A Survey on Cellular Data Offloading for Content Delivery in Heterogeneous Vehicular Networks

Arobinda Gupta

***Abstract*—Connected vehicles that enable users to access different cloud and internet based services are becoming an important part of modern transportation systems. As the use and consequent bandwidth demand of such services have grown, cellular networks have become the primary mean of providing connectivity to such external services for content transfer to and from vehicles. However, if all vehicles use the cellular network for all their content download/upload needs, this may result in bandwidth constraints and service degradation. Also, the use of cellular networks is expected to be costlier relative to other local vehicular communication technologies such as DSRC. A heterogeneous vehicular network refers to a network of connected vehicles that uses multiple communication technologies such as cellular, DSRC, wi-fi etc. for vehicle connectivity. The cellular data offload problem in a heterogeneous vehicular network addresses the problem of designing schemes for offloading part of the data transfer between vehicles and external services normally happening through cellular links to the network of vehicles and road side units (RSUs) using other non-cellular communication technologies. In this survey, we first provide a brief review of existing works in data dissemination and content delivery in non-heterogeneous vehicular networks as a background, as some of the techniques are also relevant in heterogeneous networks. We then address the main focus of this survey, cellular data offloading in heterogeneous vehicular networks for content transfer between vehicles and external services, where a vehicle may use both cellular and non-cellular links for the transfer. A classification of the existing works in this area is provided first. The works belonging to each class are then discussed in more detail. Finally, based on the survey done, some potential areas of future research are identified.**

I. Introduction

With the proliferation of cloud and internet based services, the need for ubiquitous access to data and services has become a necessary part of everyday life. It is now commonplace to expect that network-based services will be available even when a user is travelling in a vehicle. Some examples of such services are map download, streaming multimedia contents, social media access, online meetings, uploading vehicle diagnostic data, traffic information exchange etc. Vehicles that can provide connectivity to access such services on-the-move are being increasingly seen as an important part of future transportation systems. According to a recent report [1], the worldwide shipment of connected vehicles is expected to reach around 76 million by 2023.

Vehicles that can provide connectivity with other vehicles and roadside infrastructure have been extensively explored in the context of vehicular ad-hoc networks (VANET) for nearly two decades. A VANET is a special form of a MANET (Mobile Ad Hoc Networks) where the vehicles communicate with other vehicles or to fixed road-side units (RSUs) placed by the side of the road. The RSUs are usually connected to a backbone network using high speed wired or wireless links. The main motivation for the development of VANETs originated from traffic safety and convenience related applications. A global status report on road safety published by World Health Organization in 2018 shows that around 1.35 million people die in road accidents every year [2]. Reports published by National Highway Traffic Safety Administration show that in US alone, around 36000 people have died in around 33000 car crashes in 2018 [3]. While crashes are caused by a variety of reasons, it is believed that intelligent applications leveraging connected vehicles can be very effective in reducing the number of crashes. Applications on VANETs have been traditionally classified as safety-related, convenience related, and infotainment applications. Early VANET applications focused more on communication within groups of vehicles, using RSUs for internet access. Dedicated Short Range communication (DSRC) has been the predominantly proposed communication technology in VANETs for quite some time. However, cellular V2X (Vehicle-to-Everything) communication has also been standardised recently and is being explored by various stakeholders.

As the scope of the applications desired by users from moving vehicles has increased, it is now more common to talk about a connected vehicle where it is equally important to provide internet connectivity to access various services on the go. Centralized cloud based services have proliferated which collect data from the vehicular environment (traffic information, user requests for services, vehicle diagnostic information etc.), centrally process them and provide services back to the vehicular environment (traffic advisory, files, streaming media, car health information etc.). As the scope of such services have increased, so has the size of the transfers and the consequent bandwidth need. Cellular networks have been primarily used to support such services.

With the growing needs of a very large number of users, available cellular network bandwidth may not be adequate to provide services to all users. Also, cellular network usage is expected to be costlier compared to the free usage of DSRC. To address these issues, it has been argued that exploring vehicular networks with inter-vehicle and vehicle-to-RSU communication can be explored to augment the capacity of the cellular network in vehicular environments. A *heterogeneous vehicular network* refers to a vehicular environment where

Arobinda Gupta is with the Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, India 721302. E-mail: agupta@cse.iitkgp.ac.in

multiple different communication technologies are used for effective implementation of useful applications in a connected vehicle scenario involving different entities such as vehicles, RSUs, cellular base stations, wi-fi APs, remote servers etc. Some examples of such technologies are DSRC (IEEE 802.11p), wi-fi (IEEE 802.11), LTE/LTE-A, 5G etc. Among the different networking technologies, the use of DSRC and cellular technologies together in a vehicular environment has seen the most amount of work. The *cellular data offload* problem in a heterogeneous vehicular network refers to the problem of offloading part of a data transfer from cellular networks to networks using other technologies.

In this survey, we focus on existing works that have addressed the cellular data offload problem in heterogeneous vehicular networks in the context of content delivery to/from a vehicle from/to an external content server. More specifically, we consider only works in which either the source or the destination (but not both) of the data is a vehicle, and a vehicle has the option of using the cellular network for data access (as opposed to works in which all vehicles use only vehicle to vehicle and/or vehicle to RSU/AP communication). We provide a classification of the works based on the main design issues involved in data offloading from cellular networks. Finally, we discuss existing works in each of these classes in detail, and identify some potential research challenges for future work. At the same time, we also present, in a separate section, a review of general data delivery techniques in non-heterogeneous vehicular networks (i.e., involving only vehicles and RSUs communication) as a background, as some of the techniques used in data offloading algorithms in heterogeneous vehicular networks are based on data delivery schemes in non-heterogeneous networks. Thus, this survey is expected to provide a detailed exposure to works related to data offloading techniques in heterogeneous vehicular networks, while giving a good general background of data delivery in vehicular networks.

To the best of our knowledge, the works in [4], [5] and [6] are the only surveys to address heterogeneous vehicular networks in particular. The surveys in [4] and [5] are broad surveys on heterogeneous vehicular network in general, and does not focus on the data offload problem in particular. The survey by Zhou et al. in [6] addresses the data offload problem and is the closest to our work. However, their survey takes a broader view of data offload and includes many works that do not directly use cellular data offloading in vehicular networks, including works on mobile data offload in general, vehicle-to-vehicle content transfer, VANET capacity estimation, vehicle-to-RSU link scheduling, computation offloading through fog computing etc. The survey also does not cover many existing works on cellular data offloading. We provide a more focused and extensive survey on cellular data offloading for content delivery, based on a classification different from the one proposed in their survey.

The rest of this paper is organized as follows. Section II provides a general overview of connected vehicles, including application landscapes and communication technologies used. Section III introduces the data delivery and offload problems, and specifies the scope and focus of this survey. Section IV broadly surveys works that has been done in the area of data delivery in non-heterogeneous vehicular networks. Section V identifies and discusses the main issues of interest in data offloading in heterogeneous vehicular networks, and proposes a classification. Section VI and Section VII surveys in detail existing works related to data offloading based on the proposed classification. Section VIII identifies some potential future research directions in the area of cellular data offloading in heterogeneous vehicular networks Finally, Section IX presents some concluding remarks.

II. Connected Vehicles

The term *connected vehicles* broadly refers to vehicles that can communicate with other entities in their environment. The communication can be with one or more of other vehicles, roadside infrastructure, cellular base stations, or any other entity in the internet, using possibly different communication technologies. An early use of connected vehicles that has become popular appeared in the form of various services (ex. driver assistance, call center services etc.) provided by various automobile manufacturers to the customers of their high-end vehicles. Special purpose devices embedded inside the vehicles communicate through an embedded cellular link to the manufacturers' call center to avail of these services. Some notable examples of such services in use today are GM On-Star [7], BMW Connected Drive [8], Audi Connect [9] etc. As opposed to such *embedded* connectivity, the concept of *brought-in/tethered* connectivity tries to leverage the use of connections provided by smartphones to enable applications inside a vehicle. Such systems may run applications in a vehicle-integrated system using the smartphone only for connection to the internet, or may also use the smartphone for running applications. Some examples of such systems include Ford SYNC [10], Toyota Touch [11], Volvo Apple CarPlay [12], and other systems providing smartphone integration using Android Auto [13], Apple CarPlay [14] etc. Many automakers also provide hybrid versions of their vehicle connectivity products using both an embedded connectivity for accessing specialized services and interface with smartphones for integrating smartphone based services in the vehicle's infotainment system.

The most commonly used term in the area of connected vehicles is Vehicular Ad Hoc Networks (VANETs). A VANET is a special form of a MANET (Mobile Ad Hoc Networks) where the mobile entities are vehicles. A VANET typically has two important entities – vehicles and road side unit/infrastructure (RSU). Each vehicle has an on-board unit (OBU) which is used to communicate with OBUs of other nearby vehicles or with RSUs. The RSUs are usually fixed infrastructure at roadside locations (for example, at road intersections, traffic lights, gas stations etc.) and are connected to a backbone network using high speed wired or wireless links. The communication in a VANET can be vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I). Vehicles can communicate with each other using wireless links when they are within a certain range of each other or when they are in the range of a RSU (the range is defined by the wireless technology in use). RSUs can communicate with each other using the backbone network and are also connected to the internet. A vehicle may connect to the internet indirectly through the RSUs, or directly through a cellular link. Also, given the high mobility of vehicles, the connection between two vehicles or between a vehicle and a RSU can be short-lived, and hence, by themselves, are not directly suitable for providing continuous connectivity.

The term *IoV* or *Internet-of-Vehicles* has been coined more recently to refer to connected vehicles in the context of the overall ecosystem of IoT (Internet-of-Things). This term encompasses both VANET as well as communicating vehicles that act as any other internet-enabled device interconnected with all other internet-enabled devices to open up new application areas for vehicular use.

In the next two subsections, we briefly review the application landscape for connected vehicles and discuss two predominant communication technologies currently in use for vehicular communication.

A. Application Landscape

The primary motivation for the growth of interest in VANETs has been safety and traffic management applications that are expected to result in huge economic and social gains. As noted in Section I, a large number of persons are killed every year in traffic accidents, which also cause significant loss of property. It is well known that many of these accidents are caused mainly due to driver error, poor judgment, driver fatigue, or distraction. It has also been seen that an average driver spends a significant amount of time in traffic jams and at stop lights, which has significant economic effects. It is reported that the cost of extra travel time and fuel due to congestion in U.S. urban areas was around USD 179 billion in 2017, compared to USD 75 billion in 2000 [15]. In a vehicular network, vehicles, RSUs, and other entities can cooperate to perceive potentially dangerous/inconvenient situations over an extended space and time horizon, which can lead to reduced chance of accidents and better traffic management.

In early works on application specification on VANETs, the Vehicular Safety Communications (VSC) project of US Department of Transportation identified and specified a set of applications to be enabled over VANETs. The VSC consortium, comprising of seven leading automakers (BMW, DaimlerChrysler, Ford, GM, Nissan, Toyota, and VW), identified around 45 potential application scenarios and classified them into two groups – safety (34) and non-safety (11) applications, grouped into different classes such as Intersection Collision Avoidance, Public Safety, Sign Extension, Vehicle Diagnostics, Traffic Management etc. Some examples of the applications identified include Traffic Signal Violation Warning, Intersection Collision Warning, Post-Crash Warning, Curve Speed Warning, Emergency Electronic Brake Lights, Lane Change Warning, Intelligent On-Ramp Metering, Point-of-Interest Notifications, Map Downloads and Updates etc. [16]. The application specifications were based on a standard message specification for vehicular applications by SAE (Society of Automotive Engineers) [17]. The standard identifies basic message data elements (speed, position etc.), and data frames that are composed of a subset of data elements (for example, the Position3D data frame is comprised of latitude, longitude, and elevation data elements). The message set, among other messages, identifies a set of safety messages and basic information that must be sent by each vehicle in safety application messages; examples of such information include vehicle size, position, speed, acceleration, braking and throttling information etc.

While the primary goal of the VSC report was to improve safety and traffic management, users nowadays have also come to expect convenience-related applications from connected vehicles. Vehicular communication applications are now broadly classified into three classes as follows.

- *Safety applications*: the goal of these applications is to avoid accidents caused by different factors and thus enhance overall traffic safety. A list of important safety related applications has been given earlier. These class of applications usually have stringent communication latency constraints ($<$ 100ms). The destinations of the messages are usually all vehicles in a specified Region of Interest (RoI) which can vary from application to application (for example, 10s of meters for applications like Intersection Collision Warning to 100s of meters for applications like Post Crash Notification). The messages are usually broadcast to all nearby vehicles. The broadcasts can be periodic with no time limit (for example in Curve Speed Warning) or event-triggered broadcasts for specific time periods (for example, in Emergency Electronic Brake Light (EEBL)).
- *Traffic management applications*: the goal of these applications is to improve the efficiency of traffic flow in general. Some example applications falling in this class are traffic density monitoring, traffic light scheduling, road congestion notification etc. These class of applications usually have medium latency constraints. The destinations of the messages are again usually all vehicles in a specified RoI which can vary from application to application.
- *User services and convenience related applications*: The goal of these applications is to provide relevant information to the drivers to enhance their comfort and ability to do different personal/business operations, i.e. to improve the driving experience as a whole. Some example applications in this class are disseminating local information (tourist spots, gas stations etc.), map downloads, Infotainment (audio/video streaming etc.), instant messaging, online shopping, automatic toll collection, vehicle health monitoring etc. These class of applications have different latency constraints depending on the application. For example, instant messaging can tolerate relatively high latency while multimedia streaming will require low latency. Also, the destinations for messages belonging to these applications can vary, from all vehicles in a region (for example, parking availability information disseminated around a tourist spot) to unicast/multicast to one or a designated set of users (for example, instant messaging).

Many other innovative applications have been proposed for enhancing the cost-efficiency and user convenience of the vehicular eco-system. Some examples include driver monitoring and profiling for insurance applications, algorithm based vehicle pricing (used car pricing based on fine grain driving data collected), contextual advertisement, energy optimization, integration with smart homes and wearable devices, predictive maintenance, fleet management etc. Note that with internet-enabled vehicles, the whole range of services that are open to any other device should become available to users in a vehicle also.

On a related note, a number of applications have been developed that take advantage of the On-Board-Diagnostic (OBD) port that is mandated to be present in all cars since 1996 following the OBD-II specs. An OBD adapter can be plugged in to the OBD port to access the data and many applications can be built on them, including car tracking, diagnostics, driver scoring etc. Some examples of systems built for different applications using data from OBD port include Dash [18], Automatic [19], TorquePro [20], HobDrive [21], OBD Auto Doctor [22], Carista OBD2 [23] etc.

