{E}\{c_s\} = [1 - (\lambda R + 1)e^{-\lambda R}] / \lambda e^{-\lambda R}.$$

Since Gamma distribution can describe the sum of exponentially distributed random variables, we thus use a Gamma distribution to approximate the PDF of  $c_s$ , which is the sum of a geometrically distributed random number of i.i.d. exponential random variables. Hence, we approximate the PDF  $f_c(x)$  of  $c_s$  using the following form of Gamma distribution:

$$f_c(x) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)}, \quad x > 0$$

where  $k$  and  $\theta$  are two parameters to be determined, and  $\Gamma(k)$  is the Gamma function evaluated at  $k$ . Note that we have obtained the mean and variance of  $c_s$  from Theorem 2, which can be used to determine the values of the parameters  $k$  and  $\theta$  in  $f_c(x)$ . Let  $f_c(x)$  have the same mean and variance as those obtained before, we have the following equation set:

$$\begin{cases} k\theta = \mathbf{E}\{c_s\} \\ k\theta^2 = \mathbf{E}\{c_s^2\} - \mathbf{E}\{c_s\}^2. \end{cases} \quad (6)$$

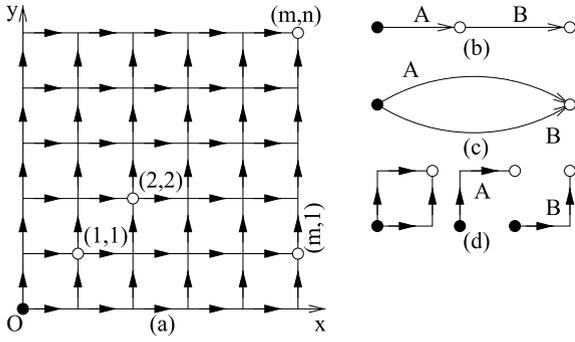
By solving this equation set, we obtain that  $k = (\mathbf{E}\{c_s\}^2) / (\mathbf{E}\{c_s^2\} - \mathbf{E}\{c_s\}^2)$  and  $(\theta = \mathbf{E}\{c_s\} / k) = (\mathbf{E}\{c_s^2\} - \mathbf{E}\{c_s\}^2) / (\mathbf{E}\{c_s\})$ .  $\square$

*Remark 1:* From the above example, we identified a general approach for the distribution estimation of  $c_s$ . Specifically, we can use (3) and (4) in Theorem 2 to obtain  $\mathbf{E}\{c_s\}$  and  $\mathbf{E}\{c_s^2\}$ , respectively. Applying the similar approach, we can obtain the higher order moments of  $c_s$ . Next, solve (6) to obtain the values of the parameters (e.g., mean and variance) in a given random distribution function of the cluster size. Note that this approach can be applied for any i.i.d. intervehicle distances, e.g., those following a log-normal distribution [36].

Given the cluster size distribution, we can directly obtain the probability that two vehicles with a given distance  $d$  are within the same cluster or not. In other words, the connected probability of the two vehicles can be obtained by  $\int_0^d f_c(x) dx$ . In a connected cluster, as long as we maintain the reliability of each single-hop relay transmission, the reliability of message delivery can be ensured.

## B. Two-Dimensional V2V Networks

Connectivity for 2-D V2V networks such as those in urban scenarios is much more complicated. The network is composed of road segments, where each road segment has a linear topology. From the above analysis, we can derive the probability that vehicles in a road segment are connected. A sample 2-D lattice network is shown in Fig. 4(a), where a message originated at the origin  $O = (0, 0)$  will be broadcast and relayed toward vehicles at



**Fig. 4.** System model and basic principles for 2-D network connectivity [37]. (a)  $m \times n$  lattice. (b) Sequential paths. (c) Parallel paths. (d) Decomposition of a  $1 \times 1$  lattice.

location  $(m, n)$ . Using a Geocast routing protocol, the message will be propagated along the lattice edges in the directions indicated by arrows if possible. Since we have solved the 1-D connectivity problem, we can easily derive the connectivity probability between two vehicles on the same road segment, and the connectivity between two neighbor intersections (vertices in the figure), called the bond probability,  $p$ . We are interested in the fundamental network property that two vertices in this 2-D network are connected, as a function of  $p$ . In general,  $p$  may vary for different road segments given the uneven vehicle density, road segment length, etc. Here, we simplify the problem and assume a homogeneous case for easy presentation, and the heterogeneous case is an extension [37].

Even with the simplified homogeneous case, the problem is still a hard one. There are  $2mn + m + n$  independent road segments between  $O$  and  $(m, n)$ , and each segment can have two possibilities, connected or disconnected. Thus, the total number of possibilities is  $2^{mn+m+n}$ , a prohibitively large number. For instance, with a medium-size network of  $7 \times 7$ , the number of possibilities is  $2^{112}$ , too large to conduct a thorough Monte Carlo simulation.

To derive the connectivity, we begin with a simple case to explain the basic principles. As shown in Fig. 4(b), the end-to-end connectivity between the source and the destination containing path  $A$  and path  $B$  sequentially is  $P(AB) = P(A)P(B)$ , as  $A$  and  $B$  are independent without common edges. Here,  $P(A)$  denotes the connectivity of path  $A$ . As shown in Fig. 4(c), if there are two parallel paths connecting the source and the destination, the connectivity becomes  $P(A + B) = P(A) + P(B) - P(AB)$  according to the principle of inclusion and exclusion (PIE). These two principles can be used to solve the  $1 \times 1$  lattice problem as shown in Fig. 4(d): given  $P(A) = P(B) = p^2$ ,  $P(1, 1) = P(A) + P(B) - P(A)P(B) = 2p^2 - p^4$  as  $A$  and  $B$  are independent.

Following the principles, in an  $m \times n$  lattice, any directed path from  $O$  to vertex  $(m, n)$  has  $m + n$  segments, so the connectivity probability of each path is simply  $p^{(m+n)}$ . The number of paths between them is  $\binom{m+n}{n}$ , and these paths are dependent on each other when they share some

common edges. Thus, if we simply apply the PIE principle, we need to calculate the connectivity probability of each subset of paths, resulting in a total complexity of  $O(2^{\binom{m+n}{n}-1})$ , not affordable even for a small size network of  $5 \times 5$ . This also explains why direct percolation (closely related to this problem) has no closed-form solution even with extensive research in physics, chemistry, and material sciences for more than half a century.

In [37], a recursive decomposition approach was developed to solve the problem of typical-size vehicle networks with affordable complexity. The key idea is to decompose the network to a smaller version with mutually exclusive cases. Fig. 5 gives a  $2 \times 2$  lattice example, where the top-leftmost path  $A$  and the union of the rest paths,  $B$ , are identified. Note that  $B$  is simpler than the original  $2 \times 2$  lattice. In general,  $A$  is a single path of  $m + n$  edges (i.e., no branches possible), and we can have a simple partition on the sample space with  $m + n$  mutually exclusive events. Define  $S_i$  as the event that the last  $i$  edges along  $A$  leading to the destination are all connected, but the last  $(i + 1)$ th one is not, so  $P(S_i) = p^i(1 - p)$  for  $0 \leq i \leq m + n - 1$ . For the origin and destination to be connected, we then have  $m + n$  mutually exclusive cases, including  $B|S_i$  and  $A$  being connected where  $P(A) = p^{m+n}$ . Defining the probability that  $B$  is connected given  $S_i$  as  $P(B|S_i)$ , we have

$$P(m, n) = P(A + B) = 1 - P(\overline{B + A})$$

$$= 1 - P(\overline{B} \overline{A}) \quad (7)$$

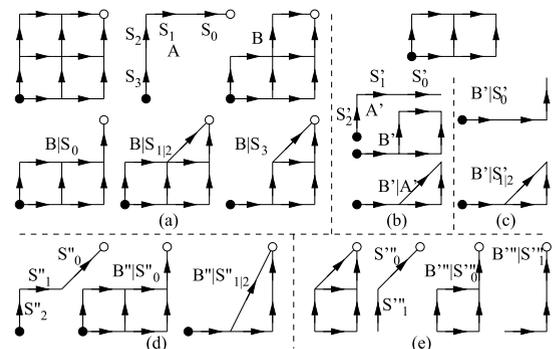
$$= 1 - P\left(\overline{B} \bigcup_{i=0}^{m+n-1} S_i\right) \quad (8)$$

$$= 1 - \sum_{i=0}^{m+n-1} P(\overline{B}|S_i)P(S_i) \quad (9)$$

$$= 1 - \sum_{i=0}^{m+n-1} (1 - P(B|S_i))P(S_i)$$

$$= P(A) + \sum_{i=0}^{m+n-1} P(B|S_i)P(S_i) \quad (10)$$

where (7) is due to De Morgan's law, (8) due to  $\bigcup_{i=0}^{m+n-1} S_i = \overline{A}$ , (9) due to  $S_i$  being mutually exclusive,



**Fig. 5.** Decomposition of a  $2 \times 2$  lattice [37] in (a) and its components recursively in (b)-(e).

and (10) due to  $\sum_{i=0}^{m+n-1} P(S_i) = P(\bar{A}) = 1 - P(A)$ , i.e.,  $A$  and  $S_i$  partition and constitute the entire event space in *total probability*. We can repeat this procedure, by decomposing the network into two parts, the top-leftmost path and the union of the rest. Then, eventually, we can decompose the lattice to smaller networks to obtain the connectivity recursively.

