A Survey of the Connected Vehicle Landscape--Architectures, Enabling Technologies, Applications, and Development Areas

Article in IEEE Transactions on Intelligent Transportation Systems · October 2017
DOI: 10.1109/TITS.2017.2749459

3 authors, including:

Josh Siegel
Michigan State University
60 PUBLICATIONS 471 CITATIONS

Sanjay E. Sarma
Massachusetts Institute of Technology
265 PUBLICATIONS 9,256 CITATIONS

Some of the authors of this publication are also working on these related projects:

Project Secure and efficient architectures for connectivity View project

Project PDMS peeling View project
A Survey of the Connected Vehicle Landscape—Architectures, Enabling Technologies, Applications, and Development Areas

Joshua E. Siegel, Dylan C. Erb, and Sanjay E. Sarma

Abstract—This paper summarizes the state of the art in connected vehicles—from the need for vehicle data and applications thereof, to enabling technologies, challenges, and identified opportunities. Connectivity is increasing around the world and its expansion to vehicles is no exception. With improvements in connectivity, sensing, and computation, the future will see vehicles used as development platforms capable of generating rich data, acting based on inference, and effecting great change in transportation, the human-vehicle dynamic, the environment, and the economy. Connected vehicle technologies have already been used to improve fleet safety and efficiency, with emerging technologies additionally allowing data to be used to inform aspects of vehicle design, ownership, and use. While the demand for connected vehicles and its enabling technology has progressed significantly in recent years, there remain challenges to connected and collaborative vehicle application deployment before the full potential of connected cars may be realized. From extensibility and scalability to privacy and security, this paper informs the reader about key enabling technologies, opportunities, and challenges in the connected vehicle landscape.

Index Terms—Connected vehicles, telematics, automotive applications, vehicular ad-hoc networks, automotive electronics, controller area network, vehicle-to-vehicle, vehicle-to-infrastructure, vehicle-to-roadside, vehicle-to-everything, V2V, V2I, V2X, V2R, V2B, intelligent transportation systems, automotive engineering, road transportation.

I. MOTIVATION AND ORGANIZATION

AUTOMOTIVE computation and networking emerged out of necessity when the mechanical engine controls used through the 1970’s were unable to meet new, stringent emission regulations [1]. Early modules were used for local vehicle control, though prescient researchers envisioned controller networks as enabling collaborative problem solving [2].

Economies of scale and demand for increased efficiency and performance paved the way for complex electronic networks. Fragmentation became an issue, and in 1986 the Controller Area Network (CAN) was introduced as a specification taming complicated wiring harnesses [3], [4]. By 1988, the California Air Resources Board had standardized on-board diagnostics (OBD) [5] and 1996 saw the introduction of OBD-II, with provisions for enhanced data and real-time diagnostics [1], [6]. By 2008, CAN was adopted as the de facto standard for OBD-II [6], [7], encouraging manufacturers to repurpose CAN for proprietary communications [8].

Vehicle sensing and actuation has since proliferated, with modern cars incorporating hundreds of sensors and dozens of computers [9], [10]. These technologies facilitate local sensing, inference, and action – proximity sensors pre-tension seatbelts in the event of an imminent collision [11], accelerometers vary shock damping to improve comfort [12], and vehicles predict common destinations [13]. However, there exists significant opportunity in connecting vehicles to one another and with infrastructure. Connected vehicles may generate pothole maps [14], [15] or predict engine idles to eliminate short shutoffs [16]. Aggregate data will improve vehicle longevity by helping to optimally time maintenance [17]. The insight and control facilitated by extra-vehicular data sharing will improve transportation safety, efficiency, comfort, convenience, and reduce operating costs by allowing distributed sensing, remote computation, and action at scale. This document summarizes connected vehicle’s enabling technologies, application needs and opportunities and research directions. It provides a high-level survey, and the reader should find helpful references throughout to gain casual familiarity with the topics considered.

II. ENABLING TECHNOLOGIES

Today’s automotive market demands software and electronic innovations. As a result, modern vehicles possess complex networks capable of sensing, wide-area connectivity, inference, and action consisting of up to 70 electronic control units (ECUs) capturing 2500 signals from the chassis, powertrain, user interfaces, and safety networks [8], [18]. These underlying technologies enable connected vehicles and are facilitated by commoditization, decreasing power and cost requirements, and scalability. This section explores these foundational technologies, with particular emphasis on connectivity.

A. Sensing

Sensors translate physical attributes into signals to measure complex inputs [19], [20]. Sensing technologies facilitate local...
vehicle optimization and cooperative transportation, including low-cost commoditized microelectromechanical sensors [21]. Commonly-used sensors relating to safety and enhanced motion control include wheel speed sensors, yaw rate and acceleration sensors, steering and driver input sensors, and powertrain outputs such as current gear selection and engine speed [22].

Nascent sensing enhances driver perception using cameras, positioning systems, and ranging equipment to provide context information about a local vehicle and its environment [23]–[25]. Range sensors like RADAR enhance perception, with its long range of up to 150m and a viewing angle of 20 degrees, while LIDAR has a shorter range but increased angular precision to support autonomous navigation. Recently, camera systems have been implemented to support lane departure warnings and automated lane holding, though the rich data generated from these imaging sensors can be difficult to process in realtime [22]. These data comprise sharable, information-dense maps [26] and provide over-the-horizon awareness [27].

Of particular relevance to collaborative navigation and autonomy are Global Navigation Satellite Systems (GNSS) including Global Positioning Systems (GPS) [22], GLONASS, BeiDuo, and Galileo. These localization technologies identify vehicles’ relative and absolute positions within networks, and may incorporate corrective technologies. These systems may fuse data from inertial-type sensors to improve estimates and noise immunity [22], [28].

\[ \text{B. Intravehicle Connectivity} \]

Intravehicle networks share data among computing modules, sensors, and actuators to facilitate the operation of a single vehicle. Such networks underly OBD’s services, reducing service costs [6], [7], [29]. Local networks also support aftermarket telematics devices, which access data through OBD’s standardized interface.

Original Equipment Manufacturers (OEMs) today implement proprietary sensors and networks sharing OBD hardware. The resulting networks may follow OEM standards [30] and carry rich information useful for local vehicle optimization and supporting future connected applications.