B. Communication Technologies

DSRC (Dedicated Short Range Communication) has been proposed as the foremost access technology for VANETs for more than a decade. However, lately LTE support for more efficient vehicular communications has also been standardised. Below we give a brief overview of the two technologies.

DSRC system is a short to medium range communication technology that operates in the 5.9 GHz band for the use of public safety and other applications. In the United States, the Federal Communications Commission (FCC) allocated 75MHz in the 5.850–5.925 GHz band for DSRC, whereas in Europe, European Telecommunications Standards Institute (ETSI) allocated 70 MHz in the 5.855–5.925 GHz band. The DSRC system supports a vehicle speed up to 200 km/h, nominal transmission range of 300 meters (up to 1000 meters), and the default data rate of 6 Mbps (range of 3 to 27 Mbps with 10 MHz channels). It is to be noted that DSRC is a general system description proposed by ASTM (standard 2213-03) which has been standardized by different standard bodies in US and Europe in different manners. For example, IEEE has adopted DSRC as IEEE 802.11p specification (a modification of IEEE 802.11) in which the 75 MHz spectrum is divided into one control channel and six service channels of 10 MHz each. The service channels are allowed to be used by any application. On top of IEEE 802.11p, IEEE 1609 standard for Wireless Access in Vehicular Environments (WAVE) has been proposed to standardise higher layer service access, network layer, and security. In contrast, in Europe, ETSI has adopted DSRC as the ITS-G5 standard in which the channels have specific use, with 5.855 MHz-5.875 MHz reserved for two service channels for non-safety applications, 5.875 to 5.905 MHz reserved for two service channels for safety and traffic efficiency applications and one control channel, and the remaining spectrum reserved for future ITS applications. In Japan, the DSRC has been adopted by ARIB (Association of Radio Industries and Broadcasters).

DSRC is well-suited for short range broadcast based applications. Its ad-hoc nature and direct communication link allows for low latency communication between nearby vehicles and RSUs. It has also been field-tested for more than a decade now and is considered as a mature and well-understood technology. While DSRC remains the predominant access technology so far, several limitations have also been pointed out for its widespread use in future. DSRC based systems, being short range, connects to the internet through RSUs; thus a widespread network of RSUs are needed to provide near-continuous connectivity to users as needed for some applications. Putting up such RSUs is costly and it is not clear if such networks of RSUs will be available in future, especially in non-urban areas. The limited bandwidth of DSRC is also a concern for providing high-end services in future ITS applications. Another problem with large scale DSRC based communication is that DSRC suffers from the problem of broadcast storm, the problem of too many collisions occurring when many nearby vehicles try to broadcast together. It is to be noted that the standardisation of DSRC have happened over basic IEEE 802.11, which was developed for unicast communication, and is a contention-based MAC protocol that does not support efficient broadcasts very well.

The use of cellular technologies has been investigated to offset some of the problems associated with the use of DSRC for vehicular communication. Unlike DSRC, cellular technologies can potentially offer high bandwidth (up to 1 Gbps for LTE-A and higher in forthcoming 5G networks), much higher than DSRC, to enable more applications. It does not require any additional infrastructure as the existing base stations can also cater to the vehicular communication. Given the wide coverage of cellular networks, it can also enable continuous network connectivity to vehicles.

Recently cellular communication has been proposed as an alternative to DSRC, commonly referred to as C-V2X (Cellular-Vehicle-to-Everything) systems. LTE-V2X has been proposed in Rel 14 of LTE in 2017 and further refined in Rel 15 in 2018. Cellular systems have the advantage of a much larger bandwidth, especially with the expected deployment of 5G to which LTE-V2X is claimed to be compatible. Also, it is useful for providing continuous connectivity for internet-based applications. However, the centralized nature of LTE networks makes all communication go through a central infrastructure, causing latencies which may not be acceptable for safety applications in vehicular environments. To address this problem, LTE-V2X allows for two types of communication, a cellular connection from the vehicle to base station as usual, and a direct V2V connection using the D2D feature of LTE for short range low-latency communication. The cellular link from the vehicle to base station works in the same way as for any other user device in a cellular network. The V2V link, also sometimes called a sidelink (as opposed to uplink/downlink), works in the same 5.9 GHz band reserved by ITS for vehicular applications, with 10MHz channels, with RRB allocations as in LTE channels. The V2V communication can be operated in two modes, a centralized-control based mode (Mode 3) in which a central controller decides on the resource allocation and scheduling (and hence the vehicle must be in the coverage area of a base station), and a decentralized mode (Mode 4) in which the nodes decide on their resource allocation and schedule locally. C-V2X is claimed to be compatible with 5G networks, with the higher bandwidth of 5G network expected to address some of the bandwidth concern. Also, the higher layers of C-V2X systems reuse the existing standards and are compatible with other DSRC-based systems such as IEEE WAVE standards.

Despite the many benefits of using cellular communication for connected vehicles, it has certain disadvantages also. Using the cellular network alone may overload the network given the high number of vehicles and applications, especially as it is shared with other non-vehicular applications also. The use of cellular network is also expected to be costlier than the use of DSRC-based communication. The mostly centralized nature of LTE-V2X increases application latency and may not meet strict latency requirements of many vehicular applications, especially safety related applications. The use of V2V (Mode 4) communication can reduce this latency; however LTE-V2X is a relatively less mature technology compared to DSRC and the effectiveness of V2V over DSRC based communication has not been clearly established so far. Finally, while cellular networks can provide continuous connectivity in principle, the high mobility of vehicles makes the problem of handover management more complex, especially in upcoming 5G networks with small cell sizes and ultra dense networks.

While DRSC has been a mature and more well-tested technology that is more widely adopted by the vehicular networking community in practice, LTE-V is also being aggressively proposed as an alternative to DSRC by many 5G network operators, automakers, and communication technology providers. It will be interesting to see how these two technologies compare in practice in future.

III. The Cellular Data Offload Problem

In this section, we first describe some characteristics of the data delivery problem in vehicular networks, and then identify the cellular data offload problem and the scope of this survey.

The data delivery problem deals with transferring data from a source to one or more destinations. Data delivery requirements in vehicular networks can be classified in several ways.

- Source and destination of data transfer:
  - Source and destinations can both be vehicles. Some example applications for this can be delivery of safety messages, instant messaging between vehicles, live meetings between vehicles, sending sensed information to other vehicles etc.
  - Only the source is a vehicle and the destination is an entity outside the vehicular network. Some example applications for this can be upload of vehicle data to remote servers, upload of files to remote servers etc.
  - Only the destination is a vehicle and the source is an entity outside the vehicular network. Some example applications for this can be web surfing, live streaming of videos etc.

  Even though there has been some works [24][25] where the source and destination are both outside the vehicular network (i.e., the vehicular network is used only as an intermediate network in between the source and the destination), we do not consider this any further as they do not involve data delivery to/from vehicles.
- Type and size of data:
  - Event-driven: Data generation occurs only when certain events happen, The size of the data can be either small or large depending on the application. Some example applications are accident information generated by a vehicle, request for a file upload/download, streaming of movies etc.
  - Periodic: Data is periodically generated. An example application is diagnostic information uploaded by vehicles. The size of the data is usually small and can be transferred in a short time.
- QoS requirements: Depending on the application, different QoS requirements can be placed on the data delivery process:
  - Bandwidth guarantee for a class of traffic.
  - Delay guarantee for a class of traffic. This can be in the form of continuous delay guarantees for components of the data as in the case of video streaming, or an overall deadline imposed on the entire data transfer as in the case of a file upload with deadline.
  - Data loss guarantee. Some applications may be able to tolerate small amount of packet losses such as movie streaming while some other applications like file downloads need to be totally loss-less.
  - Spatial guarantees such as delivery to vehicles within a region of interest

Given the proliferation of cellular networks, a simple solution to address the data delivery issues can be to enable cellular access in vehicles and use the cellular network for all data transfers. However, as discussed earlier, this approach has problems in terms of both available bandwidth and cost, especially in urban areas with a large number of vehicles. The cellular data offload problem addresses these concerns by attempting to use vehicle-to-vehicle or vehicle to RSU communication as much as possible for data delivery to offload some of the cellular traffic on to these vehicular networks. The primary goal of this offload is usually cost reduction in some form, with some works also using the offload to provide continuous connectivity when cellular network may not be available or quality of available connections may be poor.

In a scenario employing cellular data offload, a vehicle has at least two interfaces, a cellular interface and a DSRC/wi-fi interface. Vehicular networks in which vehicles employ more than one type of connections for information exchange are called heterogeneous vehicular networks. This survey primarily focuses on existing works that address data offload from cellular to vehicular networks that specifically consider content delivery scenarios between a vehicle and an external server, where a vehicle can use the cellular interface for data transfers. Thus, we do not focus on works that consider scenarios in which no vehicles access the cellular network at all (i.e., all vehicles communicate only with other vehicles or RSUs using DSRC/wi-fi), or access the cellular network only for control information such as content request (as such transfers are usually of small size and the data offload problem is not relevant in such scenarios). However, in the next section, we first briefly review works on data delivery in non-heterogeneous vehicular networks to provide a general background in data delivery in vehicular networks, as many of the techniques used in such works have been used in data offloading in heterogeneous networks as well. We then survey in detail works within the scope of this survey as mentioned above in subsequent sections.

In the rest of this survey, we will use the term *road side unit* or *RSU* to refer to any non-cellular roadside infrastructure that a vehicle communicates with such as DSRC based devices, wi-fi APs etc. Also, the term *Vehicle-to-Infrastructure* or *V2I* will refer exclusively to the communication between a vehicle and any of these RSUs only, and not for communication to any cellular base station. Similarly, the term *Vehicle-to-Vehicle* or *V2V* will refer to vehicle to vehicle communication using non-cellular technologies only. This distinction is made clearly as even though traditional VANET literature uses V2I to mean the same as our use, some recent works note (rightfully) that cellular base stations are also infrastructure and include them in V2I communications; the term V2I has also been used by a few recent works to indicate Vehicle-to-Internet communication with V2R used for communication with RSUs. Similarly, while cellular D2D technologies can be used for communication between two vehicles as per the current standard, it has still not been explored explicitly in the context of cellular data offloading in existing works, and hence we reserve the term V2V for non-cellular vehicle-to-vehicle communication only.

IV. Review of Data Delivery in Non-Heterogeneous Vehicular Networks

In this section, we give a brief review of the works that have been done in the area of data dissemination in a vehicular network setting where only DSRC and/or wi-fi based V2V and V2I communication have been used for data transfer. In these works, the vehicles do not use any cellular access for data transfer. Any data transfer to and from any back-end server is done only through RSUs or wi-fi APs. The works on data delivery in such vehicular networks can be broadly classified into two types, works that directly address effective schemes for lower level mechanisms for unicast/broadcast/multicast of a single message to one or more recipients, and works that address higher level content delivery in general (which may or may not be single message) to one or more recipients, using lower layer services provided by the basic communication schemes as stated above. In the next two subsections, we briefly review the two classes of work. More details can be found in other existing surveys on these topics [26][27][28].

A. Unicast, Broadcast, and Multicast Dissemination

The problem of data dissemination has been extensively studied as problems of broadcast, unicast, and multicast in vehicular networks. Most safety and traffic management applications employ broadcast communication, transmitting public information to all entities within a given range. Unicast communications are used primarily for user convenience applications such as messaging, video calls etc. Multicast communication is also of interest in some applications such as group chat etc.

Unicast routing protocols in VANETs are usually classified as position-based routing or non-position-based routing. Position based routing schemes use geographic information about the position of the destination and other vehicles to route the packet towards the destination. Most of these protocols are based on the well-known greedy forwarding technique, with different schemes to address the local optima problem inherent in greedy forwarding [29]. Many protocols also employ vehicular environment specific information such as intersections, location of landmarks along the way, roadmaps etc. to optimize the routing. Some of these protocols assume a dense deployment of vehicles where the network is always connected, and do not address the problem of network partitioning. Examples of such protocols are Spatially Aware Packet Routing (SAR) [30], Contention-Based Forwarding (CBF) [31], Anchor-based Street and Traffic Aware Routing (A-STAR) [32], Spatial and Traffic-Aware Routing (STAR) [33], Greedy Traffic Aware Routing (GyTAR) [34], GPSRJ+ [35], Connectivity-Aware Routing (CAR) [36], Greedy Routing with Abstract Neighbor Table (GRANT) [37], Landmark Overlays for Urban Vehicular Routing Environments (LOUVRE) [38], Intersection-based Geographical Routing (IGRP) [39], Junction-Based Routing (JBR) [40], Back-bone-assisted Hop Greedy Routing (BAHG) [41] etc. Some other protocols consider that the network may be disconnected at times and incorporate mechanisms to handle this. Some example protocols in this class are Vehicle-Assisted Data Delivery (VADD) [42], static-node assisted adaptive data-dissemination (SADV) [43], Reactive Pseudo-suboptimal-path Selection (RPS) [44], GeoSpray [45], Opportunistic routing based on Symmetrical Traffic Distribution (OSTD) [46] etc. There are also hybrid protocols that switch between schemes that work in connected environments and schemes that can handle network disconnection based on the network connectivity status sensed. Some example protocols in this class are Geographic DTN Routing with Navigator Prediction (GeoDTN+NAV) [47], Connectivity-aware Minimum-delay Geographic Routing (CMGR) [48], Roadside Units as Message Routers (ROAMER) [49] etc. Other routing protocols that are not based on position-based routing include Geographical Opportunistic Routing (GeOpps) [50] that chooses the forwarding nodes based on minimum estimated time of delivery to the destination, and Map-based Sensor-data Delivery Protocol (MSDP) [51] that also uses estimated time based on road map and programmed routes (for example routes of buses and taxis).

Broadcast based information dissemination schemes have been extensively studied in vehicular networks. Single-hop broadcasts are commonly used for sending out information periodically, such as beacon information (cooperative awareness messages (CAM)) containing different vehicle information (vehicle id, speed, position etc.), emergency information within a small region etc. Choosing the period of broadcast is important as too small an interval will cause too many broadcasts and too large an interval can give out stale information. In multi-hop broadcast, messages received by one vehicle are further broadcast/forwarded to other vehicles by the receiving vehicle. Multi-hop broadcasts thus enable extending the range of broadcast messages beyond what is allowed by the underlying technology. Multi-hop broadcast can be typically achieved by flooding. However, broadcast by flooding can cause many redundant broadcasts if all receiving vehicles broadcast the same received packet, causing wastage of limited bandwidth available, and too many collisions and packet loss (*broadcast storm problem*). Hence, multi-hop broadcast protocols have also tried to reduce redundant broadcasts in different ways. Typical approaches to reduce redundant broadcasts try to select only a small subset of the receiving nodes to rebroadcast a received packet as opposed to all receiving nodes rebroadcasting the packet. Existing works have used different techniques to select these relay nodes intelligently using many factors such as direction and speed of vehicles, distance from the sender, vehicle density, topology information etc., which can be primarily classified as delay based, probability based, distance/topology based, or counter based schemes.