For the sample  $2 \times 2$  lattice network, given  $S_0$ , no end-to-end connection is possible via vertex  $(0, 2)$  or  $(1, 2)$ , so we can discard the edges adjacent to the two vertices and have  $\mathcal{B}|S_0$  as shown in Fig. 5(a). Given  $S_1$ , it implies that  $(1, 2)$  and  $(2, 2)$  are connected, and  $\mathcal{B}$  does not include any edges from  $(0, 2)$ , so we can merge  $(1, 2)$  with  $(2, 2)$  in  $\mathcal{B}$  to further obtain  $\mathcal{B}|S_1$ . Since  $S_1$  and  $S_2$  have the same effect, they are illustrated as  $\mathcal{B}|S_{1|2}$  in Fig. 5(a). Given  $S_3$ , no connection is possible through  $(0, 1)$ , so the edges adjacent to it can be removed; it also implies that  $(0, 1)$ ,  $(0, 2)$ ,  $(1, 2)$ , and  $(2, 2)$  are connected sequentially, so they can be merged as  $\mathcal{B}|S_3$  in Fig. 5(a).

After this decomposition, we have  $\mathcal{B}|S_0$  to  $\mathcal{B}|S_3$ . Using the serial principle,  $\mathcal{B}|S_0$  can be decomposed into a  $2 \times 1$  lattice (or ladder) and an edge. Fig. 5(b) shows how we further decompose the ladder into  $A'$ ,  $B'$ , and  $B'|A'$ .

Fig. 5(c) shows the decomposition approach with  $\mathcal{B}'|S'_{0|1|2}$ , which can both be solved directly using the serial principle. Similarly for  $\mathcal{B}|S_{1|2}$ , they are decomposed in Fig. 5(d) to components of known connectivity (e.g.,  $\mathcal{B}''|S''_0$  is the same as  $\mathcal{B}|S_0$ ), and part of  $\mathcal{B}|S_3$  is decomposed in Fig. 5(e), where the serial principle and  $P(1, 1)$  can be applied. Using the total probability approach, the connectivity of the decomposed components can be reassembled, and then  $P(2, 2)$  is given by

$$\begin{aligned} P(2, 2) &= P(A) + \sum_{i=0}^3 P(\mathcal{B}|S_i)P(S_i) \\ &= p^{12} - 4p^{11} + 2p^{10} + 4p^9 + 2p^8 - 4p^7 - 6p^6 + 6p^4. \end{aligned} \quad (11)$$

Using the similar recursive partition approach, we can decompose an  $m \times n$  lattice by separating the top-leftmost path from the rest (which is a simpler network) and partitioning it into a set of mutually exclusive events and repeating this procedure till all the paths are decomposed and partitioned. This recursive decomposition approach can obtain the connectivity of an  $n \times n$  lattice network with a complexity of  $O(n^2 \cdot ((2n)!)/(n!(n+1)!))$ . The lattice connectivity (up to  $10 \times 10$  lattice) as a function of  $p$  is given in [38].

The calculation results from the derivation are compared with simulation results in Fig. 6. Obviously, for the same lattice, improving the bond probability leads to better connectivity. With the same bond probability, however, the increase of lattice size does not have the same impact on the end-to-end connectivity. For small bond probabilities, i.e., from 0.35 to 0.65, a clear drop of connectivity can be observed. This actually corresponds to the conclusion in [39], where the end-to-end connectivity

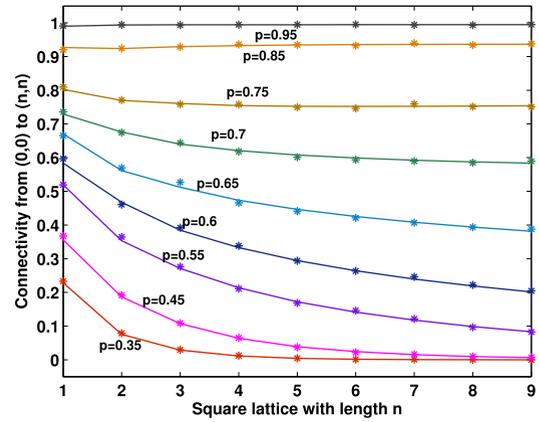


Fig. 6. Connectivity of  $n \times n$  lattices [37].

shows an exponential decay and the exponent is determined by how far away the bond probability is from the critical bond probability. The critical bond probability is around 0.6447 for the directed bond percolation on square lattices. When the bond probability is reasonably large, i.e., higher than 0.65, the connectivity remains stable with respect to the network size.

*Critical Bond Probability for Finite Lattices:* By calculating the solution of the second derivative equal to 0, we can obtain a critical bond probability, i.e.,  $p_c(m, n)$ , where the connectivity curve achieves the sharpest increase for a given finite lattice network. This critical bond probability is also the probability where the maximum value of the first derivative curve occurs. By calculation, we find that the value of  $p_c(m, n)$  is 0.645, 0.665, 0.668, and 0.669 for  $2 \times 2$ ,  $4 \times 4$ ,  $6 \times 6$ , and  $8 \times 8$  lattices, respectively, and the values are only subject to precision. An asymptotic behavior of bond probability could be observed since the difference between two consecutive critical bond probabilities reduces with the increment of the lattice size. The critical transition range can also be obtained from the second derivative, indicated by the range of the bond probability from the maximum second derivative to the minimum.

#### IV. V2V/V2I CARRY-AND-FORWARD NETWORKS

The above connectivity model and analysis show the limitation of pure V2V communications in delivering information over a long distance in sparse networks. Due to the uncertainty of vehicle locations and V2V links, the reliability of message delivery is difficult to guarantee. To have the message delivered farther away in a sparse network, the carry-and-forward mechanism has been proposed based on the predictable vehicle mobility [40]: a vehicle carrying a message will try to forward it to the next-hop relay in the data dissemination direction if possible. If not possible, the message will be stored and carried until it reaches a suitable relay. By utilizing the mobility and connectivity of bidirectional vehicle traffic, the reliability of message dissemination is largely improved, at the

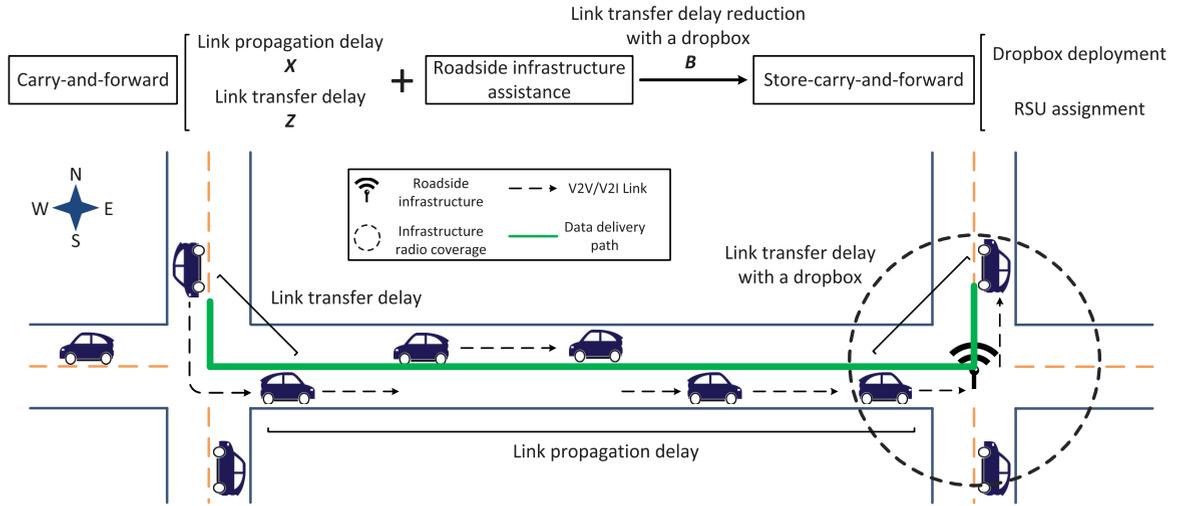


Fig. 7. Store-carry-and-forward with dropboxes.

