Solutions for Vehicular Communications: a Review

Vasco N. G. J. Soares\textsuperscript{1,2}, João N. Isento\textsuperscript{1}, João A. Dias\textsuperscript{1}, Bruno M. Silva\textsuperscript{1}, and Joel J. P. C. Rodrigues\textsuperscript{1}

\textsuperscript{1}Instituto de Telecomunicações, University of Beira Interior, Covilhã, Portugal
\textsuperscript{2}Superior School of Technology, Polytechnic Institute of Castelo Branco, Portugal

vasco.g.soares@ieee.org, joao.isento@it.ubi.pt, joao.dias@it.ubi.pt, bruno.silva@it.ubi.pt, joeljr@ieee.org

Abstract— Vehicular networks experience a number of unique challenges due to the high mobility of vehicles and highly dynamic network topology, short contact durations, disruption intermittent connectivity, significant loss rates, node density, and frequent network fragmentation. All these issues have a profound impact on routing strategies in these networks. This paper gives an insight about available solutions on related literature for vehicular communications. It overviews and compares the most relevant approaches for data communication in these networks, discussing their influence on routing strategies. It intends to stimulate research and contribute to further advances in this rapidly evolving area where many key open issues that still remain to be addressed are identified.

Index Terms— Vehicular Networks; Vehicular Ad Hoc Networks; Delay-Tolerant Networks; Vehicular Delay-Tolerant Networks.

I. INTRODUCTION

A vehicular network can be defined as a spontaneous self-organized network, where vehicles, equipped with short-range wireless communication capabilities, cooperate with each other to enable communications with other vehicles or roadside infrastructure equipment. In these networks, nodes can be located in line of sight or out of the radio range if a multi-hop network is built among several vehicles.

Vehicular networking has attracted a growing interest by the research community and industry due to its potential for application to a wide range of real-world scenarios. These networks are regarded as a key technology for improving road safety, optimizing the traffic flow and road capacity. They can also be used as monitoring networks for sensor data collection. Several commercial applications (e.g., commercial advertisements and parking space availability) and entertainment applications (e.g., Internet access and multimedia content sharing) have been envisioned. Vehicular networks can also be employed to provide connectivity to remote rural communities and regions, or to assist communication between rescue teams and other emergency services in catastrophe hit areas lacking a conventional communication infrastructure.

In a vehicular network all nodes act as information transmitter and receiver, participating in the routing and data forwarding process. Routing is a very challenging task due to the unique characteristics of this kind of networks. Most of the problems arise from the mobility and velocity of vehicles, which is responsible for a highly dynamic network topology and for short contact durations. Limited transmission ranges, radio obstacles due to physical factors (e.g., buildings, tunnels), and interferences (i.e., high congestion channels caused by high density of nodes), lead to disruption, intermittent connectivity, and significant loss rates. All these conditions turn vehicular networks object of frequent fragmentation/partition. Its node density, which is affected by location and time, can be highly variable. For example, a vehicular network can be categorized as being dense in a traffic jam, whereas in suburban traffic it can be sparse. In fact, in rural areas, the network can be extremely sparse. Vehicular networks have potential to a large-scale growing.

Among various available approaches that have been proposed in the literature for vehicular communications, vehicular ad hoc networks (VANETs), delay-tolerant VANETs, and vehicular delay-tolerant networks (VDTNs) are the most frequently studied strategies. This paper describes these solutions in detail, comparing their approaches towards the realization of effective vehicular communications.

The remainder of this paper is organized as follows. Section II describes the VANET approach towards vehicular communications. The problems caused by frequent network disconnection, partitions, or long delays, have motivated the introduction of delay tolerant networking concepts to vehicular networks. These concepts, which are presented in Section III, have led to the study of delay-tolerant VANETs. Section IV overviews the VDTN layered network architecture proposal that aims to provide innovative solutions for challenged vehicular communications. Finally, Section V summarizes the main conclusions of this work.

II. VEHICULAR AD HOC NETWORKS

Vehicular ad hoc networks (VANETs) \cite{1, 2} were proposed as a special type of mobile ad hoc network (MANET) \cite{3} with the distinguishing property that mobile nodes are vehicles, such as cars, trucks, buses and motorcycles. This implies that mobile nodes movement is restricted to roads with constraints of traffic flow and traffic regulations.