Non-OBD intravehicle networks are purpose built. Some, like drive-by-wire systems are designed for robustness and security of critical data, while others host a deluge of peripheral data. Supporting network technologies including CAN, LIN, MOST, and FlexRay are well documented [8], [31]–[33] and provide fault tolerance, determinism, and flexibility. These differed protocols are often bridged by a gateway device metering the flow of information between internal and external devices and acting as a firewall or information aggregator [34], [35].

\[ \text{C. Intervenehicle Networks} \]

Data are valuable beyond the confines of a single vehicle, so intelligent transportation systems need connectivity that works at high velocities, long range, and with dynamic peers. Intervenehicle networks share data among vehicles and infrastructure, facilitating data collection and optimization at scale. Improvements in network scalability, routing efficiency, data security and quality of service have made wireless networking tenable, allowing for the use of mesh and top-down networking approaches for data’s movement from within vehicles to remote computing devices.

1) Mesh Networks: Vehicular mesh technologies support the needs of transportation data sharing, connecting vehicles and infrastructure in transient, ad-hoc neighborhoods. This section considers mesh networks’ enabling communication standards and data routing protocols.

a) Communication standards: Standardization is a critical enabler of connected vehicles. A leading standard supporting traffic safety and efficiency is Dedicated Short Range Communication (DSRC). In the United States, the Federal Communications Commission allocates 75 MHz of bandwidth from 5.850 to 5.925 GHz for DSRC’s vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication [36]. In Europe, DSRC operates from 5.855–5.905GHz [37].

In the US, IEEE 802.11P, an amendment to 802.11, addresses the Medium Access Control (MAC) portion of DSRC and supports Wireless Access in Vehicle Environments (WAVE). It describes a robust, high-throughput communication specification that may become a leading intelligent transportation system technology [38], [39]. 802.11P is derived from ASTM E2213-03, which defines support for V2I roadside communications and specifies the Media Access Control (MAC) and Physical (PHY) layers of 802.11 and 802.11a. This specification supports line of sight and distances up to 1 KM, with provisions for authentication and privacy preservation mechanisms [31]. IEEE 1609 is a higher-layer family of standards upon which 802.11P is based, supporting ubiquitous communication for different vehicle manufacturers and facilitating secure and efficient vehicle-to-everything (V2X) applications [31], [40].

Vehicle-to-vehicle messaging elements are also standardized. In the Application Layer, V2V and V2I message types are defined by SAE J2735 [41], which offers guidelines for deploying DSRC-enabled applications. This guide aids users in meeting performance targets while maintaining interoperability [42]. Other messages are defined in the Message Set for Advanced Traveler Information Systems (ATIS) [43], which is suggested for use when implementing WAVE [44].

In Europe, ETSI and CEN are the major standards development organizations. ETSI focuses on communications standards, while CEN standardizes applications. ETSI and CEN produce European Norm (EN) standards to avoid conflict with national standardization efforts. Europe’s parallel development to 802.11P, defining the MAC and PHY layers, is termed ITS-G5. Unlike the United States’ single allocated block for DSRC, ITS-G5 is subdivided into a 30MHz spectrum for safety applications (ITS-G5A) and a 20MHz spectrum for non-safety applications (ITS-G5B) [45]. Similar to SAE’s application- and message-specific standards, ETSI standardizes message types for cooperative transportation applications and environmental notification [46], [47]. Additional standards
are compared in Festag (2015) and Härri and kenney (2015) [45], [48].

b) Broadcast types: Connected messages have varied sensitivity to timeliness, data protection, and network range. It is imperative to choose an appropriate broadcast protocol to assure application performance. For example, the Wave Short Message Protocol (WSMP) enables the use of smaller packets for time-sensitive safety and convenience applications [38].

In car-to-car or car-to-infrastructure applications, messages may broadcast openly, or receivers may subscribe to specific topics. In this case, publishers push event data to a network without a target, and recipients accept select message types [49]. This approach allows nodes to anonymously publish or subscribe to data streams, protecting identities.

Message dissemination takes place a number of ways:

Beacon messaging transmits vehicle identification and context information in high frequency, short packets, or else sends data upon an event, e.g. if a hazard must be reported. Beacons are essential to cooperative awareness applications [34], but may suffer issues related to location accuracy, latency [50], and network saturation [36].

Flooded messages extend beaconing, retransmitting messages until a “time to live” elapses. Time to live may be measured as time since first transmission, absolute clock time, physical distance (radius), or hop count [34]. When nodes join or leave a network during transmission, flooded messages may behave unpredictably [50].

GeoCast messages rely on connected vehicles knowing their own location, and uses this information to create a “neighborhood” table for all connected vehicles and their locations, trajectories, and transmission reliability [34]. This allows routing based on position, so messages may address a particular vehicle. These messages use beaconing for neighbor discovery and location collection. This protocol relies on honesty in vehicles reporting their own location. Additional issues include handshaking overhead, which can lead to network congestion [51].

Regardless of the type of message, the flow of data is critical to ensure that messages reach their destinations in a timely manner. One tactic is the direction-aware broadcast. This ensures information flows in a particular direction, perhaps using a GeoCast message. An example use case is a vehicle reporting a road hazard to all following vehicles. This form of message propagation is shown in Figure 1.

Sometimes, message propagation must radiate outward. In this case, Intelligent Broadcast with Implicit Acknowledgment may be used. It begins with a periodic broadcast; once the message originator receives the same message from another node along the direction of message propagation, it infers that this other device has received the message and allows this new node to take responsibility for future retransmission [52], [53].

This approach limits network congestion, improving reliable transmission of safety-critical messages. Broadcasting with implicit acknowledgement is shown in the series of graphics in Figure 2.

There are proactive and reactive dissemination topologies. Proactive data flow maintains a routing table of all connected nodes and relies on periodic control messages to keep this table up to date. This approach requires overhead to ensure routing information is current, but messages may be transmitted without a searching delay. Reactive topologies flood control packets to identify optimal routing only when data are sent, imparting a delay in data transmission [53], [54].