cost of random delay. The carry-and-forward mechanism has been modeled with an evolving graph, which revealed that the carry-and-forward mechanism leverages the temporal connectivity of mobile vehicles for multihop delivery and it effectively increases the reachability of network-wide data delivery [41]. The performance is affected by various factors including the vehicle density, the penetration ratio, etc. Given the reliability is the probability of the message being successfully delivered to the destination before the deadline, the delay under the carry-and-forward mechanism is essential to evaluate the corresponding reliability. The delay analysis can also be used to design routing protocols to select suitable forwarding paths in 2-D vehicular networks. In this section, we first study the performance of V2V-only networks, and then the hybrid V2V/V2I networks.

### A. V2V-Only Carry-and-Forward Delay Analysis

Considering a 2-D vehicular network shown in Fig. 7, the delivered message will go through one or multiple road segments and intersections, respectively. Thus, the delivery delay analysis can be categorized into two parts, i.e., link propagation delay along road segments and link transfer delay at intersections. Specifically, the link propagation delay measures the time of message forwarded through a link (road segment), while the link transfer delay measures the time of message forwarded to the given direction at an intersection [34], [42].

Link propagation delay considers a message to be delivered through one or multiple vehicles from one location to another location of a road segment, without crossing any intersection. With the knowledge of everyone's own cluster size and travel plan, the cluster head (CH) will seek the opportunities to communicate with the CH of the clusters in the opposite direction. Furthermore, with the help of the opposite cluster, the CH can forward the message to its neighbor cluster in front of it under the

condition that the opposite cluster connects with both clusters simultaneously.

By utilizing the opposite vehicle cluster to link the disconnected vehicle cluster on the same direction, the data delivery speed, denoted by  $v_d$ , is estimated by

$$v_d = \frac{\mathbf{E}\{d_f\}}{\mathbf{E}\{X_c^t\} + 2\delta} = v + \frac{(R + v_{ij}/\lambda_{ij}) + \mathbf{E}_{ij}\{c_s\}}{\mathbf{E}\{X_c^t\} + 2\delta} \quad (12)$$

where  $d_f$  is the total distance of the message forwarded through the opposite cluster and the neighbor cluster in front,  $R$  is the communication range,  $X_c^t$  is the time needed for a message being held by CH until it is forwarded to its front neighbor cluster,  $\lambda_{ij}$  and  $v_{ij}$  are the vehicle density and velocity from intersection  $i$  to  $j$ , respectively, and  $\delta$  is the minimum required time of two clusters to exchange the message. Apparently, the data delivery speed combines the movement speed of vehicles and the message forwarding speed, which demonstrates the advantage in improving the data delivery reliability. The following theorem estimates the expected link propagation delay.

*Theorem 3:* Let  $d_{ij}$  be the distance from intersection  $i$  to its neighbor intersection  $j$ , and  $X_{ij}$  be the corresponding propagation delay

$$\mathbf{E}\{X_{ij}\} \approx \max \left\{ \frac{d_{ij} - \mathbf{E}_{ij}\{c_s\}}{v_d}, 0 \right\} \quad (13)$$

where  $v_d$  can be obtained from (12).

In (13), the  $-\mathbf{E}_{ij}\{c_s\}$  term represents the fast-forwarded distance to a CH at the beginning. From the above results, vehicle mobility and connectivity are two key factors that influence the data dissemination delay on the road segment. Since the vehicle mobility is restricted by the traffic regulation and conditions, improving the network connectivity is a more promising way to achieve satisfying performance. In Sections IV-B and IV-C, we will discuss the

possibility of applying roadside infrastructures to enhance connectivity and reliability.

Link transfer delay describes the latency of a message being delivered to a targeted direction at an intersection purely rely on V2V transmissions, i.e., the message is relayed between vehicles only until it is forwarded to the selected direction. Specifically, when the message carrier arrives at the intersection, it needs to find a suitable vehicle to take the message to the targeted direction if not itself. However, it is possible that the carrier cannot find a suitable vehicle within its communication range. Thus, when the carrier leaves the intersection, it will pass the message to a new carrier moving back to the intersection. If the new carrier finds a suitable vehicle at the intersection, the link transfer is completed; otherwise, this procedure may repeat till success.

Without loss of generality, suppose that the message carrier is searching for a vehicle moving toward the north direction. The average link transfer delay estimation is obtained using the following theorem.

*Theorem 4:* Let  $Z_{wn}$ ,  $Z_{en}$ , and  $Z_{sn}$  be the link transfer delay from west, east, and south to north, respectively, at an intersection. Their average values can be estimated by solving the following set of equations:

$$\begin{bmatrix} \mathbf{E}\{Z_{wn}\} \\ \mathbf{E}\{Z_{en}\} \\ \mathbf{E}\{Z_{sn}\} \end{bmatrix} = \begin{bmatrix} 1 & -B_{wn}^e & -B_{wn}^s \\ -B_{en}^w & 1 & -B_{en}^s \\ -B_{sn}^w & -B_{sn}^e & 1 \end{bmatrix}^{-1} \begin{bmatrix} C_{wn} \\ C_{en} \\ C_{sn} \end{bmatrix} \quad (14)$$

where  $w, s, e$ , and  $n$  denote directions, and how to obtain the values of the parameters on the right-hand side is given in the proof in Appendix B in the Supplementary Material of [34].

The above theorem considered a four-way intersection that the message carrier may arrive from any direction except the north. Solving (14) can obtain the link transfer delays from any direction to north. The above results can also be applied to the intersections with more or fewer roads, with unidirectional or bidirectional traffic.

End-to-end path delay is the sum of all link propagation delay and link transfer delay along a given path as shown in the following theorem.

*Theorem 5:* Suppose that a message is forwarded according to a path  $i_0 \rightarrow i_1 \rightarrow \dots \rightarrow i_m$ , and let  $D_{i_0 i_m}$  be the path delay. Then, we have

$$\mathbf{E}\{D_{i_0 i_m}\} = \sum_{\ell=1}^m \mathbf{E}\{X_{i_{\ell-1} i_\ell}\} + \sum_{l=1}^{m-1} \mathbf{E}\{Z_{i_{l-1} i_l i_{l+1}}\}$$

where  $\mathbf{E}\{X_{i_{\ell-1} i_\ell}\}$  and  $\mathbf{E}\{Z_{i_{l-1} i_l i_{l+1}}\}$  can be calculated by (13) and (14), respectively.

The above theoretical findings offer an overall delay analysis of the carry-and-forward mechanism.

The accurate and detailed quantification of the delay performance is critical to guide the minimum delay routing and network planning in the following sections. The theorems are limited by the assumptions on the knowledge of the vehicle arrival process and the negligible transmission delay variations, which require further investigation.

## B. Dropbox Deployment

Given the intermittent connectivity of V2V-only solutions, it may not be able to ensure a stringent QoS; using V2I only, vehicles always use infrastructures to exchange information, which may lead to a high deployment cost and unnecessary latency, and may encounter scalability issues when the number of vehicles is large. Therefore, a hybrid V2V/V2I system is more desirable to take the advantages of both sides.

