# Intelligent Context-Aware Communication Paradigm Design for IoVs Based on Data Analytics

Feng Lyu, Nan Cheng, Hongzi Zhu, Haibo Zhou, Wenchao Xu, Minglu Li, and Xuemin (Sherman) Shen

## Abstract

IoVs have been envisioned to improve road safety and efficiency, and provide Internet access on the move, by providing a myriad of safety and infotainment applications to drivers and passengers. However, with limited spectrum resource, harsh wireless channel, and variable vehicle density, IoV communication faces severe challenges to achieve scalability, efficiency, and reliability. In this article, we propose a context-aware IoV paradigm design to enhance the communication performance, where the high-level contextual information is utilized to bring intelligence in the design. Specifically, through big data analytics on large-scale IoV communication traces collected from an extensive experiment conducted in Shanghai, we investigate the impacts of different contextual information on V2V communication performance. We reveal that among many types of contextual information, the NLoS link condition is a major one that significantly affects V2V link performance. Based on that observation, we discuss three critical but challenging communication paradigm designs with context awareness of V2V link conditions: smart medium resource allocation, efficient routing establishment, and reliable safety message broadcasting. Furthermore, we present a case study of a cooperative beaconing scheme, where machine learning methods are utilized to learn the real-time link contextual information, and vehicles in deep NLoS condition choose helpers to enhance the overall beaconing reliability.

## Introduction

The Internet of Vehicles (IoVs) constitutes the cornerstone of intelligent transportation systems (ITS) by allowing vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications, and in general, vehicle-to-everything (V2X) communications [1–3]. By efficiently exchanging information among vehicles, communication infrastructure, and the Internet, a wide spectrum of applications, ranging from road safety and real-time navigation to entertainment and self-driving, can be provided to drivers and passengers. Specifically, in safety applications, including collision avoidance, safety warnings, remote vehicle diagnostic, and so on, vehicles periodically broadcast safety messages of positional and kinematic information to their one-hop neighbors, where ultra-low delay and high reliability is required; in comfort applications like file downloading, web browsing, and video streaming, vehicles may fetch the contents from Internet servers or edge servers through IoVs, where user quality of experience (QoE) should be satisfied. To this end, IoV communication paradigms (e.g., medium resource allocation, routing, and broadcasting) should be carefully designed with efficiency, scalability, and reliability.

However, the inherent features of IoV pose great challenges to efficient communication paradigms design, which can be described as follows.

**Spectrum Resource Shortage:** The IEEE 802.11p-based dedicated short-range communication (DSRC) has been a standard for vehicular communications in recent years, in which one control channel (CCH) and multiple service channels (SCHs) with two optional bandwidths of 10 MHz and 20 MHz are set to simultaneously support safety and non-safety services. However, the Federal Communications Commission (FCC) only allocates 75 MHz 5.9 GHz licensed spectrum for DSRC, which is insufficient to support ever increasing and medium-rich applications, especially for those scenarios where the vehicle density is high [4]. Therefore, it is challenging to guarantee the required vehicular ad hoc network (VANET) performance under limited spectrum resources.

**Harsh Wireless Channel:** As vehicles move, the link conditions between them vary quickly and can be intermittent over time. For instance, urban driving environments can be highly dynamic and complex with many unpredictable factors such as time-varying traffic densities, surrounding buildings, types of roads, and trees, which stochastically degrade the V2V link performance [5]. As a result, reliable packet delivery under an unstable wireless channel is nontrivial.

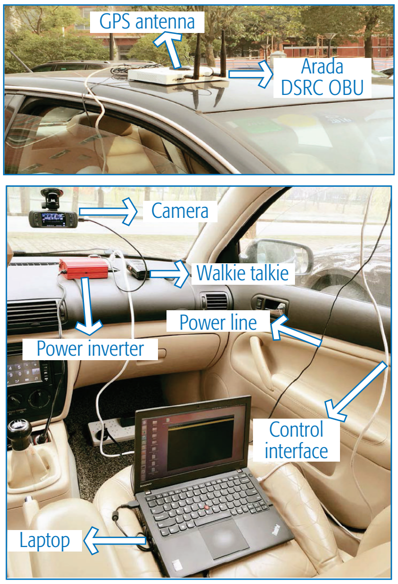
**Variable Vehicle Density:** Vehicle density dramatically varies due to vehicle mobility [6]. It could be extremely high under traffic jams in urban and highway environments, while relatively low in suburban environments. The dynamic topology makes optimal resource scheduling very hard to achieve. For example, scarce spectrum may be wasted under low-density scenarios, while channel congestion can easily be triggered under high-density scenarios.

In IoV, contextual information is usually utilized in designing efficient and intelligent protocols in order to improve the resource utilization

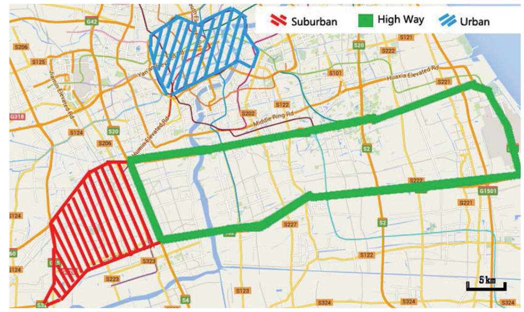
Digital Object Identifier: 10.1109/MNET.2018.1800067

*Feng Lyu, Hongzi Zhu, and Minglu Li are with Shanghai Jiao Tong University; Nan Cheng (corresponding author), Wenchao Xu, and Xuemin (Sherman) Shen are with the University of Waterloo; Haibo Zhou is with Nanjing University.*