In delay based approaches, nodes are made to wait for variable delay periods before rebroadcasting the packet. The nodes with lower delay has higher priority and broadcast ahead of others. Also, waiting nodes abort their attempt to broadcast if they receive the rebroadcasted packet. The delay can be chosen in different ways, one good approach being giving the nodes farthest from the source of the broadcast (in the direction of propagation) the lowest delay, which maximizes the chance and rate of forwarding of the packet. Several algorithms have been proposed based on this approach, with different schemes for assigning delays to nodes. Variations have also been proposed to handle network partitioning and adapting the period of transmission based on traffic congestion. Some example algorithms in this class include Urban Multi-hop Broadcast (UMB) [52], Multi-hop Vehicular Broadcast (MHVB) [53], Efficient Directional Broadcast (EDB) [54], Slotted 1-Persistence Broadcasting [55], Reliable Broadcasting of Life Safety Messages (RBLSM) [56], Reliable Method for Disseminating Safety Information (RMDSI) [57], Opportunistic Broadcast in VANETs (OB-VAN) [58], Adaptive Probability Alert Protocol (APAL) [59], Road Oriented Dissemination (ROD) [60], Link-based Distributed Multi-hop Broadcast (LDMB) [61], Adaptive Traffic Beacon (ATB) [62], Simple and Robust Dissemination (SRD) [63], Urban Vehicular broadcast (UV-CAST) [64], Autonomic Data Dissemination in Highway for VANETs (ADDHV) [65], PREemptive algorithm for DAta Transmission (PREDAT) [66], FAST-OB-VAN [67], Adaptive Data Dissemination Protocol (AddP) [68], Efficient Adaptive Probability Data Dissemination Protocol (EEAPD) [69] etc.

In probability based approaches, nodes rebroadcast the packets with different assigned probabilities. Since not all nodes will rebroadcast the packets, the number of redundant packets are reduced. Probability assignments can be static and pre-assigned for each vehicle, or they can be dynamic. In the dynamic schemes, probability assignment can take into account different factors such as distance from transmitter (higher distance gets higher/lower probability), traffic density (higher traffic around a vehicle may mean lower probability as others may rebroadcast anyway), or combinations of both. Many variations of this has been proposed. Some examples of algorithms falling in this class are Optimized Adaptive Probabilistic Broadcast (OAPB) [70], Weighted p-Persistence [55], Slotted p-persistence [55], AutoCast [71], Adaptive Probability Alert Protocol (APAL) [59], stochastic broadcast scheme (SBS) [72], Preset Delay Broadcast (PDB) [73] etc.

In distance/topology based approaches, the distance and/or the topology between the sender and the receiver is used to decide whether the receiver will rebroadcast the packet or not. For distance-based schemes, receivers very close to the sender should not be rebroadcasting the packet as it will not increase the coverage area significantly. Topology based methods also use maps and other topology information to ensure other constraints for rebroadcasting. Some example constraints/methods are the sender and receiver being on different streets, formation of a CDS (Connected Dominating Set), distance from junctions, detection of at least one new neighbor in neighborhood graph, neighbor density estimation to use broadcast suppression techniques in dense neighborhoods, choosing the neighbor with the highest number of one-hop neighbors, inter-vehicle distance estimation to form constraining zones etc. Some examples of algorithms falling in this class are The Last One (TLO) [74], enhanced Street Broadcast Reduction (eSBR) [75], Distributed Vehicular Broadcast (DV-CAST) [76], Density-Aware Reliable Broadcasting (DECA) [77], Hydi [78], enhanced Message Dissemination for Roadmaps (eMDR) [79], Nearest Junction Located (NJL) [80], Junction Store and Forward (JSF) [81], Neighbor Store and Forward (NSF) [82], Dynamic Partitioning Scheme (DPS) [83], Adaptive Data Dissemination Protocol (AddP) [68] etc.

In counter based approaches, a vehicle decides to rebroadcast a packet or not based on how many times it has received the message. Typically, such schemes have also been used along with delay-based or probabilistic schemes to rebroadcast with delay or probability, with the delay/probability adjusted or broadcast cancelled if too many duplicates are received. The threshold count can be static or adapted dynamically. Some examples of protocols in this class include Adaptive Probability Alert Protocol (APAL) [59], Adaptive approach for Information Dissemination (AID) [84] etc.

Works have also been done on multicasting in VANETs. Multicast protocols for VANETs can be broadly classified into two classes, ones that use topology information (clustering, multicast tree etc.), and ones that use location information of the group members for delivering content to multicast groups. Because of the overhead of maintaining clusters, cluster based protocols have not been used much. Some examples of protocols in this class are COIN [85] and CBDRP [86]. Location based protocols have been used more, and some examples include IVG [87], Cached Geocast [88], Abiding Geocast [89], GvGrid [90], DRG [91], ROVER [91], DG-Castor [92], Mobicast [93], DTSG [94], CBDRP [86], Constrained Geocast

[95], Geocache [96] etc. Some protocols have also built multicast trees with different properties and other overlay structures between the source and the destination vehicles for efficient multicast.

B. Content Delivery

Content delivery in vehicular networks deals with the problem of delivering requested content to and from vehicles. To deliver contents efficiently to its destination, the content has to be moved closer to the destination. In the context of a vehicular network, this may take many forms such as choosing which RSUs or other vehicles to store the content in (other than the source) so that the destination vehicle/RSU can receive it. These content stores are typically referred to as replicas in CDN (Content Delivery Network) terminology, and act analogously as caches. The two main problems addressed in any CDN are that of replica selection and content delivery. The replica selection problem decides which vehicles/RSUs to select as replicas. Note that in case of a vehicular CDN, this has a spatio-temporal aspect as the replica may change with time also as they move out of a Region of Interest (RoI). Also, given that the contact time between a vehicle and a replica may be small, multiple replicas may be needed to complete a delivery for larger content sizes. The content delivery problem deals with how a vehicle discovers a replica or requests for content, and how the content is actually delivered to the vehicles from the chosen replicas. Below, we provide a brief overview of works done in both replica selection and content delivery, with classification of the works based primarily on schemes proposed in [28].

Different strategies have been proposed for replica selection/allocation in vehicular CDNs. In centralized strategies, a central entity collects all information and decides the replicas. Typical information used by the central entity are current network topology, expected network topology, vehicle information (type, speed, direction, position etc.), content popularity, and content demand profile. Network topology information is used to ascertain expected contact time between vehicles and RSUs, which can be used to choose replicas that can be most effective in transferring the content to a large number of vehicles. Existing works have also considered keeping track of how many vehicles have already received the content to decide if new replicas are needed, popularity/demand of data in certain areas etc. Some example works in the class are MobTorrent [97], On-Time [98], OVS-OBRM [99], GO-DCR [100], SAMCDN [101] etc.

In contrast, in distributed schemes, replica placement is decided in a distributed fashion either with the help of some infrastructure (for example, RSUs) or by the vehicles without the use of any infrastructure. Some examples of Infrastructure based solutions are Figaro [102], VTube [103] etc. In infrastructure-less solutions, vehicles take decision to act as replicas or not based on local information only. In its simplest form, every vehicle that receives a content can cache it for some limited time. However, though simple, this is a blind approach that is not very effective always. Other works use a cooperative caching scheme where nodes cooperate to decide which content to keep and for how long based on the cache of their neighbors. Some works that do distributed infrastructure-less replica allocations are InfoShare [104], SPAWN [105], Abiding Geocast [89], InfoCast [106], Hamlet [107], ARM [108], LINGER [109], RADD [110] etc.

Once the replicas are chosen, the next problem is to deliver the content to the vehicles from the replicas. Delivery mechanisms can broadly be divided into two classes. In pull-based schemes, vehicles send request to content provider and receive the content in response. Various techniques have been used to find the provider and to route the content to the requester. One approach is to flood the request until it finds a provider, which then unicasts the response back using the reverse path recorded. This approach is used in different forms in InfoShare [104], SPAWN [105], and CRoWN [111]. In other approaches, expected trajectory information available from infrastructure nodes is used to schedule deliveries (MobTorrent [97]), brokers that keep content locations are used to find the replica locations and then send a request to it (Figaro [102]), network coding is used to send more content to vehicles which then share it (VANETCODE [112]) etc. Many solutions use only infrastructure for delivery, like MoPADS [113] which uses cellular links to get the request and chooses RSUs to schedule the delivery to the requesting vehicle when it passes by a RSU.

In Push-based schemes, the content provider pushes the content to vehicles, assuming all vehicles in a RoI are interested in the content. A simple scheme is periodic broadcasts of content in the RoI. Rebroadcasts, using standard schemes for redundant broadcast suppression, are used to increase the coverage area. The RoI may be fixed or may be selected by the content provider based on different factors. Vehicles may also request the content through cellular links, and the content may be pushed to nearby RSUs which then periodically broadcast it. Some examples of protocols that adopt such schemes are InfoCast [106], PrefCast [114], TBCD [115], RTAD [116], SAMCDN [101] etc. Some algorithms such as TSF [117] also use RSUs as rendezvous points for pushing content to a single vehicle.

V. Design Issues for Data Offloading

As mentioned earlier, the main goal of cellular data offloading is to use the vehicular network as much as possible to take the load off the cellular network. In order for cellular data offloading to happen, a vehicle requesting for content needs to get at least part of the requested content from other vehicles or RSUs. Thus, achieving efficient cellular data offloading requires addressing two primary problems, that of a vehicle deciding when (or for which part of the data) it should use the cellular network and when it should use the vehicle/RSUs for data transfer (the *network selection problem*), and that of a vehicle/RSU deciding what data to cache and the associated caching policies so that other vehicles may receive the data from it later without having to access the cellular network (the *caching problem*). We next look at these two problems in more detail.

The network selection problem addresses the issue of deciding when to switch networks (between cellular, V2V, and V2I) and deciding which network to use at each instant of time for the application in question. The decision to change networks and choice of a target network can be based on different parameters ranging from lower level network parameters such as signal quality, load, ease of handover etc. to higher level application parameters such as cost, QoS constraints etc. We briefly discuss these parameters below.

- Signal quality: Low signal quality may prohibit the choice of a network.
- Current load on networks: Choice of a heavily loaded network can increase delay and/or packet loss.
- Handover efficiency: Even when a network with low load and good signal strength is available, handover issues may suggest avoiding choosing the network. As an example, the choice of a particular base station may be bad for a high speed vehicle if the expected duration of stay of the vehicle in the coverage area of the base station is low (implying another horizontal handover will be needed very soon).
- QoS constraints: QoS constraints can affect the choice of the network to ensure that the constraints are met. As an example, for a file download with deadline, RSUs and other relay vehicles can be used to reduce the cost until the deadline comes nearer, at which time the cellular network may be used if it is decided that the V2V and V2I infrastructure may not be sufficient to meet the deadline.
- Cost: Different networks have different costs which affect the network choice. While V2V and V2I communication are expected to be cheaper than the use of cellular networks in general, the cost-benefit analysis may become complex in cases such as when RSUs are put by private operators to be used on a payment basis. A vehicle may also be provided with multiple cellular interfaces (similar to dual-SIM phones) to give more choice between cellular networks themselves. Within the same network also, different users may incur different costs based on the subscription status of the user with the operator.
- Caching policy: RSU or other edge caches (including other vehicles) may be employed by operators for efficient data transfer. A network selection strategy should take into consideration location and type of caches to choose a network that allows the use of the cache. Note that this is a dual problem in the sense that one of the parameters chosen for cache placements may be the availability of a particular network that enables good use of the cache.
- Incentive for use: Using a V2V network requires the cooperation of the vehicle users. Different incentive schemes may be applied to encourage vehicles to act as gateways/relays or caches for data delivery to other vehicles. The network selection strategy can consider the incentives in place to choose the appropriate network. Note that this may be considered as a subclass of cost as finally, for a single vehicle, the cost of transfer is affected by this.

The relative importance of the different parameters in the network selection problem will depend on both the application at hand and the content provider's goals. The content provider may also try to optimize each application (user need) individually or collectively. Once the network selection is done, an associated problem is that of vertical handover that addresses the actual switch between the networks to give a seamless communication experience to the user. The exact vertical handover process is technology dependent and we do not address it in this survey.

It is well known that in most data delivery problems, caching plays an important role. Note that in traditional content delivery, replicas essentially act as content caches. Caching is also an important design issue in data offloading in heterogeneous vehicular networks. Satisfying a vehicle's content request fully or partially from cached contents in caching vehicles/RSUs can allow the content provider to save on cellular link usage costs. While caching has been mentioned as one of the factors that may affect network selection, it is of independent interest of study. Even if a network is already selected, the performance of the system may be affected by how caching is used to ensure data delivery with different constraints. As an example, a data delivery system may use the simple network selection policy of selecting a V2I network whenever available over cellular network (independent of caching or any other parameter). However, while connected to a RSU, some vehicles may download popular content and cache them for some time so that other vehicles have a chance of getting that content using V2V communication even if they are not in the range of any RSU, without having to go to the cellular network. Caching has been studied in the context of data offloading in heterogeneous vehicular networks, addressing the problems of deciding the caching locations (vehicles/RSUs), contents to cache, caching duration, cache replacement policies, cost of caching etc.

In the next sections, we describe in detail works that have been done in the areas of network selection and caching in the context of cellular data offloading in heterogeneous vehicular networks. While some works have addressed both network selection and caching issues, we classify them based on which problem is the primary focus of the work.

VI. Network Selection for Data Offloading

In this section, we look at works that have primarily focused on the network selection problem. Network selection for data offloading in heterogeneous vehicular networks can be primarily classified into two types - gateway based or non-gateway based. In gateway based schemes, some RSUs and/or vehicles are chosen as gateways. Instead of all vehicles downloading or uploading their data through cellular links only, gateways are usually the only nodes that access the cellular network, with other vehicles downloading/uploading data through them using the vehicular network, except for possible direct connection to cellular network in some specific cases such as if some QoS constraint is about to be violated etc. Gateways can cache data for future downloads or aggregate data for uploads to reduce the use of the cellular link. Since RSUs are usually connected to the backbone network by high speed wired/wireless links, the choice of RSUs as gateways can also avoid the use of cellular links in most cases. Gateways can be chosen statically or dynamically, and the status of a gateway can be permanent or temporary. For non-gateway nodes, the network selection problem then reduces to choosing between the gateway connection or the cellular connection. For gateway nodes, the network selection problem is implicit in their choice as gateways. These gateways are also alternately referred to as clusterheads, seeds, downloaders, relays etc. in different works. The choice of the gateways and the role (gateway or non-gateway) of each vehicle can be made centrally by a central coordinator, in a distributed manner by coordination between the vehicles and/or RSUs, or with a combination of both central and distributed schemes. Different schemes also utilize different levels of system knowledge in making a decision. In non-gateway based schemes, there is no designated gateway and vehicles can connect with either the cellular or non-cellular network at any time based on the algorithm/policy followed.