2) Cellular Networks: Widespread Machine to Machine (M2M) connectivity has commoditized cellular bandwidth and hardware [55], lowering costs and allowing vehicles to connect to one another and to remote data sources and sinks directly and indirectly. Such direct connectivity uses a vehicle’s integrated modem to stream data to a remote server. Unlike mesh and short-range networking, direct car-to-Internet cellular connectivity is robust and capable of sharing data when traffic density is sparse [56], [57]. Direct cellular networking also facilitates parallelized data streams. In this manner, a vehicle may use a cellular connection for media consumption, freeing additional cellular or mesh networking technologies for use in safety-critical and time-sensitive applications. As 5G networking technologies roll out and costs fall, vehicles may indeed rely on direct
cellular technologies to facilitate connectivity for critical applications [58]–[60]. Such technology offers high bandwidth (1Gb/s) for vehicles, and use of femtocells allows this connection to be repurposed by vehicle occupants’ own devices [61].

In indirect cellular connectivity, Bluetooth, Wi-Fi, and other RF technologies interface in-car hardware with peripheral devices or visualization tools [62]. Indirect connectivity, common to OBD dongles, benefits from having no direct monetary cost, low local latency, and the security afforded by short-range connectivity. Once relayed from an intermediate device to a remote server, these benefits are negated. This approach also relies on people: improper configuration or forgetting to charge a gateway phone severs connectivity and causes data loss [56], [57].

Cellular connectivity may be considered a subset of V2I if the cloud is considered to be extra-vehicular support infrastructure. Other researchers suggest cellular connectivity be termed Vehicle to Broadband (V2B) [31]. Vehicle communication technologies are compared in [63].

3) Alternative Networking: Cellular and other technologies, such as hybridized vehicle-to-vehicle-to-infrastructure [64], [65] may be deployed to allow the benefits of each technology to be realized in an efficient and cost-effective manner. Hybrid approaches combine mesh, short-range, and cellular technologies. These rely on handoff mechanisms to switch communications methods [66], [67], or vehicles may communicate with one another directly with only the last node in a chain providing external connectivity [65], [68], e.g. cellularly.

LTE D2D is a promising technology for these hybrid approaches, allowing vertical handovers from DSRC networks to infrastructureless, ad-hoc networks of LTE-enabled vehicles to address network “dead ends” [69]. LTE D2D improves upon DSRC- and cellular-only approaches by facilitating high-throughput, low-latency, communications with minimum spectrum allocation and power consumption [70].

These approaches improve reliability and facilitate low-cost connectivity in rural areas and developing nations where cellular coverage and vehicle density are unpredictable and bandwidth costs are high. Several mesh and non-mesh approaches to networking appear in Figure 3.

D. Inference

Increased vehicular sensing and connectivity creates troves of useful data. A growth in in-vehicle and cloud computing power, as well as scalable data handling tools, has made gleaning critical insights from vast data sets tenable.

The decreasing price and power consumption of microcontrollers has led to the integration of powerful embedded systems in many vehicle components [18]. This computation allows for local application development and data aggregation and synthesis prior to dissemination.

An example of local vehicle analytics is on-line vehicle analysis, where failures are predicted and performance is monitored during use [29], [71]. Other in-vehicle applications apply data to identify and react to driver fatigue [31], [72]–[76], or utilize affective computing to minimize stressors and improve safety [29].

Recent data analysis extends beyond the local vehicle. Applications now are capable of applying computation to assess, learn from, and adjust vehicle operation, providing starting points for large-scale data informed applications. Improvements in cloud computing allow for scalable server-side processing and the offloading of in-car processing to remote locations.

Applications may today fuse data from multiple vehicles and external data sources. Example applications may use driver and vehicle data to deliver corrective nudges improving the safety and efficiency of vehicle use [77]. Other remote applications may collect vehicle data from third-party devices like mobile phones, using aggregate data to predict vehicle failures [17], [78]–[81] or to characterize driver behavior [82].

Local analytics demonstrate the value of data in controlling vehicle functions in realtime, while remote analytics show the potential to apply large-scale connectivity, computation and distributed information toward improving vehicle efficiency, reliability, and performance. As computing and connectivity technologies improve and reduce application latency, even aggregate data will become useful to realtime applications.

Data management and analysis is a key challenge that must be addressed to improve application latency and performance. Despite the cloud’s scalability and relative low cost of operation, increasing computing power does not address data management challenges. Traditional databases cannot
handle realtime requests, so technologies like Hadoop or other approaches to map reduction are used to distribute data storage and processing, helping to execute tasks in parallel [83]. This approach offers improved processing speed relative to conventional MySQL, and is shown in Figure 5.

Once stored properly, novel tools for data mining and visualization will allow the creation of analytics using collaboratively-generated automotive data.

E. Action and Feedback

Data-informed control is necessary to maximize the impact of intelligent transportation systems. This control may take a direct approach, e.g. using networked data to manipulate an actuator, or an indirect approach, e.g. by providing feedback to a human operator.

Direct vehicle control is facilitated by the proliferation of technologies that allow connected computers to modulate vehicle functionality directly, e.g. drive-by-wire whereas indirect control relies on in-vehicle or second-screen interfaces to provide occupant feedback. Examples of in-car visual feedback systems include examples using data to improve economy [84] and increasingly minimalistic "invisible" approaches to interaction [85]. Second-screen feedback systems, such as cellphone-based telematics systems with remote locking or maintenance timing apps, work to increase the radius of interactivity a driver has with their vehicle and allows applications to run on upgradable hardware.

III. APPLICATION LANDSCAPE

There is a latent consumer demand for vehicle connectivity [4]. Applications connecting vehicles, occupants, and infrastructure have been created to address this need. Telematics applications collect vehicle data locally to inform remotely-run algorithms, whereas V2V/V2I applications operate upon networks of vehicles and infrastructure devices.

A. Telematics Applications

Telematics applications blend telecommunications and informatics, allowing information to flow between vehicles and the world. Consumers frequently desire safety and security telematics applications, including automatic collision notification, roadside assistance, remote door unlocking, voice services, turn-by-turn direction, and hands-free phone use [4]. These applications may also enable location-based applications or the use of external data to improve local applications, e.g. aggregating diagnostic data to understand fleet-wide performance [4], [56], [86]. Another studying using aggregate data used taxi cabs as a distributed sensing system to collect information about traffic flow, aiding in travel time estimation and routing at minimal cost [87].