(a)




| Dataset | Highway | Suburban | Urban |
|---|---|---|---|
| No. transmissions | 165,631 | 114,436 | 320,078 |
| No. receptions | 136,043 | 101,891 | 292,479 |
| Mean distance (m) | 342.535 | 160.187 | 147.796 |
| Mean Tx speed (m/s) | 4.386 | 3.523 | 5.107 |
| Mean Rx speed (m/s) | 6.195 | 3.632 | 5.041 |
| LoS duration (s) | 13,659 | 10,029 | 29,290 |
| NLoS duration (s) | 2,763 | 1,001 | 3,144 |

(b)

FIGURE 1. Illustration of data collection: a) experimental devices; b) data collection environments and dataset details.

and enhance the network performance under the dynamic vehicular communication environments. Typical context information such as position, speed, direction, acceleration, road traffic, road conditions, and weather information can be effectively obtained from onboard sensors or remote servers, and shared among vehicles to design intelligent context-aware IoV protocols, such as location-aware or mobility-prediction based protocols [7, 8]. Furthermore, from such low-level context information, implicit high-level context such as vehicle density, link condition and driver behavior can be deducted, which reflects more detailed and deeper information about the environments.

Modern communication networks are generating fast escalations of data due to the proliferation of mobile devices, widespread network monitoring, and emerging data-craving applications. Such big data can be extensively analyzed to reveal useful information and insights on the network, improve the network operations, and intelligently support a variety of mobile users and services. For example, high-level contextual information such as network key performance indicators (KPIs) can be inferred or learned from the network or traffic monitoring data by artificial intelligence (AI) methods and data analytics, and employed to design intelligent network operations [9].

In this article, we propose intelligent context-aware IoV protocol design by employing big data analytics and machine learning methods, where the high-level contextual information of link conditions, that is, line of sight (LoS) and non-line of sight (NLoS) conditions, is inferred and considered. We collect large datasets in Shanghai from two experiment cars conducting V2V communications, with each mounting a commodity DSRC radio and two cameras, and we visually label all LoS and NLoS situations by checking videos across all collected data traces [10]. Through statistical analytics on the collected data, it is revealed that contextual information such as separation distance, velocity, and altitude does not significantly impact V2V performance. Instead, link condition (e.g., LoS/NLoS condition) is a major context that significantly affects V2V link performance. With such knowledge, we then discuss three V2V communication paradigm designs with context awareness of link conditions:

- Smart medium resource allocation
- Efficient routing establishment
- Reliable safety message broadcasting

Moreover, a case study is provided, where an intelligent cooperative beaconing scheme is proposed to enhance the IoV broadcast reliability. Machine learning methods are employed to detect the NLoS contextual information in real time, and helpers are chosen to rebroadcast the beaconing messages of the vehicles in harsh NLoS conditions.

As vehicular environments can be highly dynamic due to the complicated urban scenarios in terms of traffic densities, surrounding buildings, types of road, and so on, we deploy two cameras in front and rear of each experiment car to record video of surrounding environment.

## Characterizing the Context in V2V Communication

### Collecting V2V Communication Data

We collect data in two experiment cars driving normally in Shanghai, which continuously conduct V2V communications through an Arada LocoMateTM onboard unit (OBU) mounted on the roof. Specifically, we implement a *unicast* application in the OBU, in which a 300-byte packet is sent every 100 ms, based on the Wave Short Message Protocol (WSMP) in the DSRC module. Each packet includes a sequence number, and the latitude, longitude, altitude, and instantaneous speed of the transmitter vehicle. Both vehicles log the packet transmission and

We label all NLoS conditions by checking the video data, where vehicles cannot visually see each other, and we have synchronized the time between the video data and V2V communication data for in-depth data analysis.

reception, along with their own position information. As vehicular environments can be highly dynamic due to the complicated urban scenarios in terms of traffic densities, surrounding buildings, types of road, and so on, we deploy two cameras in the front and rear of each experiment car to record video of the surrounding environment, which can help us delve into the specific contextual information (e.g., the traffic density, surrounding vehicles, and NLoS/LoS conditions). Figure 1a shows the experimental devices, where the laptop and the Ethernet control interface are adopted to control the OBU and download the collected data from it due to its very limited storage capability. To cover different traffic conditions, and various road types involving a sufficiently long time period, we conduct data collection campaigns in highway, suburban, and urban areas in Shanghai, and the overall campaign lasts for over two months including both rush hour and off-peak time conditions. Three datasets from highway, suburban, and urban areas are collected, with the total traveling distance over 1500 km and file size up to 110 GB. Figure 1b demonstrates the data collection environments and key statistics of the three data traces. Notice that we label all NLoS conditions by checking the video data, where vehicles cannot visually see each other, and we have synchronized the time between the video data and V2V communication data for in-depth data analysis.

### Slight Impact of Kinematic Factors

In this subsection, we study the impacts of kinematic factors on V2V performance, that is, separation distance, velocity, and altitude. To achieve the most reliable communication, we adopt the highest transmission power, 14 dBm, and the lowest data rate, 3 Mb/s, on a channel with 10 MHz bandwidth in all experiments.

**Impact of Separation Distance:** We first investigate the impact of distance on V2V performance. Figure 2 shows the cumulative distribution functions (CDFs) of packet delivery ratio (PDR) within different distance ranges in three environments. We borrow two benchmarks from [5], *good reception* (PDR ≥ 80 percent) and *poor reception* (PDR < 20 percent), to examine the overall performance. It can be seen that across all environments PDR drops slightly as distance increases. For example, under the suburban environment, the proportion of good reception is about 99, 97, 96, and 92 percent in respective distance ranges, that is, 0–100 m, 100–200 m, 200–300 m, and 300–500 m. A more significant performance degradation can happen in highway and urban scenarios. The main reason that a supreme PDR performance can be achieved in the suburban environment is because slower speeds, shorter distances, and lighter traffic densities prevail in this environment. Nevertheless, the overall performance is rather reliable in all scenarios compared to results in [5], where *intermediate reception* (20 percent ≤ PDR > 80 percent) prevails throughout the whole communication range with a probability of 50.6 percent, while the good reception zone only has a probability of 35.2 percent. The conclusion is a little different from the reference because very likely the setup of the two experiments may be different (scenarios, transmission power, etc.). In summary, the separation distance has limited impact on V2V performance degradation, and the V2V link is reliable enough to support most vehicular applications when the separation distance is within 500 m.