Based on the above discussion, we classify the existing works in the following classes:

- *Centrally controlled gateway based schemes*: In these gateway-based schemes, the gateways are chosen centrally by a central controller.
- *Distributed gateway based schemes*: In these gateway based schemes, the gateway is chosen in a distributed manner with local coordination between the vehicles and the RSUs.
- *Hybrid gateway based schemes*: In these schemes, the choice of the gateways and their use uses a combination of central and distributed schemes.
- *Non-gateway based schemes*: In these schemes, there are no gateways chosen specifically, and vehicles make the choice of communicating with the cellular or vehicular network individually. The choice can be made centrally or by local coordination among the vehicles and/or RSUs.

Table I shows the a partition of the existing works surveyed that fall into each of these classes. The next subsections describes each of the works in more detail.

A. Centrally controlled gateway based schemes

Whitbeck et al. [118] address the problem of disseminating a content with expiry time to a group of subscribers. A subscriber may subscribe to a central server for receiving the content anytime (even after the generation of the content), but must still get it within the expiry time. In the proposed scheme, a central node chooses some vehicles as seed nodes and injects the content directly at them using cellular links. The seed nodes then disseminate the content using epidemic protocol. The central server periodically monitors the infection rate (fraction of subscribed users who have received the content) by communicating with the subscribers, and if the rate is less than a target rate (a fixed function of time till expiry time of the content), it chooses more seed vehicles to inject more copies into the network. The central server also directly pushes the content to a subscriber using cellular link if the content expiry time is close and that subscriber has not yet received the content.The method is generally referred to as *push-and-track*. The paper discusses several strategies for choosing the seed vehicles and to decide when to inject more copies. The objective is to deliver to all users within the expiry time while reducing the load on cellular infrastructure by reducing redundant messages for downloading same content by multiple users.

Rebecchi et al. [119] propose a scheme very similar to the one in [118]. However, the difference from the earlier push-and-track strategies is that while the earlier strategies used a fixed target infection rate and compared the current infection rate to it to make a decision on injecting new seeds, this paper also considers the evolution of the infection rate. This is done by keeping the infection rates for a past time window, and using properties of it including the slope of change to decide how many new copies of the content to inject and where to inject them. The intuition behind this approach is that different mobility patterns affect the epidemic dissemination process differently and the injection rate adaptation should consider the mobility pattern to decide who should get how many copies if needed. Similar to the earlier work, the central server directly pushes the content to a subscriber if the content expiry time is close and that subscriber has not yet received the content.

Jia et al. [120] propose a simple approach for upload of data generated by vehicles to a backend server, which may then be used for services like traffic management etc. Clusters of vehicles are formed and a clusterhead is selected for each cluster from nearby vehicles. The clustering is assumed to be done centrally by the cellular base station and informed to the vehicles using cellular links. The clusterhead acts as the gateway to the cellular network. All vehicles in the cluster send their data to the clusterhead, which then compresses/aggregates the data and sends to the server using cellular link. In certain scenarios when the local channel condition is poor or the mobility is high, a vehicle can also send its data directly to the base station using cellular links with some probability. The goal is to reduce the load on the cellular network to have a lower impact (low blocking probability) on other ongoing human-to-human cellular traffic.

Hong et al. [121] consider the problem of dissemination of a single content to multiple subscribers interested in it. However, the content is large and cannot be transferred in a single contact; it has K packets all of which must be transferred to the destinations within an expiry time for the content. The paper assumes that a central server schedules the offload. The server receives dynamic information from vehicles (overtaking, entering/leaving communication regions of base stations (BS) etc.), through a separate control channel (cellular link). Based on the information received, the server maintains a contact graph, and solves a flow-based problem on it to decide a set of relay vehicles that should get the data packets directly through cellular network, and which vehicles should get the packets through V2V links from the relay nodes. The server then instructs the vehicles to act accordingly. The contact graph is maintained dynamically based on new information received, which may change the current offloading decision also as the graph changes. The paper also proposes a damage-threshold based scheme to

Class	References
Centrally Controlled Gateway based	[118], [119], [120], [121], [122], [100], [123], [124], [125], [126]
Distributed Gateway based	[127], [128], [129], [130], [131]
Hybrid Gateway based	[132], [133], [134], [135]
Non-Gateway based	[136], [137], [138], [139], [140], [141], [142], [143], [144], [145], [146], [147], [148], [149]

TABLE I
CLASSIFICATION OF NETWORK SELECTION SCHEMES FOR DATA OFFLOAD

decide whether to change the offloading decision on a dynamic update or not, as changing it on each dynamic change can be costly. The goal of the proposed scheme is to minimize the number of packets transmitted through cellular links while ensuring that the content reaches all interested subscribers within its expiry time.

Munyoung et al. [122] propose a system called DOVE that assumes a central traffic manager that maintains vehicle trajectories based on GPS data sent through 4G-LTE. Based on this data, the traffic manager can predict the mobility of the vehicle. It is assumed that relay nodes (RN), with DSRC connectivity and storage but no internet, are placed at intersections. The vehicles send their content request to a central manager. Based on the requests received, the central manager forms request groups of users requesting the same content. For each request group, the central manager selects a set of offload positions (OP) out of the RNs so as to deliver the content with minimum number of OPs using a variation of the set cover problem. One vehicle in the request group is designated as a provider vehicle for each OP (usually the one that will reach the OP first). The provider vehicle downloads the content through cellular link and transfers it to the OP. The OP caches the content for disseminating to other vehicles in the request group. The goal of the scheme is to select minimum no. of OPs to download all content so as to reduce cellular link usage.

Silva et al. [100] consider a scenario where vehicles send their origin and destination points to a central server through cellular links, and the central server chooses some vehicles to act as replicas for a content to be delivered to all vehicles passing through a region of interest. The choice of the replica is made based on function of the distance travelled by a vehicle inside the region of interest as well as how well the region is covered by other replicas when the vehicle is passing through the region. However, unlike the works in [118] and [119], the central server does not track the coverage and the content has no deadline. The goal is simply to maximize the number of vehicles that can get the content from the replicas.

Yuan et al. [123] consider both space and time constraints in delivering a content, meaning that the data has to be delivered both before (or within some spatial bound) the vehicle reaches some spatial location and within a time delay. The main approach involves a central controller that creates a probabilistic contact graph based on predicted mobility of the vehicles in near-term. The graph vertices also carry additional information like vehicle location, data demand and their delivery constraints, and data already received. Based on this graph and some additional parameters, the controller selects a set of offloading nodes to directly inject the content through cellular links. These offloader vehicles then disseminate the content to other vehicles using V2V links. When a node receives a content, it informs the controller. The controller monitors the progress of this process and chooses more offloader vehicles to inject more content copies if needed. The method tries to maximize a utility function which is based on the difference between the total amount of offloaded data and the amount of offloaded data that is delivered after its expiry time.

Mezghani et al. [124] consider a scenario in which contents are classified into different topics. A set of users, where each user has a different preference for each topic, request for contents belonging to their topics of interest. Each content has a lifetime within which it has to be delivered. The paper proposes a centralized algorithm called SIEVE for delivering the content to the users. A central controller periodically collects vehicular information such as location, speed, direction etc., and predicts the potential contacts between the vehicles in a near-future time span based on this. Given the contact predictions and the requests for the users, the controller chooses seed vehicles for each content that maximizes a utility function (essentially the sum of user interest utility for the topic of the content for all vehicles that the seed vehicle will come in contact with within the lifetime of the content). The seed vehicles download the requested content using cellular links and deliver to other vehicles on contact using V2V links.

Huang et al. [125] address the problem of maintaining continuous connectivity instead of just delivery of some content; it is assumed that this will also help in better offloading. A central controller collects all information from users and RSU. For users, this involves context information such as location, speed, direction, RSUs in range etc. For RSUs, this involves network information that may affect handover, such as congestion window size, from which a measure of the network quality is derived. The controller predicts which RSUs the vehicle will encounter and the staying time of the vehicle in the RSU range. Based on the staying time (directly proportional) and the network quality (inversely proportional), a score is given to each RSU for each vehicle at each time. Based on that, it advises a vehicle to switch or not when a RSU comes, and if yes, the time of switch. If yes, the vehicle prepares for the handoff early, saving time. The cellular network is used for any communication need of the vehicle when no suitable RSUs are found.

In the work by Pescosolido et al. [126], a requester node sends request for content with a deadline to a central Content Distribution Management System (CDMS) running in a base station (BS). The CDMS is assumed to have updated information of the location and trajectory of all vehicles, from which it can predict the contacts in its area of coverage. The CDMS uses this to find a content provider vehicle that will be closest to the requester and schedules the delivery from that provider;

the requester just waits for the content. One interesting issue addressed in this paper is that in addition to choosing the provider, the schedule also chooses the exact time and position of delivery, as delivery as soon as in contact may require more power due to longer range if the vehicles are expected to come closer shortly. It assumes that distance based power control can be done. Each node receiving a content stores it for some time and can act as provider (CDMS has exact knowledge of who stores what for what time). The cellular link is only used if no providers are expected to be in contact with the requestor before the deadline, at which time the cellular link is used to directly download the content. When a vehicle moves out of range of a BS, the CDMS hands over the control of the download to the CDMS of the next BS the vehicle associates with.

B. Distributed gateway based schemes

Benslimane and Taleb [127] propose a gateway selection based on clustering. The vehicles are clustered into different clusters based on direction of movement, 3G signal strength received and other parameters, and a clusterhead is elected for each cluster. The clustering is done in a distributed manner. The clusterheads act as gateways to the cellular network, with all other vehicles communicating with the clusterheads using V2V links. The clusterheads advertise themselves for easier discovery, and a source vehicle tries to select one of the advertised clusterheads as a gateway based on different parameters such as link stability etc., which is then communicated to other nearby vehicles. Cluster maintenance is also briefly discussed. The goal of the approach is to minimize the number of gateways selected to provide continuous connectivity to all the vehicles. While the paper does not directly address data offloading, the gateways can be used to cache common content for downloads or aggregate data for upload to achieve cellular data offloading.

Gramaglia et al. [128] propose a scheme called SILVIO for using VANET to offload traffic from a 3G network. All devices requiring network access inside a vehicle connect to an on-board router, which then connects to a RSU through a tree of vehicles rooted at the RSU that is built in a distributed manner by the RSU, and maintained by the RSU and the vehicles based on relative distance, speed etc. (the tree-building uses an earlier protocol called TREEBOL [150] for this). The RSU at the root act as the gateway. Two classes of traffic are considered, critical/real-time constrained traffic and other traffic that can tolerate some delay. The first class is always served using the 3G network; for the second class, the wi-fi network is used whenever available. However, different strategies are proposed for wi-fi access such as connect only when directly in contact with a RSU, allow multi-hop connection to RSU but allow connection-disconnection only once within its coverage area (meaning if the connection to RSU is broken and the vehicle goes back to 3G, it will not try to reconnect to that RSU), allow multi-hop and multiple-time connection-disconnection with some constraints (like only when moving towards the RSU) etc. The objective is to maximize the data offloaded to the wi-fi network.

Stanica et al. [129] consider the upload of quasi real-time data generated at the vehicles to a central server. The approach is to choose a set of vehicles to collect the data from neighbors and upload it using cellular link after data fusion/aggregation. The goal is to minimize the number of upload nodes chosen while ensuring that all data is uploaded. To this end, the paper first shows that if the contact graph is known and static, then the dominating set in the contact graph is a good choice for the set of collector nodes. However, in a practical scenario, this will not be the case. So it proposes three local heuristics by which a node chooses itself to be the uploader (dominator) in a distributed manner. As an example, a node can select itself to be the uploader with a probability based on its number of neighbors, and use an acknowledgement scheme to handle the case when neither it nor any of its neighbors decide to be the uploader (which will break 100% coverage if not handled). The goal is to maximize the fraction of nodes that do not have to use the cellular link.

Salvo et al. [130] address the problem of upload of data from vehicles to a central server. Instead of all vehicles uploading their data through the LTE network, a set of vehicles are chosen as clusterheads to collect data from nearby vehicles and upload to the server through LTE link. To choose the clusterheads, a set of relay nodes are chosen first. The relay nodes are chosen periodically in a distributed manner as part of a standard broadcast dissemination process of request messages, with delay based broadcast suppression. The broadcast dissemination process is started by a request message from a centrally located RSU or vehicle, and the vehicles that rebroadcast the message (following the delay based broadcast suppression scheme) are chosen as the relay nodes. The relay nodes can directly act as clusterheads for data transfer themselves, giving a fully distributed scheme. Alternately, the relay nodes can send neighboring vehicle information to a central server, which can then choose a set of clusterhead vehicles based on this information using any other suitable policy.

Ju et al. [131] also address the problem of gateway discovery through which a vehicle will access the cellular network. Any vehicle can be a gateway or a normal node. Gateways selected are within one-hop of the vehicle. The choice of gateways is done through a distributed algorithm that uses location of the vehicles. The algorithm is run periodically in which current gateways advertise themselves and non-gateway vehicles in range that wish to transmit data associate with the nearest gateway. If no gateways are found, the vehicle assumes the role of a gateway. Similarly, a gateway changes back to a normal vehicle if it is found that two gateways are within range of each other.

C. Hybrid gateway based schemes

The work by Zhioua et al. [132] uses clustering and assumes that a set of vehicles has been chosen as clusterheads (CH) by some underlying central clustering algorithm. However, instead of only choosing clusterheads as gateways, their scheme also allows the choice of other vehicles as Gateway Candidates (GwC) if no suitable clusterheads are found. Traffic is divided into classes. When a vehicle wants to send traffic of a class, it uses several parameters (depending on traffic class) such as Received-Signal-Strength (RSS) from LTE eNBs at the vehicle and at the CHs/GwCs, eNB load, CH load, CH link stability etc. to choose a network or gateway. High delay-sensitive traffic (Class 1) is sent through cellular links if signal is high and load is below threshold. In all other cases, it first tries to find a suitable CH if available based on certain QoS parameters. If no such CH is found, the vehicle solicits new gateway choice among potential gateway candidates. An ordinary vehicle can act as a gateway candidate if it has high RSS, low load, and good link quality with the vehicle. A vehicle finally communicates with the CH or gateway candidate chosen, which then transfers the vehicle's data using the cellular link. The CH/gateway choice is updated periodically during the data transfer if needed. The CH selection is centrally controlled, but the gateway choice is done in a distributed manner by local coordination among vehicles.

