Internet of Vehicles in Big Data Era

Wenchao Xu, Haibo Zhou, *Member, IEEE*, Nan Cheng, *Member, IEEE*, Feng Lyu, Weisen Shi, Jiayin Chen, Xuemin (Sherman) Shen, *Fellow, IEEE*

Abstract—As the rapid development of automotive telematics, modern vehicles are expected to be connected through heterogeneous radio access technologies and are able to exchange massive information with their surrounding environment. By significantly expanding the network scale and conducting both real-time and long-term information processing, the traditional Vehicular Ad-Hoc Networks (VANETs) are evolving to the Internet of Vehicles (IoV), which promises efficient and intelligent prospect for the future transportation system. On the other hand, vehicles are not only consuming but also generating a huge amount and enormous types of data, which is referred to as Big Data. In this article, we first investigate the relationship between IoV and big data in vehicular environment, mainly on how IoV supports the transmission, storage, computing of the big data, and how IoV benefits from big data in terms of IoV characterization, performance evaluation and big data assisted communication protocol design. We then investigate the application of IoV big data in autonomous vehicles. Finally, the emerging issues of the big data enabled IoV are discussed.

Index Terms—Autonomous vehicles, big data, big data applications, data communication, IoV, vehicular networks.

I. INTRODUCTION

THE number of the connected vehicles is increasing rapidly, and it is predicted that one fifth vehicles on road will have Internet connection and the global vehicular traffic is expected to reach $300\,000$ Exabyte by the year of 2020 [1]. Such kind of vehicular data explosion is mainly due to the development of vehicular telematics along with desired vehicular applications, such as for driving safety, in-vehicle infotainment, and Intelligent Transportation System (ITS).

Traditional VANETs consider the vehicle as a network node to transmit or relay the data traffic among vehicles and infrastructures through the Vehicle-to-Vehicle (V2V) communication and Vehicle-to-Infrastructure (V2I) communication [2]. Various vehicular applications are taking advantage of

the expanding connectivity of modern vehicles, such as safety application, ITS services, mobile Internet access, etc. For example, the V2V communication can enable vehicles to share information directly with their neighbors for safety message dissemination, while the V2I communication can be used to collect the information from infrastructure facilities [3]. Significant industrial and research efforts have been made to satisfy the diverse requirement for these vehicular applications. Cellular networks are naturally chosen to provide wide coverage for vehicular users. In addition, small cells which are densely deployed can provide economic network access within limited coverage areas to reduce the communication cost and enhance the vehicle connectivity by adding alternative data pipes [4], [5]. In North America, the IEEE 802.11p-based Dedicated Short Range Communication (DSRC) On Board Unit (OBU) are mandatory to be installed on shipped vehicles to enable the vehicular communications [6]. However, as more and more sensors are equipped on modern vehicles, massive data are generated from monitoring the on-road and in-board status. This exponential growth of generated vehicular data, together with the increasing data demands from in-vehicle users, has led to a tremendous amount of data in VANETs [7]. Furthermore, the data types also significantly increase from various sources or terminals, which are often across large geographical areas requiring critical Quality of Service (QoS), e.g., restricted service delay, extremely high delivery rate, massive connections, etc. Such new characteristics of the data traffic in vehicular conditions have advanced the VANETs to the Big Data era, showing that both the data type and data amount are greatly expanded, and thus pose challenges in traditional VANETs [8]. On the other hand, vehicles now are equipped with powerful processing unit and large storage devices, and thus are not only a moving network node, but also a computer and storage center with intelligent process capability, which evolve the traditional VANETs to IoV, by connecting large-scale intelligent vehicles with advanced telematics [9].

The relationship between the big data and IoV is two-folders as shown in Fig. 1. First, IoV is required to support the big data acquisition, transmission, storage and computing for data owners and users. Secondly, IoV can also benefit from mining the internal value of big data, which include large scope information from internal and external environment of vehicles. Such reciprocal relationship between IoV and big data has not been well explored in the state-of-art works. In order to show how the IoV can support the big data handling in vehicular conditions, and how IoV can benefit from big data technologies, in this article, we investigate the IoV in big data era. First, the big data acquisition is introduced, including the on-board sensing, passengers collecting and spatial extensions, e.g., flying vehicles. Then the big data transmission is investi-
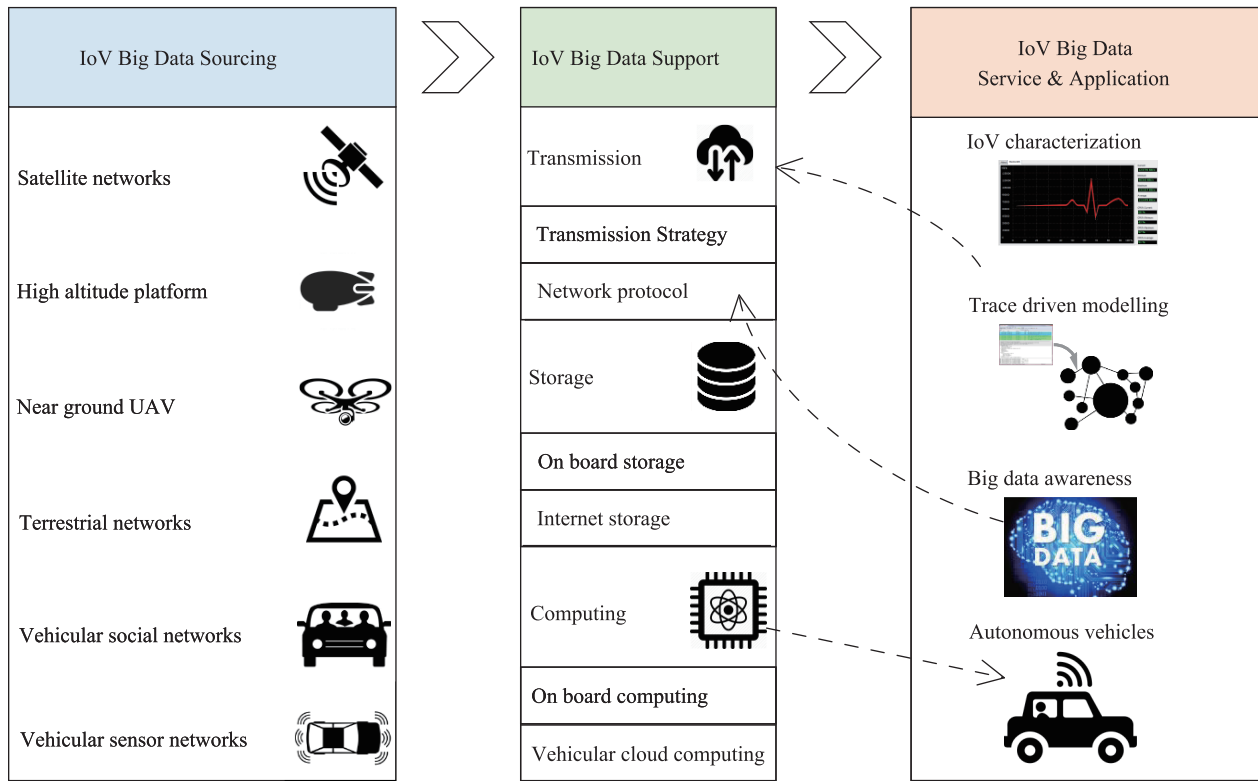
Fig. 1.   IoV big data architecture.

gated from big data aware protocol design and application aware strategy development. The storage and computing mechanism of IoV big data is further discussed, which is an essential part for future IoV big data applications. Furthermore, we also investigate the big data enabled IoV, whereby many IoV-related issues can benefit from big data, such as network characterization, performance evaluation, and intelligent protocol design. We also show the importance of IoV big data to the revolutionary autonomous vehicles, which is a crucial part of the future ITS. Finally, the emerging research issues on IoV in big data era are discussed about the necessity of big data aware IoV development and big data powered IoV services. An overview of the discussed content is shown in Fig. 2.

The remainder of this article is organized as follows. Section II investigates how IoV supports the big data acquisition, transmission, storage and computing. Section III demonstrates how the big data can be utilized for the IoV evolution including characterization, performance evaluation and protocol design. The autonomous vehicles based on big data enabled IoV are also discussed. Section IV provides the emerging research issues for big data enabled IoV, and Section V concludes the article.

II. Big Data Support in IoV

In this section, the big data support in IoV is described, including data acquisition, transmission, storage and computation.

A. Big Data Acquisition

The IoV data can be from vehicles, passengers, roadside facilities and Internet. The IoV data generated by vehicles

can be categorized into two parts: on-board data and on-road data. On-board data which monitors the vehicle status such as velocity, engine parameters, brake status, etc., can be obtained directly from in-vehicle sensors. On-road data refers to the event information happening on road, which can be from both on-board sensors (e.g., inter-vehicle distance, blind point objectives, pilot camera video, etc.) or from others via the IoV connections (e.g., traffic light status, neighboring vehicle brake notification, road map, etc.)

As shown in Table I, there are various types and bandwidth requirements of data from sensors on current vehicles. The big data generated by on-board sensors are not only of various types, but also of large amount. It is predicted that there will be more than 200 sensors on future vehicles in 2020 [7], and around 4000 GB data will be generated every day [10]. Compared to the traditional sensors such as Engine Control Unit or GPS, the Lidar and Camera generate much more data amount for applications such as autonomous driving, navigation, etc. Some of the sensor data can be acquired by other vehicles through the IoV connection to support various applications. Xu *et al.* adopted the DSRC channel to send the safety messages to other vehicles [27]. Peng *et al.* used the V2V communication to transmit the control message between platoons [11], [12]. Besides, the V2X technology can be utilized for the communication between vehicles and ITS infrastructure to predict the traffic condition and calculate the optimal navigation route [14]. Many Internet applications become popular when vehicles have the Internet access. For example, Tesla vehicles use the cellular connection to upgrade the automotive software through cellular and WiFi connections.
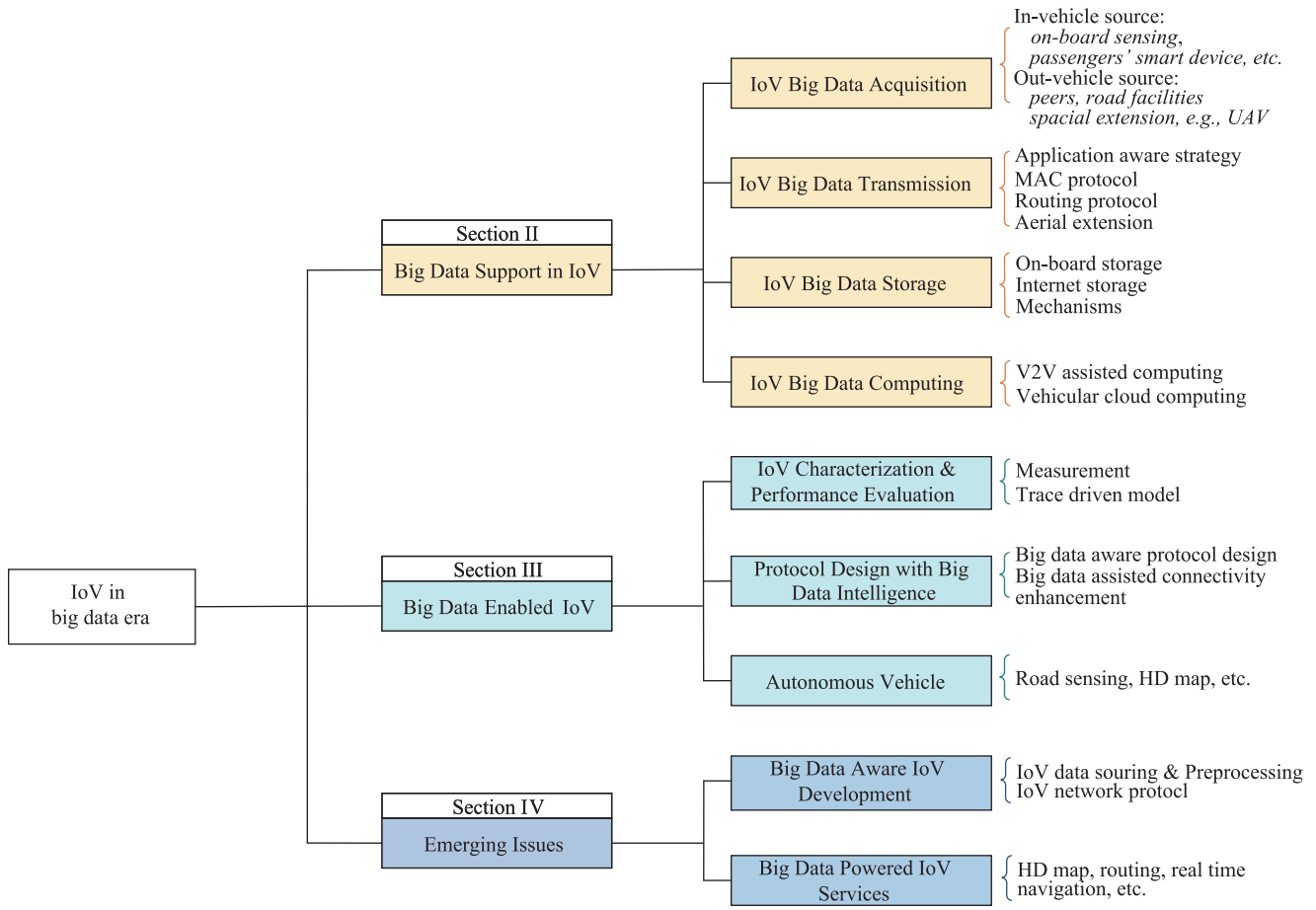
Fig. 2. IoV big data overview.