As reasoned before, in remote areas or off-peak hours, the low connectivity in sparse networks significantly affects the reliability of the data dissemination services. As a static node deployed, a dropbox functions similar to a network router with storage. Therefore, dropboxes enable the store-carry-and-forward mechanism, as they can temporarily store the message when no suitable relay is found at the intersection. In addition, deploying dropboxes with a wide coverage can increase the contact probabilities between vehicles and effectively enhance network QoS [43]. It is particularly suitable for remote or developing regions lack of communication backbone.

Considering the deployment cost, it is desirable to deploy dropboxes in the locations that are most beneficial only. To comprehensively evaluate the deployment performance, a utility function considering the trade-off between the benefit and cost is proposed, followed by the utility maximization problem as follows [43]:

$$\begin{aligned} & \max_{s \in \Theta, p_{sd}^k \in \Omega} U(\mathbf{E}\{T_{sd}(k)\}) - mC_d \\ & \text{s.t.} \quad \sum_{i=1}^n s_i = m \end{aligned} \quad (15)$$

where  $T_{sd}(k)$  is the end-to-end delivery delay from the source to the destination through path  $k$ ,  $U(\cdot)$  is the utility function related to the delay,  $m$  is the number of the deployed dropboxes, and  $C_d$  is the unit deployment cost. To solve (15), in addition to the delay analysis given in Section IV-A, the link transfer delay estimation with dropboxes should also be obtained.

Link transfer delay with dropboxes measures the time needed to forward a message to the targeted direction with the help of both vehicles and dropboxes. When the message carrier approaches the intersection and enters the communication coverage of the dropbox, it can directly forward it to the dropbox. Then, the dropbox will search for a suitable vehicle to carry the message. Assuming that

the targeted direction is north, the expectation of the link transfer delay with a dropbox, denoted by  $\mathbf{E}\{\tilde{Z}_{wn}\}$ , satisfies

$$\begin{aligned} \mathbf{E}\{\tilde{Z}_{wn}\} &= 0 \times (1 - e^{-\lambda_{esw}^n \frac{R_d}{v_n}}) + e^{-\lambda_{esw}^n \frac{R_d}{v_n}} \times \frac{1}{\lambda_{esw}^n} \\ &= \frac{1}{\lambda_{esw}^n} e^{-\lambda_{esw}^n \frac{R_d}{v_n}} \end{aligned} \quad (16)$$

where  $\lambda_{esw}^n = \lambda_{en} + \lambda_{sn} + \lambda_{wn}$ ,  $v_n$  is the vehicle speed in the north direction, and  $R_d$  is the communication range of dropboxes. Clearly,  $\mathbf{E}\{\tilde{Z}_{wn}\}$  is a decreasing function of  $(\lambda_{esw}^n)/v_n$  and  $R_d$ , which shows how a higher vehicle density and a larger communication range of the dropbox benefit the link transfer of the message.

Combining the link transfer delay analysis in Section IV-A, the benefit of dropboxes facilitating the message delivery is obtained by the following theorem.

**Theorem 6:** Suppose that a message is needed to be forwarded from direction  $w$  to direction  $n$  at an intersection. Let  $B_{wn}$  be the reduction on link transfer delay of using a dropbox. Then, we have

$$\begin{aligned} \mathbf{E}\{B_{wn}\} &= \frac{\left(1 - e^{-\lambda_{esw}^n \frac{R_d}{v_n}}\right)}{v_n^d} \\ &\times \left[ \left(1 - e^{-\lambda_{esw}^n \frac{R_d}{v_n}}\right) \left(R_d - \frac{v_n}{\lambda_{esw}^n}\right) + R_d e^{-\lambda_{esw}^n \frac{R_d}{v_n}} \right] \\ &+ (\mathbf{E}\{Z_{wn}\} - \mathbf{E}\{\tilde{Z}_{wn}\}) + \frac{R_d - R}{v_w^d} \end{aligned} \quad (17)$$

where  $v_w^d$  and  $v_n^d$  are the delivery speed in the west and north direction, respectively, which can be obtained by (12).

The above theorem depicts that the delay reduction with the dropbox comes from two aspects. As illustrated in Fig. 7, first, it increases the probability of finding a vehicle going to the targeted direction at the intersection. Second, it reduces the delivery distance on the road segment because the message carrier can send the message to the dropbox even before it arrives at the intersection.

Based on the delay analysis, the utility maximization problem can be solved by the optimal dropbox deployment algorithm (ODDA), using information dimension enlargement and dynamic programming, as depicted in Algorithm 1.

First, an information set  $\Phi_i$  for node  $i$  is defined as follows:

$$\Phi_i = \begin{bmatrix} Y_i^m & i_m^* & I_i^m \\ Y_i^{m-1} & i_{m-1}^* & I_i^{m-1} \\ \dots & \dots & \dots \\ Y_i^0 & i_0^* & I_i^0 \end{bmatrix} \quad (18)$$

where for the number of deployed dropboxes  $\ell = 0, 1, \dots, m$ ,  $Y_i^\ell$  denotes the minimum delay from  $i$  to the destination using  $\ell$  dropboxes,  $i_m^*$  is the optimal next-hop node corresponding to  $Y_i^\ell$ , and  $I_i^\ell$  is the multidimensional index vector.  $I_i^\ell$  is given by

$$\begin{aligned} I_i^\ell &= [I_i^\ell(1), \dots, I_i^\ell(\ell)] = \begin{bmatrix} S_i^\ell \\ \bar{B}_i^\ell \end{bmatrix} \\ &= \begin{bmatrix} i_\ell^1 & i_\ell^2 & \dots & i_\ell^\ell \\ \mathbf{E}\{B^{i_\ell^1}\} & \mathbf{E}\{B^{i_\ell^2}\} & \dots & \mathbf{E}\{B^{i_\ell^\ell}\} \end{bmatrix} \end{aligned} \quad (19)$$

where  $S_i^\ell = [i_\ell^1, i_\ell^2, \dots, i_\ell^\ell]$  is the set of dropbox deployment strategies, denoting the locations deployed with dropboxes, and  $\bar{B}_i^\ell = [\mathbf{E}\{B^{i_\ell^1}\}, \mathbf{E}\{B^{i_\ell^2}\}, \dots, \mathbf{E}\{B^{i_\ell^\ell}\}]$  is the expected benefit set, denoting the delay reductions with the deployed dropboxes.

Given  $\ell$ , three important notations are defined to calculate the delivery delay under different dropbox deployment strategies, considering the message forwarded from node  $i$  to node  $j$  and then to node  $d$ . Denote by  $Z_{ijj_\ell^*}^{\rightarrow}$  the link transfer delay from direction  $\vec{i}_j$  to direction  $\vec{j}_j^*$ . The delivery delay under the strategy that there is no dropbox deployed at node  $j$  and node  $j$  uses the strategy  $S_j^\ell$  is given by

$$\mathbf{E}_{ij}^\ell(0) = \mathbf{E}\{X_{ij}\} + \mathbf{E}\{Z_{ijj_\ell^*}^{\rightarrow}\} + \mathbf{E}\{Y_j^\ell\}. \quad (20)$$

Denote by  $B_{ijj_\ell^*}^{\rightarrow}$  the link transfer delay reduction from direction  $\vec{i}_j$  to direction  $\vec{j}_j^*$ . The delivery delay under the strategy that a dropbox is deployed at node  $j$  and removed from strategy  $S_j^\ell$  is given by

$$\begin{aligned} \mathbf{E}_{ij}^\ell(1) &= \mathbf{E}\{X_{ij}\} + \mathbf{E}\{Z_{ijj_\ell^*}^{\rightarrow}\} + \mathbf{E}\{Y_j^\ell\} \\ &- \mathbf{E}\{B_{ijj_\ell^*}^{\rightarrow}\} + \min\{\mathbf{E}\{B^{j_\ell^*}\} | B^{j_\ell^*} \in I_j^\ell\}. \end{aligned} \quad (21)$$

Let

$$\mathbf{E}_{ij}^\ell(2) = \mathbf{E}\{X_{ij}\} + \mathbf{E}\{Z_{ijj_{\ell-1}^*}^{\rightarrow}\} + \mathbf{E}\{Y_j^{\ell-1}\} - \mathbf{E}\{B_{ijj_{\ell-1}^*}^{\rightarrow}\} \quad (22)$$

be the delivery delay under the strategy that a dropbox is deployed at node  $j$  and node  $j$  uses strategy  $S_j^{\ell-1}$ . In the above,  $\mathbf{E}\{X_{ij}\}$ ,  $\mathbf{E}\{Z_{ijj_\ell^*}^{\rightarrow}\}$ , and  $\mathbf{E}\{B_{ijj_\ell^*}^{\rightarrow}\}$  can be calculated by (13), (14), and (17), respectively.