Malandrino et al. [133] address the problem of downloading contents with specified deadlines by vehicles. A downloader vehicle sends its request to a central query manager using cellular link, which then forwards pending requests to RSUs in the region of travel of the downloader vehicle. The RSUs decide which data to download/prefetch when, and how to get it to the downloader. It is assumed that a central traffic manager periodically gives a prediction of future V2I and V2V contacts based on a probabilistic contact graph built to predict mobility based on past information. Based on this, RSUs locally decide based on a local optimization solved at each time step what contents to download, when to prefetch a content and how to transfer the content to a downloader vehicle. The content download can be directly to the vehicle when it is in range, or to relay vehicles which are expected to meet the downloader (the relays are given the downloaders ids, the content they want and the expected contacts predicted, and schedule the transfers accordingly). RSUs and vehicles are used as much as possible to get the content; however, if the downloader does not receive its full content within a deadline, it uses the cellular network directly for rest of the download. The objective is to maximize the amount of data offloaded from the cellular network. Thus central control is used to get mobility prediction, while inter-RSU distributed algorithms are used to schedule the transfer and prefetch.

El Mouna Zhioua et al. [134] address the problem where a downloader vehicle may want to download different classes of traffic like video, data etc. The vehicles transmit periodic beacons, based on which the RSUs form and maintain local connectivity graphs. Downloader vehicles can be directly connected to the RSU in this graph or can connect to a RSU in a multi-hop path using other vehicles as intermediate nodes. The problem is to maximize the number of flows that can be assigned to the VANET based on path availability (based on the graphs formed by the RSUs) and data volume. The rest of the data is sent through the cellular network. The graph is built by the RSUs locally, but the optimization problem formed is solved centrally using a max-flow based formulation assuming all flows are known a-priori.

De Felice et al. [135] propose a scheme for incidence detection based on analysis of vehicle data uploaded by vehicles to a central server periodically. The data upload can be done in one of three modes, VANET only, VANET and LTE, and LTE only. Instead of each vehicle sending their data to the central server through LTE always, the scheme uses two gateway RSUs, a source RSU and a sink RSU along the path of the vehicles. The source forwards a message periodically which is relayed by the vehicles after adding their own information, which is then collected by the sink RSU which sends the information to the server. VANET mode uses standard V2V communication and delay-based broadcast suppression. The switch between the modes is done in a distributed manner based on several local timers to detect connection/disconnection events based on receive of some message from the source RSU. Network selection policy is simple, use VANET to relay, use LTE only if VANET is not available. Thus, the source and the sink RSUs are fixed a-priori centrally; however, the mode selection is done in a distributed manner.

D. Non-gateway based schemes

Balasubramanian et al. [136] address a data offload problem between cellular and fixed wi-fi AP for requests coming from vehicles. Every vehicle independently uses both a dynamic AP encounter prediction based on the number of APs that the vehicle has encountered in the recent past, and a throughput estimate of each AP to compute how much data can be offloaded through wi-fi APs in the vehicle's path. This is used to decide whether to wait for APs or use the cellular network. If the request is not met within a delay threshold, cellular network is used anyway. For more delay sensitive traffic like VoIP etc., cellular network is used if the link layer packets are not delivered within a delay threshold (indicating high latency of the wi-fi network). The paper proposes a system called Wiffler based on the above scheme.

Xiaoxiao et al. [137] defines a utility function that gives the benefit of using a particular network (as function of throughput achieved) and a cost function for using any network with some throughput. It then frames an optimization problem that tries to maximize the difference between the two. An analytical solution of the problem shows that wi-fi should be used with the maximum throughput whenever available, using cellular to bridge the gap if the maximum wi-fi throughput is below a threshold parameter. Each vehicle uses a wi-fi throughput predictor for future APs based on measurements over a past time window, and uses the above formulation to decide what throughput to aim for from wi-fi and cellular in a future window. The network selection problem is addressed in a somewhat different manner here in that instead of always choosing one network or the other as is commonly done, it allows for choosing both networks together, with targeted throughputs through each. The objective of the optimization is to maximize the difference between utility gained and cost incurred.

The work by Siris et al. [138] assumes that a large content has to be downloaded by a vehicle, and that the mobility information of the vehicle is known. Wi-fi APs along the path of the vehicle are used for data offloading from the cellular network. It assumes that the location and average throughput of APs along the path are known. Based on this information, APs that the vehicle will encounter on its path are found. Two types of traffic are considered, and the vehicle makes different decisions for each. For delay-tolerant traffic, the APs are used as much as possible, and prefetching is done at the next AP to improve its throughput when the vehicle comes in range and data is transferred to it. In this case, the cellular network is used only for the traffic that cannot be downloaded from the APs (based on throughput calculated), or if the deadline of the content is close. Issues of some uncertainties in prediction are also handled (like a vehicle arriving at the next AP before the prefetching is complete etc.). For delay-sensitive traffic, the goal is to reduce transfer delay, and hence cellular network is used with maximum throughput whenever wi-fi connection to APs are not available. However, when APs are available, they are still used as much as possible, including the use of prefetching, to reduce the data that needs to be transferred through the cellular network. The goal is to reduce the amount of data transferred through cellular network while satisfying all delay-tolerant requests within deadline and minimizing the transfer delays of the delay-sensitive traffic.

Lee and Lee [139] use prediction of future AP details and application usage pattern to decide on offload to APs during movement of vehicles. A vehicle keeps detailed information of past AP encounters, including APs encountered and bandwidth/data rates obtained, history of transition from one AP to another, time, direction of movement and speed etc. Based on this, a vehicle makes a local decision of the next best AP to use. Detailed information is also kept on past application usage pattern of the user, including which class of applications were used, data rates needed etc. Applications are divided into different classes with different delay requirements. Based on the above, the vehicle decides whether to use the cellular network or switch to the next best AP.

Bruno et al. [140] present a simple offloading scheme in which a requesting node requests a central dissemination manager (CDM) for data that needs to be received within a deadline. The CDM checks if the content is already present in some other mobile nodes; if so, it waits till deadline to receive an ACK from the requesting node. If no ACK is received, the CDM sends the content directly to the node through the cellular network. The requesting node, after sending the request to the CDM, also asks neighboring vehicles for the content. If any of them has it, it is sent to the node. A node receiving the content caches it for a fixed duration to give to other node that may request for the same content. The CDM keeps track of who has the content and for how long. While the work is described in the context of mobile nodes in general, it is mentioned that the goal is focused towards vehicular networks, with experimental evaluation also done on vehicular network scenarios.

Wang and Wu [141] model the utility gained by a user for every bit as a function of download of every bit (gain), delay (loss), and cost if cellular network is used (loss). It then tries to schedule RSU and cellular usage to maximize this utility over total size of data. At discrete times, each vehicle predicts the RSUs to be encountered and solves the optimization problem to schedule the download. However, if past predictions are wrong (ex. vehicle arrives earlier at a RSU), the offloading strategy is dynamically adapted again at those time points.

In the work by Si et al. [142], vehicles are assumed to be moving in N-S/E-W roads only, with known velocity and some probabilistic values for wait times at intersections etc. The paper considers the transfer of delay-tolerant traffic primarily, with each data having a lifetime. The source and the destination of the data can be inside or outside the vehicular network; it is assumed that the destination position is known, though no details are given as to how this may be updated correctly for moving vehicles. The paper proposes an offloading architecture called DaVe in which delay-tolerant data is broken into data blocks (DBs), and all DBs of the data has to be sent to the destination for a successful transfer. Each vehicle with a DB makes a decision in a distributed manner at each step whether to forward the DB (and if yes, to who) or hold it in its storage, based on current storage, delay, and distance and angle to destination. If the deadline is close, the cellular network is used to send the remaining data directly to the destination. The goal is to maximize a reward function which is based on a positive reward component for the transfer and a negative component for cost of using the cellular network.

He et al. [143] address the problem of downloading video streams encoded with scalable video coding (SVC) through a heterogeneous cognitive vehicular network consisting of cellular base stations (BS) and cognitive radio enabled RSUs. A SVC coded video has a base layer that must be downloaded for minimum video quality, and additional enhancement layers which incrementally add to the video quality. Instead of downloading all requested layers from the BS, a vehicle tries to download only the base layer of the video from a BS. The BS, on getting a request from the vehicle, decides whether to admit the vehicle's request or not based on its current load of other background users in its coverage area. If a vehicle is admitted, it gets the base layer of the video from the BS, and can request the RSUs on its path for channel allocations to download additional layers for enhanced video quality. The RSUs can allocate additional channels to vehicles for downloading further enhancement of the video. If the BS is not able to admit the request, the vehicle may request the RSUs for radio resources for downloading both the base and enhancement video layers. The paper proposes a method for joint call admission control at BS and channel allocation at RSU. The RSU works independently but considers the result of the admission control decision of the BS in making its decision. If a RSU does not have sufficient channels available for serving a vehicle's request for a certain number of layers of a video, an algorithm is proposed that progressively degrades the video quality of other vehicles to receive less number of layers subject to certain minimum quality restrictions till enough channels are freed to admit the new vehicle's request. The goal is to maximize a reward function defined based on a tradeoff between the video quality available to a user and the quality loss suffered by users due to degradation caused by loss of layers.

In the work by Yu et al. [145], vehicles send requests to a central content server by cellular network. Requests may or may not have deadlines, and a content can be requested by multiple vehicles. The proposed solution assumes contacts with RSUs and future requests (up to a time) are known. It proposes a simple greedy algorithms for scheduling RSU broadcasts to satisfy user requests, which is sent to all requesting vehicles. Knowing the schedule, and deadlines if any, vehicles can decide whether to wait for receiving the content from the RSUs or to download directly from the content server using the cellular network. The goal is to minimize cellular network usage and maximize the number of user requests satisfied.

Xu et al. [144] propose a system called EcoMD to address the problem of multimedia content delivery in an information-centric network where contents are addressed by name. The content is assumed to be available from multiple sources, including multiple backend servers accessed through cellular links, RSUs, and other vehicles. Each of these types of providers have different service rates and cost of access. The paper proposes strategies for choosing a path (vehicles, RSUs, or backend servers) for downloading content, choosing appropriate provider nodes along the path, and caching. The choice of the path and the provider is based on computation of path delays derived from a queueing model to ensure a minimum rate for video streaming quality, and cost of access. The number of caches is decided based on the request rate of a content. The problem is formulated as Mixed Integer Programming problem that tries to minimize the cost of streaming while satisfying the supply-demand, available bandwidth, and caching constraints. The problem is shown to be NP-hard and a heuristic solution is proposed.

Sun et al. [146] consider data offloading in the context of video streaming applications. Vehicles can act as downloaders (the end user), relays (from RSU to downloaders), or carrier (to downloaders in opposite direction). The paper frames an optimization problem for choosing a transmission scheme (who should transmit what to who at what time). A greedy algorithm based on a max-flow formulation is then proposed. It assumes that a single service channel is time-multiplexed for the transfers. A vehicle switches to cellular if no RSU or assisting vehicles are in range. RSUs also collaborate with nearby RSUs when a vehicle is about to leave the range of a RSU or its relay (carriers in opposite direction can carry this information to the RSU) to allow the next RSU to prepare and prefetch content if possible. The work assumes all mobility information and requests are known in advance. The goal is to minimize the video stalling times at the downloaders.

Dai et al. [147] consider delivery of data services to vehicles where the data has temporal properties (ex. versions updated periodically). Vehicles upload requests to the central controller. Requested data can be of different sizes, have delay requirements, and may have temporal order of download. Wireless interfaces are also considered to be heterogeneous in terms of transmission rate. The goal is to schedule the transmissions across the available interfaces (DSRC/wi-fi, cellular etc.) so that maximum no. of requests are satisfied within their delay bound. It is assumed that the central controller gets mobility information from the vehicles to accurately predict the mobility and hence contact with the RSUs etc. Based on these, a centralized scheduling algorithm is proposed and network selection is done as per the schedule output by the algorithm.

Rhaiem et al. [148] consider the problem of cooperative download of SVC-coded video. Vehicles are dynamically classified as requestors, helpers, or forwarders. A requester vehicle uses 3G to download a desired content if available. If 3G connection is not available or if connection is broken in the middle, it takes the help of helper vehicles nearby. The helper vehicles are chosen dynamically in a distributed manner based on a simple distance-based policy for each request. Helper vehicles can download the remaining part of the video using 3G connection and send to the requestor directly or through forwarder vehicles. Forwarder vehicles, also chosen dynamically in a similar manner as the helper vehicles, only forward the content, and do not directly participate in the download. A scheme to choose a helper-forwarder path that reduces the content delivery delay to the requestor is also proposed. Since a SVC-coded video requires the base layer to be present for any enhancement layer to be useful, helpers also cache the base layer of the downloaded content for future demand. The primary goal of this work is to use the vehicular network for good quality video download with small delay even when cellular connection from some vehicles are of poor quality. However, the proposed schemes can be also be used for cellular data offloading if the 3G connection, even if available, is used only if the content cannot be obtained through the vehicular network.

Al-Hilo et al. [149] consider a data offload model for video streaming in which each RSU tries to put enough content in a vehicle's buffer (directly or through vehicle relay) to allow the vehicle to keep playing the video till it reaches the next RSU. If a RSU cannot put enough content, the vehicle downloads the remaining portion from the streaming server directly using cellular link, and continues on to the next RSU. The amount of data needed till the time of reaching the next RSU is calculated with a simple model considering a straight highway and equidistant RSUs, and constant playback rate. The RSU resource is scheduled periodically taking into account the vehicles in range, their requests, and current buffer position. V2V communication is also utilized assuming some content overlap. The paper proposes a revenue model between the content provider and the RSU operators, using which the RSU operator tries to maximize its revenue gain while still keeping a good user experience.

VII. Caching for Data Offloading

In this section, we look at works on data offloading that have used caching at vehicles and/or RSUs to enable a requesting vehicle to get the desired content from caches without having to use the cellular network. Caching has been extensively investigated for data delivery in mobile networks in different contexts, including in vehicular networks. However, given the focus of this survey, we only look at works where the goal of the caching is to reduce the load on the cellular network, in contrast to works which try to use caching for other reasons such as to optimize the use of intermittent RSU access available to vehicles in non-heterogeneous vehicular networks.