TABLE I
IoV SENSING & APPLICATION BANDWIDTH

On-board Sensing	Bandwidth	On-road Sensing	Bandwidth	IoV Application	Quality of Service
Vehicle position (GPS)	~ 50 KBps	Radar	$\sim 10-100$ KBps	Vehicle management	Ubiquitous connection
Driving state (Velocity, tire presser, etc.)	~ 100 KBps	Sonar	$\sim 10-100$ KBps	In-vehicle HD video streaming	Large bandwidth
ECU (coolant temperature, RPM, etc)	~ 100 KBps	Lidar	$\sim 10-70$ MBps	Autonomous driving	High delivery rate with delay constraint
Accidental event (braker, accident, collision, etc.)	Small	Camera	$\sim 20-40$ MBps	HD map	High data volumn

The smart devices of the passengers or the vehicle driver can also generate massive data. First, the sensors on the smart devices, e.g., GPS chip, gyroscope built in passengers' smart phone can provide a lot of sensing data such as vehicle trajectory data, accelerometer data, etc., which can be used for various vehicular applications. For example, Handel *et al.* used the smart phone inside the vehicles as the probes to monitor the road traffic [15]. Ghose *et al.* let the in-vehicle smart phone to be a connected sensor and report the road condition for peers on Internet [16]. Besides, the vehicular social networks also generate the social data belonging to the mobile vehicles and passengers [17].

Apart from acquiring data from on-board sensors, passengers and their smart devices, other vehicles and RSUs on

ground, the space and aerial platforms can also be the data source in IoV. Satellites, High Altitude Platforms (HAPs) and drones (i.e. unmanned aerial vehicles, UAVs) are three promising platforms, which have two main advantages to better acquire the data for IoV.

1) Enhanced Coverage: Comparing with ground data sources, the space and aerial platforms can help vehicles acquire data from wider geographic areas. The satellites based platforms, such as GPS and remote sensing system, have already proven its capability of collecting data in worldwide. Working at 20 km altitude with 20−30 km coverage radius [18], the HAP is leveraged for city-wide data collections and communication services provisions. It is proved that the drone flying hundred meters high can monitor or access more

ground devices in Line-of-Sight (LoS) way than vehicles on the ground [19]. Such an extended coverage capability can significantly help IoV to acquire data more efficiently both in global and regional areas.

2) Specific Data Acquisition: In IoV, some specific data can only be collected by dedicated space and aerial platforms. For instance, the positioning of each vehicle is highly depended on the GPS system enabled by satellites. And in rural areas lacking WiFi or cellular coverages, HAP is a feasible solution for data acquisition in IoV. Besides, equipped with cameras or specific sensors, drones can integrate with traffic surveillance system to monitor vehicle traffic dynamically. Based on the specific data collected by space and aerial platforms, the IoV controller can have a comprehensive view of the network and make accurate decisions.

As the development of heterogeneous radio technologies and aerial extension of communication platform, the data acquirement in IoV is becoming more and more complex but ubiquitous, which helps the IoV to collect even bigger data featured in various data types and huge data volume.

B. Big Data Transmission in IoV

The transmission of big data in IoV is challenging due to various data types and large data volume as well as the strict QoS, as shown in Table I. In this section, we present the support of big data transmission in IoV, as shown in Fig. 3. First, the transmission requirements of different IoV applications are introduced as well as the challenges for current VANET. Then, the transmission strategies fitted for different applications are classified and summarized. After that both the MAC and routing protocols for current IoV are investigated.
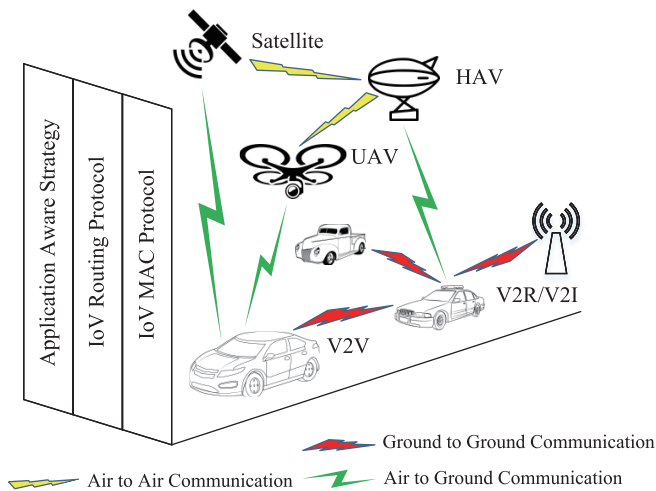

Fig. 3. IoV big data transmission support.

1) Applications and Transmission Requirements: Two major categories of applications are extensively researched in IoV, namely, road safety and travel comfort. Road safety applications such as collision avoidance, emergency movement warning, are supported by keeping up-to-date knowledge of neighboring vehicles [20], [21]. Each vehicle needs to periodically report its status information, including position,

velocity, heading direction, acceleration and turn signal status, to all surrounding vehicles. Due to the small size of safety messages, normally 200−500 bytes [22], these applications require low data rate. However, to keep accurate knowledge of neighboring vehicles in high mobility condition, strict latency and reliability constraints should be guaranteed. The travel comfort applications, however, are of relaxed delay constraints and some are known as delay-tolerant applications [23]. They can help improve driving experience and road efficiency by enabling the delivery of road administration announcements and advertisements, such as traffic signal status, traffic condition of the road ahead, the information of available parking lots nearby, etc. Compared to safety message, such announcements or advertisements may need to be transmitted through a longer distance to inform more vehicular users on road. Qin *et al.* chose a small number of public vehicles such as taxis and buses as seed vehicles to propagate digital advertisements to other vehicles in a mobile advertising application. The advertisement information can gradually spread out within the network via the contacts between seed vehicles and others in a wide area [24].

2) Challenges of IoV Big Data Transmission: In vehicular scenario, there are some unique characteristics that need to be taken into account compared to normal mobile ad-hoc networks.

a) Harsh wireless channel conditions: In driving scenarios, environments such as tall buildings, tunnels and overhead bridges elevated roads and time-varying traffic conditions would have heavy and dynamic impact on V2V link performance. These obstacles and multi-path fading result in unstable wireless channel [25].

b) Spectrum resource shortage: In DSRC standard, one Control Channel (CCH) and multiple Service Channels (SCHs) with optional bandwidths of 10 MHz and 20 MHz are supported. However, the Federal Communications Commission (FCC) only allocates 75 MHz licensed spectrum for DSRC, which is insufficient to support the IoV transmission under high-density condition for media-rich applications [112], [113].

c) High mobility: Due to the high velocity of vehicles on road, the wireless connections between moving vehicles and limited-coverage roadside infrastructure will be interrupted frequently. And vehicles may come close or leave at any moment, which might cause random signal interferences to each other. Besides, the network topology varies dramatically with diversified mobilities caused by speed, acceleration, road layout, traffic lights, driving behaviors and traffic rules, which affect channel resource allocation, routing protocols design and message receptions. Furthermore, the Internet management may be requested frequently as mobile vehicles may switch the access infrastructure frequently and thus increases the Internet access overhead [26].

d) Dynamic vehicle density: The vehicle density can be very high in the case of a traffic jam in urban and highway, and very low in suburban or highways. The open and important issue is how protocols can efficiently adapt to the variable vehicle densities such that channel resource would not be wasted in low vehicle density and channel congestion can be alleviated

in high vehicle density.

e) *Absence of global coordination:* It is difficult to deploy central coordinator for IoV since IoV comprises of heterogeneous access network and expands to wide areas. The transmission protocols in IoV need to consider such constraint and work in a distributed way.

3) IoV Big Data Transmission Methodology: To satisfy the big data transmission requirements and deal with the challenges, the transmission mechanisms from application layer strategies to MAC layer protocols are reviewed.

Application-aware transmission strategies: To convey the target messages among the big data in IoV, the transmission strategies should be designed carefully to improve the data transmission efficiently. Current research about the target message dissemination strategy can be categorized into two models, i.e., *push model* and *pull model*. In push model, data is periodically broadcasted regardless of users' requests, while in pull model, data is transmitted upon users' requests. Safety applications should use push model to synchronize up-to-date knowledge of neighbors, while delay-tolerant applications usually adopt pull model to save the communication resources. In real implementations, the two models are often combined together, namely, hybrid model, which supports multiple applications simultaneously [28].

a) *Push-based strategies:* The push-based strategies choose to broadcast the target messages, e.g., safety messages, to reach as many as possible destinations. Plain flooding is the simplest way of broadcasting, where each vehicle broadcasts data packets to all its one-hop neighbors. And all receiving neighbors will then broadcast the message again and repeat the procedure in order that multi-hop neighbors can also receive the target message. However, the plain flooding can easily lead to broadcast storm that many vehicles within the communication range broadcast simultaneously and then a considerable number of packets will be dropped due to transmission collision. Besides, it is common that some packets may be excessively broadcasted many times, which results in the waste of the channel resources in IoV [29]. Several strategies are designed to specifically reduce such kind of collision and reduce packet transmission overhead [30]. For example, the weighted *p*-persistence strategy assigns vehicles with different rebroadcast probability. It allows vehicles away from the sender to rebroadcast with higher probability. In addition, vehicles only rebroadcast the packets that are received for the first time and will drop repetitive ones. Another strategy called slotted 1-persistence scheme assigns vehicles with different waiting delays where vehicles with shorter transmission delay should rebroadcast first, which can suppress the broadcast storm problem by deescalating redundant rebroadcasts. The slotted *p*-persistence strategy combines the above probability-based and delay-based schemes together and assigns the highest priority receiving vehicle with the shortest delay to rebroadcast the target messages to improve the efficiency for push-based strategies.

b) *Pull-based strategies:* Pull-based strategies follow the request-response mechanism to disseminate target messages [23]. Compared with push-based strategies, pull-based strategies require less overhead, but extra latency may be introduced due to the request-response negotiation prior to transmission. Pull-based strategies are mainly applied for the IoV data transmission between vehicles and backbone connected entities, e.g., remote Internet servers or peers. To reduce the delay of pull-based strategies, data caching is often used to improve the data dissemination efficiency by storing the required data at infrastructures close to the vehicular users [31]−[33]. We will discuss related caching scheme in Section II-C.

