Computation Offloading for Vehicular Environments: A Survey

ALISSON B. DE SOUZA\textsuperscript{1}, PAULO A. L. REGO\textsuperscript{1}, (MEMBER, IEEE), TIAGO CARNEIRO\textsuperscript{2}, JARDEL DAS C. RODRIGUES\textsuperscript{3}, P. P. REBOUÇAS FILHO\textsuperscript{3}, (MEMBER, IEEE), JOSÉ N. DE SOUZA\textsuperscript{1}, (SENIOR MEMBER, IEEE), VINAY CHAMOLA\textsuperscript{5}, (SENIOR MEMBER, IEEE), VICTOR HUGO C. DE ALBUQUERQUE\textsuperscript{4}, (SENIOR MEMBER, IEEE), BIPLAB SIKDAR\textsuperscript{6}, (SENIOR MEMBER, IEEE)

\textsuperscript{1}GREat, Federal University of Ceará, Fortaleza, CE, Brazil. (e-mails: alisson@ufc.br, pauloalr@ufc.br, neuman@ufc.br)
\textsuperscript{2}INRIA Lille - Nord Europe, Lille, France. (e-mail: tiago.carneiro-pessoa@inria.fr)
\textsuperscript{3}Lapisco, IFCE, Fortaleza, CE, Brazil. (e-mails: Rodrigues.jardel@ifce.edu.br, pedrosarf@ifce.edu.br)
\textsuperscript{4}Universidade de Fortaleza, Fortaleza, CE, Brazil. (e-mail: victor.albuquerque@unifor.br)
\textsuperscript{5}Department of Electrical and Electronics Engineering, Birla Institute of Technology & Science (BITS), Pilani, Rajasthan, India. (e-mail: vinay.chamola@pilani.bits-pilani.ac.in)
\textsuperscript{6}Department of Electrical and Computer Engineering, National University of Singapore, Singapore. (e-mail: bsikdar@nus.edu.sg)

Corresponding author: Biplab Sikdar (e-mail: bsikdar@nus.edu.sg).

**ABSTRACT** With significant advances in communication and computing, modern day vehicles are becoming increasingly intelligent. This gives them the ability to contribute to safer roads and passenger comfort through network devices, cameras, sensors, and computational storage and processing capabilities. However, to run new and popular applications, and to enable vehicles operating autonomously requires massive computational resources. Computational resources available with the current day vehicles are not sufficient to process all these demands. In this situation, other vehicles, edge servers, and servers in remote data centers can help the vehicles by lending their computing resources. However, to take advantage of these computing resources, computation offloading techniques have to be leveraged to transfer tasks or entire applications to run on other devices. Such computation offloading can lead to improved performance and Quality of Service (QoS) for applications and for the network. However, computation offloading in a highly dynamic environment such as vehicular networks is a major challenge. Therefore, this survey aims to review and organize the computation offloading literature in vehicular environments. In addition, we demystify some concepts, propose a taxonomy with the most important aspects and classify most works in the area according to each category. We also present the main tools, scenarios, subjects, strategies, objectives, etc., used in the works. Finally, we present the main challenges and future directions to guide future research in this active research area.


I. INTRODUCTION

In the latest update from the World Health Organization (WHO), in January 2020, there were more than 2 billion registered vehicles around the world [1]. That number continues to increase, theoretically generating more comfort, convenience and efficiency for people, but also generating traffic congestion, accidents and pollution [2]. Significant efforts have been made to improve the situation. Among these efforts, the Vehicular Ad Hoc Networks (VANETs) that are used in Intelligent Transportation Systems (ITS) stand out. VANETs provide connectivity to vehicles through vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I) and vehicle-to-everything (V2X) communications. VANETs allow the exchange of information even in places without infrastructure [3], [4], [5]. Through these connections, vehicles can send alerts of collisions, accidents, overtaking, congestion, etc., and improve road safety [2], [6], [7]. In this way, the National Highway Traffic Safety Administration (NHTSA) of the United States estimates that adoption of only two V2X safety applications would save about 1,000 lives and prevent about half a million crashes a year [8]. In addition, these connections allow vehicle journeys to be more pleasurable for drivers and passengers by allowing the execution of entertainment applications [9]. Thus, different players have developed...
and adopted vehicular communications into academia and manufacturing of devices and vehicles. In fact, this market has been moving millions of dollars and is expected to move billions of dollars in the coming years [10], [11].

In addition to advancing communications, vehicles have also become intelligent and evolved in cameras, embedded systems, sensors, Advanced Driver-Assistance Systems (ADAS) and computing power. In fact, vehicles are increasingly being produced with powerful on-board computing capabilities. In this way, they can act as servers to assist in the processing of other vehicles, mobile devices of passengers or of people walking on the sidewalks, and even to relieve the computational load of other devices on the network [2].