Li et al. [151] consider a scenario where contents are tagged with interest keywords and users/subscribers interest in different content are modeled by their degree of interest in these keywords. The popularity of a content is modeled by the distribution of user interests in the keywords associated with the content. A central controller receives all subscriber requests over cellular links, and chooses a set of helper vehicles that can act as replicas/caches. It is assumed that these helpers are already chosen, possibly using different incentive schemes. The data is transferred to the helpers by the central server using cellular links. The helpers then transfer the content to the subscribers requesting the content using V2V links when they are in range. It is assumed that all subscribers have the same contact rates with a helper. A subscriber vehicle tries to get the requested data from helpers. If the data is not obtained within a tolerable delay, the subscriber gets it directly from the central server using cellular links. The helpers acting as caches have limited storage. The paper investigates the use of erasure coding for effective transfer. In erasure coding, the data is broken up into smaller segments and coded, and only a subset of the segments are required to retrieve the original data. However, this increases the total buffer space requirement in the helpers. The central controller decides the replication and coding policy for the different contents based on the content popularity and user interests, i.e. which helper should cache which contents at what time and whether in coded form or not. An optimization problem is formulated and solved to get an optimal solution. The goal is to maximize a function of expected satisfaction of user interests.

Ahn et al. [152] consider the data offload problem when the vehicles are grouped into different groups. Similarly, the contents are grouped into different types, with each type having a delivery deadline associated with it. It is assumed that the i-th group of vehicles are interested in the i-th type of content. The paper proposes an optimization framework for choosing an initial set of seed vehicles for each group that can receive the contents for the respective types directly through cellular links at time 0 from a central server, and then disseminate to other vehicles of the group on contact with them using V2V links. Vehicles receiving a content cache it, the caching policy being that a vehicle only receives and caches content that it is interested in. The paper formulates an optimization problem under certain assumptions that tries to maximize the expected number of users whose requests can be satisfied subject to cost and delay constraints, and proposes an analytical solution for deciding the optimum number of seeds that tries to maximize the defined utility function.

Vigneri et al. [153] propose a scheme in which a central server seeds a set of vehicles with contents, which then act as caches for disseminating the content to other vehicles. Contents have popularity and requests have a deadline. When a vehicle needs a content, it continues to ask nearby vehicles for it until the deadline is near, at which time it downloads the content from the central content server using the cellular link. The paper proposes a scheme for initial seeding of the content based on content popularity under certain assumptions, and incremental seeding as needed when popularity changes. The scheme decides the vehicles and the number of copies of each content to cache under both when there is no cache size constraint in the vehicles and when the vehicles have a fixed cache size. It also handles some other cases like contact breaks during full transfers (repeat or resume). The goal is to minimize the access to the cellular network, which is modeled as a function of cache miss (vehicle has to download the content from cellular eventually) and the additional number of seeds needed over initial seeding (incremental seeding). Incremental seeding is used as even though both seeding and cache miss cause access to cellular network, minimizing one in absolute terms can cause the other to increase. Note that incremental seeding may need some other content to be removed from cache in case of fixed buffer size. A cache replacement policy is proposed that replaces a content only if the potential gain (number of cache misses saved) from the new content copy is larger than the potential loss (additional cache misses incurred) from the replaced content.

Bian et al. [154] consider both caching at RSUs provided by a central caching server, and caching at vehicles determined by a distributed network of vehicles, to reduce the load on cellular networks. For caching provided by the central caching server, vehicles are assumed to have predetermined routes. However, the QoS demand of the users may be different than the QoS of caching service provided. Contract games are proposed as a tool by which a caching service provider can decide on pricing strategies for different users to maximize its own payoff. Similarly, for caching at vehicles, evolutionary games between vehicles are proposed as a tool by which stable caching strategies can be evolved in which performance can be maintained in spite of a small number of vehicles leaving or joining the system. This paper does not actually propose any caching strategy for data offload; rather, it discusses game-theoretic tools that can be used to develop such strategies, which is illustrated with a small example.

Zhu et al. [155] consider a scenario similar to that in [151]. Contents are tagged with interest keywords and subscribers interest in different content are modeled by their degree of interest in these keywords. The popularity of a content is modeled by the distribution of user interests in the keywords associated with the content. However, the data requested by the nodes may be of large size that cannot be transferred in one contact. Similar to the work in [151], a central controller receives all subscriber requests over cellular links, and uses a set of helper vehicles that act as caches and do the actual transfer to subscribers using V2V links. It is also assumed that the central controller knows the distributions of the contact rates and the contact duration between vehicles and RSUs, with the parameters of the distribution assumed to be known from past data. Based on this information, the central controller decides the caching policy, to decide which helper should cache which content at what time subject to the content popularity and the fact that requesting vehicles may need to contact multiple helpers to get the entire content. It also schedules the transfers from the helpers to the subscribers. The goal is to maximize a function of expected satisfaction of user interests. A vehicle tries to get the requested data from helpers. The central controller monitors the transfers and the cellular network is used to push a content to a subscriber if a deadline is reached and the helpers are unable to transfer the content.Check the goal, Doesn't it have a cost issue?

Zhao et al. [156] address the use of vehicles as cache for delivery of content to offset the unavailability of cellular network at times in a Information Centric Network (ICN). The paper considers the privacy of vehicles, and groups vehicles into public (low privacy, high chance of caching) and private (high privacy, low chance of caching) vehicles. The private vehicles are further divided into three subclasses. Users can set their initial privacy ratings, which is then dynamically adjusted depending on the number of times a vehicle participates in caching. When a vehicle receives a content, the caching probability is determined based on three factors, a utility based on the privacy rating of the vehicle, a content similarity based on the name of the new content and the names of contents accessed recently, and a moving similarity based on the link stability between the two vehicles. A mechanism to reduce redundant caching is also introduced to enable caching only every so many hops. A popularity-prediction based cache replacement policy is also proposed. The goal is to reduce the delay and distance travelled by a vehicle to receive the desired content while utilizing the limited storage space of the vehicles.

In the work by Wu et al. [157], contents can be cached in vehicles, RSUs, and TV White space stations (TVWSS). Files that are cached are coded using MDS (Maximum Distance Separable) code so that only a subset of the coded packets are needed to reconstruct the entire content. The cellular base stations are supposed to have access to servers that contain all contents, coded content for ones that are cached elsewhere (so that if a subset is obtained from the RSU/TWSSS, the rest can be obtained from BS) and in non-coded form for the rest. A vehicle first checks nearby vehicles for content. Vehicles receiving a content cache it with a probability equal to its request rate. If the content is not available, the vehicle goes to RSUs and TWSSSs. Finally, if the entire content is not available from the caches within a deadline, it is downloaded directly from the base stations. A bipartite matching problem, with contents to be cached on one side and the caching infrastructure on the other, is formulated to allocate content to caches, and the well-known Gale-Shapely algorithm for stable matching is adopted to find the actual matching. The preference for each file to the infrastructure is based on average delay of access. On the other hand, RSU/TWSSS set their preference for the files based on the chance of request coming to them (files that have low probability of delivery through V2V and higher request rate are preferred). The goal is to minimize average content download delay while satisfying the storage constraints of the caches.

In the work by Deng et al. [158], there is a set of requesters and a set of helpers. Vehicles that want to be helpers advertise their available storage capacity to a central network operator, which chooses a subset of them as helpers based on some policy (the network operator may give some incentive to helpers also). A requester needing a content tries to get it from a nearby helper; if it cannot get the content within a delay bound, it is directly downloaded from the server using cellular links. A content is stored in a helper for a certain time, called its retention time. There is a cost associated with the storage at the helpers. The goal is to decide which content to cache in which helper and for how long so that a cost function comprising of the download cost of requesters (cost of downloading remaining content directly) and the storage cost of the helpers is minimized. An optimization problem is formulated, which is shown to be NP-hard. A polynomial time dynamic programming algorithm is proposed for a restricted special case and a heuristic algorithm is proposed and evaluated for the general case.

Zhang et al. [159] present a scenario where there are three entities, the base station, vehicles acting as caches, and users requesting for content. By default, a user always requests nearby caching vehicles for content. The caching vehicle may be able to serve the content itself from its cache, or inform the requesting vehicle if it cannot, which then goes to the base station to directly download the content. Future user demands are predicted at every time slot based on past demands using a neural network based learning scheme. This is used to decide on caching decisions of the vehicles at future time slots to minimize an objective function that tries to achieve high total throughput and low energy consumption.

Wang et al. [160] propose a scheme for disseminating dynamic maps among vehicles. Dynamic maps are maps of a region that are generated and updated continuously in response to certain events such as road work, accident related diversions etc. These are usually dynamic information overlaid on a static map, which the vehicle is assumed to already possess. Dynamic maps have a spatio-temporal validity, which means the updates have a region of interest (RoI) and must be delivered within a time interval. Instead of every vehicle downloading the map through cellular links, the paper proposes a scheme by which a RSU may generate the map or its updates and all vehicles may get it through other RSUs/vehicles as cache. A vehicle requiring a dynamic map first looks at nearby vehicles, then RSUs, and if still not obtained within a tolerable delay, pulls it from a central map application server (MAS) using cellular link. The tolerable delay is a function of the vehicle's distance from the region of interest. A vehicle caches a map received in its cache subject to availability of space. The cache replacement policy is based on three factors, the size of the map, the remaining time till expiry of the map, and the request probability of the map as calculated by the vehicle which is a simple function of the distance to the RoI. Based on these, a caching value is computed for each item in the cache which is a measure of how much cellular bandwidth can be saved by caching it. When a new item has to be cached and there is not sufficient space, the items with the lowest caching values whose total size exceeds the size of the new item to be cached are evicted from the cache.

VIII. Future Research Directions

From the discussion in the previous sections, it can be seen that the problem of data offload from cellular to vehicular networks has seen significant research efforts in recent times.

With the advent of 5G and with more and more new applications enabled, the demand and consequent load on the cellular network is expected to increase, and the data offload techniques will play a significant role in efficient use and cost reduction in accessing cellular networks. Looking at the research done so far, we highlight some potential research directions for future works in this area.

Most of the existing works on data offload address the network selection problem by considering higher level objectives such as cost saving, user request satisfaction etc. This gives rise to assumptions that a cellular or V2V network is always available for seamless use if selected, which may not always be true. However, network selection is also dependent on lower level network issues such as available channels/bandwidth, signal strength, ease of handover etc., which have been considered only in a few existing works. Ignoring such practical aspects in network selection may result in a poor choice. Investigating network selection for data offload keeping in consideration both higher level objectives and lower level networking constraints will be important for effective use of cellular data offloading.

We have identified network selection and caching as two major components in data offload techniques. However, it is seen that existing works have primarily focused on one or the other of the two, with very little work being done that focus on study of issues in integration of the two. As an example, in gateway based network selection schemes, a gateway can be considered as a cache for disseminating content to other vehicles. However, most works in this area assume that all gateways cache all content as long as needed, ignoring caching issues such as deciding which contents to cache in which vehicles, cache replacement etc. which are necessary to account for practical issues such as storage constraints etc. Similarly, most works that address caching focus more on caching policy, with little consideration of how effective network selection can affect caching decisions. We feel that studying the interdependence of network selection and caching and designing effective offload schemes that leverage both in an integrated manner will be an important area of research.

It is seen that all existing works consider the cellular network as a flat network of base stations. While this is true for most current generation cellular networks, more widespread use of Ultra Dense Networks (UDNs) in which a hierarchy of base stations (Macrocells, Femtocells etc.) with different ranges and other parameters will be present. The small range of some of the base stations (Femtocells) that vehicles may pass through may not make network switching viable in some cases. On the other hand, availability of a larger number of base stations with smaller ranges on the way may make for pre-planning network selection strategies more interesting. Designing data offload techniques that effectively leverage future UDNs can be an interesting area of future work.

Finally, existing works on cellular data offloading use only DSRC based vehicle to vehicle communications. It is discussed earlier that Mode 4 of LTE-V2X can be used for vehicle to vehicle communication without any central intervention, and hence, without any additional load on the base stations. The operating frequency range is also similar to the DSRC range, and hence has the same advantages of low cost and not interfering with the normal cellular communication. However, it has not been explicitly explored so far in the area of cellular data offloading. It will be interesting to see if the use of this cellular mode for vehicle to vehicle to communication can give any additional gains in performance or cost.

IX. Conclusions

In this paper, we have presented a survey of cellular data offload techniques in heterogeneous vehicular networks. In particular, we have focused on techniques to offload cellular traffic to vehicular networks in content upload/download scenarios where exactly one of source and destination is a vehicle and the other is a content server outside the vehicular network, and the vehicle can potentially use both cellular and vehicular access for content transfer. We have presented a classification of existing works in this area, and discussed each of the works in more details. At the same time, we have presented a brief review of existing work on data delivery in non-heterogeneous vehicular networks as a background, covering both lower level schemes for broadcasting/unicasting/multicasting of single messages as well as higher level schemes for content delivery. Finally we have identified some potential future research directions in cellular data offloading in heterogeneous vehicular networks. It is hoped that this survey will act as a good foundation for research in data offloading in heterogeneous vehicular networks.

References

[1] "Worldwide connected vehicle forecast, 2019–2023," *IDC Report US45050319*, 2019.
[2] "Global status report on road safety 2018," *World Health Organization*, 2018.
[3] "Fatality analysis reporting system (FARS) encyclopedia," *National Highway Traffic Safety Administration: Department of Transportation USA*.
[4] K. Zheng, Q. Zheng, P. Chatzimisios, W. Xiang, and Y. Zhou, "Heterogeneous vehicular networking: A survey on architecture, challenges, and solutions," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2377–2396, 2015.
[5] A. Zekri and W. Jia, "Heterogeneous vehicular communications: A comprehensive study," *Ad Hoc Networks*, vol. 75-76, pp. 52–79, 2018.
[6] H. Zhou, H. Wang, X. Chen, X. Li, and S. Xu, "Data offloading techniques through vehicular ad hoc networks: A survey," *IEEE Access*, vol. 6, pp. 65 250–65 259, 2018.
[7] https://www.onstar.com.
[8] https://www.bmw-connecteddrive.com.
[9] https://www.audi.com/en/experience-audi/models-and-technology/digital-services/audi-connect.html.
[10] https://www.ford.com/technology/sync/.
[11] https://www.toyota.com.
[12] https://www.volvocars.com/us/own/connected-car/apple-carplay.
[13] https://www.android.com/auto/.
[14] https://www.apple.com/in/ios/carplay/.
[15] "2019 urban mobility report," *The Texas A&M Transportation Institute, Texas A&M University, USA*, 2019.
[16] "Vehicle safety communications project: Task 3 final report," *Crash Avoidance Metric Partnership (CAMP) and US National Highway Traffic Safety Administration, USA*, 2005.
[17] "Dedicated short range communications (dsrc) message set dictionary," https://www.sae.org/standards/content/j2735/.
[18] https://www.dash.by/.
[19] https://automatic.com/.
[20] https://torque-bhp.com/.
[21] http://hobdrive.com.
[22] https://www.obdautodoctor.com/.