MAC layer designs: It is critical for vehicles to access the transmission media efficiently for big data transmission to satisfy the QoS of IoV applications. To deal with the challenges such as high mobility, dynamic topology and varying vehicle densities, wide range of MAC protocols have been proposed to reduce the access delay and transmission collisions, and improve the throughput, robustness, scalability, etc. The MAC layer protocols can be divided into two categories, i.e., contention-based and contention-free protocols.

a) *Contention-based MAC protocols:* IEEE 802.11p [6] standard as an amendment to the existing IEEE 802.11a-2007 or Wi-Fi [34] standard has been dedicated by the Federal Communications Commission (FCC) as the MAC layer for V2V and V2I communications in United States. The basic MAC method of IEEE 802.11p adopts the Distributed Coordination Function (DCF) of IEEE 802.11 standard, which is based on CSMA/CA mechanism. In CSMA/CA scheme, when a node wants to access the channel, it has to sense the channel status first. If the channel is idle, the node is allowed to access; otherwise, the node has to perform random back-off procedure. Such kind of contention-based MAC may result in unbounded delays, which does not satisfy the real-time requirement of safety applications in IoV. Many researchers conducted studies on the delay and reliability evaluation of 802.11p beaconing and showed that the 802.11p MAC has serious issues with unbounded delay and channel congestion under high vehicle density conditions [35]. Another issue is that to facilitate real-time response, in broadcast mode of 802.11p protocol, RTS/CTS/ACK schemes are not employed, which may cause hidden terminal problem and transmission tracing difficulties.

b) *Contention-free MAC protocols:* Unlike contention-based MAC protocols, contention-free MAC protocols require a channel access scheduling prior to transmission. Several contention-free MAC protocols have been proposed for big data transmission in IoV, i.e., Time Division Multiple Access (TDMA), Space Division Multiple Access (SDMA) and Code Division Multiple Access (CDMA).

In TDMA-based MAC protocols, time is partitioned into frames consisting of a constant number of equal-length time slots and synchronized among vehicle nodes. Each vehicle node is guaranteed to access the channel at least once in each frame [36]. Moreover, each vehicle not only can transmit its message but also can report the status of time slots used by its neighbors. The slotted channel can guarantee the stringent time requirement of safety-related applications, and extra information exchange can be used to mitigate the hidden terminal problem. In ADHOC MAC [37], a slotted/framed structure is applied and coordinated with the well-known Reservation ALOHA (R-ALOHA) protocol [39]. ADHOC MAC can detect transmission collision by checking the time slot reservation

status from neighbors' broadcasted messages. To support efficient one-hop and multi-hop broadcast services with reduced access collisions and merging collisions, VeMAC [38] assigns disjoint sets of time slots to vehicles moving in opposite directions and to Road Side Units (RSUs). Other TDMA-based MACs [40]−[42] are also proposed to solve problems in different scenarios. The benefits of utilizing TDMA-based MACs are considerable: First, short delivery delay for real-time safety applications can be achieved and transmission collisions can be largely avoided for reliable communication. Besides, all vehicles are guaranteed to access the channel with equal chances. Furthermore, the channel resource can be utilized efficiently due to less packet collision. However, there are still some issues to be solved. For example, when vehicles move, merging collisions can happen randomly and are difficult to avoid.

In SDMA-based MAC protocols [43], [44], each vehicle acquires channel resources according to its own location. Such MAC schemes usually contain three steps: 1) discrete process: dividing the road into several smaller road units; 2) mapping function: assigning unique time slot sets to each road unit; 3) rules of assignment: based on the road unit where the node resides, determining which time slots are available for the node to use. SDMA-based MACs require the vehicle's location information, while the inaccuracy of such information may result in frequent transmission collisions. Moreover, when the traffic density is low or uneven, the bandwidth resources are likely wasted; when the traffic density is high, it is then difficult to ensure the fair access for all the vehicles within short duration.

In CDMA-based MAC protocols [45], [46], channel resources are distinguished by different pseudo random code. The receivers use the same Pseudo Noise (PN) code to demodulate the received signals. As this mechanism is resistant to interference and noise, and also can improve the bandwidth utilization, many CDMA based MAC protocols are designed for IoV message dissemination. However, in IoV where the vehicle density is high, the different PN code for each vehicle can be very long, which reduces communication resource utilization. Moreover, to demodulate packets, receivers must save the PN code for every sender. This is not realistic since it is difficult for vehicles to save all PN codes of neighbors within the communication range.

Routing protocols: Routing protocol is important for efficiently relaying the data from source to destination. Due to the dynamics nature of IoV, it is challenging to make the optimal routing strategy for big data transmissions. The IoV routing protocols can be classified into two categories, i.e., topology-based and position-based routing protocols [47], [48].

a) *Topology-based routing protocols:* The network topology information is utilized to select the relay nodes. Source node uses link's status which is listed in the routing table to calculate the optimal routing path. It is difficult to maintain and update the routing table in vehicular conditions due to the highly dynamic network topology. Hence, proactive, reactive or hybrid approaches are proposed in literatures about the routing protocols in IoV.

Proactive routing protocols: In proactive schemes [49], [50],

each node uses the routing table to store the information for all other nodes and in the table, each entry contains the next hop node used in the path to the destination. The routing information is stored regardless of whether the information is actually in need or not. To reflect the network topology changes, the table must be updated frequently and periodically to the neighboring vehicles. In this routing protocols, the optimal route will be chosen from available paths according to algorithms to guarantee shortest path or least transmission delay. In high density scenarios, such scheme will cause significant overhead to synchronize the network information and route calculation.

Reactive routing protocols: Reactive routing protocols [51], [52] are also referred to as on-demand protocols, which is designed to reduce the routing overhead, where the routing information is maintained only when the source node starts a route discovery process. If there is no existing route to the destination, the source floods the route request messages to the whole network. The network nodes that can receive the route request message will send a route reply message back to the source node to indicate the routing path between the source and the node. Such protocols are applicable to the high density and frequent topology changes networks since they require less synchronization overhead.

Hybrid routing protocols. Hybrid protocol [53] is a combination of proactive and reactive protocols. It aims to minimize the control overhead of proactive routing protocol and reduce the delay of the route discovery process in reactive routing protocols. Each node divides the network into two regions: inside region and outside region. Proactive routing scheme is used to maintain routes to nodes within inside region while reactive route discovery mechanism is used to for nodes in outside region. Under this scheme, the network can be divided into many zones for reliable route discovery and maintenance process.

b) *Position-based routing protocols:* Position-based protocols utilize the vehicle position information to make the route decision. Under vehicular condition, the mobility of a vehicle is usually restricted in just bidirectional movements along roads and streets. Together with the geographical location information obtained from street maps, traffic models and some prevalent navigational systems, position-based protocols can reduce the route overhead, especially in high mobility conditions [54], [55]. In position-based protocols, the node identifies its neighbors by checking if they are located within the communication range. When the source node sends a packet, the position information is added in the header of the packet, and then the packet is forwarded to the nodes who are closer to the destination by calculating the positions of peers. Following such process, there is no need for route discovery, maintenance and even awareness of the network topology. However, the intermediate nodes are often required to forward packets, which is hard to find under sparse condition. The position-based routing protocols may cause unbounded delay due to the unknown topology and lack of connectivity to all vehicles, which may be applied in scenarios with tolerant service restriction, e.g., carry-and-forward scheme, data mule, etc.

Aerial extension: Space and aerial platforms are not only an important supplement for big data acquisition methods for IoV, but also can improve the network performance [56], [57]. The benefits of aerial extension for IoV are studied and summarized as follows.

a) Ubiquitous connectivity: The IoV connections supported by infrastructures such as RSUs, cellular base stations, may be unavailable in rural places. And in dense vehicle places, it is hard to satisfy the IoV requirements due to the heavy traffic demand and limited wireless resource. By leveraging the aerial platform, the IoV connectivity can be enhanced based on the additional communication choice. GPS and other successful Geo-location based applications have already proven satellites' capability to fulfill communication tasks in worldwide [58]. Previous works also show that HAP is an effective and economical approach to provide broadband network services for sparse vehicle scenarios [59]. Satellite can provide full coverage for vehicles including rural areas, while the HAP can provide provisional connection according to the dynamic IoV transmission demand. In dense vehicle scenarios, HAP and drone platforms can provide additional spectrum resources, which not only enhances the communication capacity, but also enables the dynamic resources provision to adapt to the spatial and temporal changes of IoV by the high flexibility of drones [60].

b) LoS A2G connections: Lyu *et al.* conducted empirical experiments and showed that the dominant reason for connection outage and throughput deterioration in IoV is the absence of the Non Line-of-Sight (NLoS) channel conditions between vehicles [25]. Since the one-hop data link of the V2V and V2R connections are often blocked, the aerial platforms, especially the HAP and drone, have great potential to provide the LoS A2G communication and thus enhance the coverage and connectivity [19].

c) Dynamic deployment: Traditional infrastructure construction and upgrade are impractical due to geographically limited and cost restriction. And since the facilities are fixed on ground, it is difficult to adapt to the dynamic data traffic demands of IoV, e.g., in peak hours, more data traffic task will burst and thus drain up the communication capacities. Alternative methods such as deploying small cells or redundancy capacity design may lead to waste of wireless resources due to inflexibility. Such problem can be solved by the flexible UAV with re-location ability, whereby the dynamic and unpredictable traffic requirements can be satisfied via the cost-effective on-demand deployments.

The applications of the space and aerial platforms in IoV have been investigated and researched extensively, mainly for satellites, HAPs and flying drones.

a) Satellite systems, such as GPS, are mainly used for collecting and providing position information of the vehicles in current IoV. Related works integrating IoV with satellite systems leverage various technologies to improve positioning accuracy [61]. Due to the ubiquitous coverage and high satellite communication cost, in most scenarios, the communication between vehicles and satellites is conducted in one-way method (i.e., vehicles receive downlink signal from satellites) with relatively small throughput. However, it is expected that satellites can provide contingent assistance in emergency cases.

b) HAPs can enhance the performance of IoV in two aspects: 1) Acting as network infrastructures for vehicles to access in rural scenarios; 2) Providing additional spectrum resources for IoV in dense scenarios. It is shown that HAPs are capable to support broadband communications [62]. In [63], the authors prove that the HAP working on 47/48 GHz bands can achieve 25 Mbps data rate. Currently the spectrum resources allocated to HAP start from 2 GHz to 48 GHz according to different agreements [18], further standardization works are expected for HAP communication integration with IoV.

c) Drone communication are integrated with IoV and focus on mainly two cases: 1) Monitoring vehicular traffics through drones embedded cameras or sensors; 2) Assisting IoV communication through drones equipped with dedicated communication modules. The research works on ITS have already considered to apply drones as the surveillance devices to monitor road traffic. In [64], Motlagh *et al.* introduced various drone assisted airborne traffic surveillance systems. With the maturity of flying control technology and commercialized drone products, using drones or drone swarm networks to enhance connectivity of ground networks attracts increasing number of researchers in recent years. In [65], Goddemeier *et al.* assigned different roles (e.g. sensing agent, relay node, articulation point and returnee) to drones to facilitate self-optimized air-to-ground connectivity. The role assignment and switching strategies is investigated based on current network status. Dhekne *et al.* conducted a field experiment to prove drone-cell's capability of assisting base stations to improve the signal strength in coverage holes [66]. In [67], Al-Hourani *et al.* proposed a closed-form expression of A2G path-loss model for drone communication which considers the probabilities of both LoS and NLoS A2G links in different scenarios. Based on the path-loss model in [67], Mozaffari *et al.* explored the optimal trajectories and locations for drones collecting IoT information based on clustering approach [68]. Some heuristic algorithms are also proposed to find the optimal deployment of drones to assist future RAN in 5G era [69]−[71]. An aerial-ground cooperative vehicular networking architecture is proposed by Zhou *et al.* with road tests validating its performance [72]. In this architecture, the aerial sub-network formed by drones aids the ground IoV through Air-to-Air (A2A) and Air-to-Ground (A2G) communications. Not only collects network and road information for vehicles, the aerial sub-network also performs as intermediate relay for V2V communications. In [73], Oubbati *et al.* proposed a drone assisted routing protocol (UAVR) for IoV. Drones monitor traffic density and connectivity status, and exchange collected information with vehicles through dedicated messages. Based on those information, UAVR can guide vehicles to find the best multi-hop routing path for V2V communications.

C. Big Data Storage in IoV

In this section, the IoV big data storage technologies are investigated from the aspects of storage types and storage mechanisms, respectively.