The algorithm design considers three important facts. First, since the number of the deployed dropboxes is given, the optimal strategy should choose the locations with the highest benefit. Second, if node  $i$  is included in the optimal path  $p_{sd}^*$  and  $\tilde{m}$  dropboxes are deployed between  $i$  and

**Algorithm 1** ODDA

**Input:** Given the nodes  $s$  and  $d$  and graph  $G = (\mathcal{V}, E, \mathcal{W})$ .  
Let the initial information set satisfy

$$\Phi_d(0) = \begin{bmatrix} \emptyset & \emptyset & \emptyset \\ \dots & \dots & \dots \\ 0 & \emptyset & \emptyset \end{bmatrix} \quad (23)$$

and set  $\Phi_i(0) = \emptyset$  (all elements are  $\emptyset$ ) for  $i \in \mathcal{V}$  and  $i \neq d$ .

- 1: **while** the iteration time  $t \leq n - 1$  **do**
- 2: Each node sends its information set to its neighbor nodes.
- 3: **if** node  $i$  receives the information set from its neighbors **then**
- 4: **for**  $\ell = m, m - 1, \dots, 0$  **do**

$$\mathbf{E}\{Y_i^\ell\} = \min\{\mathbf{E}_{i_j}^\ell(0), \mathbf{E}_{i_j}^\ell(1), \mathbf{E}_{i_j}^\ell(2) | j \in N_i\} \quad (24)$$

where  $N_i$  is the neighbor set of node  $i$ , and then finds the optimal next-hop node  $i_\ell^*$  by

$$i_\ell^* = \arg \min\{\mathbf{E}_{i_j}^\ell(0), \mathbf{E}_{i_j}^\ell(1), \mathbf{E}_{i_j}^\ell(2) | j \in N_i\}.$$

- 5: **end for**
- 6: **end if**
- 7: **if**  $\mathbf{E}_{i_i^*}^\ell(0) = \mathbf{E}\{Y_i^\ell\}$  **then**

$$I_i^\ell = I_{i_i^*}^\ell;$$

- 8: **else**
- 9: **if**  $\mathbf{E}_{i_i^*}^\ell(1) = \mathbf{E}\{Y_i^\ell\}$  **then**
- 10:

$$I_i^\ell = I_{i_i^*}^\ell; I_i^\ell(l) = \left[ \begin{array}{c} i_\ell^* \\ \mathbf{E}\{B_{i_i^* [i_\ell^*]_\ell^*}\} \end{array} \right];$$

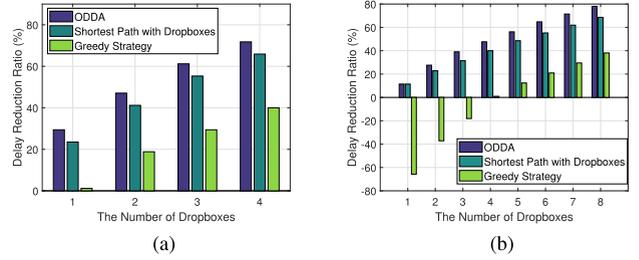
- 11: **else**

$$I_i^\ell(l) = \left[ \begin{array}{c} I_{i_i^*}^{\ell-1}(l), \left[ \begin{array}{c} i_\ell^* \\ \mathbf{E}\{B_{i_i^* [i_\ell^*]_{\ell-1}^*}\} \end{array} \right] \end{array} \right]$$

- 12: **end if**
- 13: **end if**
- 14: Node  $i$  updates its information set  $\Phi_i, t = t + 1$ .
- 15: **end while**

**Output:**  $Y_s^m$  and  $I_s^m$ . The nodes in the strategy set  $S_s^m$  are where the  $m$  dropboxes should be deployed.

destination  $d$ , then these  $\tilde{m}$  dropboxes should be optimally deployed to maximize the delay reduction benefit. Last, the benefit of deploying dropboxes is independent at each



**Fig. 8. Delay reduction ratio comparison [43]. (a)  $3 \times 4$  road network. (b)  $5 \times 6$  road network.**

intersection. The optimality and complexity of ODDA were analyzed in [43].

To verify the benefit of the dropbox deployment, ODDA is applied in 2-D networks and compared with the shortest path deployment strategy and the greedy strategy. The shortest path deployment strategy first finds the path with the minimum expected delay without dropboxes. Then, given the number of dropboxes, the locations with the highest benefit are selected to deploy. For the greedy strategy, among all directions toward the destination, the message carrier always selects the one with the largest vehicle density. Fig. 8 presents the delay reduction ratio using the above methods in  $3 \times 4$  and  $5 \times 6$  road networks, respectively.

Here, the delay reduction ratio shows the improvement compared with the delay over the shortest path without any dropboxes. It is observed that ODDA achieves a higher reduction in delay than the other methods, as it combines the information dimension enlargement and dynamic programming to optimize the deployment strategy and path selection simultaneously. It is interesting to note that when the number of deployed dropboxes is small, the greedy strategy even has a higher delay compared with the shortest path without dropboxes. The reason is that the path chosen by the greedy strategy is much longer than the shortest path. Simulation results show that with the optimally deployed dropboxes, the delay and reliability performance can be substantially improved.

### C. RSU Deployment

When Internet connectivity is feasible, RSUs can exchange information through the Internet and serve drive-through vehicles effectively. Thus, how many RSUs to deploy and their locations are key issues to enhance network performance. Owing to the correlation between the RSU deployment and task assignment (i.e., which RSU is assigned to serve a task), it is necessary to jointly consider the optimization of these two coupled problems.

We here still use the delivery delay as the performance metric, given the delay-location-coupled reliability requirements of IoV applications. The 2-D IoV is modeled as a graph  $G = \{\mathcal{V}, \mathcal{E}\}$ , where intersections are abstracted as nodes, and road segments between intersections are edges. Let  $\mathcal{V} = \{v_i : i = 1, 2, \dots, N\}$  be the node set and

$\mathcal{E} = \{e_j : j = 1, 2, \dots, M\}$  be the edge set.  $N$  and  $M$  are the number of nodes and edges, respectively. Then, a comprehensive utility function considering the service benefit and deployment cost has been proposed to make the tradeoff [44]

$$F_U(X, Y) = F_B(X) - F_C(Y) = \sum_{i=1}^N \sum_{j=1}^M \tilde{b}_{ij} x_{ij} - \sum_{i=1}^N f_i y_i \quad (25)$$

where  $\tilde{b}_{ij}$  is the benefit of deploying an RSU at intersection  $v_i$  in providing services to vehicles on edge  $e_j$ , and  $f_i$  is the RSU deployment cost at  $v_i$ . Considering both the installation and maintenance cost,  $f_i$  is the depreciation cost plus the maintenance cost of RSU  $v_i$  over time period  $T$ . It is noted that both benefit and cost are measured over  $T$ .

The deployment and task assignment should satisfy the constraints of both service requirements and RSU capacity. To guide the strategy design, a utility maximization problem is formulated as shown in

$$\begin{aligned} \max_{x,y} \quad & \sum_j \sum_i r_j b_{ij} x_{ij} - \sum_i f_i y_i, \\ \text{s.t.} \quad & \sum_i x_{ij} = 1, \quad \forall j, \\ & x_{ij} \leq y_i, \quad \forall i, j, \\ & \sum_j r_j x_{ij} \leq u_i y_i, \quad \forall i, \\ & 0 \leq x_{ij}, y_i \leq 1, \quad \forall i, j. \end{aligned} \quad (26)$$

Here, it is noted that the feasible region of  $y_i$  has been relaxed from  $\{0, 1\}$  into  $[0, 1]$ , and the dual problem is

$$\begin{aligned} \min_{\alpha, \beta, \gamma, z} \quad & \sum_i z_i - \sum_j \alpha_j \\ \text{s.t.} \quad & \alpha_j \leq -r_j b_{ij} + \beta_{ij} + r_j \gamma_i \quad \forall j \\ & \sum_j \beta_{ij} \leq f_i + z_i - u_i \gamma_i \quad \forall i, j \\ & \beta_{ij}, \gamma_i, z_i \geq 0 \quad \forall i, j \end{aligned} \quad (27)$$

where  $r_j$  and  $u_i$  are the tasks of  $e_j$  and the capacity of  $v_i$ , respectively. Here the utility is measured by the service benefits minus cost. The problem is about how to deploy RSUs to maintain the expected message delivery delay between RSUs and vehicles below the required threshold with the maximum utility [44].