However, new and popular applications have emerged, such as applications based on artificial intelligence, augmented reality (AR), image-aided navigation, intelligent vehicle control, gaming, etc. Such applications demand massive computation and storage resources to handle complicated data processing and storage operations, and still have critical latency requirements [12], [13]. In addition, it is expected that with hundreds of sensors in future vehicles generating an enormous amount of data, there will be a lot of pressure on the computational resources of the vehicles. The processing power required to process data from future vehicles will easily deplete the vehicles’ on-board resources. For example, a study predicts that each vehicle will need approximately $10^6$ DMIPS of computational power and must be able to process real-time traffic conditions within a latency of 100 ms to enable autonomous vehicle steering [14], [2]. Also, approximately 1 GB of data will need to be processed every second on real-time operating systems of self-driving cars [15].

Unfortunately, vehicles do not yet have enough on-board computing resources to handle all of these requirements and are still limited compared to the scale of data that needs to be stored and processed. In fact, it is a challenge to guarantee the proper quality of service (QoS). Even if more powerful processors such as Graphics Processing Units (GPUs) are installed, it can lead to high energy consumption due to the greater power and the need for cooling to meet thermal restrictions. In this way, fuel efficiency of the vehicle and driving range can be significantly affected [2].

In this sense, computing paradigms such as Cloud Computing will certainly be an important way to help intelligent vehicles, but may not be enough. While computing power, cost and power consumption are the main limiting factors for on-board computing, long latency and massive data transmission are the bottlenecks of cloud-based processing. In addition, this latency depends on the condition of the wireless channel, network bandwidth, and traffic congestion. Therefore, real-time processing and reliability cannot be guaranteed. In this way, cloud resources can be used for heavier processing, for long-term storage and for non-real-time processing [2].

To overcome the limited resources of in-vehicle computing, communication, storage and power while avoiding excessive latency in cloud computing, deploying computing resources at the edge of the wireless network has received significant attention from academia and industry. In fact, it is possible to deliver cloud services directly from the edge of the network and support delay-sensitive applications and also meet the low-latency requirement for mission-critical tasks [16], [17]. For example, highly complicated tasks such as the powerful Convolutional Neural Network (CNN) for feature extraction and vision-based perception tasks can run on nearby edge servers with more computing resources to help intelligent vehicles [2]. Another way to provide cloud services at the edge of the network is through the vehicles themselves, i.e., a cloud composed of vehicles. In this way, groups of vehicles can have better computing power and can provide services such as processing, connectivity, data collection, and storage, among others. The resources of these vehicles can be used for data preprocessing to reduce the use of bandwidth [18], [19].

With these computational resources available at different levels of the network, in the traditional cloud, in groups of vehicles, or in edge servers coupled to roadside infrastructures or base stations, it is necessary to take advantage of them to meet the requirements of the applications. In this sense, a technique that can help is called computation offloading, also known as computing offloading, computational offloading, computation as a service, computing as a service, task offloading, workload offloading, or cyber foraging [20], [21]. This technique consists of partitioning tasks of an application and sending them to be performed on other devices, as seen in Figure 1. In this case, these device have accessible and possibly idle or underutilized computing resources [22]. Computation offloading can be used to increase the computational capabilities of devices, for performance gains, or to decrease the processing load on an overloaded device by migrating parts of an application (or the whole application) to a remote computing device that may lend their computing power [23], [24], [25], [26], [27], [28], [29].

Existing computation offloading solutions have applied various strategies and mechanisms to manipulate steps in the offloading process, such as device discovery, resource profiling, application partitioning, and offloading decision [30]. Computation offloading can be classified as static when the developer or system defines before execution (in the project or at the time of installation) which parts of the application should be downloaded and where. It can also be classified as dynamic when the framework/system decides at run-time.
which parts of the application should be downloaded and where to download, based on metrics related to the current network condition, mobile devices, and remote server.

However, it is very important to use computational offloading systems that can handle the challenges, of vehicular networks, have reliable communications and improve the performance of applications [31], [32], [33]. Therefore, the integration of computation offloading techniques, computing paradigms, and vehicular networks will play a vital role in the development of intelligent vehicles.

Thus, our survey aims to demystify concepts, aggregate and classify state-of-the-art research on computation offloading for vehicular environments and present its challenges. Important subjects are covered and we propose a complete taxonomy of the works in this domain.

A. EXISTING SURVEYS AND TUTORIALS

Although there are several surveys that deal with computation offloading and others that deal with vehicular networks, there is still a lack of a work that completely analyzes the integration of the two themes.