1) Storage Types: a) *On-board storage:* The OBU installed on modern vehicles usually has rich extensible interfaces, e.g., universal serial bus, by which the on board storage space can be largely extended. NVIDIA's self-driving learning data collection system adopts Solid-State Drive (SSD) as the external storage, which can be extended to several Terabyte (TB) [13]. Besides, the smart devices of the passengers or the drivers, e.g., smart phones, can also be used to store the vehicular data. For instance, nowadays smart phones normally have tens of GigaBytes storage spaces, which can be used for data collecting and caching.

b) *Roadside storage:* Roadside units are usually installed with on-board storage and extension interfaces for external storage. For example, the commercial RSU product from *Cohda Wireless* has 10 GigaBytes on-board storage, and also integrated with USB interfaces [76]. It is the same way for roadside Wi-Fi hotspot, which can also be installed with massive storage spaces. Besides, He *et al.* used the dropboxes deployed along the road to store the drive-by vehicles' contents and forward to other vehicles to disseminate the data contents [74].

c) *Internet storage:* As more vehicles are connected to the Internet, it is possible that vehicles can access to the storage devices remotely via Internet. For example, vehicles can access to most cloud storage service providers, such as onedrive, dropbox, etc., and upload provisional data to Internet storage to save local resources and download target contents on demand. Besides, the storage space from other vehicles and roadside facilities can also be accessed via Internet under authorization.

2) Storage Mechanism: According to the access delay of the data, the storage mechanisms mentioned above can be categorized as fast storage, medium storage and slow storage. Fast storage includes on-board storage devices and the storage of the in-vehicle passengers' smart devices, which can be access within a guaranteed delay. Delay sensitive applications should use the fast storage, such as vehicle safety, autonomous driving, etc. Medium storage includes the external storage connected via reliable connection. Vehicles can access to medium storage with a relative larger delay when Internet connection is available. For example, two vehicles can share each others' storage space when both of them are connected to cellular networks, which involve multi-hop communication and backhaul transmission delay. Slow storage includes the external storage via opportunistic connection. For example, vehicles can access to the storage space of neighboring vehicle or drive-by roadside facilities within the V2V or V2R communication range. Due to the mobility of the vehicle, the slow storage are accessed opportunistically and the delay might be unbounded.

On-board storage adopts the proven technologies such as computer file system or query platform for vehicular applications to access the data. The Apollo project from Baidu Inc. adopts an Industrial PC with 1 TB hard drive storage and supports various operating systems & file systems [75]. Cohda OBU uses Linux operating system and employed with the fourth extended file system (ext4) [76]. To store the vehicular data at external places such as other vehicles or roadside units in IoV, the storage devices of vehicles on roads or parking areas are utilized through the V2V assistance [77]. Zhao *et al.* proposed a V2V assisted content delivery scheme, where the storage memory of a forwarding vehicle caches the popular content temporally until it comes across another forwarder [78]. Zheng *et al.* adopted the Vehicular Cloud Computing (VCC) diagram for IoV and proposed a semi-Markov decision process to allocate the cloud resource, including storage space from vehicle equipments in the VCC domain, to handle the situation where the data task is beyond the local storage capacities and improve the utilization of overall vehicular resources [79]. For slow storage that relies on the opportunistic connection, the mobility of the vehicles are utilized to required contents from a vehicle coming close [80]. The roadside infrastructures can also be used as storage for vehicles. Mahmood *et al.* let the vehicle cache the poplar content at roadside edge node and capture the mobility of vehicles to deliver the content to target vehicles [81]. Ding *et al.* treated the storage space of the RSU as network cache to deliver the popular contents required by the nearby vehicles [82]. Other similar examples can be found in [83], [84].

D. Big Data Computing in IoV

The history of on-board computer for vehicles goes back to decades ago, when Volkswagen applied the computer based Electronic Fuel Injection (EFI) system on vehicles [85]. In 1970s, the Engine Control Unit (ECU) becomes a standard on most vehicles to control/manage the emission, fuel and engine. Before the connected vehicle era, the on-board computing devices mainly focus on processing sensor data for engine and vehicle controls. And for connected vehicles, they are required to support extra computation tasks such as in-vehicle infotainment, V2V/V2I communication protocols, etc. In big data era, vehicles are expected to have magnificent computing capability to process various types of big data. The NVIDIA DRIVE PX platform integrates the computation capabilities of deep learning, sensor fusion and real-time vision processing for autonomous driving and can be scaled to support 24 trillion deep learning operations in one second [135]. It is considered that the vehicles in IoV are the "networked computing centers" [143]. Besides, the roadside units normally adopt powerful multi-core CPU and massive storage running modern operating systems. For example, the cohda RSU MK5 has the cortex-A8 based ARMv7 Processor and runs Linux 3.10 kernel, which provides high flexibility for cooperative computing applications in IoV [76]. Furthermore, the enormous computing resources on ground facilities can also be utilized for big data computing through the IoV connections [141].

Vehicles can assist their neighbors with their computing resources and thus the tasks that beyond single vehicle's computing resource limitation can be fulfilled via the cooperation of neighboring vehicles through the V2V request and response [140]. Bauza *et. al* proposed a scheme which relies on the computing resources on multiple vehicles to calculate the road conditions via the cooperation among vehicles [142]. And in the routing protocol, the computing resources on each vehicle are utilized to calculate the next hop or routing directions for the big data relaying in IoV [47], [48]. Self-organized vehicle

formation control, such as vehicle platooning and clustering, relies on the computation of mobility status, e.g., velocity, acceleration, etc., of adjacent vehicles [86]. Such kind of formation control helps the IoV to better support the big data performance. Zhang *et al.* proposed a platoon aware data access scheme to improve the data availability and reduce access overhead [87]. The computing resource of vehicles, both on road and in parking lots, are the important part of the mobile computing cloud, which provide the computing-as-a-service for big data processing in IoV. Arif *et al.* utilized the computing resources on vehicles parking in the airport to build a data center, which provides the cloud computing services for customers in IoV [88]. Such kind of computing resources from static vehicles can also be found in shopping malls and other gathering areas [144]. Zhang *et al.* analyzed the computing resource within a roadway segment considering both the free flow and queueing up traffic types [89]. Zheng *et al.* adopted a semi Markov decision process model to optimize the computing resource allocation from the vehicle cloud to achieve the best income level in the long term by selling these resources to the IoV cloud customers [79]. Zhang *et al.* offloaded the computing tasks from smart phones to vehicle cloud and showed that the performance of the mobile applications can be improved via experiments [145]. As the development of the on-board computing ability and edge facilities, it is envisioned that more and more big data computing tasks will be assigned flexibly among moving and parking vehicles, roadside units and data centers to develop the value of big data for IoV benefits.

III. BIG DATA ENABLED IoV

Although the ever-growing big data are causing challenges to IoV, mining the big data can also bring great value to the development of IoV. In this section, how the big data can evolve the IoV are investigated, from aspects of IoV characterization, performance evaluation, and protocol design, respectively. The relation between connected and autonomous vehicles and big data is also discussed in this section.

A. IoV Characterization and Performance Evaluation

To better understand the characteristics and evaluate the performance of IoV, the big data like the ground measurements can be utilized to study the wireless transmission and generate precise models for theory verification.

1) Big Data for IoV Measurement: The big data in IoV can be used as a reflection of the IoV, which plays a vital role in IoV characterization that involves many influencing factors and are difficult to model. Some ground experiments have already shown that the measurement data can help to characterize IoV in urban, suburban, rural, open fields and freeways.

Bai *et al.* proposed a black box framework to analyze the impacts of various factors on key V2V communication characteristics [90]. Packet Delivery Ratio (PDR) is analyzed in [3], [25], [90] as the main indicator of the reliability of V2V communication. In addition, the correlation functions are also investigated in [90], such as temporal correlation, spatial correlation and symmetric correlation. In [25], the distribution of packet inter-reception (PIR) time and Packet Inter-Loss (PIL) time is studied. Based on the distribution, the relationship between channel conditions and packet loss is investigated and showed that LoS and NLoS channel conditions have significant impacts on V2V communication. In [91], Karedal *et al.* focused on application-level reliability, where the T-window reliability was proposed to study the measurement data to evaluate the network capability of successfully transmitting at least once during the tolerance time window T, where parameter T is application related.

Besides the reliability performance, the V2V communication in IoV also pays attention on the network performance like throughput, connection duration and time overhead of connection establishment, etc. [92]. Cheng *et al.* used the measurement data to evaluate the offloading capability of different strategies reduce the cellular load [4]. The measurement data can also be used to analyze the channel characteristics in IoV, such as path loss, Doppler spectrum and coherence time, etc. Cheng *et al.* applied the positioning system to enable large-scale measurement and obtain the big data to model such kind of channel characteristics [93]. The discussion in [94], [95] shows that the packet size and modulation scheme can affect the retransmission possibility and communication range. In addition, channel measurement results prove that the 802.11p standard is well designed to account for Doppler spreads and delay spreads in vehicular channels [95].

2) Big Data Enabled Trace-Driven Models: In communication protocol research and design, simulation is an important verification method to save time and reduce economic cost from performing real experiments. A realistic IoV simulation environment is essential to obtain reliable results. The IoV big data can be used to provide accurate vehicular mobility model and V2X channel model for research and development of IoV protocol design.

For example, the GPS data which reflects the movement of vehicles can be traced to simulate the mobility model in IoV, which determines the topology of vehicular networks [96]. In [97], [114]−[116], the trace data are collected from vehicles in California, Shanghai or San Francisco to build realistic mobility models. Akhtar *et al.* integrated the trace data to the microscopic mobility model generated by Simulation of Urban Mobility (SUMO) and showed that the accuracy of the mobility model can be improved [97]. Zhu *et al.* utilized trace data and formulate the inter-contact time of any two vehicles to follow the tail distribution in the mobility model [114]. Celes *et al.* exploited the vehicle trajectory data and by predicting the route the gap in the trace data can be filled [116].

As a prerequisite for communication protocol development, channel models have been adopted in IoV analysis, such as log-normal shadowing model [97] and geometry-based stochastic model [94]. The parameters of the channel models in different scenarios can be adjusted by comparing the big trace data with the theoretical output, such as urban areas [99], [101], suburban areas [93], [98], rural areas and highway [100]. Besides, the big data trace can also be used to study the influence of antenna pattern variations for passing vehicles

[100]. A Speed-separation diagram is introduced in [93] to predict coherence time and channel Doppler spread in V2V propagation environment. Paier *et al.* showed that the big trace data under a typical traffic situation demonstrated that a constant delay and zero Doppler spectrum can be found [101].

B. IoV Protocol Design with Big Data Intelligence

With the knowledge of the IoV characteristics and performance from the big data based on measurement and field data collection, communication protocols for IoV can be designed more efficiently and accurately.

1) Big Data Aware Protocol Design: As stated above, the big data can provide convincing guidance for the network topology modeling and channel modeling, which is important for designing and testing efficient communication protocols. Besides, with the big data assistance, such as providing position information, the MAC and route protocols can better adapt to the high mobility and frequently changing topology of IoV [102]. Hartenstein *et al.* compared the position-based and topology-based routing methods, and showed that position-based routing performs better in terms of adaptivity and scalability. Besides, the big data can be used to predict the tendency of the IoV, such as traffic flow, short-term route, transmission possibility of specific direction and node position, etc. The transmission can be optimized based on the prediction information in terms of transmission delay, hop count and delivery ratio [103], [106]. Yao *et al.* proposed a Prediction based Routing protocol via a Hidden Markov Model (PRHMM) to decrease traffic overhead and data buffer occupancy [104]. The lifetime of the routing is predicted in [106], which chooses the links with longer connectivity time and lower path loss. The position of the neighboring vehicles can also be employed for protocol development. Huang *et al.* proposed a distance aware epidemic routing (DAER) algorithm to get better bundle delivery ratio [115]. A Greedy Perimeter Stateless Routing (GPSR) scheme is discussed in [105] to adapt to the dynamic changing of the IoV network topology. Jiang *et al.* utilized the position information in the MAC protocol design and proposed Prediction-based TDMA MAC protocol (PTMAC) that predicts encounter collision to avoid packet collision [40]. In order to achieve minimal delay and reduce control messages exchanges, Relative Position Based Collision-free MAC Nucleus mechanism (RPBMACN) is proposed to adjust the directional antenna according to the relative positions of vehicles using dedicated channel pairs [107].