An advantage of telematics over "local" applications is that data are allowed to exit a vehicle, expanding the radius of customer engagement, e.g. texting a cell phone about an upcoming maintenance need. These convenience-adding telemetry applications using vehicle or peripheral device data also include pay-as-you-drive insurance [88] and vehicle miles traveled tracking [56], which charge drivers based on true vehicle use [89], along with range prediction [90], [91] and electric vehicle battery state of health monitoring [92].

Telematics applications may utilize a number of enabling communications technologies, including Vehicle to Broadband (V2B). In V2B, data may be transmitted from car directly to the cloud, allowing simplified data storage, aggregation, and processing [31]. This approach offers reliable, if relatively costly, connectivity [56]. A secondary advantage of cloud data storage is simplified interaction between vehicles and other networked devices and services – this approach allows the multiple use of data for varied applications, e.g. using a car’s location data to control a home’s heating, whereas non-cloud telematics applications commonly discard information after use.

Telematics applications exist today in consumer and research forms. OnStar and Tesla offer telemetry services, as do platforms like CarTel [14], which relies on data buffering and opportunistic WiFi, or the CloudThink platform, which uses direct cellular vehicle mirroring to create vehicle “avacars” for open application development [57]. Similar commercial management systems exist to improve fleet utilization and uptime while reducing operating costs [93]–[95].

B. V2V and V2I Applications

Another class of application connects nodes directly. These collaborative applications have found use in non-automotive industries – for example, swarm robots operate based on consensus [96] and mobile phones distribute data generation [97]. In vehicles, vehicle-to-vehicle and vehicle-to-infrastructure applications tend to employ DSRC and Vehicular Ad-Hoc Networks (VANETs) to provide rich, real-time fleet data to improve safety, efficiency, reliability, comfort, and convenience that would be infeasible to collect within the confines of a single vehicle. Many collaborative automotive applications revolve around information gathering and dissemination [27] to facilitate the knowledge-based safety improvements driving connected vehicle adoption [98].

Several projects, consortia, research institutions, and government initiatives work to enable V2V and V2I
Fig. 6. Connected vehicle applications may be categorized in several different ways. Here, we show information services, safety services, individual motion control, and group motion control. Applications may fit into several of these categories.

applications [31], [34], [53]. Research examples include FleetNet’s car-based, real-world field trials of position-based forwarding, MineFleet, with its aftermarket on-board unit for performance analytics, and CarTALK 2000, which explored communication-based adaptive-cruise-control and the economic and congestion impacts of assistive connectivity [99]. This section explores typical applications of V2V and V2I data.

V2V and V2I’s application space is broad, ranging from vehicle tracking for cooperative safety systems [100] to enhanced driver information [101]. Additional categories have been proposed, from road security, fleet management [102], navigation, tolling [103], and multimedia sharing [104] to parking location and payment [105], [106].

Applications may be divided into four categories based on a modified Wilke’s taxonomy [107]. These categories are “Information Services,” “Safety Services,” “Individual Motion Control,” and “Group Motion Control,” though applications may transcend these categories. A broader classification schema instead may consider “Operation Critical” versus “Non-Critical.” Example applications in these categories are shown in Figure 6. Other taxonomies include those from the Connected Vehicle Reference Implementation Architecture (CVRIA) [108] and ETSI [46], [47] standards.

1) Information Services: Information Service applications generally tolerate transmission delays and errors. These applications include remote vehicle dashboards or diagnostic services [109]. Many of these services enhance users’ comfort and ability to perform other tasks while driving, or allow viewing vehicle parameters remotely [34].

a) Fault prediction and response: Fault prediction and response services include diagnostics, prognostics, and driver-aware technologies. Example applications may extend OBD to predict remaining component life based on physical or machine-learned models and trends from aggregated vehicle data [17], [57], [110].

b) Data collection and generation: Information Generation applications create data useful for shared decision making, digital mapping, collision avoidance, and path prediction to allow driving behavior optimization [111].

Map generation [112] is critical to autonomous vehicle development. Instrumented vehicles connect and synthesize localization data to improve positioning accuracy while simultaneously generating maps [113]–[115], or otherwise apply cooperative sensing to facilitate decision making [116].

Beyond mapping, V2V applications may use sensing to enhance perception. Cameras, LIDAR, and RADAR identify and report hidden obstacles to the vehicle and/or driver [117]. Other sensors may improve knowledge of a nearby vehicle’s trajectory to improve road safety and efficiency. For example, in vehicles today, a driver has no indication of the throttle pedal position of another vehicle. V2V can generate a map possessing information about congestion, road hazards, or throttle pedal positions in real-time, enriching driver perception [26], [27].

These applications are challenging by semantics. Vehicles must be able to communicate with one another and understand the contents of each message. Translation platforms like Open-VSeSeMe partially address this problem using a common reporting language across several applications [118, pg. 186].

c) Data dissemination and distribution: Data dissemination applications share information between vehicles. This content may include media generated by other vehicles, such as a video stream from a windshield-mounted camera [101]; other content may be generated from non-vehicle sources.

Proposed applications use communication for driver entertainment and to improve productivity by allowing occupants to access external content [38]. Transient car-to-car chat systems have been envisioned [119], as have systems for music sharing, interactive gaming, and Internet sharing [53].

d) Efficiency improvement: Efficiency Improvement applications share information to optimize fleet-wide fuel efficiency and minimize congestion. Example applications estimate the density of traffic and build data-informed models to improve traffic flow and reduce congestion [38], [120], [121]. These models may be used to optimally route a vehicle to a destination while maintaining smooth traffic flow and reducing drive time [53], [122]. In simulations with human-piloted and autonomous operation, connected intersection models substantially increase an intersection’s vehicle processing rate and average vehicle velocity while reducing the percentage of time a vehicle is stopped [123].

e) Convenience services: Convenience services improve occupant comfort and convenience by freeing their cognitive facilities and hands to focus on driving or other productive tasks. Services include automated traffic routing, tolling, and vehicle tracking [26]. Some applications may help occupants even after they leave their vehicles, e.g. in the case of drivers using data to locate a parked vehicle [124]. Increasingly, convenience services are beginning to include actuation, e.g. tuning radios or moving seats automatically based on driver preferences. In this vein, platooning or automated braking may also be viewed as convenience services capable of freeing drivers for other tasks.
2) Safety Services: In the U.S., the National Transportation Safety Board reports 16,000 vehicle crashes daily, often caused by driver distraction or fatigue [31]. Advanced Driver Assistance Systems (ADAS) allow vehicles to use local data to minimize the risk of incidents [34] through lane keeping and adaptive cruise control [31].