Vehicle communication in a VANET can be classified as either vehicle to vehicle (V2V) or vehicle to roadside infrastructure (V2I). Roadside infrastructure units (RSUs) are static nodes deployed along the road, which are used to
improve connectivity and service provision. It is possible for roadside units to be connected to a core network and to the Internet. These concepts are illustrated in Figure 1.

![Illustration of a vehicular ad hoc network (VANET).](image)

Several approaches and architectures have been considered to implement the communication links among vehicles [4]. Examples include a pure V2V ad-hoc network, a wired backbone with wireless last-hop, or a hybrid architecture combining both.

Traditional routing protocols proposed for MANETs aim to establish end-to-end paths between network nodes [5]. Chennikara-Varghese et al. [6], Li and Wang [7], and Lee et al. [8] state that these protocols can not be directly applied to VANETs due to their difficulties in dealing with rapid topology changes and frequent fragmentation. Therefore, these routing protocols must be adapted to suit VANETs’ unique characteristics, or new protocols must be designed for VANETs. This has been a topic of interest for many researchers over the years and has resulted in a large number of routing protocol proposals. The interested reader may refer to [6-9] for detailed theoretical background and surveys of these protocols.

It is important to recall that different VANET applications have distinct requirements. A single routing protocol is not capable to efficiently handle all the inherent characteristics of the multiplicity of the above presented applications, as they may use unicast, broadcast, or multicast transmission facilities. Hence, several attempts have been made to develop routing protocols specifically designed for particular applications.

This observation was used by Lin et al. [9] to classify recent VANET routing protocols according to a taxonomy that considers three categories: unicast, multicast/geocast, and broadcast. Unicast routing constructs a source-to-destination path. Multicast routing is used to deliver data from one source to many interested recipients. Geocast routing is used to deliver data to a predefined geographic region. Finally, broadcast routing is used to deliver data to all nodes in the network. Figure 2 illustrates these different routing principles.

![Illustration of unicast, multicast/geocast, and broadcast routing schemes.](image)

Regarding unicast, a taxonomy for these routing protocols is proposed by Lee et al. [8], which divides them into two broad categories: topology-based and position-based. Topology-based routing protocols use network information about links to perform packet forwarding. This type of routing protocols can be further divided into proactive and reactive protocols. Reactive routing protocols determine routes on a demand or need basis. Proactive routing protocols propagate topology information periodically and find routes continuously between any pair of nodes in the network, regardless of whether they are needed or not.

Contrary to the previous protocols, position-based routing protocols, also called geographic routing protocols, do not exchange link state information and do not maintain established routes. They make forwarding decisions based on the geographic location of the destination node and the location of neighboring nodes. Hence, it is required that nodes have location capabilities, which can be provided by Global Positioning System (GPS) devices or location services.

Zhang and Wolff [10] observed that most routing protocol research studies for VANETs consider scenarios like highways and urban areas, which are characterized by high node densities. However, rural and sparse areas present significantly different conditions, resulting from low node densities, little or no fixed roadside infrastructure available, and terrain effects. These conditions lead to long periods of time where V2V or V2I communications is infrequent, interrupted, or simply not possible. Similar observations have also been made by many other authors, including Little and Agarwal [11], Jakubiak and Koucheryavy [1], Abuelela and Olariu [12], and Yousefi et al. [13], who state that vehicular networks can frequently form partitions, and thus prevent end-to-end communication strategies.

Routing protocols designed for fully connected networks...
are not suitable for data delivery in sparse/intermittent or partially connected vehicular networks. Hence, different routing techniques need to be designed from the perspective that vehicular networks are disconnected by default. To address these issues, researchers incorporated the store-carry-and-forward model of routing proposed for delay tolerant networks (DTNs) [14] into VANETs [15, 16]. The idea behind this is to exploit node mobility to physically carry data between disconnected parts of the network. This approach circumvents the lack of an end-to-end path, enabling non real time (i.e., delay-tolerant) applications. Main DTN concepts are explained in the next section.

III. DELAY-TOLERANT NETWORKS

Delay-/disruption-tolerant networking (DTNs) focuses on the design, implementation, evaluation, and application of architectures and protocols that intend to enable data communication among heterogeneous networks in extreme environments. Examples include interplanetary networks, underwater networks, wildlife tracking networks, sparse wireless sensor networks, people networks, military tactical networks, transient networks, disaster recovery networks, and vehicular networks.