2) Big Data Assisted Connectivity Enhancement: The network connectivity can be improved with the aid of big data in IoV. For example, the vehicle movement data can be used to predict the imminent cell handover when a vehicle passes from one to another and thus can be applied in the development of seamless handoff schemes. Wanalertlak *et al.* conducted prediction based on location information, group, time-of-day, and duration characteristics of mobile users to realize seamless handoff [108], [109]. Nadembega *et al.* introduced a time estimation model oriented from the movement data to estimate the handoff time window [110]. Yao *et al.* applied a hidden Markov model to predict a vehicle's future locations with the

knowledge of past traces, which enables seamless handoff between V2V and V2I communications [104]. In [110] a bandwidth reservation scheme with the awareness of the mobility prediction was proposed to reduce the packet dropping rate during the handoff procedure, combining Handoff time estimation (HTE) scheme and Available Bandwidth Estimation (ABE) scheme. Oubbati *et al.* used the UAVs to collect the traffic information and guide the ground vehicles to route the data traffic efficiently [111].

C. IoV Big Data for Autonomous Vehicles

The IoV big data is a key enabling technology of the revolutionary autonomous vehicles. To make the self-driving come to life, a convergence of big data is required, including the data from on-board sensors, e.g., cameras, radar, Lidar, GPS, and information shared from other connected vehicles, e.g., road condition, traffic information. etc. It is predicted that the self-driving vehicle can generate over 1 Tera Bytes data per hour [139]. With all the data available, efficient learning schemes, and massive computing and storage power, the vehicles will be capable of perceiving the environments, and make actuation decisions to drive safely and efficiently. A figure showing the big data applications in autonomous driving can be seen in Fig. 4.

One of the distinguished application of IoV big data in self-driving is the environment perception. The outfitted advanced sensors, including high-quality cameras, radars, and Lidars, will provide abundant but usually redundant environmental information. The task of environment perception is to abstract useful information from the large amount of data accurately and timely, in order to comprehensively understand the environment in proximity, including the road conditions, signs and traffic lights, pedestrians and other vehicles, etc. Computer vision has played an important role in self-driving related tasks, such as lane detection, vehicle tracking, pedestrian detection, etc. The usage of classic computer vision technologies in autonomous driving can be found in [118]−[120]. Classic computer vision technologies, such as feature descriptors (e.g., scale-invariant feature transform (SIFT)) and machine learning algorithms (e.g., support vector machines (SVM)) fail to perform well in more complex tasks in self-driving [117]. In addition, these methods need a huge amount of engineering efforts to extract and compute the features used in learning. Recently, the large and labeled data sets have been developed rapidly, such as ILSVRC [121]. Based on the enlarged data sets, and the increasing computing capability using Graphic Process Units (GPUs), the deep learning methods with Deep Neural Networks (DNNs) have been widely used for computer vision tasks. Many works have been conducted on the road vision data sets, such as KITTI [121] and CityScapes Dataset [122]. In [123], Mur-Artal *et al.* developed a real-time feature-based monocular Simultaneous Localization and Mapping (SLAM) system for estimation of camera localizations and geometrical reconstruction. In [124], a novel method named FlowNet was proposed to learn and predict the optical flow using Convolutional Neural Network (CNN). In [125], the appearance patterns of vehicles were learned for detection and
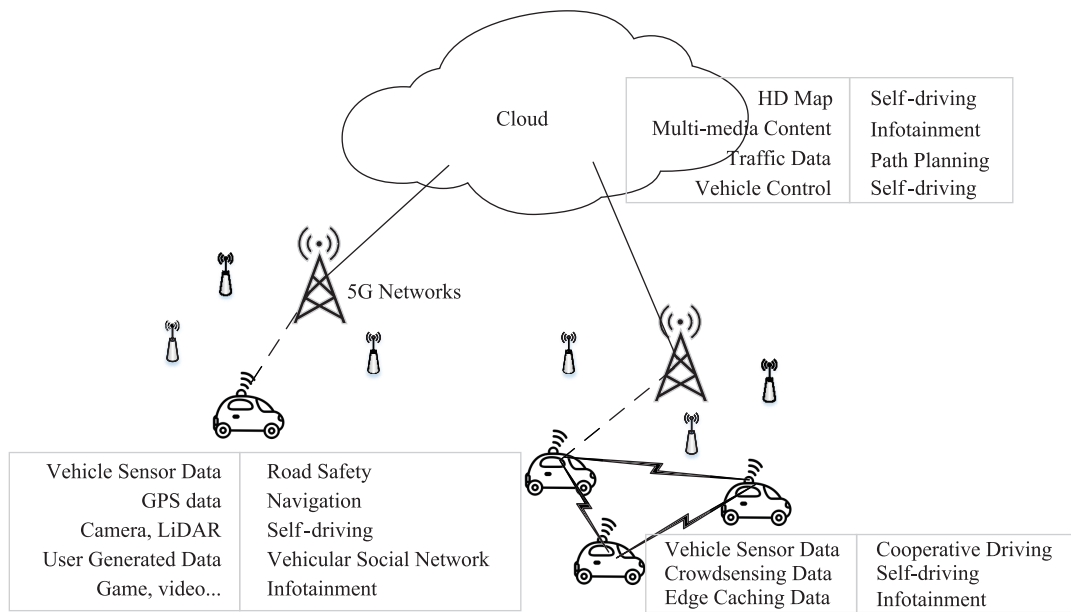
Fig. 4. Overview of IoV big data in autonomous vehicle.

orientation estimation. In [126], Premebida *et al.* trained a detector combining RGB and LiDAR data to detect pedestrians. It is found that the best performance can be obtained when a late re-scoring strategy is designed sensitive to geometric context. In [127], Li *et al.* proposed a semantic diagram to define the driving intelligence combining scenario-based and functionality-based test methods. In addition, the real view data collected by outfitted sensors is mapped into the virtual spaces, such that the facticity of the tests is enhanced by the data-driven method. Current self-driving solutions mainly rely on parsing the entire scene for making actuation decisions, and this is the choice of most existing and testing self-driving vehicles, such as Google's self-driving vehicle, Tesla Autopilot, and Intel Mobileye autonomous driving system [128]. However, there are other methods for applying IoV big data and deep learning algorithms in self-driving. An end-to-end self-driving system that directly maps the image to the steering control can be found in [13], [129], [130], where a mediate method between the two, mapping images to perception indicators that reflect road/traffic state, is proposed in [131].

High-Definition map (HD map) is a prerequisite of the driverless vehicles. Fully automated vehicles require very accurate positioning and a virtual image of the vehicle's vicinity. The LiDAT and HD map technologies are employed to provide useful data for it [132]. LiDAR is an advanced type of sensor that uses rotating laser beams to scan the vehicle proximate environment, and provide data called "point cloud" containing high-resolution information. Unlike the traditional digital map, HD map contains much more detailed information about the roads, including number and location of lanes, road signs, traffic lights, and even small bumps and holes on the road. Therefore, HD maps have a much larger volume than general maps, at about several Giga Bytes per kilometer. Combining the sensed data from LiDAR and the HD map, vehicles will be able to estimate the scene in 3D and localize itself, which

is a key to the success of self-driving solutions. However, the generation and maintenance of the HD map is a heavy task, involving a fleet of vehicles outfitted with LiDAR measuring while moving along the road, which might be very costly [133]. To reduce such expense, Máttyus *et al.* proposed a road segmentation mechanism by parsing ground and aerial images to infer the semantics of roads in [133]. In [134], a precision vehicle localization method in urban environment is proposed based on high-resolution maps. LiDAR measurements are correlated with the HD map in order to localize moving vehicles, and the results show that the method can outperform the conventional GPS-IMU-odometry-based methods by over an order of magnitude.

Although applying the big data techniques is potential and critical to self-driving, there are several challenges worth being noted. Firstly, self-driving algorithms need extremely large storage and high computing capability to deal with the related big data. If the self-driving decisions are made locally at vehicles, on one hand, the semi-static HD map, and real-time sensing data from high-quality cameras and LiDARs should be stored, which will require a massive and high-speed storage; on the other hand, fusion, processing, and deep learning algorithms applied to the big sensed data demand very high computation capability. Although there are existing solutions of equipping vehicles with such advanced devices, such as Nvidia DRIVE PX [135], it is still very expensive and results in the increasing price of the self-driving vehicles. Other solutions rely on communication facilities to offload the storage and computation tasks to clouds. However, this will pose challenges on the communication and network technologies, such as massive connections, very high data rates and low latency [132], etc. Besides, there is also another challenge about ethic problem in self-driving [120]. The IoV data security and privacy issues can also be a big concern for autonomous vehicles, as investigated in [136]−[138].

IV. Emerging Issues

In this section, the emerging issues related to the IoV big data researches are discussed. The high dynamics in IoV and ever-growing big data requirement provide great opportunities for researchers to leverage the IoV resources for big data processing. As mentioned in previous sections, the emerging issues on how future IoV could handle the big data, as well as on how the big data enabled IoV could better support the ITS applications and improve the driving experiences are discussed.

A. Big Data Aware IoV Development

1) IoV Big Data Sourcing and Preprocessing: Future IoV is expected to gather various types of big data, process them and/or relay to destinations. The more data the IoV can handle, the more information that IoV can provide for applications such as ITS and vehicular safety enhancement [146]. Thus, the data sourcing schemes are urged to acquire even bigger data from larger scope. For example, a crowd-sourcing mechanism can be introduced in vehicular scenarios and obtain road information from different geographical positions and time moments. The big data in IoV can be of great redundancy, either from the correlation of the data source or caused by the mobility of the vehicles since they might sense the same area at different drive-by moments. Thus, it is necessary to design efficient schemes to prune the redundancy of the raw data, and thus to reduce the overhead of the transmission, storage and processing. Besides, it is also important for IoV to extract the required information from data of different types and various sources. For example, the road condition information can be monitored by the on-road vehicles, roadside units, and the UAVs. The monitoring data from these entities can be integrated near the scene before transmitted to remote management centers.

2) IoV Network Protocol: The IoV network protocols are desired to reliably carry the data from source to destination. It is important that for different types of data, the network protocol should be designed accordingly using one or several radio technologies. Take the safety message as an example, which should be delivered within short delay threshold. TDMA based MAC protocol is often designed accordingly to guarantee the message deliver delay. Since IoV network is highly dynamic, the network protocol should be scalable to the network size and changing topology. Furthermore, the network protocol design should also be aware of the data properties, e.g., be aware of the data type, data source and data users. For instance, the data oriented network protocol can help users to recognize the data source as an appropriate provider or malicious attacker, and thus the security level of the transmission can be improved if the transmission protocol is designed properly. And the QoS can also be improved if the network protocol is aware of the requirement from upper layer applications and configured accordingly based on flat network architectures, e.g., software defined IoV [147]. Besides, the standardization of the IoV network protocol is also important to the popularization of the big data enabled IoV.

B. Big Data Powered IoV Services

By mining the instincts of big data, IoV can provide better support for ITS, safety applications, etc. For example, the road data collected by the vehicle can be aggregated by roadside units to form a HD map, which can then be shared for the vehicles passing the region. And the traffic condition that is oriented from the vehicle mobility data can be shared via IoV and then infer the best route and provide real time navigation for vehicles. Based on the ubiquitous sensing ability in IoV, the big data can also be queried according to customers interests [148]. The IoV big data service can also support and expand the scope of the social network on wheels [17], [149]. In addition, the IoV big data can also be used for the convergence of the ITS and smart grid, which promises a electrified prospect for the future transportation ecosystem [150]. However, the research and implementation details of such IoV services still require further investigation, as both industry and academia have not fully realize the power of the big data and thus not ready to develop the internal value of big data enabled IoV. The big data powered IoV service can be further boosted by developing proper business model for IoV service market. The IoV services related to driving safety, road management, and autonomous vehicles are expected to be the important scenarios for big data powered IoV.

V. Conclusion

In this article, we have surveyed the IoV in big data era, and investigated the relationship between the IoV and big data. We have shown that on one hand IoV can support the big data acquisition, transmission, storage and computing, and on the other hand the big data can enhance IoV in terms of network characterization, performance analysis and protocol design. Such reciprocal relationship is very useful to follow up the trend of the IoV development with the mutualism to big data. Besides, we have highlighted and elaborated the key role of IoV big data to autonomous vehicles, and pinpointed the emerging issues to show some desired directions of the future IoV in big data era.