[23] https://caristaapp.com/.
[24] B. Baron, P. Spathis, H. Rivano, and M. D. de Amorim, “Offloading massive data onto passenger vehicles: Topology simplification and traffic assignment,” *IEEE/ACM Transactions on Networking*, vol. 24, no. 6, p. 3248–3261, 2016.
[25] J. Feng and Z. Feng, “A vehicle-assisted offloading scheme for hotspot base stations on metropolitan streets,” in *IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017.
[26] 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, 2014.
[27] J. Sanguesa, M. Fogue, P. Garrido, F. Martinez, J.-C. Cano, and C. Calafate, “A survey and comparative study of broadcast warning message dissemination schemes for VANETs,” *Mobile Information Systems*, 2016.
[28] F. A. Silva, A. Boukerche, T. R. M. B. Silva, L. B. Ruiz, E. Cerqueira, and A. A. F. Loureiro, “Vehicular networks: A new challenge for content-delivery-based applications,” *ACM Computing Surveys*, vol. 49, no. 1, 2016.
[29] G. G. Finn, “Routing and addressing problems in large metropolitan-scale internetworks,” *Technical Report ISI/RR-87-180, University of Southern California*, 1987.
[30] J. Tian, L. Han, and K. Rothermel, “Spatially aware packet routing for mobile ad hoc inter-vehicle radio networks,” in *IEEE International Conference on Intelligent Transportation Systems*, 2003.
[31] H. Fuller, J. Widmer, M. Kasemann, M. Mauve, and H. Hartenstein, “Contention-based forwarding for mobile ad hoc networks,” *Ad Hoc Networks*, vol. 1, no. 4, pp. 351–369, 2003.
[32] B.-C. Seet, G. Liu, B.-S. Lee, C.-H. Foh, K.-J. Wong, and K.-K. Lee, “A-STAR: A mobile ad hoc routing strategy for metropolis vehicular communications,” in *IFIP Networking*, 2004.
[33] F. Giudici and E. Pagani, “Spatial and traffic-aware routing (STAR) for vehicular systems,” in *High Performance Computing and Communications*, L. T. Yang, O. F. Rana, B. Di Martino, and J. Dongarra, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 77–86.
[34] M. Jerbi, R. Meraihi, S.-M. Senouci, and Y. Ghamri-Doudane, “GyTAR: Improved greedy traffic aware routing protocol for vehicular ad hoc networks in city environments,” in *3rd International Workshop on Vehicular Ad Hoc Networks (VANET’06)*, 2006.
[35] K. C. Lee, J. Haerri, U. Lee, and M. Gerla, “Enhanced perimeter routing for geographic forwarding protocols in urban vehicular scenarios,” in *IEEE Globecom Workshops*, 2007.
[36] Q. Yang, A. Lim, and P. Agrawal, “Connectivity aware routing in vehicular networks,” in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2008.
[37] S. Schnaufer and W. Effelsberg, “Position-based unicast routing for city scenarios,” in *International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2008.
[38] K. C. Lee, M. Le, J. Harri, and M. Gerla, “LOUVRE: Landmark overlays for urban vehicular routing environments,” in *IEEE 68th Vehicular Technology Conference (VTC-Fall)*, 2008.
[39] H. Saleet, R. Langar, K. Naik, R. Boutaba, A. Nayak, and N. Goel, “Intersection-based geographical routing protocol for VANETs: A proposal and analysis,” *IEEE Transactions on Vehicular Technology*, vol. 60, no. 9, pp. 4560–4574, 2011.
[40] S. Tsiachris, S. Koltsidas, and F. Pavlidou, “Junction-based geographic routing algorithm for vehicular ad hoc networks.” *Wireless Personal Communications*, vol. 71, pp. 955–973, 2013.
[41] P. K. Sahu, E. H. Wu, J. Sahoo, and M. Gerla, “BAHG: Back-bone-assisted hop greedy routing for VANET’s city environments,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 199–213, 2013.
[42] J. Zhao and G. Cao, “VADD: Vehicle-assisted data delivery in vehicular ad hoc networks,” *IEEE Transactions on Vehicular Technology*, vol. 57, no. 3, pp. 1910–1922, 2008.
[43] Y. Ding and L. Xiao, “SADV: Static-node-assisted adaptive data dissemination in vehicular networks,” *IEEE Transactions on Vehicular Technology*, vol. 59, no. 5, pp. 2445–2455, 2010.
[44] Xin Wang, Changle Li, Lina Zhu, and Chunchun Zhao, “An effective routing protocol for intermittently connected vehicular ad hoc networks,” in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2013.
[45] V. N. Soares, J. J. Rodrigues, and F. Farahmand, “GeoSpray: A geographic routing protocol for vehicular delay-tolerant networks,” *Information Fusion*, vol. 15, pp. 102–113, 2014.
[46] N. Mirjazaee and N. Moghim, “An opportunistic routing based on symmetrical traffic distribution in vehicular networks,” *Computer and Electrical Engineering*, vol. 47, no. C, pp. 1–12, 2015.
[47] P. Cheng, K. Lee, M. Gerla, and et al., “GeoDTN+NAV: Geographic DTN routing with navigator prediction for urban vehicular environments,” *Mobile Network and Applications*, vol. 15, pp. 61–82, 2010.
[48] K. Shafiee and V. C. Leung, “Connectivity-aware minimum-delay geographic routing with vehicle tracking in VANETs,” *Ad Hoc Networks*, vol. 9, no. 2, pp. 131–141, 2011.
[49] K. Mershad, H. Artail, and M. Gerla, “ROAMER: Roadside units as message routers in VANETs,” *Ad Hoc Networks*, vol. 10, no. 3, pp. 479–496, 2012.
[50] I. Leontiadis and C. Mascolo, “GeOpps: Geographical opportunistic routing for vehicular networks,” in *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2007.
[51] S. Martínez Tornell, C. T. Calafate, J. Cano, and P. Manzoni, “A map-based sensor data delivery protocol for vehicular networks,” in *The 11th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, 2012.
[52] G. Korkmaz, E. Ekici, F. Özgüner, and U. Özgüner, “Urban multi-hop broadcast protocol for inter-vehicle communication systems,” in *1st ACM International Workshop on Vehicular Ad Hoc Networks (VANET’04)*, 2004, pp. 76–85.
[53] T. Osafune, L. Lin, and M. Lenardi, “Multi-hop vehicular broadcast (MHVB),” in *6th International Conference on ITS Telecommunications*, 2006.
[54] D. Li, H. Huang, X. Li, M. Li, and F. Tang, “A distance-based directional broadcast protocol for urban vehicular ad hoc network,” in *International Conference on Wireless Communications, Networking and Mobile Computing (WICOM)*, 2007.
[55] N. Wisitpongphan, O. K. Tonguz, J. S. Parikh, P. Mudalige, F. Bai, and V. Sadekar, “Broadcast storm mitigation techniques in vehicular ad hoc networks,” *Wireless Communications*, vol. 14, no. 6, p. 84–94, 2007.
[56] M. M. I. Taha and Y. M. Y. Hasan, “VANET-DSRC protocol for reliable broadcasting of life safety messages,” in *IEEE International Symposium on Signal Processing and Information Technology*, 2007.
[57] S. Khakbaz and M. Fathy, “A reliable method for disseminating safety information in vehicular ad hoc networks considering fragmentation problem,” in *The Fourth International Conference on Wireless and Mobile Communications*, 2008.
[58] B. Blaszczyszyn, A. Laouiti, P. Muhlethaler, and Y. Toor, “Opportunistic broadcast in VANETs (OB-VAN) using active signaling for relays selection,” in *8th International Conference on ITS Telecommunications*, 2008.
[59] K. Suriyapaiboonwattana, C. Pornavalai, and G. Chakraborty, “An adaptive alert message dissemination protocol for VANET to improve road safety,” in *IEEE International Conference on Fuzzy Systems*, 2009.
[60] M. O. Cherif, S. Secouci, and B. Ducourthial, “How to disseminate vehicular data efficiently in both highway and urban environments?” in *IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications (WICOM)*, 2010.
[61] Qiong Yang and Lianfeng Shen, “A multi-hop broadcast scheme for propagation of emergency messages in VANET,” in *IEEE 12th International Conference on Communication Technology*, 2010.
[62] C. Sommer, O. K. Tonguz, and F. Dressler, “Adaptive beaconing for delay-sensitive and congestion-aware traffic information systems,” in *IEEE Vehicular Networking Conference*, 2010.
[63] R. S. Schwartz, R. R. R. Barbosa, N. Meratnia, G. Heijenk, and H. Scholten, “A simple and robust dissemination protocol for VANETs,” in *European Wireless Conference (EW)*, 2010.
[64] W. Viriyasitavat, O. K. Tonguz, and F. Bai, “UV-CAST: an urban vehicular broadcast protocol,” *IEEE Communications Magazine*, vol. 49, no. 11, pp. 116–124, 2011.
[65] R. I. Meneguette, G. Maia, E. R. M. Madeira, A. A. F. Loureiro, and L. A. Villas, “Autonomic data dissemination in highway vehicular ad hoc networks with diverse traffic conditions,” in *IEEE Symposium on Computers and Communications (ISCC)*, 2014.
[66] R. I. Meneguette, A. Boukerche, G. Maia, A. A. Loureiro, and L. A. Villas, “A self-adaptive data dissemination solution for intelligent transportation systems,” in *11th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, Ubiquitous Networks*, 2014.
[67] S. Gonzalez and V. Ramos, “Fast-OB-VAN: A fast opportunistic broadcast protocol for VANETs,” in *9th International Conference on Next Generation Mobile Applications, Services and Technologies*, 2015.

[68] R. Oliveira, C. Montez, A. Boukerche, and M. S. Wangham, "Reliable data dissemination protocol for VANET traffic safety applications," *Ad Hoc Networks*, vol. 63, pp. 30–44, 2017.
[69] J. Sospeter, D. Wu, S. Hussain, and T. Tesfa, "An effective and efficient adaptive probability data dissemination protocol in VANET," *Data*, vol. 4, p. 1, 12 2018.
[70] H. Alshaer and E. Horlait, "An optimized adaptive broadcast scheme for inter-vehicle communication," in *IEEE 61st Vehicular Technology Conference (VTC-Spring)*, vol. 5, 2005.
[71] A. Wegener, H. Hellbrück, S. Fischer, C. Schmidt, and S. Fekete, "AutoCast: An adaptive data dissemination protocol for traffic information systems," in *IEEE 66th Vehicular Technology Conference (VTC-Fall)*, 2007.
[72] M. Slavik and I. Mahgoub, "Stochastic broadcast for VANET," in *7th IEEE Conference on Consumer Communications and Networking Conference (CCNC)*, 2010.
[73] S. Gonzalez and V. Ramos, "Preset delay broadcast: a protocol for fast information dissemination in vehicular ad hoc networks (VANETs)," *EURASIP Journal on Wireless Communications and Networking*, 2016.
[74] K. Suriyapaibonwattana and C. Pomavalai, "An effective safety alert broadcast algorithm for VANET," in *International Symposium on Communications and Information Technologies*, 2008.
[75] F. J. Martinez, M. Fogue, M. Coll, J.-C. Cano, C. T. Calafate, and P. Manzoni, "Evaluating the impact of a novel warning message dissemination scheme for VANETs using real city maps," in *IFIP Networking*, 2010.
[76] O. K. Tonguz, N. Wisitpongphan, and F. Bai, "DV-CAST: A distributed vehicular broadcast protocol for vehicular ad hoc networks," *IEEE Wireless Communications*, vol. 17, no. 2, pp. 47–57, 2010.
[77] Nawut Na Nakorn and Kultida Rojviboonchai, "DECA: Density-aware reliable broadcasting in vehicular ad hoc networks," in *ECTI-CON2010: The 2010 ECTI International Confernce on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 2010.
[78] G. Maia, A. L. Aquino, A. Viana, A. Boukerche, and A. A. Loureiro, "HyDi: A hybrid data dissemination protocol for highway scenarios in vehicular ad hoc networks," in *Second ACM International Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications (DIVANet)*, 2012.
[79] M. Fogue, P. Garrido, F. J. Martinez, J.-C. Cano, C. T. Calafate, and P. Manzoni, "Evaluating the impact of a novel message dissemination scheme for vehicular networks using real maps," *Transportation Research Part C: Emerging Technologies*, vol. 25, pp. 61–80, 2012.
[80] J. Sanguesa, M. Fogue, P. Garrido, F. J. Martinez, J.-C. Cano, C. T. Calafate, and P. Manzoni, "On the selection of optimal broadcast schemes in VANETs," in *16th ACM International Conference on Modeling, Analysis Simulation of Wireless and Mobile Systems (MSWiM)*, 2013, p. 411–418.
[81] J. A. Sanguesa, M. Fogue, P. Garrido, F. J. Martinez, J. Cano, and C. T. Calafate, "Topology-based broadcast schemes for urban scenarios targeting adverse density conditions," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2014.
[82] J. A. Sanguesa, M. Fogue, P. Garrido, F. J. Martinez, J.-C. Cano, and C. T. Calafate, "Using topology and neighbor information to overcome adverse vehicle density conditions," *Transportation Research Part C: Emerging Technologies*, vol. 42, pp. 1–13, 2014.
[83] M. S. Rayeni, A. Hafid, and P. K. Sahu, "Dynamic spatial partition density-based emergency message dissemination in VANETs," *Vehicular Communications*, vol. 2, no. 4, pp. 208–222, 2015.
[84] M. Bakhouya, J. Gaber, and P. Lorenz, "An adaptive approach for information dissemination in vehicular ad hoc networks," *Journal of Network and Computer Applications*, vol. 34, no. 6, pp. 1971–1978, 2011.
[85] J. Blum, A. Eskandarian, and L. Hoffman, "Mobility management in IVC networks," in *IEEE Intelligent Vehicles Symposium*, 2003.
[86] T. Song, T. Song, and L. Shen, "A cluster-based directional routing protocol in VANET," in *IEEE 12th International Conference on Communication Technology*, 2010.
[87] A. Benslimane and A. Bachir, "Inter-vehicle geocast protocol supporting non-equipped GPS vehicles," in *Ad-Hoc, Mobile, and Wireless Networks (ADHOC-NOW)*, 2003.
[88] C. Maihöfer, R. Eberhardt, and E. Schoch, "CGGC: cached greedy geocast," in *International Conference on Wired/Wireless Internet Communications (WWIC)*, 02 2004.
[89] C. Maihöfer, T. Leinmüller, and E. Schoch, "Abiding geocast: Time–stable geocast for ad hoc networks," in *2nd ACM International Workshop on Vehicular Ad Hoc Networks (VANET'05)*, 2005.
[90] W. Sun, H. Yamaguchi, K. Yukimasa, and S. Kusumoto, "GVGrid: A QoS routing protocol for vehicular ad hoc networks," in *14th IEEE International Workshop on Quality of Service (IWQoS)*, 2006.
[91] M. Kihl, M. Sichitiu, and H. Joshi, "Design and evaluation of two geocast protocols for vehicular ad-hoc networks," *Journal of Internet Engineering*, vol. 2, 2008.
[92] T. Atechian and L. Brunie, "DG-CastoR for query packets dissemination in vanet," in *5th IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, 2008.
[93] Y. Chen, Y. Lin, and S. Lee, "A mobicast routing protocol in vehicular ad-hoc networks," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 2009.
[94] H. Rahbar, K. Naik, and A. Nayak, "DTSG: Dynamic time-stable geocast routing in vehicular ad hoc networks," in *9th IFIP Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, 2010.
[95] W. K. Wolterink, G. Heijenk, and G. Karagiannis, "Constrained geocast to support cooperative adaptive cruise control (CACC) merging," in *IEEE Vehicular Networking Conference*, 2010.
[96] A. Lakas and M. Shaqfa, "Geocache: Sharing and exchanging road traffic information using peer-to-peer vehicular communication," in *IEEE 73rd Vehicular Technology Conference (VTC-Spring)*, 2011.
[97] B. B. Chen and M. C. Chan, "MobTorrent: A framework for mobile internet access from vehicles," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2009.
[98] U. Acer, P. Giaccone, D. Hay, G. Neglia, and S. Tarapiah, "Timely data delivery in a realistic bus network," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2011.
[99] M. J. Khabbaz, W. F. Fawaz, and C. M. Assi, "Which vehicle to select?" *IEEE Communications Letters*, vol. 16, no. 6, pp. 812–815, June 2012.
[100] F. A. Silva, A. Boukerche, T. R. B. Silva, L. B. Ruiz, and A. A. Loureiro, "Geo-localized content availability in VANETs," *Ad Hoc Networks*, vol. 36, no. 2, pp. 425–434, 2016.
[101] R. I. Meneguette, A. Boukerche, F. A. Silva, L. Villas, L. B. Ruiz, and A. A. F. Loureiro, "A novel self-adaptive content delivery protocol for vehicular networks," *Ad Hoc Networks*, vol. 73, pp. 1–13, 2018.
[102] F. Malandrino, C. Casetti, and C.-F. Chiasserini, "Content discovery and caching in mobile networks with infrastructure," *IEEE Transactions on Computers*, vol. 61, pp. 1507–1520, 10 2012.
[103] T. H. Luan, L. X. Cai, J. Chen, X. Shen, and F. Bai, "VTube: Towards the media rich city life with autonomous vehicular content distribution," in *8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, 2011.
[104] M. Fiore, C. Casetti, and C.-F. Chiasserini, "On-demand content delivery in vehicular wireless networks," in *8th ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, 2005.
[105] A. Nandan, S. Das, G. Pau, M. Gerla, and M. Y. Sanadidi, "Co-operative downloading in vehicular ad-hoc wireless networks," in *Second Annual Conference on Wireless On-Demand Network Systems and Services (WONS)*, 2005.
[106] M. Sardari, F. Hendessi, and F. Fekri, "Infocast: A new paradigm for collaborative content distribution from roadside units to vehicular networks," in *6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, 2009.
[107] M. Fiore, F. Mininni, C. Casetti, and C. . Chiasserini, "To cache or not to cache?" in *IEEE International Conference on Computer Communications (INFOCOM)*, 2009.
[108] D. Borsetti, M. Fiore, C. Casetti, and C.-F. Chiasserini, "An application-level framework for information dissemination and collection in vehicular networks," *Performance Evaluation*, vol. 68, pp. 876–896, 2011.
[109] M. Fiore, C. Casetti, C. Chiasserini, and D. Borsetti, "Persistent localized broadcasting in VANETs," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 480–490, 2013.
[110] N. Kumar, J. Rodrigueaz, J. Lloret, and et al., "Replication-aware data dissemination for vehicular ad hoc networks using location determination," *Mobile Networks and Applications*, vol. 20, pp. 251–267, 2015.
[111] C. Campolo, M. Amadeo, and A. Molinaro, "CRoWN: Content-centric networking in vehicular ad hoc networks," *IEEE Communications Letters*, vol. 16, no. 9, pp. 1380–1383, 2012.
[112] S. Ahmed and S. Kanhere, "VANETCODE: network coding to enhance cooperative downloading in vehicular ad-hoc networks," in *International Conference on Wireless Communications and Mobile Computing*, 2006.
[113] S. Yoon, D. Ha, H. Ngo, and C. Qiao, "MoPADS: A mobility profile aided file downloading service in vehicular networks," *IEEE*