A comprehensive review of opportunistic offloading is presented in [26]. Such type of offloading can refer to both traffic/data offloading for content distribution and computation offloading for task distribution. In computation offloading, the classification is subdivided into with wired networks, with/without node selection, single/multi-objective(s) and tasks from inside/outside of the cluster. Discussions about the subject are also presented, as well as research directions, offloading strategies and future research problems. However, this survey shows opportunistic offloading in general and, in most cases, focuses on mobile devices and both traffic and computation offloading. In this way, it is not specific for computation offloading or for vehicular scenarios.

The survey by Zhou et al. [34] investigated data offloading techniques through Vehicular Ad Hoc Networks. Based on communications standards between vehicles and infrastructures, the authors classify data offloading through vehicle-to-vehicle communications, vehicle-to-infrastructure communications, and vehicle-to-everything communications. The pros and cons for some data offloading techniques in VANETs are also discussed. Then, challenges and open research problems are presented. Nonetheless, although the paper is specific to VANETs, it does not deal with computation offloading.

Lin et al. [35] analyzed computation offloading for edge computing. An insight into the architecture and types of edge computing nodes was given. Based on this overview, the paper reviews the challenges of computing offloading in terms of application partitioning, task allocation, and task execution, focusing on features for edge computing. Then, some application scenarios for edge computing are presented, such as real-time video analytics, smart “things”, vehicular applications, and cloud gaming. Subsequent, opportunities and future research directions are discussed. Nevertheless, the work presents computation offloading for edge computing in general and not specific for vehicular networks.

In 2019, Jiang et al. [36] discussed computation offloading in edge computing. Some aspects of computation offloading related to edge computing are surveyed such as: partitioning, what/when/where to offload, energy consumption minimization, Quality of Services guarantee, and Quality of Experiences enhancement. Case studies of cooperation between edge and cloud, resource scheduling approaches, gaming and trade-off among system performance and overheads for computation offloading decision making are also reviewed. Even so, the paper focuses on computation offloading related to mobile devices and edge computing, and not to vehicular scenarios.

In 2019, the authors of [2] provided partial coverage of the subject. The survey analyzes the latest developments in edge computing for Internet of Vehicles. It also presents important design problems, methodologies, typical use cases, hardware platforms and open research problems for computing in vehicular networks. Cases where vehicles act as clients and servers, and cases where edge servers coupled to base stations are used are described. Edge caching cases for content distribution and artificial intelligence models for use on the edge are also investigated. However, the article does not focus on the various aspects of computation offloading for vehicular scenarios and presents several issues not directly related to this.

In 2020, a survey focusing on computation offloading and vehicular networks was presented at [12]. The analyzed scenarios are those where the vehicles act as edge nodes for mobile devices with restriction of computational resources. In addition, the work presents some challenges in the area and focuses on solutions of computation offloading and classifying them according to the techniques of partitioning, scheduling and data retrieval. However, this survey focuses only on the part of algorithms and data retrieval, in cases where mobile devices (smartphones, tablets, wearables, etc.) act as clients, in cases where the servers are on the edge of the network, and cases where the task is completely offloaded. Our survey shows several types of clients and different servers, presents a complete taxonomy, and focuses on all aspects of offloading works, from technology to the types of experiments.

Therefore, our survey investigates, deepens and classifies in a complete way all topics related to the intersection between computation offloading and vehicular networks. A taxonomy is also proposed to classify these topics in several categories. In Table 1, we show the summary comparison of the above-mentioned surveys based on features, such as year, focus on computation offloading and vehicular networks, and complete deepening on the intersection between computation offloading and vehicular networks.

B. CONTRIBUTIONS

According to Table 1, although some surveys deal with parts of the subject addressed here in this paper, they do not focus on the intersection between computation offloading and vehicular networks, or when they do, they do not present a
complete comprehensive outlook of all aspects of computation offloading. To the best of our knowledge, this work is the first to present and classify all the most important details of the computation offloading works in VANETs, from the technologies used, to details of the experiments used to validate proposals. Thus, we highlight our main contributions below.

- Demystification of important related concepts that are often confused in literature papers.
- Taxonomy and complete classification of the main aspects of computation offloading for vehicular environments, including communication standards, problems and experiments.
- Organization, description and identification of the main tools, subjects, scenarios, strategies, objectives, etc., used in most works in the area, including graphics to facilitate visualization.
- Presentation of the main challenges and problems in the area to guide future research.
- Tables with the categorization of most articles in the area of computation offloading in VANETs.