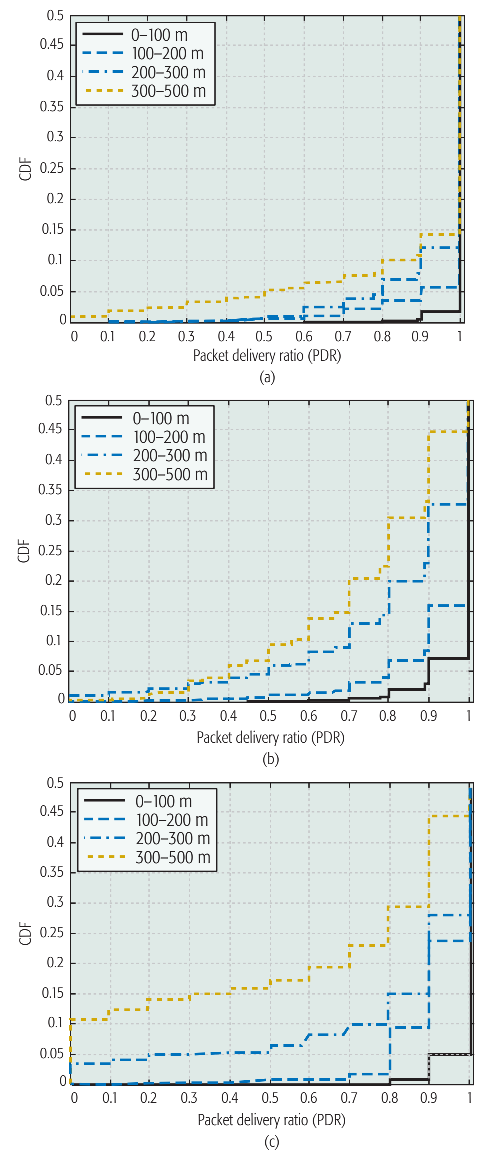



FIGURE 2. CDFs of PDR within different separation distance ranges: a) suburban; b) highway; c) urban.

**Impacts of Effective Velocity and Relative Altitude:** To investigate how velocity affects PDR performance, we adopt effective velocity, that is,

$$V_{eff} = \sqrt{V_{tx}^2 = V_{rx}^2},$$

to represent the variation of velocities of two vehicles. Figures 3a and 3b show CDFs of PDR at different effective velocity ranges in highway and

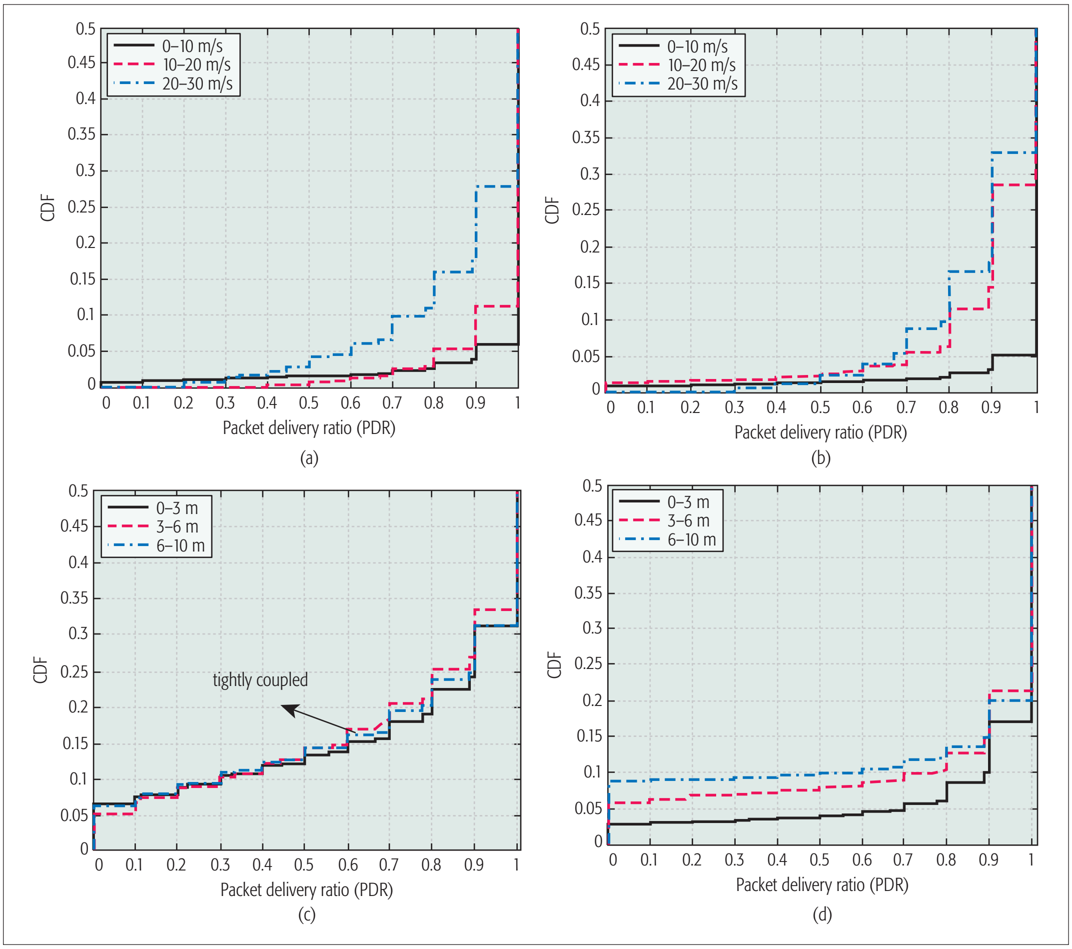



FIGURE 3. CDFs of PDR at different effect velocity and relative altitude ranges: a) highway: at different effect velocity ranges; b) urban: at different effect velocity ranges; c) highway: at different relative altitude ranges; d) urban: at different relative altitude ranges