DTNs experience any combination of the following aspects: sparse connectivity, frequent partitioning, intermittent connectivity, large or variable delays, asymmetric data rates, and low transmission reliability. More importantly, end-to-end connection cannot be assumed to be available. In order to answer to these challenges the DTN Research Group (DTNRG), proposed an architecture (RFC 4838) [14] and a communication protocol (RFC 5050) [17] for DTNs. The DTN architecture [14], illustrated in Figure 3, introduces a store-carry-and-forward paradigm by overlaying a protocol layer, called bundle layer, above the transport layer, which provides internetworking on heterogeneous networks (regions) operating on different transmission media. The bundle protocol [17] is end-to-end, strongly asynchronous, message (bundle) oriented.

Application data units are aggregated into one or more protocol data units called “bundles” by the bundle layer. The idea is to “bundle” together all the information required for a transaction (entire blocks of application-program data and metadata). This minimizes the number of round-trip exchanges, which is useful when the round-trip time is very large. To help routing and scheduling decisions, bundles contain an originating timestamp, a useful life indicator, a class of service assignment, and a length indicator. The bundle protocol also offers an optional hop-by-hop transfer of reliable delivery responsibility, called bundle custody transfer, and an optional end-to-end acknowledgement functionality (i.e., “return receipt”). When nodes accept custody of a bundle, they commit to retain a copy of the bundle until such responsibility is transferred to another node.

The store-carry-and-forward paradigm avoids the need for constant connectivity. It is used to move bundles across a region, exploiting node mobility. This paradigm, which is illustrated in Figure 4, can be described as follows. A source node originates a bundle and stores it using some form of persistent storage, until an appropriate communication opportunity becomes available. The bundle will be forwarded when the source node is in contact with an intermediate node that will be more close to the destination node. Afterwards, the intermediate node stores the bundle and carries it until a suitable contact opportunity occurs. This process is repeated and the bundle will be relayed hop by hop until (eventually) reaching its destination.

Routing is a challenging task in these networks due to the lack of contemporaneous end-to-end paths. Furthermore, information and resource shortage accentuate this challenge. It is important to note the importance of node mobility, which is exploited to carry data around the network, and thus to overcome network partitions. Numerous proposals of DTN routing protocols have been reported in the literature. Surveys of these protocols have been discussed in [18-22].

DTN routing can be defined as a sequence of independent, local forwarding decisions that make bundles “progress in steps” towards their destination. The source of knowledge that
is used to take these decisions often differs and can be used to classify routing protocols. While some routing approaches assume that there is not any knowledge available, others consider and eventually combine information about historical data (e.g., recent encounters, contact time, contact frequency, or contact location), location (e.g., past, present, future location data), or movement patterns.

DTN routing strategies can also be classified as single-copy schemes (i.e., forwarding-based) or multiple-copy schemes (i.e., flooding-based) [19, 20, 23]. Single-copy schemes maintain a single copy of a bundle in the network that is forwarded between network nodes. These routing schemes have low resource requirements (e.g., storage, bandwidth, energy), however they suffer from low delivery ratios and large delays. On the contrary, multiple-copy schemes replicate bundles at contact opportunities. The copies of the same bundle can be routed independently to increase security [24] and robustness (i.e., the chances of delivery via different paths). Bundle replication improves the probability of delivery and minimizes the delivery latency. The downside is that it consumes a high amount of energy, and increases the contention for network resources like bandwidth and storage. Therefore, it potentially can lead to poor overall network performance, as discussed in [23, 25]. These shortcomings often make multiple-copy routing strategies improper for energy-constrained and bandwidth-constrained DTN applications.

IV. VEHICULAR DELAY-TOLERANT NETWORKS

Vehicular delay-tolerant networking, or VDTN, was proposed in [26] as a novel form of a delay-tolerant network designed to provide low-cost asynchronous communications in sparse and disconnected vehicular network scenarios.