Safety Services use connectivity and ADAS to enhance human perception and mitigate the risk of hazards [38]. In Safety Services, delayed or corrupted data transmission may lead to harm. These applications, including automatic braking and extended-horizon hazard reporting, address a critical need [125]. Wide-area connectivity allows rich, wide-area context sharing to facilitate the “zero accident” vehicle capable of sensing hazards at a distance and executing appropriate avoidance maneuvers [29].

a) Collision avoidance: Human reaction time contributes to many collisions. Though visual cues like brake lights indicate potential hazards, slow response leads to incidents. Long-distance data sharing increases the reaction window to reduce the risk of a rear-end accident [52]. When packets are sized properly, latency is kept low, and broadcast power is sufficiently high, this type of collision avoidance is highly effective – up to a 99% incident reduction in some simulations [126]. Similar systems may eliminate blindspots and enhance intersection navigation by creating 360 degree situational awareness for the vehicle itself [127].

Rather than full autonomy, V2V may augment human operation to improve efficiency or safety. An example application intelligently adapts vehicle speed using signals from infrastructure devices to ensure that vehicles operate at traffic- and environment-appropriate speeds. A 2000 Swedish study with more than 5,000 vehicles tested this form of augmented control and identified the potential to reduce road injuries by 20% without increasing travel time, and suggested the potential for positive influence on surrounding traffic [24], [128].

b) Hazard reporting: Enhanced perception and long-range reporting allow lesser-equipped vehicles to take advantage of safety-improving ADAS systems present in nearby vehicles, enhancing fleet-wide safety. Connected vehicles may sense and report hazardous road conditions to other vehicles [31]. Weather-related hazards may be mapped and shared, while disabled vehicles may notify others of their location to protect themselves when line of sight visibility is not possible [31], [118, p. 88].

c) Driver monitoring: Driver monitoring applications use connected data to monitor drivers within vehicles. Applications monitoring driver impairment have been adopted in heavy-duty vehicle fleets to reduce accident frequency, with implementation costs falling and increasing deployability in the passenger fleet [24]. This technology reduces the likelihood of drowsy or impaired driving by notifying drivers, fleet managers, or nearby vehicles directly.

3) Individual Motion Control: Individual motion control applications apply connectivity to issue warnings to a vehicle operator or to directly control a single vehicle’s actuators. These applications may improve safety, as in the case of collision avoidance, or efficiency, such as automated drafting. Still other applications may aim to ameliorate traffic, as with assisted lane switching [129] and dynamic routing.

4) Group Motion Control: Group motion control uses vehicle sensors and external data to influence or control the behavior of vehicles and drivers in aggregate [107]. The purpose is to maximize a series of objective functions, e.g. to save maximal fuel or to reduce transit time. These applications offer the opportunity for increased impact relative to individual vehicle control due to scale, and will become more feasible with increasing vehicle autonomy and broader connectivity.

a) Platooning: Platooning dynamically chains vehicles to maximize fuel efficiency and was tested as early as 2000 [130]. Platooning involves vehicles traversing a route in a “pack” so as to benefit economy, safety, or comfort. Enhanced car to car reporting has enabled platooning with intelligent stop and go, high-speed merging, and obstacle avoidance. These applications maximize the use of available roadway space to increase packing density, ease congestion, and improve traffic flow [130]. Other types of platooning may facilitate a common task, such as clearing a road surface from snow [131].

Platooning’s fuel savings depend on a number of factors including velocity, follow distance, vehicle aerodynamics, number of linked vehicles, and environmental conditions. Despite this variability, fuel consumption and emissions reductions tend to be significant. A 2000 study conducted on a test track yielded a 21% fuel economy reduction for the second truck in a convoy driving at 80kph with an electronically-controlled spacing of 10m [132]. A 2011 three-truck platoon conducted at 85kph with constant 6m spacing resulted in an average fuel savings of 4.3% for the first truck, 10% for the second truck, and 13-14.5% for the third truck. This test may have underestimated the potential savings, due to the need for a 30cm lateral offset on the middle truck to facilitate radio propagation as well as operating at increased elevation over sea level [132].

b) Intersection control: Connected data may be used to control vehicles and intersection signals to maximize vehicle throughput and efficiency. Simulations show that automatic maneuvering results in reduced traffic, CO2 emissions, and fuel consumption. One study shows a 99% stop delay reduction, 33% total travel time reduction and 44% reduction in CO2 emission and fuel consumption relative to business as usual [133]. The flow of information is not only vehicle to vehicle – SafeCop’s V2I solution for signaling may eliminate 80% of vehicle collisions, reduce fuel consumption by 17%, and cut time spent in traffic by 10% [134].

IV. APPLICATION AND TECHNOLOGY CHALLENGES

While there is latent demand for connected applications, vehicle architectures, privacy and security, and other challenges inhibit deployment today. This section considers common connectivity pitfalls and barriers to application deployment.

A. Connectivity Considerations

The wide-area connectivity used in connected vehicle possesses challenges in assuring data reliability, node density,
bandwidth constraints, and latency targets. Developers must balance an application’s need for latency, bandwidth, and data density against technological limits and privacy and security policy issues.

1) Network Performance: Connected vehicle applications may require realtime data. For time critical applications such as safety applications or vehicle control, messaging must be low latency and delay bounded [135], [136]. Stale data must be pruned to avoid having a net-negative effect on the system [137].

For some applications, today’s network performance may be sufficiently enabling [138]. However, applications requiring richer information may struggle to provide information as rapidly and reliably as necessary due to heightened computation time, network saturation, and packet loss. Beyond network type, deciding what data to collect, process, and transmit may cause delays [31] and drive sensor resource use [139]. Emergency event detection, for example, must be fast (milliseconds) and certain (no false positives) in the face of rapidly changing road dynamics to minimize the latency between event detection and receipt at a remote vehicle.

High vehicle density is required to enable many connected applications [140], making rural deployment challenging. It is not enough to have multiple vehicles physically proximate – a significant percent must be communication-enabled [127]. Realtime changing network conditions further challenge decentralized information flow, requiring the use of complex network architectures utilizing up-to-date routing information.