urban environments. Note that the result under suburban environments is omitted due to similar observations and space limitations. It can be seen that as effective velocity increases, the PDR performance gradually decreases. Particularly, in the highway environment, the probability of good reception decreases from 97 to 89 percent when the effective velocity range increases from 0–10 m/s to 20–30 m/s. A similar observation also holds in the urban environment. In addition, at the relatively high velocity ranges (e.g., 20–30 m/s), the overall performance is still reliable (with a probability of about 90 percent in a good reception zone), which indicates that high velocity has little impact on V2V performance.

We then examine the impact of altitude on V2V performance, and adopt relative altitude (i.e., $A_{rel} = |A_{tx} - A_{rx}|$) to represent the variation of altitudes of two vehicles. Figures 3c and 3d show CDFs of PDR at different relative altitude ranges under highway and urban environments. It can be observed that relative altitude has very small effect on V2V link performance. For example, in highway environments, CDF results at different relative altitude ranges are analogous, all of which can achieve good reception (about 80 percent); under the urban environments, the proportion of good reception is about 94, 90, and 88 percent in relative altitude of 0–3, 3–6, and 6–10 m, respectively, which are still very close.

During the data collection experiments, we surprisingly find that packet losses could be very common when the sending vehicle cannot see the receiving vehicle directly, that is, two vehicles are being blocked by big obstacles (NLoS), including big trucks and buses, slopes, turns, and so on.

## Key Factors in Performance Degradation

Normally, good reception can broadly happen throughout our overall trace; however, during the data collection experiments, we surprisingly find that packet losses could be very common when the sending vehicle cannot see the receiving vehicle directly, that is, two vehicles are being blocked by big obstacles (NLoS), including big trucks and buses, slopes, turns, and so on. On the contrary, when two vehicles are within sight of each other (LoS), packets are always perfectly received. Therefore, we then investigate the factor of link conditions in terms of LoS and NLoS, which may impact the V2V performance. By checking videos, we label all NLoS situations when two vehicles cannot visually see each other. It should be noted that although NLoS conditions found by cameras are not necessarily NLoS for RF radios, those visually NLoS conditions are still good approximations of real radio NLoS conditions and valuable for analysis.

**LoS/NLoS Conditions Are Key Factors in V2V Performance Degradation:** According to found

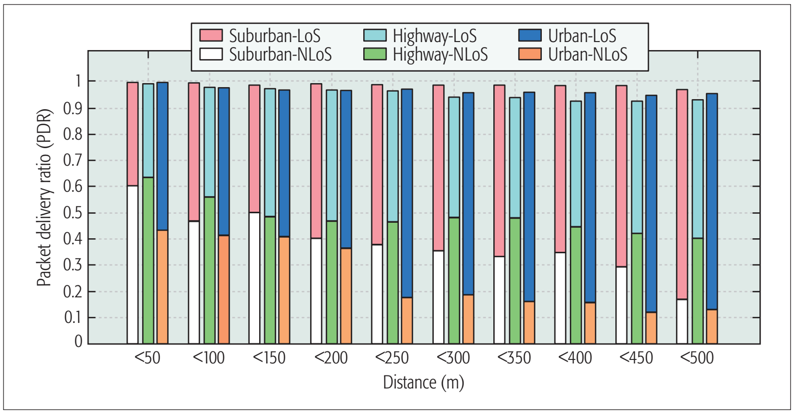



FIGURE 4. Average PDR vs. distance under LoS/NLoS conditions.

LoS and NLoS conditions, we divide the trace into two parts and obtain the average PDR vs. distance under LoS/NLoS conditions, respectively (Fig. 4). We can easily observe that under LoS conditions, most packets are well received, while under NLoS conditions, packet reception failures can happen frequently. For example, under LoS conditions, all PDRs are above 95 percent, whereas most PDRs under NLoS conditions are below 50 percent regardless of the distance variation. We can conclude that it is the NLoS condition that significantly affects link performance, and this can explain well why PDR can be very high even at relative long distance, high effective speed, and large relative altitude. Note that even though the distance is not the direct reason for poor PDR, long distance very likely leads to a higher probability of encountering an NLoS condition.

## Context-Aware Communication Paradigms Design

Many communication paradigms could benefit from context-aware design if the link condition can be detected in real time. In this section, we elaborate the context-aware design of three IoV paradigms, that is, medium resource allocation, routing, and broadcasting safety messages. In the following section, we provide a case study to show online NLoS detection using machine learning methods and a safety message beaconing scheme based on the link condition context.