### C. SURVEY ORGANIZATION

The rest of this survey is organized as follows: Section II presents some concepts widely used in the field of computing in vehicular environments and their differences; in the Section III and in the Tables at the end of this paper, we provide a taxonomy, classify several articles according to the proposed categories and approach the main subjects used in the works; Section IV identifies the open issues and challenges to guide future research; finally, in Section V, we present the conclusion.

### II. DEMYSTIFYING RELATED CONCEPTS

In the literature, there are several concepts related to cloud, edge and fog and computing in vehicles and other devices that have subtle differences from each other. Therefore, below, we analyze the six main concepts. The first three deal with types of clouds (or remote execution environments) called the Vehicular Cloud (VC), Edge and Traditional Cloud (TC). The other three concepts are the main paradigms related to the subject of our survey: Vehicular Cloud Computing (VCC), Vehicular Edge Computing (VEC) and Vehicular Fog Computing (VFC).

### A. VEHICULAR CLOUD (VC)

The concept of Vehicular Cloud (VC) emerged to make better use of the computing, communication and storage resources of vehicles [37]. Although there are several similar terms as Vehicular Cloudlet (VCL) [40], [41], Vehicular Cloudlet (Vc) [41], V-Cloud [42], and Vehicular Ad hoc Cloud (VACloud) [22], we use the term Vehicular Cloud. In this type of cloud, vehicle owners can lend or rent the surplus computing resources on board, similar to what traditional cloud providers do. The computational resources of two or more vehicles, stationary or in motion, can be gathered, coordinated, allocated and offered, through V2X (Vehicle-to-Everything) connections, as a real-time service for other vehicles or customers to use [43], [31], [44].

A Vehicular Cloud (VC), as seen in Figure 2, is a pool of vehicular computing resources that can be dynamically coordinated and to offer services on demand, through V2V connections (orange lines), as in the cloud computing model. These VCs can be integrated with remote clouds, as well as be isolated, self-organized, autonomous, smaller, mobile, and ad hoc clouds, based on the availability of neighboring vehicles, and can be formed anywhere using the computational capabilities of on-board vehicles. In addition, VCs have advantages such as low-cost computing, support for decision-making in scenarios without infrastructure, use without energy limitations (with vehicles with large capacity batteries and recharged by engine operation) and guarantee of real-time services removing network delays involved in accessing traditional clouds. However, this type of cloud suffers from extreme mobility and dynamism of the environment and resources [45], [43], [18], [42], [31], [44], [46].

### B. EDGE

In the context of vehicular networks, we use the term Edge to refer to the set of edge servers deployed in the vicinity of streets, avenues and roads (e.g., servers attached to RSUs or base stations) by service providers. These servers provide processing and storage for one or a few hops of vehicles through isolated servers or small data centers. This set of servers offers advantages such as greater computing, communication and storage capacities, reduced latency due to faster processing and being close to vehicles (eliminating excessive network hops), reduced traffic congestion in links between the core and the periphery network, and QoS improvement of vehicular applications with strict time requirements. However, the computing resources available at Edge are moderate compared to the traditional cloud [47], [32], [38], [48], [49], [50], [51].

### C. TRADITIONAL CLOUD (TC)

We use the term Traditional Cloud (TC) to refer to a set of large-scale centralized data centers that are fixed on the facil-
ities of a cloud provider. This cloud infrastructure is made available to the general public and offers advantages such as greater accessibility and availability to computational resources, aid in the execution of computationally intensive applications, unlimited and powerful computational resources to serve different customers, and absence of restriction of energy consumption due to the constant supply of energy in the data center. However, the time required or the latency to access the TC can be very high and may not be practical for some mission critical applications or applications with ultra-low latency requirements, in addition to requiring an internet connection at all times, presenting challenges of connectivity [52], [38].

D. VEHICULAR CLOUD COMPUTING (VCC)

Despite the fact that there are several similar terms such as Vehicle-Assisted Cloud Computing (VACC) [53], Cloud Computing in VANETs (CC-V) [54], Vehicle using Cloud (VuC) [54], Hybrid Vehicular Cloud (HVC) [54], Vehicle-to-Cloud (V2C) [42], Vehicular Cloud Networking (VCN) [55], and VANET-Cloud [42], we use the term Vehicular Cloud Computing (VCC) to refer to the paradigm that allows two types of cloud to be used in an isolated or integrated way: VC and TC, as shown in Figure 2. In this way, computation offloading can be done for both vehicles nearby and for edge servers. In this paradigm, TC can be used to manage the resources of edge servers [59]. However, in general, TC is not used to perform services requested by vehicles [47], [32], [48]. Thus, computing tends to be limited at most to edge servers with in a few hops from vehicles. In VEC, these servers are computing platforms isolated from the rest of the network available only to users within the Radio Access Network (RAN), can operate with little or no Internet connectivity, and can be in a remote location connected to or disconnected from the data centers of the traditional cloud [38], [60].