Realtime data is necessary but not sufficient. Data transmission quality impacts application performance. Many issues in DSRC applications stem from low redundancy and non-receipt-acknowledgement of messages [141]. These applications require improved Quality of Service (QoS) to ensure that data arrive to a recipient rapidly and without corruption [142]. Retransmission can solve these issues at the cost of network congestion.

Network congestion must be managed, as significant loading can result in interference with the transmission and receipt of data, impacting safety applications. Other factors such as data richness impact network loading, with Hitachi estimating an hourly data generation rate of 25GB per vehicle [143]. Wireless networks have limited capacity, and interference, congestion, self-competition, and redundant transmission affect performance of applications by causing data collisions, transmission delays, and packet loss [52], [144]. Increasing available bandwidth is not feasible as only a small portion of wireless spectrum is licensed to connected vehicles [36].

Several message dissemination techniques addressing congestion have been explored through simulation [145]. Other techniques, like Data Proxies [146], apply models to minimize initial data transmission thereby reducing network loading, bandwidth, and storage needs. Alternatively, data transmission may be split by technology, with DSRC reserved for low-latency applications and cellular service used for media applications. Such approaches optimize costs, network loading, and performance for various application suites.

2) Security, Privacy and Authentication: Networks with rich data and actuation are attractive hacking targets. Without appropriate protection, vehicle control may be commandeered, data stolen, and user privacy compromised [147]–[149]. These challenges are inherent to OEM and aftermarket connectivity systems [150]–[156]. The danger of compromise is especially significant in automated vehicles with electronic actuators. Petit (2015) examines common attack surfaces and risks in such vehicles [149].

Implicitly-trusting architectures allow attackers to inject false information to divert traffic or cause emergency braking [139]. Without authentication, malicious agents cannot easily be blocked, making denial of service (DoS) and jamming a threat [98], [148], [157]. Often, authentication systems are at odds with privacy. V2X technologies inherently rely upon sharing private data: beacon messages, for example, share location histories that can expose a driver’s identity [158], [159].

Many network security approaches compromise vehicle safety by worsening network and computation delays [160]. Low overhead is required for authentication, anonymity, and certificate validation and revocation to ensure data are accurate and that malicious senders are removed from a network or ignored [161], [162].

3) Protocol Design: Designing scalable and extensible standards for V2V and V2I networking, data structuring and sharing, and neighborhood management is a challenge. While traffic improves data density, it simultaneously challenges network stability due to transmission collisions and other errors [163]. Existing protocols like TCP are also of no use – these were often designed for wired networks and therefore have high overhead, or were designed for systems where data transmission had no cost.

Further, maintaining routing and connectivity is a challenge in vehicular networks. Vehicles move rapidly, resulting in node transience and evolving vehicle neighborhoods as vehicles join and leave an area. Changes in node density can overwhelm networks, while data retransmission due to broken links causes delays and reduces network capacity [164].

In common routing protocols, proactive schemes require significant overhead [54] while “greedy” algorithms for location-based packet forwarding lead to gaps and data retransmission. Vehicle-to-vehicle networks suffer data loss stemming from quick update rates which worsens when moving. Some of these issues are partially addressed with different network topologies [130]. Challenges in routing efficiency and maintaining connectivity vary across application categories, further complicating matters [53]. Further testing to ensure performance across varied regions is necessary [165].

B. Vehicle Design

In-vehicle network electronic platforms must evolve to support connectivity-enhanced applications. The widespread adoption of ADAS provides richer data for connected applications, but systems today are not designed with the development of these and future applications in mind.

1) Security: Security and car hacking is a challenge [150]–[156]. Vehicle software complexity has
increased dramatically [18] and hurried development has led to vulnerable hardware and firmware. A review of security challenges is discussed in Parkinson et al. (2017) [148].

A common approach to harden vehicles is to shield critical actuators and memory from external networks, using gateway devices and software to meter inter-network data flow [34], [139] or anomaly detection software [31]. Authenticated gateways can prevent commodity OBD and telematics devices from exploiting physical access, limiting masquerading, impersonation, flooding, jamming, spamming, and the spread of malware [118, p. 61–62]. SAE formalizes a secure design and testing approach in their guidebook [166], and methods of applying cognition to vehicle supervision and protection have been proposed [146].

In the context of wide-area network security, ETSI has drafted number of technical specifications for secure communications architecture and security management, including public key infrastructure [167]. Standardized security implementations will speed time-to-market and help ensure the safety and scalability of connected transportation applications.

2) Sensing: Though vehicle sensing has proliferated, modules are often black-boxed and mask raw data. For example, vehicles with Adaptive Cruise Control incorporate RADAR, but the authors have determined that sensors often filter data and share only “clear/not clear” messages with the broader network, limiting future utility.

Beyond availability, accuracy also matters. Safety services and even direction-aware message broadcasts rely on location data. However, satellite localization technologies suffer from imprecision and signal loss [26], leading applications to perform poorly [168]. Inertial navigation is now a mainstream technology improving GNSS precision, though common MEMS sensors drift over time. Referencing data-rich internal digital maps for location correction addresses this drift problem, while lane-level maps may present a space- and computation-efficient solution for map-relative localization [169].

3) Communications: Modern vehicles have cellular, mesh networking, WiFi and even Bluetooth radios to interface at scale [170]. Each technology has different range, latency, bandwidth, security and cost constraints, along with different market penetration. Additionally, radio systems have differing robustness to motion, line-of-sight obstruction, and antenna design. Not all technologies will be suited for every application.

Architecture design impacts radio feasibility, with thick and thin clients varying the location of data processing and transmission. In a thin client, data are transmitted exactly as captured, while a thick client transmits preprocessed information. These architectures strike different balances of power, computation, data accuracy and feasibility [56].

Similar network architecture choices must be made for intravehicle networks. CAN, LIN, MOST, FlexRay, Ethernet and other technologies have different cabling requirements, noise immunity, maximum throughput, provisions for encryption, and more. Wireless technologies can also be used for in-vehicle data sharing, though these can present new attack surfaces. As vehicle data generation increases, network capacity will become an increasing concern. Emerging technologies such as Ethernet, which improves network resilience, timing, supported node count and speed, are discussed in Neumann et al. (2017) [171].