### Smart Medium Resource Allocation

Medium access control (MAC) protocol is of great importance for applications in VANETs, and distinct applications have their own particular medium access requirements. Specifically, for safety applications, ultra-low delay and high reliability should be guaranteed. In contrast, comfort applications can tolerate a certain delay but need large numbers of media for file downloading or video streaming. In current VANET MAC protocols like contention-based MAC *802.11p* [11] and time-division multiple access (TDMA)-based MACs (e.g., *VeMAC* [12]), vehicles are treated equally to access the channel without considering link conditions. However, due to the medium scarcity, this equal fairness scheme may degrade the system performance, especially for high-density scenarios, as transmissions under NLoS conditions can hardly succeed but incur interference to neighboring vehicles. A smarter strategy is to allocate the medium resources to those vehicles having good links with their receivers. For example, in Fig. 5a, each vehicle wants to access the channel and transmit data to vehicle *D*. As heavy traffic, like big buses or trucks, will follow vehicle *D*, links are under NLoS conditions between vehicle *D* and behind vehicles *a*, *b*, and *c*. In contrast, links between vehicle *D* and ahead vehicles *A*, *B*, and *C* are under LoS conditions. Obviously, vehicles *A*, *B*, and *C* should have higher priority to access the channel. In 802.11p, vehicles negotiate the channel usage under distributed coordination function (DCF), which relies on the carrier sens multiple access with collision avoidance (CSMA/CA) mechanism: when a vehicle wants to access the medium, it has to sense the channel first; if the channel is idle, the vehicle can access the medium; otherwise, it has to perform random back-off. Under this case, back-off time could be unequally set considering link qualities. By doing this, when in light traffic scenarios, vehicles with good link conditions can access the channel with small delays, and in heavy traffic scenarios, the resource utilization can be enhanced. On the other hand, in TDMA-based MAC, time is partitioned into frames consisting of equal-length time slots and synchronized among vehicles. During each time slot, every vehicle decides whether to transmit a packet with a probability $p$. To achieve fairer time slot acquisition, the $p$ value of each vehicle can be dynamically set based on its link conditions. In the broadcast paradigm for safety applications, the number of LoS links $|LoS|$ and NLoS links $|NLoS|$ of each vehicle to its one-hop neighbors can also be an effective indicator for medium resource allocation. Beyond that, the congestion control framework can also take the context-aware factor into consideration to maximize the optimal performance.

### Efficient Routing Establishment

In VANETs, for two remote vehicles, data packets can be exchanged through multiple hops. However, routing in vehicular networks is challenging due to highly variable network topologies and intermittent wireless link connections. On one hand, if broadcasting is used to discover routes and each vehicle blindly retransmits broadcasted packets (flooding), the channel will witness an explosive growth of traffic (broadcast storm problem) and could be congested. As a result, transmissions can hardly succeed due to frequent collisions. On the other hand, if a routing table is maintained, the source node can query the routing table when it starts a route discovery process. However, it is hard to efficiently update the table and guarantee the correctness of routing information due to the high speed of moving vehicles. Therefore, the packets are very likely to be dropped due to the error routes or unreliable wireless link conditions. Consequently, taking link conditions into consideration can mitigate this problem, that is, routing packets to those links with good conditions, which could improve the routing performance. For instance, in Fig. 5b, vehicle *A* wants to transmit packets to vehicle *B*, which is out of communication range of vehicle *A*. Vehicle *C* is within one-hop range of both vehicles *A* and *B*, while vehicle *D* and *E* are within

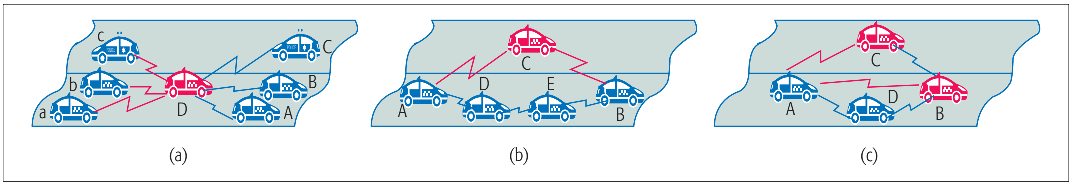



FIGURE 5. Context-aware communication paradigm designs: a) smart medium resource allocation; b) efficient routing establishment; c) broadcasting beacons reliably.

one-hop range of the sender $A$ and receiver $B$, respectively. To route a packet from $A$ to $B$, one simple way is to choose vehicle $C$ as a relay due to the minimum hops. However, since link conditions between $C$ and both $A$ and $B$ are under NLoS, which means that packet reception failures would commonly happen, $C$ can hardly receive the transmitted packet, and even if it receives the packet, the retransmission is also very likely to fail. Another route is from $A \rightarrow D$, $D \rightarrow E$, then $E \rightarrow B$. Although one more hop of transmission is required, the packet can be well received by the receiver due to LoS links. Considering the probability of successfully receiving a packet under LoS and NLoS conditions to be $P_L$ and $P_N$ and with the value of 0.95 and 0.2,[1] respectively, the probability of successfully routing the packet by vehicle $C$ is $P_N^2 = 0.04$, while the probability could reach $P_L^3 = 0.86$ when relayed by vehicles $D$ and $E$. Therefore, link conditions can heavily affect route performance and should be considered as a significant factor in designing efficient routing strategies.

### Reliable Safety Messages Broadcasting

To enhance driving safety, IEEE 802.11p-based DSRC standards provide broadcast services (i.e., periodically broadcasting safety beacons) to let each vehicle keep up-to-date knowledge of surrounding environments. Based on constantly updated information, upper-layer safety applications can be supported, and thus any loss of beacons might result in potential danger. In the literature, many works have studied how to transmit safety beacons and guarantee transmission requirements to support reliable one-hop broadcast communication. For example, in [13], Zhang *et al*. designed adaptive beaconing schemes with safety awareness to control beacon congestion on the control channel. In all these research works, a common assumption is that if a beacon is sent out, it can be received well as long as two vehicles are within communication range. However, through our experiment, we find that V2V communication can be intermittent and heavily influenced by driving context (e.g., NLoS conditions). Moreover, in broadcast mode of 802.11p protocol, request to send/clear to send/acknoledgment (RTS/CTS/ACK) packets are removed to facilitate response, and without ACKs, it is very hard to confirm whether safety beacons are received well. Therefore, it is nontrivial to achieve reliable safety beacons. With the knowledge that V2V communications could be reliable in LoS conditions but hardly succeed in NLoS conditions, one possible beaconing strategy is to find a proper neighbor vehicle to cooperatively help rebroadcast beacons when encountering harsh NLoS conditions. For example, in Fig. 5c, since vehicles $A$ and $B$ are blocked by obstacles, vehicle $B$ can hardly receive safety beacons from vehicle $A$. To enhance the reliability of beaconing, vehicle $A$ can ask vehicle $D$ to rebroadcast the beacon, and we define this vehicle as a *helper*. The sender should select a helper with the best link quality with both the sender and the receiver among all the optional helpers. Therefore, vehicle $C$ should not be chosen as a helper due to the NLoS link condition between itself and the sender. This context-aware beaconing strategy can improve the reliability in safety beaconing without costing much rebroadcast overhead. Since beacon exchange in real time is highly related to driving safety, ultra-high reliability is normally required. To this end, the space-air-ground integrated network architecture can also be utilized, in which drone assisted communication can easily provide LoS links to relay the beacons [14, 15].