F. VEHICULAR FOG COMPUTING (VFC)

Though there are some similar terms such as Fog Vehicle Computing or Fog Vehicular Computing (FVC) [61], we use Vehicular Fog Computing (VFC) to refer to the paradigm that allows to use three cloud types in an isolated or integrated way: VC, edge and TC, as shown in Figure 2. In this way, computation offloading can be done for vehicles nearby, edge servers, and for servers in a remote data center. VFC extends the fog computing paradigm to vehicular networks. Thus, VFC is also hierarchical and provides computational resources and a continuity of traditional cloud services anywhere, from the cloud, through the core and the edge of the network, to the end devices, instead of performing computations only at the edge of the network. VFC is capable of handling applications with a variety of QoS requirements, as applications can be run at a hierarchy level that provides adequate processing capacity and meets latency requirements [38], [60], [62], [63], [64], [65], [66], [67], [68].
III. TAXONOMY

In this section, we present a taxonomy of research in computation offloading in vehicular environments, that is the core of this survey. This taxonomy allows a classification of the different research articles in this area. The taxonomy is presented in Figure 3 and the articles of computation offloading in vehicular scenarios are classified in different categories in Tables 2-4 on the final pages of this survey. In addition, in this section, we describe the main subjects covered in the articles and categorize the works according to the proposed taxonomy. We also synthesize the classification data for these articles and present them through graphics.

According to Figure 3 and Tables 2-4, the main taxonomy categories are: "Communication Standard", "Problem", and "Experiment". "Communication Standard" categorizes communication technologies, client and server, type of communication in relation to uploading tasks and number of wireless hops. In "Problem", the objectives sought by the works are presented as well as strategies for formulating the problem, algorithms and techniques to solve the problem. Finally, in "Experiment", the details of the experiments carried out by the works are described as type of network, mobility, scenario and application, in addition to a classification in relation to vehicle density. Each of these categories has subsections that further detail the subject and present examples of articles that are in each of the subcategories.

Below, we detail each category and subcategories of computation offloading in vehicular environments.

A. COMMUNICATION STANDARDS

In computation offloading on vehicular networks, choosing how to communicate is an important decision. In this section, we summarize the details of the communication standards that previous works used to provide computation offloading on VANETs. For further details on the works, see Table 2 at the end of the paper.

1) Technology

The Internet of Vehicles (IoV) is a paradigm that has attracted growing interest from the scientific community, government agencies and the automotive industries. This paradigm has been proposed to provide collaboration between vehicles and reliable Internet services, and to improve the experience of drivers and passengers [2], [69], [70]. To enable these communications, different technologies have been proposed, as seen in Figure 4(a). These communication technologies are presented below.

a: Wireless in the Vehicular Environments (WAVE)

The WAVE architecture is a family of protocols standardized by the IEEE for communications in vehicular networks. This architecture was created to offer safe and convenient communications in Intelligent Transportation Systems (ITS) and to provide vehicles with direct connectivity to other vehicles (V2V) or infrastructures (V2I) [71], [72], [73], [74]. Some protocols of the WAVE architecture stand out. For example, IEEE 1609.1 is responsible by synchronization of the On-board Units (OBUs) and Roadside Units (RSUs). IEEE 1609.3 is responsible for network and transport layers. Through the Logical Link Control (LLC) layer, it can choose to use WAVE Short Message Protocol (WSMP) (to provide lower latency) or the TCP/UDP/IP stack. IEEE 1609.4 enables the multi channel operation, and prioritization of packets. IEEE 802.11p defines the Physical Layer (PHY), and the Media Access Control (MAC) layer. Some of its features are: operating ranges of up to 1000 meters, random MAC address and wildcard Service Set Identifications (SSIDs) [72].

With respect to the band spectrum of the physical layer of WAVE, a well accepted standard is called Dedicated Short
Range Communications (DSRC). DSRC has an allocation of spectrum in the range of 5.85-5.925GHz. It is structured in 7 channels of 10 MHz, having one control channel (for security and control through WSMP messages) and the others being service channels available for different uses [72].

As the WAVE architecture connects vehicles with other devices via wireless communication, some works use it to enable computation offloading in vehicular scenarios, such as: [75], [76], [77], and [78].

b: Cellular networks

Cellular networks are high-capacity and high-speed communication networks and can be defined as a radio network in which the coverage area is divided into cells. Each cell contains a base station (BS) comprising transceivers and control units, and operates in its own frequencies. Each BS can serve several user equipments (UEs) operating within that cell [79]. Over the years, different generations and technologies of cellular networks have been proposed. Although third generation networks (3G) have been used in some works (as in [80]), below we will focus on the most used cellular networks in computing offloading in vehicular environments.