References

[1] "Israeli startup looks to secure connected cars from hackers", Report from Business Insider, 2016. [Online]. Available: http://www.businessinsider.com/isreali-startup-looks-to-secure-connected-cars- from-hackers-2016-4

[2] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: Solutions and challenges," *IEEE Internet of Things Journal*, vol. 1, no. 4, pp. 289−299, 2014.

[3] F. Bai, H. Krishnan. "Reliability analysis of DSRC wireless communication for vehicle safety applications,"*In proc. of Intelligent Transportation Systems Conference,* pp. 355−362. IEEE, 2006.

[4] N. Cheng, N. Lu, N. Zhang, X. Shen, and J. W. Mark,"Vehicular WiFi offloading: Challenges and solutions," *Vehicular Communications*, vol. 1, no. 1, pp. 13−21, 2014.

[5] H. Zhou, N. Cheng, N. Lu, L. Gui, D. Zhang, Q. Yu, F. Bai, and X. Shen,"WhiteFi Infostation: Engineering Vehicular Media Streaming with Geolocation Database," *IEEE J. Selected Areas of Communications,* vol. 34, no. 8, pp. 2260−2274, 2016.

[6] IEEE Std 802.11p-2010, "Standard for Information Technology-Telecommunications and Information Exchange between Systems-Local and Metropolitan Area Networks-Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments," pp. 1−51, Jul. 2010.

[7] Automotive Sensors 2017: http://www.automotivesensors2017.com/

[8] N. Zhang, P. Yang, J. Ren, D. Chen, L. Yu, and X. Shen, "Synergy of Big Data and 5G Wireless Networks: Opportunities, Approaches, and Challenges," *IEEE Wireless Communications Magazine - SI on Wireless Big Data*, to appear.

[9] F. Yang, S. Wang, J. Li, Z. Liu, and Q. Sun, "An overview of internet of vehicles," *China Communications*, vol. 11, no. 10, pp. 1−15, 2014.

[10] Intel Autonomous driving: https://www.driverlessguru.com/blog/intel-to-invest-250-million-into-autonomous-driving

[11] H. Peng, D. Li, K. Abboud, H. Zhou, W. Zhuang, X. Shen, and H. Zhao, "Performance analysis of IEEE 802.11 p DCF for inter-platoon communications with autonomous vehicles," *In Proc. of Global Communications Conference (GLOBECOM)*, pp. 1−6. IEEE, 2015.

[12] H. Peng, D. Li, K. Abboud, H. Zhou, H. Zhao, W. Zhuang, and X. Shen, "Performance Analysis of IEEE 802.11 p DCF for Multiplatooning Communications With Autonomous Vehicles," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 3, pp. 2485−2498, 2017.

[13] M. Bojarski, D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, "End to end learning for self-driving cars," arXiv preprint arXiv:1604.07316, 2016.

[14] J. Wedel, B. Schünemann, and I. Radusch. "V2X-based traffic congestion recognition and avoidance." *In Pervasive Systems, Algorithms, and Networks (ISPAN),* In Proc. 10th International Symposium on, pp. 637−641. IEEE, 2009.

[15] P. Händel, J. Ohlsson, M. Ohlsson, I. Skog, and E. Nygren, "Smartphone-based measurement systems for road vehicle traffic monitoring and usage-based insurance," *IEEE Systems Journal*, vol. 8, no. 4, pp. 1238−1248, 2014.

[16] A. Ghose, P. Biswas, C. Bhaumik, M. Sharma, A. Pal, and A. Jha. "Road condition monitoring and alert application: Using in-vehicle smartphone as internet-connected sensor," *In Proc. IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops),* pp. 489−491. IEEE, 2012.

[17] X. Liang, X. Li, T. H. Luan, R. Lu, X. Lin, X. Shen, "Morality-driven Data Forwarding with Privacy Preservation in Mobile Social Networks," *IEEE Trans. on Vehicular Technology,* vol. 61, no. 7, pp. 3209−3221, 2012.

[18] S. Karapantazis and F.-N. Pavlidou, "The role of high altitude platforms in beyond 3G networks," *IEEE Wireless Communications,* vol. 12, no. 6, pp. 33−41, 2016.

[19] L. Gupta, R. Jain, and G. Vaszkun, "Survey of important issues in UAV communication networks," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1123−1152, 2016.

[20] CAMP Vehicle Safety Communications Consortium, "Vehicle safety communications project: Task 3 final report: identify intelligent vehicle safety applications enabled by DSRC," *National Highway Traffic Safety Administration, US Department of Transportation, Washington DC*, Mar. 2005.

[21] H. Zhou, B. Liu, T.H. Luan, F. Hou, L. Gui, Y. Li, G. Yu, and X. Shen, "ChainCluster: Engineering a Cooperative Content Distribution Framework for Highway Vehicular Communications," *IEEE Trans. on Intelligent Transportation Systems,* vol. 15, no. 6, pp. 2644−2657, 2014.

[22] DSRC Committee, "Dedicated Short Range Communications (DSRC) Message Set Dictionary," *Soc. Automotive Eng., Warrendale, PA, USA, Tech. Rep. J2735_200911*, Nov. 2009.

[23] M. Chaqfeh, A. Lakas and I. Jawhar, "A survey on data dissemination in vehicular ad hoc networks," *Vehicular Communications*, vol. 1, no. 4, pp. 214−225, Oct. 2014.

[24] J. Qin, H. Zhu, Y. Zhu, L. Lu, G. Xue, and M. Li. "POST: Exploiting dynamic sociality for mobile advertising in vehicular networks." *IEEE Transactions on Parallel and Distributed Systems,* vol. 27, no. 6, pp. 1770−1782, 2016.

[25] F. Lv, H, Zhu, H. Xue, Y, Zhu, S, Chang, M, Dong and M. Li, "An Empirical Study on Urban IEEE 802.11p Vehicle-to-Vehicle Communication," *In Proc. 13th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pp. 1−9, Jun. 2016.

[26] W. Xu, H. A. Omar, W. Zhuang, and X. Shen, "Delay Analysis of In-Vehicle Internet Access via On-Road WiFi Access Points," *IEEE Access,* vol. 5, no. 1, pp. 2736−2746, 2017.

[27] Q. Xu, T. Mak, J. Ko, and R. Sengupta, "Vehicle-to-vehicle safety messaging in DSRC," *In Proc. of the 1st ACM international workshop on Vehicular ad hoc networks,* pp. 19−28, ACM, 2004.

[28] M. Dikaiakos, S. Iqbal, T. Nadeem, and L. Iftode, "VITP: an information transfer protocol for vehicular computing," *In Proc. of the 2nd ACM international workshop on Vehicular ad hoc networks,* pp. 30−39. ACM, 2005.

[29] J. Lipman, H. Liu, I. Stojmenovic, "Broadcasting in ad hoc networks," *Guide to Wireless Ad Hoc Networks*, Springer-Verlag, London, UK, pp.121−150 (Chapter 6), 2009.

[30] N. Wisitpongphan, O. K. Tonguz, J. S. Parikh, P. Mudalige, F. Bai, V. Sadekar, "Broadcast storm mitigation techniques in vehicular ad hoc networks," *IEEE Wireless Communication*, vol. 14, no. 6, pp. 84−94, 2006.

[31] L. Yin and G Cao. "Supporting cooperative caching in ad hoc networks," *IEEE transactions on mobile computing*, vol. 5, no. 1, pp. 77−89, 2006.

[32] M. Fiore, F. Mininni, C. Casetti, C.F. Chiasserini, "To cache or not to cache?" *IEEE INFOCOM*, pp. 235−243, 2009.

[33] G. Y. Chang, J. P. Sheu, T. Y. Lin, K. Y. Hsieh, "Cache-based routing for vehicular ad hoc networks in city environments," *In Proc. IEEE Wireless Communications and Networking Conference, WCNC*, pp.1−6, 2010.

[34] IEEE Std 802.11-2007 (Revision of IEEE Std. 802.11-1999), "Standard for Information Technology-Telecommunications and Information Exchange between Systems-Local and Metropolitan Area Networks-Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," pp. 1−1184, June 2007.

[35] M. I. Hassan, H. L. Vu, and T. Sakurai, "Performance Analysis of the IEEE 802.11 MAC Protocol for DSRC Safety Applications," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 8, pp. 3882−3896, Oct. 2011.

[36] M. Hadded, P. Muhlethaler, A. Laouiti, R. Zagrouba, and L. A. Saidane, "TDMA-Based MAC Protocols for Vehicular Ad Hoc Networks: A Survey, Qualitative Analysis, and Open Research Issues," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2461−2492, Jun. 2015.

[37] F. Borgonovo, A. Capone, M. Cesana, and L. Fratta, "ADHOC MAC: new MAC architecture for ad hoc networks providing efficient and reliable point-to-point and broadcast services," *Wireless Networks*, vol. 10, no. 4, pp. 359−366, 2004.

[38] H. A. Omar, W. Zhuang, and L. Li, "VeMAC: a TDMA-based MAC protocol for reliable broadcast in VANETs," *IEEE Transactions on Mobile Computing*, vol. 12, no. 9, pp. 1724−1736, 2013.

[39] F. Borgonovo, A. Capone, M. Cesana, and L. Fratta, "Rr-aloha, a reliable r-aloha broadcast channel for ad-hoc inter-vehicle communication networks," *Proc. Med-Hoc-Net*, pp. 1−5.

[40] X. Jiang and D. H. C. Du, "PTMAC: A Prediction-Based TDMA MAC

Protocol for Reducing Packet Collisions in VANET," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 11, pp. 9209−9223, Nov. 2016.

[41] W. Yang, P. Li, Y. Liu, and H. Zhu, "Adaptive TDMA slot assignment protocol for vehicular ad-hoc networks," *The Journal of China Universities of Posts and Telecommunications*, vol. 20, no. 1, pp. 11−25, Feb. 2013.

[42] R. Zhang, X. Cheng, L. Yang, X. Shen, and B. Jiao, "A Novel Centralized TDMA-Based Scheduling Protocol for Vehicular Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 411−416, Feb. 2015.

[43] J. J. Blum and A. Eskandarian, "A reliable link-layer protocol for robust and scalable intervehicle communications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 1, pp. 4−13, 2007.

[44] S. V. Bana and P. Varaiya, "Space division multiple access (SDMA) for robust ad hoc vehicle communication networks," *IEEE Intelligent Transportation Systems*, pp. 962−967, 2001.

[45] F. Watanabe, M. Fujii, M. Itami, and K. Itoh, "An analysis of incident information transmission performance using MCS/CDMA scheme," *IEEE Intelligent Vehicles Symposium*, pp. 249−254, 2005.

[46] T. Inoue, H. Nakata, M. Itami, and K. Itoh, "An analysis of incident information transmission performance using an IVC system that assigns PN codes to the locations on the road," *IEEE Intelligent Vehicles Symposium*, pp. 115−120, 2004.

[47] M. Altayeb and I. Mahgoub, "A survey of vehicular ad hoc networks routing protocols," *International Journal of Innovation and Applied Studies*, vol. 3, no. 3, pp. 829−846, 2013.

[48] F. Li and Y. Wang, "Routing in vehicular ad hoc networks: A survey," *IEEE Vehicular technology magazine*, vol. 2, no. 2, 2007.

[49] K. C. Lee, U. Lee, M. Gerla, "Survey of Routing Protocols in Vehicular Ad Hoc Networks," *Advances in Vehicular Ad-Hoc Networks: Developments and Challenges*, pp. 149−170, 2010.

[50] J. Toutouh and E. Alba, "Optimizing OLSR in VANETS with Differential Evolution: A Comprehensive Study," *ACM international symposium on Design and analysis of intelligent vehicular networks and applications*, Miami, Florida, USA, 2011.

[51] V. Naumov, R. Baumann, T. Gross, "An evaluation of Inter-Vehicle Ad Hoc Networks Based on Realistic Vehicular Traces," *In Proc. of the 7th ACM international symposium on Mobile ad hoc networking and computing*, May 2006.

[52] M. K. Marina, S. R. Das, "On-demand multipath distance vector routing in Ad Hoc networks," *In Proc. 9th IEEE International Conference on Network Protocols*, pp. 14−23, Nov. 2001.