C. Data Handling

Data accuracy and reliability are critical to connected applications, with inaccurate data negating the benefits of connectivity. For example, given a false positive of a hazard ahead, a connected car might unnecessarily stop and be rear-ended by a human operated vehicle. Networks and applications must be able to eliminate false positives and duplicate warning messages. Additionally, temporal data are important. For example, a stalled vehicle poses a hazard when it is stopped and on a main road, but even after the car has been removed the related congestion may remain [172].

Even with accurate data, processing is a challenge. With growing data, the need arises for specialized handling tools. Data sets are already too large to transmit in a cost-effective manner, though emerging technologies like 5G may make full vehicle mirroring possible [173].

Storing raw data is feasible but costly, and map reduction techniques are necessary to allow scalable data analysis and pattern identification [83]. Preprocessing metrics may make data useful in a compact form, reducing the size of digitally duplicating vehicles [56]. These metrics may be aggregated from multiple vehicles to minimize storage and analysis cost and complexity.

Determining where and how to process data for maximum efficacy is a significant research area, with particular applicability when considering large fleets [31], [174].

D. Infrastructure

Building V2I infrastructure is time-intensive and costly. The existing environment must be considered when placing radios [175], [176] and computation. There are costs associated with the deployment of nodes, leasing of space, support requirements such as power and bandwidth, and maintenance. Infrastructure must also be designed for long lifespan, requiring the use of extensible standards.

Connected vehicle applications require high density of radio-equipped vehicles, in particular for safety-critical services. Applications like collaborative collision warning work best when more than 60% of the local fleet is connected [26]. Though V2V and V2I lack significant penetration today, rapid growth is expected – with studies anticipating between 40% and 62% automotive market penetration no later than 2030 [177], [178].

V. OPPORTUNITIES FOR TECHNOLOGY IMPROVEMENT

This section identifies areas of improvement for future connected vehicle development based on a review of contemporary literature.

A. Communications Design

Mesh networking challenges must be addressed, including data minimization and packet collision reduction.
Context-aware routing will reduce network overhead, e.g. by sending hazard alerts to vehicles only on the same side of a divided highway. Trajectory-based routing will help with optimal path mapping (identifying the shortest path or lowest cost network vector) based on distance, latency, hops, vehicle speed, or another cost function cost, [179]. Latency, redundancy, and reliability can be improved by choosing which vehicles are best chosen to rebroadcast [180]. Eventually, these and other protocols such as the DSRC radio standards used in North America, Europe, and Japan must also be harmonized to simplify the production and support of extravehicular networks and to reduce redundant engineering efforts [118, p. 76].

Soon, hybridized networks will combine radio technologies to enable low-latency and long range connectivity through parallel use and handoff mechanisms [66]. Other approaches may use computation to emulate “impulse” reactions into the car, so time-sensitive applications can operate in spite of high-latency connectivity [181]. Eventually, 5G cellular will support high bandwidth, long range and low latency direct-to-cloud connectivity, minimizing requisite vehicle density and improving the richness of data stored in vehicle mirrors [182]. Feasibility assessments show promise using this technology in the context of real-time vehicle teleoperation or for safety services today using V2V technology [60], [182], [183].

While development timelines have been presented, adoption remains unpredictable. The best course of action for developers is to remain technology agnostic and focus on standardizing semantics. Though sharing data across vehicle models is a challenge [118, pg. 186], semantics facilitate communication of map, hazard and other real-time data by creating a unified vocabulary improving interoperability of data across platforms and applications. Cooperative ITS messages [184] are an example.

B. Platform and Application Design

Platform design shapes vehicle connectivity. The creation of a scalable, extensible, and hardware-agnostic platform will facilitate the best possible collaborative vehicle applications [57]. The cloud provides the ideal backdrop, with an infinitely scalable architecture supporting rich data mirroring, interoperability, and application development.

Openness improves data sharing transparency and system security, while interoperability reduces the risk of vendor lock and ensures that connected applications have access to data from the largest possible fleet. Extensible platforms allow applications to integrate other devices, services, and people into the connected vehicle experience. For example, data and displays from mobile devices may be incorporated into connected vehicle applications in order to augment the sensor and communication payload of a vehicle. Platforms themselves may have intelligence, and be used to minimize the data inputs needed to mirror a vehicle remotely [146].

With an open, interoperable and intelligent platform, applications will have an increasing number of data inputs and actuators available for use. “Application locality” must be considered during design to ensure applications take advantage of the appropriate resources in-car or remotely [146], [174]. Design choices like sensor input, communication method, location of computation, and more determine the implementation cost and performance of an application. In an example pathfinding application, use of V2V versus V2I produced similar results with 10% of the network-wide bandwidth use [185].

C. Vehicle Design

Vehicle improvements are driven by consumer willingness to pay. Bundling comfort and convenience features with safety and efficiency features incentivizes vehicle owners to pay for connectivity. For example, traffic mitigation applications may motivate drivers to purchase radio hardware enabling of collision avoidance.

An issue with today’s vehicle sensors is that they lack accuracy data. Extending sensor data with metadata including the time of acquisition or a unique sensor ID, will allow applications to make appropriate inferences based on data freshness and trustworthiness. Still other sensors are black-boxed and should be “unlocked” to provide richer data supporting increasingly-complex applications over their service lives. Over-the-air software updates [186] will allow current hardware to support these future applications.

Vehicle networks and diagnostics must also be updated to better support application development. “OBD III” should include extensibility improvements such as standardized advertising conventions for new sensors and protected access to memory and actuators. An open, centralized web repository for diagnostic trouble codes and sensor parameters, as well as a common sensor gateway architecture, would enhance On-Board Diagnostic’s viability as a data source for connected and collaborative automotive applications. Finally, inclusion of improved freeze-frame data storage, multi-PID (Parameter IDentifier) request handling, and automatic responses sent at regular intervals would unlock potential for enhanced diagnostic and prognostic applications.

D. Network Security

Connecting vehicles to the outside world exposes a host of vulnerabilities. Security must be improved during data generation, in-vehicle use, transmission, and at remote servers. In-car computers must assure that sensor data are accurate and authentic from acquisition to transmission. In use and transmission, data must be encrypted to protect against interception. At a server, credentials must be validated and information anonymized. This must occur with low computational, bandwidth, and energetic overhead. Recent work has shown that efficient and secure credentialing and key management techniques are feasible for use in vehicular networks [187].