- Fourth generation networks (4G) are cellular networks based mainly on standards developed by the 3GPP and codified in ITU called Long Term Evolution (LTE) and LTE Advanced. These networks use Orthogonal Frequency Division Multiple Access (OFDMA) and provide low latency, higher throughput, and improved QoS. Typical data rates of 4G systems are 3-5 Mbps. Enhanced versions of 4G networks incorporate new technologies such as Multiple Input Multiple Output (MIMO) and carrier aggregation that allow higher data rates [81], [82]. Some works use 4G to provide computation offloading in vehicular environments as [83] and [84].

- Fifth generation networks (5G) are systems for cellular wireless networks standardized by 3GPP and ITU-R. Some features of 5G are: uses new radio technologies such as millimeter wave (mmWave), Massive MIMO channels, and beamforming, has small cells with lower coverages and high frequencies, allows device-to-device (D2D) communications and suffers with obstacles on the way. 5G systems have the potential to improve current systems and achieve massive data rates (up to 20 Gbps) with lower latency, better mobility, and include new sets of application such as connected vehicles and Internet of Things [85], [81], [86], [87], [88], [89]. 5G was used for computation offloading on VANETs in papers such as [90].

c: Hybrid

Other papers used more than one technology at the same time. For example, [91] used WAVE and 4G to perform computational offloading; [92] used WAVE, 4G and 5G; while [93] used WAVE and 5G.

d: Other

Other works used different technologies to execute computation offloading in VANETs. For example, [94] used Worldwide interoperability for Microwave Access (WiMAX). WiMAX is a technology standardized as IEEE 802.16 and provides wireless internet access with a reasonable data rate through base stations with a range of up to a few kilometers [95]. Other case is in [96] that used a standard for broadband access employing the TV white space (TVWS) band in low population density regions called IEEE 802.22 as technology to provide offloading.

2) Client

In VANETs, a device may need to run a computationally complex application and not have enough resources to run it in a viable time. In this case, the device can apply the computation offloading technique, that allows devices to send tasks to other devices to be executed in order to reduce the task’s execution time and save energy [26], [23]. In this approach, such communication generally follows a client-server model, in which the client is the device that sends the offloading request, and the server is the device that receives the request and processes the task. It is also possible that the same device may act as a client and as a server. Several works have chosen different types of clients to request computing offloading, as seen in Figure 4(b). In the presented taxonomy, we identify the offloading clients most commonly found in the literature.
a: Vehicle
The most common approach is when a vehicle, (e.g., bus, car or truck) starts the offloading process, both in ad hoc and infrastructure mode. In this case, the vehicle may be overloaded, not have enough resources to run the application, or want to speed up the response time. Many research works carry out offloading starting in vehicles [58], [22], [97], [98]. The vehicle can initiate the offloading process in various ways. In one of them, the vehicle sends a broadcast request to find servers willing to perform the tasks [99], [75]. Another approach is to take advantage of beacon messages, exchanged periodically between devices, to know in advance who can perform tasks [100], [92].

b: Infrastructure
In this approach, the infrastructure (RSU or BS) starts the offloading process. The requests can come from a service running in the remote cloud or on an edge server. For instance, in cases of missing children (e.g., Amber Alert), vehicles in a certain region may receive offloading requests of several photos to perform face recognition tasks and collaborate to identify the missing child [101]. In other cases, the infrastructure may be overloaded and act as client needing the help of vehicular clouds or other devices [102], [103], [104].

c: Pedestrian
In this case, a pedestrian initiates the offloading process acting as clients. Pedestrians leverage unused resources of nearby vehicles to speed up the execution of tasks and save the battery life of their mobile devices. They can be inside vehicles [105], [106], or be stationary, or be standing/running/walking on sidewalks, roads, and squares [107], [108]. This type of scenario brings challenges. For example, a pedestrian sitting by the roadside can choose to offload to a vehicle that is passing on the road, and depending on the speed of the vehicle, it may be far away when the task processing is finished.

d: Hybrid
This is when the work considers more than one of the types of clients mentioned above. For instance, the works in [109] and [110] use vehicles and infrastructures as clients, while in [111] the authors consider vehicles and pedestrians as offloading clients.