VDTN follows the principle of store-carry-and-forward routing proposed for DTNs to cope with the problems caused by intermittency, disconnection, and long delays in vehicular networks. However, on contrary to DTN architecture proposal, which introduces the overlay bundle layer between the transport and application layer to allow the interconnection of highly heterogeneous networks, VDTN architecture places the bundle layer over the data link layer introducing an IP over VDTN approach (Figure 5). The protocol data unit at the VDTN bundle layer is the above-mentioned bundle, which aggregates several IP packets with common characteristics, such as the same destination node or generated with data from the same application.

Another important characteristic of VDTN architecture is the out-of-band signaling with separation of the control and data planes (Figure 5). The VDTN bundle layer is divided into two layers: the bundle signaling control layer (BSC) and the bundle aggregation and de-aggregation layer (BAD). BSC layer executes the control plane functions, such as signaling messages exchange, node localization, resources reservation (at the data plane) and routing, among others. The signaling messages include information such as, but not limited to, node type, geographical location, route, velocity, data plane link range, power status, storage status, bundle format and size, delivery options, and security requirements, among others. BAD executes the data plane functions that deal with data bundles. These functions include, among others, buffer management (queuing) and scheduling, traffic classification, data aggregation/de-aggregation, and forwarding.

Out-of-band signaling allows the control plane to exchange signaling information through a separate, dedicated, low-power, low bandwidth, and long-range link. This link is always active to allow node discovery. On the contrary, the data plane can use a high-power, high bandwidth, and short-range link to exchange data bundles. The data plane link connection is active only during the estimated contact duration time and if there are data bundles to be exchanged between the network nodes. Otherwise, it is not activated. This approach, described in [26, 27], is very important because it not only ensures the optimization of the available data plane resources (e.g., storage and bandwidth) [27], but also allows to save power, which is very important for energy-constrained fixed network nodes [26, 28].

The principle of out-of-band signaling is illustrated in Figure 6. At the time $t+t_0$, two network nodes detect each other and start exchanging signaling messages through the control plane link connection. Based on this information, the data plane connection is configured and activated on both nodes at the time $t+t_1$. Then, the data bundles are forwarded until the time $t+t_2$. The data plane connection is deactivated after this instant since the nodes are no longer in the data plane link range of each other.

The frequency and the number of contact opportunities play an important role in the performance of any DTN-based network like a VDTN. In fact, in extremely sparse scenarios with low node density, direct contacts between nodes can be so infrequent that even the store-carry-and-forward paradigm
is insufficient, by itself, to accomplish data delivery. It is interesting to note that, in scenarios like a sparse vehicular network, mobile nodes (e.g., vehicles) may not come to direct contact with each other, however they may pass in the same location, in different times, one after the other. This motivates the introduction of stationary relay nodes as extra-infrastructure elements that can be strategically placed to increase contact opportunities between mobile nodes.

In a VDTN, stationary relay nodes are defined as fixed wireless nodes with store-and-forward capabilities that are installed on road intersections, allowing passing-by vehicles to collect and leave data bundles on them. Figure 7 illustrates an example where a stationary relay node is deployed on a crossroad, creating an additional contact opportunity that would not exist before since vehicles would not meet each other. When passing along the crossroad, vehicle A exchanges bundles with the stationary relay node at time \( t+1 \). Following a distinct trajectory, vehicle B passes along the stationary relay node at a later time \( t+t_1 \), collecting bundles left there by vehicle A. Previous studies [29, 30] have demonstrated the importance of stationary relay nodes to improve the delivery ratio and reduce the delivery delay in VDTNs.

V. CONCLUSIONS

There has been an increasing research interest by the area of vehicular communications taking into account the potential of these networks to enable various applications including safety, monitoring, driving assistance, entertainment, and delivering connectivity to rural/remote communities or catastrophe-hit areas. Vehicular networks have specific characteristics that raise a number of technical challenges due to the nature of vehicular environments and to a variety of factors including node heterogeneity, node interactions, node cooperation, and limited network resources. This paper has overviewed recent research on this topic and identified several open issues related with it. It has presented and discussed different paradigms towards vehicular communication, ranging from vehicular ad hoc networks to vehicular delay-tolerant networks.

ACKNOWLEDGMENTS

Part of this work has been supported by the Instituto de Telecomunicações, Next Generation Networks and Applications Group (NetGNA), Portugal, in the framework of the Project VDTN@Lab, and by the Euro-NF Network of Excellence of the Seventh Framework Programme of EU, in the framework of the Specific Joint Research Project VDTN.

REFERENCES