In vehicle security focuses on preventing unauthorized actuator or sensor access. Intelligence in the vehicle gateway may authenticate incoming requests and limit data access by parameter, requestor, and more. This may be accompanied by an improved seed/key challenge and response system, or authentication making use of “trusted components” for verification [53].
Increased connectivity can bring about safety concerns, as incorrect or malicious messages from RoadSide Units (RSUs) or other vehicles may cause system failures. For example, a malfunctioning RSU could direct an autonomous vehicle the wrong direction down a one way street, or indicate that a traffic light is red when it is green, causing congestion. For this reason, message integrity data are necessary, as well as vehicle intelligence capable of simulating the impact of incoming commands and ensuring their intent and outcome is benign [146].

A V2X link may employ traditional communications security strategies, such as public key infrastructure, revocable certificates, encryption, or distance-bounding to make data interception a challenge. If one assumes that the majority of devices on the network are honest, these devices may be used in aggregate to police the network for systems behaving badly, sharing this information with a central server capable of revoking certificates [53]. This mode of transmission may also be structured to ensure that the intended recipient is the only node with the key to decode the message, locking intermediate nodes out [98].

These certificates and signatures may use pseudonyms to complicate deanonymization [53], [188], though computation presents a challenge [148]. Where feasible, time-limited pseudonyms, conditional anonymity, distributed resolution authority, and passive pseudonym revocation minimize the risk of adversaries extracting identifying information from a network or pushing malicious data [159], though for some applications network scale alone may provide a means of hiding within a crowd [189].

Recent research leverages security advances to balance trust, privacy, and security. One study developed privacy-preserving systems capable of authentication, message integrity verification, data nonrepudiation, and application-specific confidentiality, preventing malicious entities from modifying, discarding, or delaying messages [190]. Another approach uses message linkable group signatures to reveal attackers who double-sign messages, helping placate drivers concerned about their privacy [191].

Cloud security must be improved as well. One way to improve the security of the car-to-Cloud link is to abstract the digital duplicate of a vehicle from the physical vehicle, allowing applications to interact only with the digital duplicate. Using “Data Proxies” will allow for abstraction while shifting data handling to the cloud, where computation is abundant and certificate and credential validation is feasible. An intelligent “Cognitive Firewall” may relay approved commands to the vehicle after checking to ensure these commands will not violate predefined or learned rules [146]. Combined with clear guidelines for appropriate data use, protection, and anonymization [57], it will be possible to build a secure remote repository for vehicle data service.

E. Social Issues and Public Perception

Connected system cost remains a challenge, both for initial hardware cost and for ongoing bandwidth costs as required by cellular communication [192]. There are additional costs associated with deploying wireless technology, enhanced sensing, and computation in vehicles. While sensors like RADAR provide valuable information, many are prohibitively expensive at the volumes sold today [26].

With some radios, high bandwidth prices may drive consumers away. These costs may be offset by added comfort and convenience features, though it is unclear how significant this effect will be. Recent surveys have applied adaptive choice-based conjoint analysis to show the relative rankings of connected vehicle technology importance and found contradictory results for consumer willingness to pay, with consumers demonstrating a preference for safety-centric features [193].

Despite the high price of connectivity, a 2012 University of Michigan study shows a positive perception of connected vehicles, with drivers having significant faith that connectivity and enabled autonomous technology will improve safety, reduce congestion, and reduce emissions. People perceive connected and automated cars as leading to fewer crashes, reduced severity of crashes, improved emergency response, lower congestion, shorter transit time, reduced emissions, better fuel economy, reduced insurance costs, and reduced distraction [194]. In this survey, 66.4% of respondents indicated that they would like autonomous driving and related safety services in their vehicles. In the U.S., approximately 25% of drivers surveyed stated that they would pay up to $2,000 to have access to this technology [194].

Cost aside, drivers increasingly demand data security and privacy [57], [194] and have concerns about location and speed tracking, technology over-reliance, the coexistence of connected and non-connected vehicles, and risks of system failures [194]. These issues may be perceived to be more significant than the reality; clear messaging and policy definition are needed to minimize perception-related hurdles.

Though consumers benefit from vehicle connectivity, many are unwilling to share personal information. Anonymization may improve consumer acceptance, though many applications require invariant identifying information to function. In these instances, pseudonyms can help improve privacy and security while allowing applications to access requisite data histories [159].

While some drivers willingly share data in exchange for financial or other incentives [195], ownership remains an issue and telematics providers claiming to own vehicle data face scrutiny [196]. To that end, connected vehicles must provide clear guidelines for opt-in data ownership and privacy controls [57], as well as systems to visualize and modulate information flow [197]. Such systems will accelerate adoption of connected vehicles. A four billion dollar government program to subsidize vehicular autonomy and connectivity development announced in January 2016 [198] will further drive growth and acceptance of V2V and V2I technologies by reducing consumer cost burden, as will the shift toward a vehicle-sharing economy.

VI. Conclusions

The Connected Car industry is growing rapidly, but it is by no means new. The primary enabling technologies are
in place today – computation, sensing, and networking – as is a consumer demand for the sort of applications only connectivity can facilitate. Many applications witnessed to date – like platooning, collision avoidance, and hazard warning – are research focused but demonstrate technology with far-reaching practical implications. There additionally exists an untapped opportunity in collaborative, consumer-facing applications with a focus on optimizing user experience, human factors, and improving the total cost of ownership of a vehicle through fuel, time, and maintenance savings.

Reviewing the connected vehicle landscape shows great growth and future potential, though growth to date has been occasionally haphazard and fragmented. We identified several challenges in the form of privacy, security, scalability, and extensibility concerns.

To further the growth of Connected Car applications, we propose a set of recommendations for next-generation automotive technologies. These range from On-Board Diagnostics to in-vehicle and full system network architectures. With improved technologies and consumer messaging, we will accelerate the growth of the connected vehicle industry and start to more completely realize the transformative potential connectivity has on mobility.

ACKNOWLEDGMENTS

We thank MIT’s Dajiang Suo for generating these graphics and are appreciative of Jaguar Land Rover for funding our research.

REFERENCES


SIEGEL et al.