3) Server
As with offloading clients, several entities may act as offloading servers. Several works have chosen different types of servers to execute tasks of computing offloading, as seen in Figure 4(c). Below, we provide a description of the types of servers used.

a: Vehicular Cloud
In the present case, offloading requests are received and executed in vehicles that are part of Vehicular Clouds. In this approach, as seen in Section II-A, the resources of one or more vehicles, whether moving or parked, are grouped and treated as computational resources that can be used to provide services, just like in the cloud computing paradigm. Several works create VC and divide the offloading tasks among participating vehicles in a cooperative way such as [112], [113], and [114].

b: Edge
This is when offloading requests are executed on servers located at the edge of the network, close to base stations and RSUs. As seen in Section II-B, due to the proximity of client devices, the edge typically has lower latency than the traditional cloud, but it also has less resources available. The edge can contain one or more servers or even mini data centers to serve clients of the most varied types. As seen in the Figure 4(c), most works have chosen to offload to the edge servers such as [58], [47], and [115].

c: Traditional Cloud
This case is when offloading requests are sent and processed on the traditional cloud. As seen in Section II-C, the traditional cloud allows applications to leverage features such as elasticity, availability, and unlimited resources to speed up execution of tasks. Besides, remote clouds are supposed to have enough resources to attend to requests from several clients and are used in several works as the offloading destination as in [116], [117], and [118].

d: Hybrid
The hybrid approach happens when more than one of the server types, previously mentioned, act as offloading servers, usually with a hierarchical environment. For instance, an edge server can expand its computing resources by using remote cloud servers when the demand is high, and there are not enough resources on edge. In [119], [120], and [121], the edge was used together with the vehicular cloud to provide more computing resources to clients. In [122] and [123], the vehicular cloud was used in conjunction with the traditional cloud. In [116] and [124], the traditional cloud was used along with the edge to increase its computational power. Lastly, in [125] and [92], the three types of servers were used: vehicular cloud, edge and traditional cloud.

4) Type
As mentioned earlier, VANETs have two main types of communications:- V2V and V2I. However, given the variety of entities that can start an offloading request, we argued that offloading may be initiated by vehicles (e.g., bus, car, taxi, etc.), infrastructures (e.g., RSU, base station, etc.), or pedestrians (e.g., person carrying a smartphone, tablet, notebook). Likewise, a variety of entities can act as offloading servers and therefore receive and perform offloaded tasks. Thus, as seen in Figures 4(d) and 5, we observed various types of recommendations in the literature, from the point of view of uploading tasks. We describe these types below.
a: Vehicle to Vehicle (V2V)
In the case of direct transmission, ad hoc communication is used by the vehicles to offload tasks onto other vehicles using WAVE or other technologies that allow D2D communication. If it is an indirect transmission, when there are several hops in the path, V2V can be used as a way to forward the task to the final destination. The following works used V2V communication: [78], [127], and [128].

b: Vehicle to Infrastructure (V2I)
This type of communication allows vehicles to communicate directly with RSUs or base stations to send tasks of offloading to edge servers, micro data centers, traditional clouds or even other vehicles (passing through the infrastructure) via cellular, WAVE or other networks. Several works use this approach to upload a offloading task as in [129], [130], and [76].

c: Infrastructure to Vehicle (I2V)
In this type of communication, traditional cloud servers or edge servers offload tasks to vehicles through infrastructures such as base stations and RSUs. As seen in Figure 4(d), few works use this approach. Some papers that use this type of communication are [102], [103], and [131].

d: Pedestrian to Vehicle (P2V)
This is when a pedestrian, carrying smartphones, tablets or notebooks, offloads tasks directly from its mobile device or user equipment (UE) to vehicles passing on the road or parked. This type of communication can have the vehicle as the final destination or just as a task relay. Some works that use this communication are: [119] and [108].

e: Pedestrian inside of Vehicle to Vehicle (PV2V)
In this case, a pedestrian, through its mobile device, offloads tasks to the vehicle that he/she is inside. This vehicle can process the task or forward it via ad hoc network, to another vehicle which can either process the task or forward it via cellular or other network, to a RSU or base station and then to an edge or cloud server to process the task. This type of communication has been used in some papers as in [132] and [105].

f: Hybrid
This approach happens when more than one of the aforementioned types of communications are used. For instance, several works have developed approaches based on V2V and V2I communications [133], [91], [134], while other works consider I2V and P2V communications [107], and others consider V2V, V2I and PV2V [122], [123] and so on (for more details refer to Table 2 at the end of this paper).