Considering the irregular service area of RSU, first, the overlapping service area will affect the task assignment, as the vehicles in these areas may receive services from multiple RSUs. How to properly select the server and client has a direct influence on the delivery delay, which is critical to the capability of providing a stable service quality. Second, there are up to  $2^N$  possible deployment strategies given  $N$ , the number of location candidates. To obtain the optimal strategy, the task assignments for all the feasible solutions should be considered. Last, the algorithm should comprehensively consider the trade-off between the benefit and cost, and maximize the utility while meeting the delay requirement simultaneously.

---

## Algorithm 2 URDA

---

**Input:**  $\mathcal{C} = \emptyset, \mathcal{S} = E$

```

1: Solve (26) and (27)
2: while  $\mathcal{S} \neq \emptyset$  do
3:    $e_{j^*} = \arg \min_{e_j \in \mathcal{S}} \alpha_j, N_{j^*} = B_{j^*},$ 
4:    $\mathcal{C} = \mathcal{C} \cup e_{j^*}, \mathcal{S} = \mathcal{S} \setminus e_{j^*},$  Update  $B_j, \forall e_j \in \mathcal{S}.$ 
5: end while
6: if  $U(= F - \cup_{e_k \in \mathcal{C}} N_k) \neq \emptyset$  then
7:   for  $\forall v_i \in U$  do
8:      $e_{j^*} = \arg \max_{e_j \in \mathcal{C}} b_{ij}, N_{j^*} = N_{j^*} \cup v_i,$ 
9:   end for
10: end if
11: for  $e_k \in \mathcal{C}$  do
12:   for each  $v_i \in N_k$  do
13:     Open  $v_i$  with  $y_i = 1$ 
14:   end for
15:    $L_k = \{v_i \in N_k : y_i < 1\}, R_k = \sum_{v_i \in L_k} \sum_j x_{ij}$ 
16:   Obtain  $(x^{(k)}, y^{(k)})$  using Algorithm. 2 in [44] with
      $(L_k, R_k)$ 
17: end for
18:  $Y = \{y^{(k)}\}_{e_k \in \mathcal{C}}$ 
19: Obtain  $X$  by solving (P0) in [44] given  $Y$ 
Output:  $(X, Y)$ 

```

---

Based on the linear-program (LP) clustering algorithm [45], a Utility-based RSU Deployment Algorithm (URDA) has been proposed in [44]. As shown in Algorithm 2, based on the solution obtained by solving (26) and (27),<sup>1</sup> URDA first divides the candidate RSUs into clusters, where the set of current cluster centers is denoted by  $\mathcal{C}$  and  $N_k$  is the RSUs clustered around the  $e_k \in \mathcal{C}$ . Let  $B_j$  be the set of unclustered RSUs that are more beneficial to  $e_j$  than any other cluster centers, and  $\mathcal{S}$  be the set of edges that could be chosen as cluster centers. Then, it utilizes the property of the single-node capacitated facility location (SNCFLL) problem to obtain the optimal solution for each subproblem. By combining the solutions of each subproblem, a feasible solution of the original problem is obtained. Through the theoretical analysis, the obtained solution has a provable bound to the optimal one. With the proposed solution, a certain number of RSUs are deployed to maintain the expected message delivery delay between RSUs and vehicles below the required threshold. The results reveal that RSUs effectively improve the reliability of the service performance regardless of vehicle density and traffic estimation errors. Furthermore, it is expected that the proposed algorithm can adaptively adjust the corresponding strategy based on the delay requirements, topology, etc. Thus, the network planning is scalable under various conditions.

The above joint RSU deployment and task assignment strategy can not only improve the delay performance, but

<sup>1</sup>Let  $(x, y)$  and  $(\alpha, \beta, \gamma, z)$  be the optimal solutions to (26) and (27), respectively. Let  $F = \{v_i : y_i > 0\}$  be the opened facilities in  $(x, y)$  and  $F_j = \{v_i : x_{ij} > 0\}$  be the facilities in  $F$  that fractionally serve  $e_j$ .

also be extended to other performance metric as long as it fits the RSU-centric principle. However, for future heterogeneous network structures or service types, how to optimize network planning is much more complicated. For example, the delay performance modeling of multilayer networks is difficult. The RSU service area definition is also challenging if considering various service requirement metrics. How to properly handle these factors beckons further research efforts.

#### D. RSU Variants

The low number of vehicles or insufficient penetration ratio of vehicular communication devices will affect the network connectivity in the early stage of IoV deployment. On the other hand, the high costs of RSUs slow down the deployment progress as well. Thus, Reis *et al.* [46] investigated the feasibility of making part of the parked cars to act as RSUs. A self-organizing mechanism maximizing the network coverage while minimizing the number of the enabled vehicles was proposed to let each parked vehicle decide locally whether to act as an RSU.

In addition to the static RSUs, the idea of mobile RSUs (mRSUs) has been proposed considering the reduced maintenance costs and longer contact time with vehicles. Buses with the high route and schedule regularity are used as a virtual backbone to carry and forward message between taxis [47]. The trace-driven simulations verify the improvement of both end-to-end delay and delivery ratio. Reference [48] utilizes buses as mRSUs to replace some of the static RSUs while still maintaining the equivalent network performance. The trade-off between the performance and the cost is obtained through both mathematical analysis and experiments, which provides the in-depth understanding of the mRSUs and thus is beneficial to the network planning.

For different infrastructures, their functions and capabilities would be different. Selecting the proper infrastructure to facilitate specific applications is significant to the implementation of IoV. Due to the wide coverage and static characteristic, the dropbox is applied in the long-range end-to-end message delivery when pure V2V cannot achieve a satisfying performance. The performance gain

brought by dropboxes mainly comes from the advantage of searching for a suitable next-hop relay. For the information dissemination required in a given area, RSUs offer both the direct Internet access and necessary coverage which are critical to the timely spread. With the development of technology and the decrease of the cost, the future infrastructure is expected to be multifunctional and configurable for the applications.

#### V. CONCLUSION AND OPEN ISSUES

Considering typical applications in IoV, to ensure time and location-coupled reliability, fundamental network properties including connectivity and delay should be quantified. We introduced the modeling and analysis efforts in understanding the performance of V2V and hybrid V2V/V2I networks. Based on the analysis, effective resource management and routing solutions have been proposed. We also studied how to deploy dropboxes or RSUs to enhance network performance in hybrid V2V/V2I networks. To address the scalability issue, we may rely on the distributed reservation and DCC. However, even the state-of-the-art solutions in the literature cannot fully address the challenges due to congestion, hidden terminals, and delay-sensitive QoS requirements, which beckon further research. For example, to scale to the high-density network and the rapid-changing environment, mmWave communications require the high directionality of antennas, which also brings challenges to MAC to detect the angle to steer the communications [49], [50]. To better evaluate the IoV performance including the scalability and reliability, an open, extensive and realistic simulator is also important. For example, Segata *et al.* [51] presented PLEXE which integrated mixed scenarios, wireless networking, and vehicle movement together for the large-scale analysis and exploration. Furthermore, the semiphysical simulation and field experiments still beckon future efforts to realize the solutions in the article. Both the industry and the standards body are looking for effective solutions to address the scalability and reliability issues for large-scale, highly dynamic V2X networks. A timely innovation now will have a profound impact on the coming IoV era. ■