Transactions on Vehicular Technology, vol. 58, no. 9, pp. 5235–5246, 2009.

[114] H. Cheng and K. C. Lin, "Source selection and content dissemination for preference-aware traffic offloading," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 11, pp. 3160–3174, 2015.

[115] Yi Cao, Jinhua Guo, and Yue Wu, "SDN enabled content distribution in vehicular networks," in *Fourth International Conference on the Innovative Computing Technology (INTECH)*, 2014.

[116] J. Sanguesa, M. Fogue, P. Garrido, F. Martinez, J.-C. Cano, C. Calafate, and P. Manzoni, "RTAD: a real-time adaptive dissemination system for vanets," *Computer Communications*, vol. 60, pp. 53–70, 2015.

[117] J. Jeong, S. Guo, Y. Gu, T. He, and D. Du, "Trajectory-based statistical forwarding for multihop infrastructure-to-vehicle data delivery," vol. 11, no. 10, pp. 1523–1537, 2011.

[118] J. Whitbeck, Y. Lopez, J. Leguay, V. Conan, and M. D. de Amorim, "Push-and-track: Saving infrastructure bandwidth through opportunistic forwarding," *Pervasive and Mobile Computing*, vol. 8, no. 5, pp. 682–697, 2012.

[119] F. Rebecchi, M. deAmorim, and V. Conan, "DROid: Adapting to individual mobility pays off in mobile data offloading," in *IFIP Networking*, 2014.

[120] S. Jia, S. Hao, X. Gu, and L. Zhang, "Analyzing and relieving the impact of FCD traffic in LTE-VANET heterogeneous network,," in *21st International Conference on Telecommunications (ICT)*, 2014.

[121] Y. Hong, Z. Deze, H. Huawei, G. Song, A. Barnawi, and I. Stojmenovic, "Opportunistic offloading of deadline-constrained bulk cellular traffic in vehicular DTNs," *IEEE Transactions on Computers*, vol. 64, no. 12, pp. 3515–3527, 2015.

[122] M. Lee, J. Song, J. Jeong, and T. Kwon, "DOVE: Data offloading through spatio-temporal rendezvous in vehicular networks," in *24th International Conference on Computer Communication and Networks (ICCCN)*, 2015.

[123] Q. Yuan, J. Li, Z. Liu, and F. Yang, "Space and time constrained data offloading in vehicular network," in *IEEE 18th International Conference on High-Performance Computing and Communications, IEEE 14th International Conference on Smart City, and IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, 2016.

[124] F. Mezghani, R. Dhaou, M. Nogueira, and A. Beylot, "Offloading cellular networks through V2V communications how to select the seed-vehicles?" in *IEEE International Conference on Communications (ICC)*, 2016.

[125] C. Huang, M. Chiang, D. Dao, H. Pai, and S. Huang, "Vehicle-to-infrastructure (V2I) offloading from cellular network to 802.11p Wi-Fi network based on the software-defined network (SDN) architecture," *Vehicular Communication*, vol. 9, pp. 288–300, 2017.

[126] L. Pescosolido, M. Conti, and A. Passarella, "D2D data offloading in vehicular environments with optimal delivery time selection,," *Computer Communication*, vol. 146, pp. 63–84, 2019.

[127] A. Benslimane and T. Taleb, "Dynamic clustering-based adaptive mobile gateway management in integrated VANET-3G heterogeneous wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 3, pp. 559–570, 2011.

[128] M. Gramaglia, C. Bernardos, and M. Calderon, "Seamless internet 3G and opportunistic WLAN vehicular connectivity," *EURASIP Journal on Wireless Communications and Networking*, pp. 1–20, 2011.

[129] R. Stanica, M. Fiore, and F. Malandrino, "Offloading floating car data," in *IEEE 14th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2013.

[130] P. Salvo, I. Turcanu, F. Cuomo, A. Baiocchi, and I. Rubin, "LTE floating car data application offloading via VANET driven clustering formation," in *12th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, 2016.

[131] K. Ju, L. Chen, H. Wei, and K. Chen, "An efficient gateway discovery algorithm with delay guarantee for VANET-3G heterogeneous networks," *Wireless Personal Communications*, vol. 77, pp. 2019–2036, 2014.

[132] G. Zhioua, H. Labiod, N. Tabbane, and S. Tabbane, "An efficient QoS based gateway selection algorithm for VANET to LTE advanced hybrid cellular network," in *15th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, 2012.

[133] F. Malandrino, C. Casetti, C. Chiasserini, and M. Fiore, "Content download in vehicular networks in presence of noisy mobility prediction," *IEEE Transactions on Mobile Computing*, vol. 13, no. 5, pp. 1007–1021, 2014.

[134] G. M. Zhioua, H. Labiod, N. Tabbane, and S. Tabbane, "A traffic QoS aware approach for cellular infrastructure offloading using VANETs," in *IEEE 22nd International Symposium of Quality of Service (IWQoS)*, 2014.

[135] M. Felice, F. Cuomo, A. Baiocchi, I. Turcanu, and S. Zennaro, "Traffic monitoring and incident detection using cellular and early stage VANET technology deployment," in *First International Workshop on Internet of Vehicles and Vehicles of Internet*, 2016.

[136] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3G using WiFi," in *8th International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2010.

[137] H. Xiaoxiao, P. Deshpande, and S. Das, "Moving bits from 3G to metro-scale wifi for vehicular network access: An integrated transport layer solution," in *19th IEEE International Conference on Network Protocols (ICNP)*, 2011.

[138] V. Siris and D. Kalyvas, "Enhancing mobile data offloading with mobility prediction and prefetching," in *Seventh ACM International Workshop on Mobility in the Evolving Internet Architecture (MobiArch)*, 2012.

[139] S. Lee and S. Lee, "User-centric offloading to WLAN in WLAN/3G vehicular networks,," *Wireless Personal Communications*, vol. 70, pp. 1925–1940, 2013.

[140] R. Bruno, A. Masaracchia, and A. Passarella, "Offloading through opportunistic networks with dynamic content requests," in *IEEE 11th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, 2014.

[141] N. Wang and J. Wu, "Opportunistic wifi offloading in a vehicular environment: Waiting or downloading now?" in *The 35th Annual IEEE International Conference on Computer Communications (INFOCOM)*, 2016.

[142] P. Si, Y. He, H. Yao, R. Yang, and Y. Zhang, "DaVe: Offloading delay tolerant data traffic to connected vehicle networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 3941–3953, 2016.

[143] H. He, H. Shan, A. Huang, and S. Long, "Resource allocation for video streaming in heteregoeneous cognitive vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 7917–7930, 2016.

[144] C. Xu, W. Quan, A. Vasilakos, H. Zhang, and G.-M. Muntean, "Information-centric cost-efficient optimization for multimedia content delivery in mobile vehicular networks," *Computer Communication*, vol. 99, pp. 93–106, 2017.

[145] T. Yu, X. Zhu, H. Chen, and M. Maheswaran, "Hetcast: Cooperative data delivery on cellular and road side network," in *IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017.

[146] Y. Sun, L. Xu, Y. Tang, and W. Zhuang, "Traffic offloading for online video service in vehicular networks: A cooperative approach," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7630–7642, 2018.

[147] P. Dai, K. Liu, X. Wu, Z. Yu, H. Xing, and V. Lee, "Cooperative temporal data dissemination in SDN-based heterogeneous vehicular networks,," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 72–83, 2019.

[148] O. Rhaiem, L. Fourati, and W. Ajib, "Multi-segment cooperative transmission of scalable video streaming over vehicular networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2019.

[149] A. Al-Hilo, D. Ebrahimi, S. Sharafeddine, and C. Assi, "Revenue-driven video delivery in vehicular networks with optimal resource scheduling," *Vehicular Communication*, vol. 23, 2020.

[150] M. Gramaglia, M. Calderon, and C. J. Bernardos, "Trebol: Tree-based routing and address autoconfiguration for vehicle-to-internet communications," in *IEEE 73rd Vehicular Technology Conference (VTC-Spring)*, 2011, pp. 1–5.

[151] Y. Li, D. Jin, Z. Wang, L. Zeng, and S. Chen, "Coding or not: Optimal mobile data offloading in opportunistic vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 318–333, 2014.

[152] J. Ahn, M. Sathiamoorthy, B. Krishnamachari, F. Bai, and L. Zhang, "Optimizing content dissemination in vehicular networks with radio heterogeneity," *IEEE Transactions on Mobile Computing*, vol. 13, no. 6, pp. 1312–1325, 2014.

[153] L. Vigneri, T. Spyropoulos, and C. Barakat, "Storage on wheels: Offloading popular contents through a vehicular cloud," in *2016 IEEE 17th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2016.

[154] C. Bian, T. Zhao, X. Li, X. Du, M. Guizani, and W. Yan, "Quantifying caching effects in urban VANETs," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2016.

[155] X. Zhu, Y. Li, D. Jin, and J. Lu, "Contact-aware optimal resource allocation for mobile data offloading in opportunistic vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7384–7399, 2017.

[156] W. Zhao, Y. Qin, D. Gao, C. H. Foh, and H. Chao, "An efficient cache strategy in information centric networking vehicle-to-vehicle scenario," *IEEE Access*, vol. 5, pp. 12 657–12 667, 2017.

[157] H. Wu, W. Xu, J. Chen, L. Wang, and X. Shen, "Matching-based content caching in heterogeneous vehicular networks," in *IEEE Global Communications Conference (GLOBECOM)*, 2018.

[158] T. Deng, P. Fan, and D. Yuan, "Optimizing retention-aware caching in vehicular networks," *IEEE Transactions on Communications*, vol. 67, no. 9, pp. 6139–6152, 2019.

[159] Y. Zhang, C. Li, and T. H. Huan, "Prediction based vehicular caching: Where and what to cache?" *Mobile Networks and Applications*, 2019.

[160] R. Wang, S. Biswas, S. Das, and J. Rao, "Collaborative caching for dynamic map dissemination in vehicular networks," in *IEEE ComSoc International Communications Quality and Reliability Workshop (CQR)*, 2019, pp. 1–6.