[53] M. Abolhasan, T. Wysocki, and E. Dutkiewicz, "A review of routing protocols for mobile ad hoc networks," *Ad Hoc Networks, Elsevier*, pp. 1−22, 2004.

[54] G. Liu, B.-S. Lee, B.-C. Seet, C.H. Foh, K.J. Wong, and K.-K. Lee, "A routing strategy for metropolis vehicular communications," *in Proc. International Conference on Information Networking (ICOIN)*, pp. 134−143, 2004.

[55] H. Fubler, M. Mauve, H. Hartenstein, M. Kaemann, and D. Vollmer, "Mobicom poster: location-based routing for vehicular ad-hoc networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 7, no. 1, pp. 47−49, 2003.

[56] W. Shi, H. Zhou, J. Li, W. Xu, N. Zhang, and X. Shen, "Drone Assisted Vehicular Networks: Architecture, Challenges and Opportunities," *IEEE Network,* to appear.

[57] N. Zhang, S. Zhang, P. Yang, O. Alhussein, W. Zhuang, and X. Shen, "Software Defined Space-Air-Ground Integrated Vehicular Networks: Challenges and Solutions," *IEEE Communications Magazine,* vol. 55, no. 7, pp. 101−109, 2017.

[58] A. Mukhtar, L. Xia, and T. Tang, "Vehicle detection techniques for collision avoidance systems: A review," *IEEE Transactions on Intelligent Transportation Systems,* vol. 16, no. 5, pp. 2318−2338, 2015.

[59] Eric Mack. "Meet Google's Project Loon: Balloon-powered Net Access," *CNET. Retrieved*, 15, 2013.

[60] I. Bor-Yaliniz, and H. Yanikomeroglu, "The new frontier in RAN heterogeneity: Multi-tier drone-cell," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 48−55, 2016.

[61] N. Alam, A. Balaei, and A. G. Dempster, "Relative positioning enhancement in VANETs: A tight integration approach," *IEEE Transactions on Intelligent Transportation Systems,* vol. 14, no. 1, pp. 47−55, 2013.

[62] A. Widiawan, and R. Tafazolli, "High altitude platform station (HAPS): A review of new infrastructure development for future wireless communications," *Wireless Personal Communications,* vol. 42, no. 3, pp. 387−404, 2007.

[63] D. Grace, J. Thornton, G. Chen, G. P. White, and T. C. Tozer, "Improving the system capacity of broadband services using multiple high-altitude platforms," *IEEE Transactions on Wireless Communications,* vol. 4, no. 2, pp. 700−709, 2005.

[64] N. Motlagh, T. Taleb, and O. Arouk, "Low-altitude unmanned aerial vehicles-based Internet of things services: Comprehensive survey and future perspectives," *IEEE Internet of Things Journal,* vol. 3, no. 6, pp. 899−922, 2016.

[65] N. Goddemeier, K. Daniel, and C. Wietfeld, "Role-based connectivity management with realistic air-to-ground channels for cooperative UAVs," *IEEE Journal on Selected Areas in Communications,* vol. 30, no. 5, pp. 951−963, 2012.

[66] A. Dhekne, M. Gowda, and R. R. Choudhury, "Cell tower extension through drones: Poster," *In Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking,* pp. 456−457. ACM, 2016.

[67] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569−572, 2014.

[68] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Mobile Internet of Things: Can UAVs provide an energy-efficient mobile architecture?," *In Proceeding of Global Communications Conference (GLOBECOM),* pp. 1−6. IEEE, 2016.

[69] P. Yang, X. Cao, C. Yin, Z. Xiao, X. Xi, and D. Wu, "Proactive Drone-Cell Deployment: Overload Relief for a Cellular Network Under Flash Crowd Traffic," *IEEE Transactions on Intelligent Transportation Systems,* 2017.

[70] Bor-Yaliniz, R. Irem, A. El-Keyi, and H. Yanikomeroglu, "Efficient 3-D placement of an aerial base station in next generation cellular networks," *In Proceeding of International Conference on Communications (ICC),* pp. 1−5. IEEE, 2016.

[71] E. Kalantari, H. Yanikomeroglu, and A. Yongacoglu, "On the number and 3D placement of drone base stations in wireless cellular networks," *In Proc. 84th Vehicular Technology Conference (VTC-Fall),* pp. 1−6. IEEE, 2016.

[72] Y. Zhou, N. Cheng, N. Lu, and X. Shen, "Multi-UAV-aided networks: aerial-ground cooperative vehicular networking architecture," *IEEE vehicular technology magazine,* vol. 10, no. 4, pp. 36−44, 2015.

[73] O. S. Oubbati, A. Lakas, N. Lagraa, and M. B. Yagoubi, "UVAR: An intersection UAV-assisted VANET routing protocol," *In Proc. Wireless Communications and Networking Conference (WCNC),* pp. 1−6. IEEE, 2016.

[74] J. He, Y. Ni, L. Cai, J. Pan, and C. Chen, "Optimal dropbox deployment algorithm for data dissemination in vehicular networks," *IEEE Transactions on Mobile Computing*, 2017.

[75] BaiDu Apollo Project: http://apollo.auto/platform/hardware.html

[76] Cohda Wireless: http://www.cohdawireless.com/solutions/hardware/

[77] S. Olariu, T. Hristov, and G. Yan, "The next paradigm shift: from vehicular networks to vehicular clouds," *Mobile Ad Hoc Networking: Cutting Edge Directions, Second Edition*, pp. 645−700, 2013.

[78] J. Zhao, G. Cao, "VADD: Vehicle-assisted data delivery in vehicular ad hoc networks," *IEEE Transactions on Vehicular Technology.* vol. 57, no.3, pp. 1910−22, 2008.

[79] K. Zheng, H. Meng, P. Chatzimisios, L. Lei and X. Shen, "An SMDP-based resource allocation in vehicular cloud computing systems," *IEEE Transactions on Industrial Electronics,* vol. 62, no. 12, pp. 7920−7928, 2015.

[80] V. N. Soares, F. Farahmand, J.J. Rodrigues, "A layered architecture for vehicular delay-tolerant networks," *In Proc. IEEE Symposium on Computers and Communications.* ISCC 2009.

[81] A. Mahmood, C. Casetti, C.F. Chiasserini, P. Giaccone, and J. Harr, "Mobility-aware edge caching for connected cars," *In Proc.12th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, pp. 1−8. IEEE, 2016.

[82] R. Ding, T. Wang, L. Song, Z. Han, and J. Wu,"Roadside-unit caching in vehicular ad hoc networks for efficient popular content delivery," *In Proc. Wireless Communications and Networking Conference (WCNC)*, pp. 1207−1212. IEEE, 2015.

[83] Z. Hu, Z. Zheng, T. Wang, and L. Song, "Poster: Roadside Unit Caching Mechanism for Multi-Service Providers," *In Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 387−388. ACM, 2015.

[84] S. Zhang, N. Zhang, X. Fang, P. Yang, and X. Shen,"Self-Sustaining Caching Stations: Toward Cost-Effective 5G-Enabled Vehicular Networks," *IEEE Communications Magazine*, 2017.

[85] The ECU Computer (Engine Control Unit) [Online], Available: http://www.chipsetc.com/computer-chips-inside-the-car.html

[86] W. B. Dunbar and D. S. Caveney, "Distributed receding horizon control of vehicle platoons: Stability and string stability," *IEEE Transactions on Automatic Control*, vol. 57, no. 3, pp. 620−633, 2012.

[87] Y. Zhang, G. Cao. "V-PADA: Vehicle-platoon-aware data access in VANETs," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 5, pp. 2326−2339, 2011.

[88] S. Arif, S. Olariu, J. Wang, G. Yan, W. Yang, and I. Khalil, "Datacenter at the airport: Reasoning about time-dependent parking lot occupancy," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 11, pp. 2067−2080, 2012.

[89] T. Zhang, R. E. D. Grande, and A. Boukerche, "Vehicular cloud: Stochastic analysis of computing resources in a road segment," *In Proceedings of the 12th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks,* pp. 9−16. ACM, 2015.

[90] F. Bai, D. D. Stancil, and H. Krishnan,"Toward understanding characteristics of dedicated short range communications (DSRC) from a perspective of vehicular network engineers," *In Proceedings of the 6th Annual International Conference on Mobile Computing and Networking,* pp. 329−340. ACM, 2010.

[91] J. Karedal, N. Czink, A. Paier, F. Tufvesson, and A. F. Molisch, "Path loss modeling for vehicle-to-vehicle communications," *IEEE transactions on vehicular technology*, vol. 60, no. 1, pp. 323−328, 2011.

[92] J. Eriksson, H. Balakrishnan, and S. Madden. "Cabernet: vehicular content delivery using WiFi," *In Proceedings of the 14th ACM International Conference on Mobile Computing and Networking,* pp. 199−210. ACM, 2008.

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

[94] C. F. Mecklenbrauker, A. F. Molisch, J. Karedal, F. Tufvesson, A. Paier, L. Bernadó, T. Zemen, O. Klemp, and N. Czink, "Vehicular channel characterization and its implications for wireless system design and performance," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1189−1212, 2011.

[95] I. Tan, W. Tang, K. Laberteaux, and A. Bahai, "Measurement and analysis of wireless channel impairments in DSRC vehicular communications,"*In Proceeding of International Conference on Communications (ICC),* pp. 4882−4888. IEEE, 2008.

[96] C. M. Martinez, M. Heucke, F. Y. Wang, B. Gao, and D. Cao, "Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey,"*IEEE Transactions on Intelligent Transportation Systems,* 2017.

[97] N. Akhtar, S. C. Ergen, and O. Ozkasap, "Vehicle mobility and communication channel models for realistic and efficient highway VANET simulation," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 1, pp. 248−262, 2015.

[98] G. Setiwan, S. Iskander, Q. Chen, S. S. Kanhere, and K. Lan,"The Effect Of Radio Models On Vehicle Network Simulations," *In Proceedings of the 14th World Congress on Intelligent Transport Systems,* pp. 1−10. 2007.

[99] O. Renaudin, V. Kolmonen, P. Vainikainen, and C. Oestges, "Non-stationary narrowband MIMO inter-vehicle channel characterization in the 5-GHz band," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 2007−2015, 2010.

[100] L. Cheng, B. E. Henty, F. Bai, and D. D. Stancil, "Highway and rural propagation channel modeling for vehicle-to-vehicle communications at 5.9 GHz,"*In Antennas and Propagation Society International Symposium,*pp. 1−4. IEEE, 2008.

[101] A. Paier, J. Karedal, N. Czink, H. Hofstetter, C. Dumard, T. Zemen, F. Tufvesson, C. F. Mecklenbrauker, and A. F. Molisch, "First results from car-to-car and car-to-infrastructure radio channel measurements at 5.2 GHz," *In Proc. IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications,* pp. 1−5. IEEE, 2007.

[102] S. Bharati, H. A. Omar, and W. Zhuang, "Enhancing Transmission Collision Detection for Distributed TDMA in Vehicular Networks," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM),* vol. 13, no. 3, pp. 37, 2017.

[103] T. Cao, X. Zhang, L. Kong, X. Liu, W. Shu, and M. Wu, "Traffic aware routing in urban vehicular networks," *In Proc. Wireless Communications and Networking Conference (WCNC),* pp. 2004−2009. IEEE, 2013.

[104] L. Yao, J. Wang, X. Wang, A. Chen, and Y. Wang, "V2X Routing in a VANET Based on the Hidden Markov Model," *IEEE Transactions on Intelligent Transportation Systems*, 2017.

[105] B. Karp, H. Kung, "GPSR: Greedy perimeter stateless routing for wireless networks," *In Proceedings of the 6th Annual International Conference on Mobile Computing and Networking,* pp. 243−254. ACM, 2000.

[106] R. J. Ravi, R. PonLakshmi, "A New Lifetime Prediction Algorithm Based Routing For VANETs," *International Journal of Computer Science & Applications (TIJCSA)*, vol. 1, no. 12, 2013.

[107] C. Chigan, V. Oberoi, and J. Li, "Wsn07-3: Rpb-macn: A relative position based collision-free mac nucleus for vehicular ad hoc networks," *In Proc. Global Telecommunications Conference,* pp. 1−6. IEEE, 2006.