Many communication paradigms could benefit from context-aware design if the link condition can be detected in real time. In this section, we elaborate the context-aware design of three IoV paradigms, that is, medium resource allocation, routing, and broadcasting safety messages.

## Case Study

In this section, we study the case where contextual information of LoS/NLoS conditions is detected and applied in the V2V beaconing design. A cooperative beaconing scheme, named Co-beacon, is proposed, which integrates the following three novel techniques:
- Online NLoS detection
- Link status exchange
- Beaconing with helpers

### Online NLoS Detection

To employ the contextual information of link condition, we first propose a supervised machine-learning-based online NLoS detection method. As we have labeled all NLoS/LoS conditions by checking videos manually, that is, respective 16,425, 10,033, and 27,439 NLoS/LoS samples under highway, suburban, and urban environments, we utilize one part of the samples for model training and the other part for model verification. From Fig. 4, it can be seen that the PDR values have latent relation to the LoS/NLoS condition, since all visually identified NLoS conditions tend to have a lower PDR, and all LoS conditions tend to have a higher PDR. As explained in [10], the PDR values have memories, and therefore we employ historical values of PDR as input features to train a classification model to detect NLoS condition online. Specifically, we choose three historical PDR values (e.g., previous 1 s, 5 s, and 10 s) as the feature, and the labeled LoS/

[1] The values are empirically set based on the average PDRs under LoS and NLoS conditions.

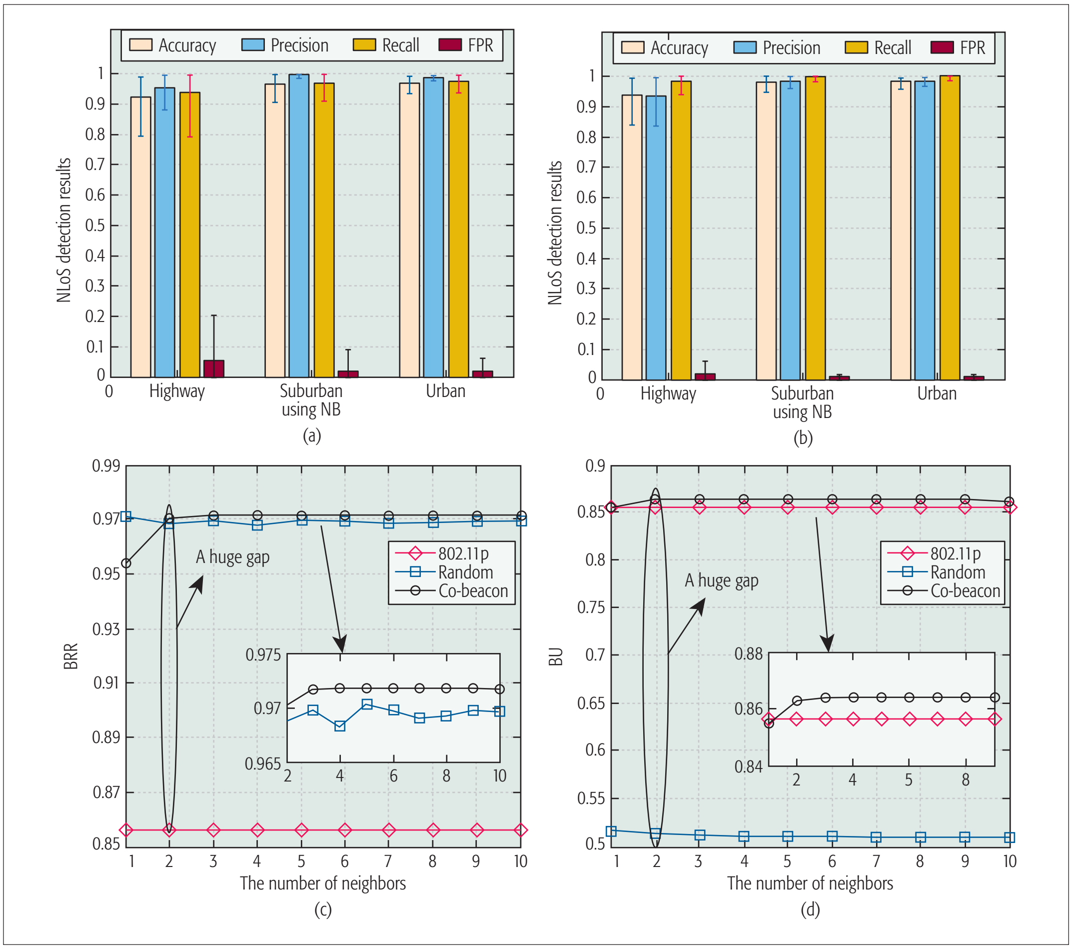



FIGURE 6. Simulation results: a) NLoS detection results with NB; b) NLoS detection results with SVM; c) BRR vs. the number of neighbors; d) BU vs. the number of neighbors.

It can be seen that the simple methods perform well with accuracy, precision, and recall higher than 90 percent, and false positive rate lower than 10 percent. Using the trained model, the vehicles can accurately detect the NLoS condition in real time by tracking the historical PDR values.