#### REFERENCES

- [1] W. Lewis, "The significance of transportation to civilization," *Ann. Amer. Acad. Polit. Social Sci.*, vol. 187, pp. 1–6, Sep. 1936.
- [2] Department of Economic and Social Affairs, United Nations. (Jun. 2019). *World Population Prospects 2019: Highlights*. Accessed: Sep. 24, 2019. [Online]. Available: <https://www.un.org/development/desa/publications/world-population-prospects-2019-highlights.html>
- [3] M. Heron, "Deaths: Leading causes for 2015," Nat. Center Health Statist., Hyattsville, MD, USA, Nat. Vital Statist. Rep., 2017, pp. 1–75, vol. 66, no. 5.
- [4] Wikipedia. *List of Self-Driving Car Fatalities*. Accessed: Sep. 24, 2019. [Online]. Available: [https://en.wikipedia.org/wiki/List\\_of\\_self-driving\\_car\\_fatalities](https://en.wikipedia.org/wiki/List_of_self-driving_car_fatalities)
- [5] The Tesla Team. *Q3 2018 Vehicle Safety Report*. Accessed: Oct. 24, 2019. [Online]. Available: <https://www.tesla.com/blog/q3-2018-vehicle-safety-report>
- [6] Z. Ning, F. Xia, N. Ullah, X. Kong, and X. Hu, "Vehicular social networks: Enabling smart mobility," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 49–55, May 2017.
- [7] X. Wan *et al.*, "Privacy-preserving content dissemination for vehicular social networks: Challenges and solutions," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1314–1345, 2nd Quart., 2018.
- [8] Z. Ning *et al.*, "A cooperative quality-aware service access system for social Internet of vehicles," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2506–2517, Aug. 2018.
- [9] J. B. Kenney, "Dedicated short-range communications (DSRC) standards in the United States," *Proc. IEEE*, vol. 99, no. 7, pp. 1162–1182, Jul. 2011.
- [10] K. Abboud, H. A. Omar, and W. Zhuang, "Interworking of DSRC and cellular network technologies for V2X communications: A survey," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9457–9470, Dec. 2016.
- [11] X. Wang, Z. Ning, and L. Wang, "Offloading in Internet of vehicles: A fog-enabled real-time traffic management system," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4568–4578, Oct. 2018.
- [12] E. Khorov, A. Kiryanov, A. Lyakhov, and G. Bianchi, "A tutorial on IEEE 802.11ax high efficiency WLANs," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 197–216, 1st Quart., 2019.
- [13] A. Vinel, L. Lan, and N. Lyamin, "Vehicle-to-vehicle communication in C-ACC/platooning scenarios," *IEEE Commun. Mag.*, vol. 53, no. 8, pp. 192–197, Aug. 2015.

- [14] A. Vinel, Y. Koucheryavy, S. Andreev, and D. Staehle, "Estimation of a successful beacon reception probability in vehicular ad-hoc networks," in *Proc. Int. Conf. Wireless Commun. Mobile Comput., Connecting World Wirelessly*, 2009, pp. 416–420.
- [15] A. Abdullah, L. Cai, and F. Gebali, "DSDMAC: Dual sensing directional MAC protocol for ad hoc networks with directional antennas," *IEEE Trans. Veh. Technol.*, vol. 61, no. 3, pp. 1266–1275, Mar. 2012.
- [16] N. Lyamin, A. Vinel, M. Jonsson, and B. Bellalta, "Cooperative awareness in VANETs: On ETSI EN 302 637-2 performance," *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 17–28, Jan. 2018.
- [17] *Decentralized Congestion Control Mechanisms for Intelligent Transport Systems Operating in the 5 GHz Range; Access Layer Part*, document ETSI TS 102 687, 2011.
- [18] G. Bansal, J. B. Kenney, and C. E. Rohrs, "LIMERIC: A linear adaptive message rate algorithm for DSRC congestion control," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4182–4197, Nov. 2013.
- [19] G. Bansal, B. Cheng, A. Rostami, K. Sjoberg, J. B. Kenney, and M. Gruteser, "Comparing LIMERIC and DCC approaches for VANET channel congestion control," in *Proc. IEEE 6th Int. Symp. Wireless Veh. Commun. (WiVeC)*, Vancouver, BC, Sep. 2014, pp. 1–7.
- [20] B. Cheng, A. Rostami, M. Gruteser, J. B. Kenney, G. Bansal, and K. Sjoberg, "Performance evaluation of a mixed vehicular network with CAM-DCC and LIMERIC vehicles," in *Proc. IEEE 16th Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Boston, MA, USA, Jun. 2015, pp. 1–6.
- [21] S. Subramanian et al., "Congestion control for vehicular safety: Synchronous and asynchronous MAC algorithms," in *Proc. 9th ACM Int. Workshop Veh. Inter-Netw., Syst., Appl. (VANET)*, New York, NY, USA, 2012, pp. 63–72.
- [22] *Physical Layer Procedures (Release 14)*, document TS 36.212, V15.2.1, 3GPP, Jul. 2018.
- [23] *Overall Description (Release 14)*, document TS 36.300, V15.2.0, 3GPP, Jul. 2018.
- [24] *Radio Resource Control (Release 14)*, document TS 36.331, V15.2.2, 3GPP, Jul. 2017.
- [25] R. Molina-Masegosa and J. Gozalvez, "System level evaluation of LTE-V2V mode 4 communications and its distributed scheduling," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC-Spring)*, Jun. 2017, pp. 1–5.
- [26] *New WID on 3GPP V2X Phase 2, RP-170798*, Huawei, CAT, LG Electron., HiSilicon, China Unicom, Dubrovnik, Croatia, 2018.
- [27] M. Gonzalez-Mart et al., "Analytical models of the performance of C-V2X mode-4 vehicular communications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1155–1166, Feb. 2019.
- [28] 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 Veh. Technol. Mag.*, vol. 12, no. 4, pp. 30–39, Dec. 2017.
- [29] 5G Automotive Association, "The case for cellular V2X for safety and cooperative driving," White Paper, Nov. 2016. Accessed: Sep. 24, 2019. [Online]. Available: <https://5gaa.org/wp-content/uploads/2017/10/5GAA-whitepaper-23-Nov-2016.pdf>
- [30] X. Wang, S. Mao, and M. X. Gong, "An overview of 3GPP cellular vehicle-to-everything standards," *GetMobile, Mobile Comput. Commun.*, vol. 21, no. 3, pp. 19–25, Nov. 2017.
- [31] N. Bonjorn, "Cooperative resource allocation and scheduling for 5G eV2X services," *IEEE Access*, vol. 7, pp. 58212–58220, Jan. 2019.
- [32] F. Bai and B. Krishnamachari, "Spatio-temporal variations of vehicle traffic in VANETs: Facts and implications," in *Proc. 6th ACM Int. Workshop Veh. Internetwork. (VANET)*, Beijing, China, 2009, pp. 43–52.
- [33] Y. Zhuang, J. Pan, Y. Luo, and L. Cai, "Time and location-critical emergency message dissemination for vehicular ad-hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 1, pp. 187–196, Jan. 2011.
- [34] J. He, L. Cai, J. Pan, and P. Cheng, "Delay analysis and routing for two-dimensional VANETs using carry-and-forward mechanism," *IEEE Trans. Mobile Comput.*, vol. 16, no. 7, pp. 1830–1841, Jul. 2017.
- [35] Y. Zhuang, J. Pan, and L. Cai, "A probabilistic model for message propagation in two-dimensional vehicular ad-hoc networks," in *Proc. 7th ACM Int. Workshop Veh. Internetwork. (VANET)*, Chicago, IL, USA, 2010, pp. 31–40.
- [36] M. Ni, J. Pan, L. Cai, J. Yu, H. Wu, and Z. Zhong, "Interference-based capacity analysis for vehicular ad hoc networks," *IEEE Commun. Lett.*, vol. 19, no. 4, pp. 621–624, Apr. 2015.
- [37] L. Zhang, L. Cai, and J. Pan, "Connectivity in two-dimensional lattice networks," in *Proc. IEEE Infocom*, Turin, Italy, Apr. 2013, pp. 2814–2822.
- [38] L. Zhang, *Square Lattice Network Directed Connectivity Calculator*. Accessed: Dec. 15, 2019. [Online]. Available: <http://lab.pan.uvic.ca/lattice/>
- [39] E. Domany and W. Kinzel, "Directed percolation in two dimensions: Numerical analysis and an exact solution," *Phys. Rev. Lett.*, vol. 47, no. 1, pp. 5–8, 1981.
- [40] J. Zhao and G. Cao, "VADD: Vehicle-assisted data delivery in vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 3, pp. 1910–1922, May 2008.
- [41] C. Glacet, M. Fiore, and M. Gramaglia, "Temporal connectivity of vehicular networks: The power of store-carry-and-forward," in *Proc. IEEE Veh. Netw. Conf. (VNC)*, Kyoto, Japan, Dec. 2015, pp. 52–59.
- [42] J. He, L. Cai, P. Cheng, and J. Pan, "Delay minimization for data dissemination in large-scale VANETs with buses and taxis," *IEEE Trans. Mobile Comput.*, vol. 15, no. 8, pp. 1939–1950, Aug. 2016.
- [43] J. He, Y. Ni, L. Cai, J. Pan, and C. Chen, "Optimal dropbox deployment algorithm for data dissemination in vehicular networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 3, pp. 632–645, Jul. 2017.
- [44] Y. Ni, J. He, L. Cai, J. Pan, and Y. Bo, "Joint roadside unit deployment and service task assignment for Internet of Vehicles (IoV)," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3271–3283, 2019.
- [45] R. Levi, D. B. Shmoys, and C. Swamy, "LP-based approximation algorithms for capacitated facility location," *Math. Program.*, vol. 131, no. 1, pp. 365–379, 2012.
- [46] A. B. Reis, S. Sargento, and O. K. Tonguz, "Parked cars are excellent roadside units," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2490–2502, Sep. 2017.
- [47] L. Zhang, Y. Zhuang, J. Pan, L. Kaur, and H. Zhu, "Multi-modal message dissemination in vehicular ad-hoc networks," in *Proc. 1st IEEE Int. Conf. Commun. China (ICCC)*, Aug. 2012, pp. 670–675.
- [48] J. Heo, B. Kang, J. M. Yang, J. Paek, and S. Bahk, "Performance-cost trade-off of using mobile roadside units for V2X communication," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 9049–9059, Sep. 2019, doi: [10.1109/TVT.2019.2925849](https://doi.org/10.1109/TVT.2019.2925849).
- [49] I. Mavromatis, A. Tassi, R. J. Piechocki, and A. Nix, "MmWave system for future ITS: A MAC-layer approach for V2X beam steering," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Toronto, ON, Canada, Sep. 2017, pp. 1–6.
- [50] C. Fiandrino, H. Assasa, P. Casari, and J. Widmer, "Scaling millimeter-wave networks to dense deployments and dynamic environments," *Proc. IEEE*, vol. 107, no. 4, pp. 732–745, Feb. 2019.
- [51] M. Segata, S. Joerer, B. Bloessl, C. Sommer, F. Dressler, and R. L. Cigno, "PLEXE: A platooning extension for Veins," in *Proc. 6th IEEE Veh. Netw. Conf. (VNC)*, Paderborn, Germany, Dec. 2014, pp. 53–60.