[108] W. Wanalertlak, B. Lee, C. Yu, M. Kim, S. Park, and W. Kim, "Behavior-based mobility prediction for seamless handoffs in mobile wireless networks,"*Wireless Networks*, vol. 17, no. 3, pp. 645−658, 2011.

[109] Y. Bi, H. Zhou, W. Xu, X. Shen, and H. Zhao, "An Efficient PMIPv6-Based Handoff Scheme for Urban Vehicular Networks," *IEEE Trans.*
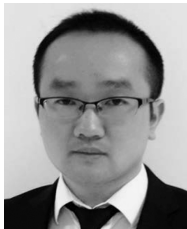
on Intelligent Transportation Systems, vol. 17, no. 12, pp. 3613−3628, 2016.

[110] A. Nadembega, A. Hafid, and T. Taleb, "Handoff time estimation model for vehicular communications," In Proceeding of International Conference on Communications (ICC), pp. 1715−1719, IEEE, 2013.

[111] O. S. Oubbati, A. Lakas, F. Zhou, M. Güneş, N. Lagraa, and M. B. Yagoubi, "Intelligent UAV-assisted routing protocol for urban VANETs," Computer Communications, vol. 107, pp. 93−111, 2017.

[112] Y. J. Li, "An overview of the DSRC/WAVE technology," In Proc. International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, pp. 544−558. Springer, Berlin, Heidelberg, 2010.

[113] H. Zhou, W. Xu, Y. Bi, J. Chen, Q. Yu, and X. Shen, "Toward 5G Spectrum Sharing for Immersive Experience Driven Vehicular Communications," IEEE Wireless Communications Magazine, to appear.

[114] H. Zhu, M. Li, L. Fu, G. Xue, Y. Zhu, and L. M. Ni,"Impact of traffic influxes: Revealing exponential intercontact time in urban vanets," IEEE Transactions on Parallel and Distributed Systems, vol. 22, no. 8, pp. 1258−1266, 2011.

[115] H. Huang, P. Luo, M. Li, D. Li, X. Li, W. Shu, and M. Wu, "Performance evaluation of SUVnet with real-time traffic data," IEEE Transactions on Vehicular Technology, vol. 56, no. 6, pp. 3381−3396, 2007.

[116] C. Celes, F. Silva, A. Boukerche, R. Andrade, and A. Loureiro, "Improving VANET Simulation with Calibrated Vehicular Mobility Traces," IEEE Transactions on Mobile Computing, 2017.

[117] R. Szeliski, "Computer vision: algorithms and applications," Springer Science & Business Media, 2010.

[118] H. Cho, Y. Seo, B. V. Kumar, and R. R. Rajkumar, "A multi-sensor fusion system for moving object detection and tracking in urban driving environments," In IEEE International Conference on Robotics and Automation (ICRA), pp. 1836−1843. IEEE, 2014.

[119] D. Held, J. Levinson, and S. Thrun, "A probabilistic framework for car detection in images using context and scale," In IEEE International Conference on Robotics and Automation (ICRA), pp. 1628−1634. IEEE, 2012.

[120] C. Caraffi, T. Vojir, J. Trefny, J. Šochman, and J. Matas, "A system for real-time detection and tracking of vehicles from a single car-mounted camera,"In Proc. 15th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp. 975−982. IEEE, 2012.

[121] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," The International Journal of Robotics Research, vol. 32, no. 11, pp. 1231−1237, 2013.

[122] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele,"The cityscapes dataset for semantic urban scene understanding," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3213−3223. 2016.

[123] R. Mur-Artal, J.M.M., and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147−1163, 2015.

[124] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. V. D. Smagt, D. Cremers, and T. Brox,"Flownet: Learning optical flow with convolutional networks,"In Proceedings of the IEEE International Conference on Computer Vision, pp. 2758−2766. 2015.

[125] C. Premebida, J. Carreira, J. Batista, and U. Nunes,"Pedestrian detection combining rgb and dense lidar data,"In Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014), pp. 4112−4117. IEEE, 2014.

[126] Google's self-driving car [online], Available: "https://waymo.com/"

[127] L. Li, W. L. Huang, Y. Liu, N. N. Zheng, and F. Y. Wang, "Intelligence testing for autonomous vehicles: a new approach," IEEE Transactions on Intelligent Vehicles, vol. 1, no. 2 pp. 158−166, 2016.

[128] Mobileye autonomous driving [online], Available: https://www.mobileye.com/future-of-mobility/history-autonomous-driving/

[129] J. Koutník, G. Cuccu, J. Schmidhuber, and F. Gomez, "Evolving large-scale neural networks for vision-based torcs," pp. 206−212, 2013.

[130] U. Muller, J. Ben, E. Cosatto, B. Flepp, and Y. L. Cun, "Off-road obstacle avoidance through end-to-end learning," In Advances in Neural Information Processing Systems, pp. 739−746. 2006.

[131] C. Chen, A. Seff, A. Kornhauser, and J. Xiao, "Deepdriving: Learning affordance for direct perception in autonomous driving," In Proceedings of the IEEE International Conference on Computer Vision, pp. 2722−2730. IEEE, 2015.

[132] H. G. Seif, and X. Hu, "Autonomous Driving in the iCity−HD Maps as a Key Challenge of the Automotive Industry," Engineering, vol. 2, no. 2, pp. 159−162, 2016.

[133] G. Máttyus, S. Wang, S. Fidler, and R. Urtasun, "Hd maps: Fine-grained road segmentation by parsing ground and aerial images," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3611−3619. 2016.

[134] J. Levinson, M. Montemerlo, and S. Thrun, "Map-Based Precision Vehicle Localization in Urban Environments," In Robotics: Science and Systems, vol. 4, 2007.

[135] Nvidia DRIVE PX: http://www.nvidia.ca/object/drive-px.html

[136] B. Kehoe, S. Patil, P. Abbeel, and K. Goldberg, "A survey of research on cloud robotics and automation," IEEE Transactions on Automation Science and Engineering, vol. 12, no. 2, pp. 398-409, 2015.

[137] Autonomous Automotive Cybersecurity [online], Available: https://karambasecurity.com/ static/pdf/Autonomous-Automotive-Cybersecurity-Report.pdf

[138] J. Petit, B. Stottelaar, M. Feiri, and F. Kargl, "Remote attacks on automated vehicles sensors: Experiments on camera and lidar" Black Hat Europe, vol. 11, 2015.

[139] Self-driving cars could create 1GB of data a second [online], Available: http://www.computerworld.com/article/2484219/emerging-technology/self-driving-cars-could-create-1gb-of-data-a-second.html

[140] M. Gerla, "Vehicular cloud computing," In Ad Hoc Networking Workshop (Med-Hoc-Net), In Proc. the 11th Annual Mediterranean, pp. 152−155. IEEE, 2012.

[141] J. Wang, J. Cho, S. Lee, and T. Ma, "Real time services for future cloud computing enabled vehicle networks," In Proc. International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1−5. IEEE, 2011.

[142] R. Bauza, J. Gozalvez, and J. Sanchez-Soriano, "Road traffic congestion detection through cooperative vehicle-to-vehicle communications," In Proc. IEEE 35th Conference on Local Computer Networks (LCN), pp. 606−612. IEEE, 2010.

[143] S. Olariu, and M. C. Weigle, eds. Vehicular networks: from theory to practice. Crc Press, 2009.

[144] M. Whaiduzzaman, M. Sookhak, A. Gani, and R. Buyya,"A survey on vehicular cloud computing,"Journal of Network and Computer Applications, vol. 40, pp. 325−344, 2014.

[145] H. Zhang, Q. Zhang, and X. Du, "Toward vehicle-assisted cloud computing for smartphones,"IEEE Transactions on Vehicular Technology, vol. 64, no. 12, pp. 5610−5618, 2015.

[146] J. Zhang, F. Y. Wang, K. Wang, W. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: a survey," IEEE Transactions

on Intelligent Transportation Systems, vol. 12, no. 4, pp. 1624−1639, 2011.

[147] J. Cheng, H. Zhou, N. Zhang, W. Xu, Q. Yu, L. Gui, and X. Shen, "Service-Oriented Dynamic Connection Management for Software Defined Internet of Vehicles," *IEEE Trans. on Intelligent Transportation Systems,* vol. 18, no. 10, pp. 2826−2837, 2017.

[148] H. Li, Y. Yang, X. Liang, T.H. Luan, L. Zhou, and X. Shen, "Enabling fine-grained multi-keyword search supporting classified sub-dictionaries over encrypted cloud data," *IEEE Trans. on Dependable and Secure Computing,* vol. 13, no. 3, pp. 312−325, 2016.

[149] T. H. Luan, R. Lu, X. Shen, and F. Bai, "Social on the road: enabling secure and efficient social networking on highways," *IEEE Wireless Communications,* vol. 22, no. 1, pp. 44−51, 2015.

[150] X. Cheng, X. Hu, L. Yang, I. Husain, F. Y. Wang, "Electrified vehicles and the smart grid: The ITS perspective," *IEEE Transactions on Intelligent Transportation Systems,* vol. 15, no. 4, pp. 1388−1404, 2014.

Feng Lyu received the BS degree in software engineering from Central South University in 2013. He is pursuing his Ph.D degree in the Department of Computer Science and Engineering at Shanghai Jiao Tong University. He is currently a visiting scholar with the Department of Electrical and Computer Engineering, University of Waterloo, ON, Canada. His research interests include vehicular ad hoc networks and cloud computing.

Weisen Shi received the B.S. degree from Tianjin University, Tianjin, China, in 2013 and received the M.S. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2016. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His interests include drone communication and networking, software defined networking and vehicular networks.

Wenchao Xu received the B.E. and M.E. degrees from Zhejiang University, Hangzhou, China, in 2008 and 2011, respectively. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. 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.

Jiayin Chen received the B.E. degree and the M.S. degree in the School of Electronics and Information Engineering from Harbin Institute of Technology, Harbin, China, in 2014 and 2016, respectively. She is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. Her research interests are in the area of vehicular networks and machine learning, with current focus on Intelligent Transport System and big data.

Haibo Zhou (M'14) received the Ph.D. degree ininformation and communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014. From 2014 to 2017, he has worked a Post-Doctoral 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.

Xuemin (Sherman) Shen (IEEE M'97-SM'02-F09) received the B.Sc.(1982) degree from Dalian Maritime University (China) and the M.Sc. (1987) and Ph.D. degrees (1990) from Rutgers University, New Jersey (USA), all in electrical engineering. He is a University Professor and the Associate Chair for Graduate Studies, Department of Electrical and Computer Engineering, University of Waterloo, Canada. Dr. Shen's research focuses on resource management, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor networks. Dr. Shen served as the Technical Program Committee Chair/Co-Chair for IEEE Globecom'16, Infocom'14, IEEE VTC'10 Fall, and Globecom'07, the Symposia Chair for IEEE ICC'10, the Tutorial Chair for IEEE VTC'11 Spring and IEEE ICC'08, the General Co-Chair for ACM Mobihoc'15, Chinacom'07 and the Chair for IEEE Communications Society Technical Committee on Wireless Communications. He also serves/served as the Editor-in-Chief for IEEE Internet of Things Journal, IEEE Network, Peer-to-Peer Networking and Application, and IET Communications; a Founding Area Editor for IEEE Transactions on Wireless Communications; an Associate Editor for IEEE Transactions on Vehicular Technology, Computer Networks, and ACM/Wireless Networks, etc.; and the Guest Editor for IEEE JSAC, IEEE Wireless Communications, and IEEE Communications Magazine, etc. Dr. Shen received the Excellent Graduate Supervision Award in 2006, and the Premier's Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada. He is a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Distinguished Lecturer of IEEE Vehicular Technology Society and Communications Society.

Nan Cheng (M'16) received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo in 2015, and B.E. degree and the M.S. degree from the Department of Electronics and Information Engineering, Tongji University, Shanghai, China, in 2009 and 2012, respectively. He is currently working as a Postdoctoral fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, under the supervision of Prof. Ben Liang. His current research focuses on big data in vehicular networks and self-driving system. His research interests also include performance analysis, MAC, opportunistic communication for vehicular networks, cognitive radio, WiFi, smart grid, and cellular traffic offloading.