NLoS conditions are used as the training target. Simple but effective machine learning methods such as *Naive bayes* (NB) and *support vector machine* (SVM) are employed, and the results are shown in Figs. 6a and 6b. It can be seen that the simple methods perform well with accuracy, precision, and recall higher than 90 percent, and false positive rate lower than 10 percent. Using the trained model, the vehicles can accurately detect the NLoS condition in real time by tracking the historical PDR values.

### Link Condition Exchange

In Co-beacon, in each beacon, in addition to application data, each vehicle also includes the information of its link conditions (LoS/NLoS) between itself and its one-hop neighbors. Specifically, for each vehicle, it keeps updating all link conditions with its one-hop neighbors, and then broadcasts all the link condition information with the application data together to its one-hop neighbors. By receiving beacons, each vehicle can not only perceive the link status information between itself and its one-hop neighbors, but also achieve the link status information between its one-hop neighbors and the neighbors of its neighbors, that is, perceiving the link status information within two hops.

### Beaconing with Helpers

Upon identifying NLoS conditions, a sender should seek helper vehicles from its neighbors to forward beacons. Such a helper would be selected if it has an LoS condition with both the sender and the receiver. For instance, when sender $x$ detects an NLoS link condition with receiver $z$, Co-beacon will select a helper $y$ from the neighbors $N(x)$ of $x$, when the receiver $z \in N(y)$ and the two links $x \rightarrow y$ and $y \rightarrow z$ are both in LoS conditions. If no such helper $y$ exists, Co-beacon will randomly choose a helper $y$ from the set $N(x)$. After helper $y$ is selected, the ID of $y$ will be added in the header of the beacon; when vehicle $y$ receives the beacon, it will rebroadcast the beacon immediately.

### Performance Evaluation

**Simulation Setup:** To investigate the impact of the number of neighbors on helper choosing, we first synthesize 100 links of data using our previous work of synthesizing V2V communication traces [10], in which each trace lasts for 100 minutes. In each simulation round, we first randomly choose two links of data mimicking the sending/receiving process from the sender to the receiver. To add a neighbor, another two pairs of links are randomly chosen. One pair is for mimicking the sending/receiving from the sender to the neighbor, and the other pair is for the sending/receiving from the neighbor to the receiver. The number of neighbors *n* ranges from 1 to 10, and the simulation is conducted over each value of *n*. All simulation results are achieved on average by computing the result in 30 simulation rounds.

**Performance Metrics:** We consider the following metrics to evaluate the performance of Co-beacon.

***Beacon Reception Ratio (BRR):*** is defined to measure the reliability, which refers to the ratio of the number of neighbors having received a beacon to the total number of one-hop neighbors. For example, consider that there are 10 neighboring vehicles in the radio range of a vehicle: if 8 of them receive the beacon broadcast by the vehicle, the BRR is 0.8.

***Broadcast Utility (BU):*** is defined to measure the communication cost, which refers to the ratio of the BRR to the total number of broadcasts of a beacon. As in the above example, if the BRR of 0.8 is achieved using no helper (i.e., the beacon is only broadcast once by the sender vehicle), the BU is 0.8. The BU becomes 0.4 if one helper is used (the beacon is broadcast twice, that is, by the sender and by the helper).

We compare Co-beacon with two candidate schemes, that is, 802.11p (without any rebroadcast) and random forwarding (always randomly select one of its neighbors to rebroadcast its beacons).

**Performance Comparison:** Figure 6c shows the average BRRs of the three beacon strategies with different numbers of neighbors, from which the following two main observations can be obtained. First, with a helper to retransmit packets, the reliability of the link can be greatly enhanced. For instance, Co-beacon and the random approach always surpass 802.11p and can increase the BRR from 85 to about 97.1 and 96.8 percent, respectively, when there are two neighbors near the sender and receiver. Second, with more neighbors in the environment, Co-beacon can achieve better performance, while 802.11p and random approach cannot react well to the environment. Specifically, the values of BRR will not change with 802.11p due to the lack of a retransmission scheme, the values fluctuate within a small range with random approach, and the values increase gradually with Co-beacon. Moreover, the value in Co-beacon can reach higher than 97 percent, outperforming about 96.8 percent in the random approach when there are more than two neighbors.

Figure 6d shows the average BUs of three beacon strategies with different numbers of neighbors. It can be seen that the random approach can achieve the lowest packet utility of below 52 percent compared to the other two strategies of more than 85 percent. Considering a large number of vehicles in the network, the low BU of each vehicle in the random approach would bring a huge burden to the network. Another key insight is that with more neighbors in the environment, Co-beacon can achieve better BU performance. Specifically, the BU values in Co-beacon gradually increase as the number of neighbors

Considering a large number of vehicles in the network, the low broadcast utility of each vehicle in the random approach would bring a huge burden to the network. Another key insight is that, with more neighbors in the environment, Co-beacon can achieve better BU performance.

increases and can reach up to 86.2 percent when there are more than two neighbors in the environment, which surpasses the value of 85.6 percent in 802.11p.

In summary, compared to the other two strategies, Co-beacon can significantly enhance the BRR without obviously degrading the BU, especially when there are more than two neighbors in the network.

## Conclusion

In this article, we have proposed a context-aware communication paradigm design for IoV by employing big data analytics and machine learning methods. We have analyzed the V2V communication performance in urban environments based on real-world data traces, and found that an NLoS condition could be the major factor that significantly affects link performance. Three important V2V communication paradigms have been discussed, that is, smart medium resource allocation, efficient routing establishment, and reliable safety message broadcasting. Moreover, a case study has been provided to demonstrate the effectiveness of the machine-learning-based NLoS detection scheme and the context-aware beaconing scheme.

### Acknowledgments

This research was supported in part by the National Natural Science Foundation of China (Grants No. 61420106010, 61373157, 61472255, 61471236, 91638204, 61772340) and the Natural Sciences and Engineering Research Council of Canada.