## ABOUT THE AUTHORS

**Yuanzhi Ni** (Member, IEEE) received the B.Sc. degree in automation and the Ph.D. degree in control science and engineering from the Nanjing University of Science and Technology, Nanjing, China, in 2012 and 2019, respectively.

From 2015 to 2017, he was a jointly supervised Ph.D. Student with the Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada. He is currently a Lecturer with the Department of Automation, Jiangnan University, Wuxi, China. His current research interests include Internet of Vehicles and networked control.



**Lin Cai** (Senior Member, IEEE) received the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2002 and 2005, respectively.

Since 2005, she has been with the Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada, where she is currently a Professor. Her research interests span several areas in communications and networking, with a focus on network protocol and architecture design supporting emerging multimedia traffic and Internet of Things.



Dr. Cai was a member of the Steering Committee of the IEEE TRANSACTIONS ON BIG DATA and the IEEE TRANSACTIONS ON CLOUD COMPUTING. She was a recipient of the NSERC Steacie Memorial Fellowship in 2019, the NSERC Discovery Accelerator Supplement (DAS) Grants in 2010 and 2015, and the Best Paper Awards of the IEEE ICC 2008 and the IEEE WCNC 2011. She has served as the TPC Symposium Co-Chair for the IEEE Globecom 2010 and Globecom 2013. She has funded and chaired the IEEE Victoria

Section Vehicular Technology and Communications Joint Societies Chapter. She has been elected to serve the IEEE Vehicular Technology Society Board of Governors from 2019 to 2021. She has served as an Area Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, an Associate Editor for the IEEE INTERNET OF THINGS JOURNAL, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE TRANSACTIONS ON COMMUNICATIONS, the *EURASIP Journal on Wireless Communications and Networking*, the *International Journal of Sensor Networks*, and the *Journal of Communications and Networks*, and as a Distinguished Lecturer of the IEEE VTS Society. She is a Registered Professional Engineer of British Columbia, Canada.

**Jianping He** (Member, IEEE) received the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2013.

He was a Research Fellow with the Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada, from December 2013 to March 2017. He is currently an Associate Professor with the Department of Automation, Shanghai Jiao Tong University, Shanghai, China. His research interests mainly include the distributed learning, control and optimization, and security and privacy in network systems.

Dr. He received the Outstanding Thesis Award, Chinese Association of Automation, in 2015. He received the Best Paper Award from the IEEE WCSP'17, the Best Conference Paper Award from the IEEE PESGM'17, and was a Finalist for the Best Student Paper Award from the IEEE ICCA'17. He was also a Guest Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, the *International Journal of Robust and Nonlinear Control*, and so on. He serves as an Associate Editor for the IEEE OPEN JOURNAL OF VEHICULAR TECHNOLOGY and the *KSII Transactions on Internet and Information Systems*.

**Alexey Vinel** (Senior Member, IEEE) received the Ph.D. degrees from the Institute for Information Transmission Problems, Moscow, Russia, in 2007, and from the Tampere University of Technology, Tampere, Finland, in 2013.

He has been a Professor with the School of Information Technology, Halmstad University, Halmstad, Sweden, since 2015, and a Professor II with the Department of Electrical Engineering, Western Norway University of Applied Sciences, Bergen, Norway, since 2018. His area of interests includes wireless communications, vehicular networking, and cooperative autonomous driving.



**Yue Li** received the Ph.D. degree in electrical and computer engineering from the University of Victoria, Victoria, BC, Canada, in 2018.

From 2008 to 2013, he worked as a Standard Preresearch Engineer with the Wireless Research Department, Huawei, Shenzhen, China. He has been closely involved in 3GPP standards evolution and has held numerous patents in WCDMA, LTE-A, and 5G systems. He is currently a Postdoctoral Research Fellow with the Department of Electrical and Computer Engineering, University of Victoria. His research interests include next-generation cellular systems, wireless network design and optimization, wireless system modeling, and performance analysis.



**Hamed Mosavat-Jahromi** (Student Member, IEEE) received the B.Sc. degree in electrical engineering from the Iran University of Science and Technology, Tehran, Iran, in 2012, and the M.Sc. degree in electrical engineering from the University of Tehran, Tehran, in 2015. He is working toward the Ph.D. degree in electrical engineering at the Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada.

His research interests include vehicular networks, Internet of Things, machine learning, and optimization with applications in networking.



**Jianping Pan** (Senior Member, IEEE) received the bachelor's and Ph.D. degrees in computer science from Southeast University, Nanjing, Jiangsu, China, in 1994 and 1998, respectively.

He did his postdoctoral research with the University of Waterloo, Waterloo, ON, Canada. He also worked with the Fujitsu Laboratories, Sunnyvale, CA, USA, and NTT Multimedia Communications Laboratories, Palo Alto, CA, USA. His area of specialization is computer networks and distributed systems. He is currently a Professor of computer science with the University of Victoria, Victoria, BC, Canada. His current research interests include protocols for advanced networking, performance analysis of networked systems, and applied network security.

Dr. Pan is a Senior Member of the ACM. He has been serving on the Technical Program Committees of major computer communications and networking conferences, including the IEEE INFOCOM, ICC, Globecom, WCNC, and CCNC. He received the IEICE Best Paper Award in 2009, the Telecommunications Advancement Foundation's Telesys Award in 2010, the WCSP 2011 Best Paper Award, the IEEE Globecom 2011 Best Paper Award, the JSPS Invitation Fellowship in 2012, the IEEE ICC 2013 Best Paper Award, and the NSERC DAS Award in 2016. He is the Ad Hoc and Sensor Networking Symposium Co-Chair of the IEEE Globecom 2012. He is an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