### References

[1] N. Cheng *et al.*, "Big Data Driven Vehicular Networks," *IEEE Network*, Early Access; https://ieeexplore.ieee.org/document/8450539. DOI: 10.1109/MNET.2018.1700460.
[2] M. A. Salahuddin, A. Al-Fuqaha, and M. Guizani, "Software-Defined Networking for RSU Clouds in Support of the Internet of Vehicles," *IEEE Internet of Things J.*, vol. 2, no. 2, Apr. 2015, pp. 133–44.
[3] W. Quan *et al.*, "Enhancing Crowd Collaborations for Software Defined Vehicular Networks," *IEEE Commun. Mag.*, vol. 55, no. 8, Aug. 2017, pp. 80–86.
[4] H. Zhou *et al.*, "TV White Space Enabled Connected Vehicle Networks: Challenges and Solutions," *IEEE Network*, vol. 31, no. 3, May/June 2017, pp. 6–13.
[5] F. Bai, D. D. Stancil, and H. Krishnan, "Toward Understanding Characteristics of Dedicated Short Range Communications (DSRC) from a Perspective of Vehicular Network Engineers," *Proc. ACM MobiCom*, 2010, pp. 329–40.
[6] B. Di *et al.*, "V2X Meets NOMA: Non-Orthogonal Multiple Access for 5G-Enabled Vehicular Networks," *IEEE Wireless Commun.*, vol. 24, no. 6, Dec. 2017, pp. 14–21.
[7] P. Shankar *et al.*, "Cars: Context-Aware Rate Selection for Vehicular Networks," *Proc. IEEE ICNP*, 2008, pp. 1–12.
[8] F. J. Ros, P. M. Ruiz, and I. Stojmenovic, "Acknowledgment-Based Broadcast Protocol for Reliable and Efficient Data Dissemination in Vehicular Ad Hoc Networks," *IEEE Trans. Mobile Comp.*, vol. 11, no. 1, Jan. 2012, pp. 33–46.
[9] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5G: How to Empower SON with Big Data for Enabling 5G," *IEEE Network*, vol. 28, no. 6, 2014, pp. 27–33.
[10] F. Lv *et al.*, "Synthesizing Vehicle-to- Vehicle Communication Trace for VANET Research," *Proc. IEEE SMARTCOMP*, 2017, pp. 1–3.

[11] C. Han *et al.*, "Analytical Study of the IEEE 802.11p MAC Sublayer in Vehicular Networks," *IEEE Trans. Intell. Transp. Sys.*, vol. 13, no. 2, June 2012, pp. 873–86.
[12] H. A. Omar, W. Zhuang, and L. Li, "VeMAC: A TDMA-Based MAC Protocol for Reliable Broadcast in VANETs," *IEEE Trans. Mobile Comp.*, vol. 12, no. 9, June 2013, pp. 1724–36.
[13] L. Zhang and S. Valaee, "Congestion Control for Vehicular Networks With Safety-Awareness," *IEEE/ACM Trans. Networking*, vol. 24, no. 6, Dec. 2016, pp. 3290–99.
[14] J. Liu *et al.*, "Space-Air- Ground Integrated Network: A Survey," to appear, *IEEE Commun. Surveys &Tutorials*.
[15] N. Zhang *et al.*, "Software Defined Space-Air-Ground Integrated Vehicular Networks: Challenges and Solutions," *IEEE Commun. Mag.*, vol. 55, no. 7, July 2017, pp. 101–09.

## Biographies

Feng Lyu [M'18] received his Ph.D degree from the Department of Computer Science and Engineering, Shanghai Jiao Tong University (SJTU), China, in 2018, and his B.S. degree in software engineering from Central South University, Changsha, China, in 2013. His research interests include vehicular ad hoc networks, cloud/edge computing, and big-data-driven application design.

Nan Cheng [S'12, M'16] received his Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, Ontario, Canada. He is currently working as a postdoctoral fellow with the Department of Electrical and Computer Engineering, University of Toronto, and the Department of Electrical and Computer Engineering, University of Waterloo. His research interests include performance analysis and opportunistic communications for vehicular networks, unmanned aerial vehicles, and cellular traffic offloading.

Hongzi Zhu received his Ph.D. degree from SJTU in 2009, and his M.Eng. and B.S. degrees from Jilin University in 2001 and 2004, respectively. He was a postdoctoral fellow at the University of Waterloo and Hong Kong University of Science and Technology in 2009 and 2010, respectively. He is now an associate professor at the Department of Computer Science and Engineering, SJTU. His research interests include wireless networks, mobile sensing and mobile computing, and vehicular ad hoc networks. For more information, please visit http://lion.sjtu.edu.cn.

Haibo Zhou [M'14, SM'18] received his Ph.D. degree in information and communication engineering from SJTU in 2014. From 2014 to 2017, he worked as a postdoctoral fellow with the Broadband Communications Research Group, ECE Department, University of Waterloo. Currently, he is an associate professor with the School of Electronic Science and Engineering, Nanjing University. His research interests include resource management and protocol design in cognitive radio networks and vehicular networks.

Wenchao Xu received his B.E. and M.E. degrees from Zhejiang University, Hangzhou, China, in 2008 and 2011, respectively. He is currently working toward a Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo. In 2011, he joined Alcatel Lucent Shanghai Bell Co. Ltd., where he was a software engineer for telecom virtualization. His interests include wireless communications with emphasis on resource allocation, network modeling, and mobile data offloading.

Minglu Li [M'08, SM'10] received his Ph.D. degree in computer software from SJTU in 1996. He is a full professor and the director of the Network Computing Center at SJTU. His research interests include grid computing, services computing, and sensor networks.

Xuemin (Sherman) Shen [F] is a university professor, Department of Electrical and Computer Engineering, University of Waterloo. His research focuses on resource management, wireless network security, social networks, smart grid, and vehicular ad hoc networks. He is an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, and a Royal Society of Canada Fellow. He was a Distinguished Lecturer of the IEEE Vehicular Technology Society and the IEEE Communications Society.