5) Wireless Hop
To further extend the possibilities of communications, we can also classify the number of wireless hops between clients and servers during offloading operations. We divided the works into two types: those that use only one wireless hop and those that use more than one wireless hop. We can see the number of works that use this in Figure 4(e). Below we describe this division.

a: One-hop
This is when the offloading task goes directly (one-hop) from clients to servers. This one-hop communication can be V2I, V2V, I2V, P2V, and PV2V, for example. This approach avoids unnecessary delays with multiple successive task forwardings and is less susceptible to offloading failures. This is the most commonly used approach and several works use it such as [135], [136], and [137].
b: Multi-hop
This case is when more than one hop (multi-hop) is required for the offloading to go from clients to servers, requiring the use of forwarder or relay nodes/devices. This approach increases the range of communication between client vehicles and servers (vehicles or infrastructure) and can use more powerful computational resources that are further away. For instance, an offloading request that needs to reach the remote cloud may pass through several vehicles until it arrives at a vehicle that has a direct connection with a base station or RSU to forward the task to the desired server. Some disadvantages can be, a higher latency and a higher rate of offloading failures. Some works that used this approach are [90], [121], and [126].

6) Discussion
The communication standards described in this section greatly affect the computation offloading performance of VANETs. For example, on the issue of technology, it may be a good option to use more than one technology to increase the data rate and bandwidth of devices. In addition, it is important to ensure a good adaptation of the new technological trends of communication (e.g. 5G and IEEE 802.11bd [138]) to the vehicle environments. Different technologies are also important for better integration with other network environments such as Internet of Things (IoT), Internet of Healthcare Things (IoHT), Industrial Internet of Things (IIoT), Smart Cities, etc. [139], [140], [141]. However, using more than one technology on devices can make them more expensive and cause problems of heterogeneity. On the issue of clients, it is very important to create solutions that meet and adapt to the specifics of each type, such as: high mobility of vehicles, lower battery capacity of mobile devices of pedestrians, and fixed and limited geographical range of base stations and RSUs. In the matter of servers, there is an important trade-off between latency and computational power. In the case of devices closer to the edge, such as vehicles and pedestrians, there is less communication latency, but also less computational power. In the case of traditional cloud servers, the communication latency is greater, but the computational power is also greater and can compensate. So it is important to carefully consider where to perform the tasks, whether in VC, Edge or TC. In the case of types of communication, as there are few studies that perform P2V, PV2V and I2V, we believe that these cases can be better explored and studied. Finally, in relation to wireless hops, there is also an interesting trade-off. Using one-hop is a more conservative approach and less likely to fail, but it may not exploit the full potential of the network’s computing resources. On the other hand, using multi-hop can exploit this potential, but it can generate more failures and delays in the offloading process. After all, if ensuring that two devices are connected is already difficult, ensuring that three or more are connected is even more difficult.

B. PROBLEM
The computation offloading works in vehicular networks propose algorithms, solutions or improvements for the problems with only one objective or several different objectives simultaneously. The objectives range from reducing the response time, improving the usefulness of the system, to reducing financial costs. The proposed solutions or strategies also range from simple algorithms and heuristics to complex mathematical modeling and use of intricate machine learning and metaheuristic algorithms. Below we present the objectives and strategies of the algorithms used in the computation offloading works in VANETs.

1) Objective
In offloading works on VANETs, several algorithms have been proposed for different objectives. Most works have more than one purpose, so they are multi-objective. Others use only one objective, but they all aim to improve some of the offloading systems and applications while dealing with the challenging scenarios of vehicular wireless networks. The next subsections present the main objectives studied by the works listed in this survey, as shown in Figure 6(a).

a: Decrease Response Time
Some vehicular applications, although they have become popular, are computationally complex, intensive, and of real time [93]. Taking too long to process an application’s tasks can compromise its performance, data validity and even the safety of humans in a vehicle. Thus, reducing response time of applications (also called task processing time and computation overhead) is the main objective of the computation offloading technique. However, sending tasks to be processed on other devices on the network can be quite challenging in vehicular scenarios and it can also have its delays of transmission/reception and processing (e.g., a bad decision would be to send tasks to already overloaded devices). Thus, in fact, according to Figure 6(a), this is the most researched objective in this area. Several studies have proposed algorithms to decrease the response time of applications [93], [100], [121], [90].

b: Decrease Energy Consumption
Computationally intensive applications with critical time constraints also pose the challenge of excessive energy consumption [142]. This can happen with user equipment (UE) such as notebooks, wearable devices, tablets and smartphones as well as with vehicles. In the case of vehicles, large amounts of computational operations are performed on their on-board computers. The expectation is that the demand for vehicle computing resources will continue to increase exponentially with the development of autonomous vehicles [143]. Thus, it is important to analyze whether energy consumption for complex applications can compromise the maximum distance traveled by vehicles, especially for electric vehicles [142]. For this reason, some works perform
computation offloading in order to reduce the energy consumption of vehicles [130], [144], [142].

In the case of in-vehicle UEs, they may have limited battery capacity and may be able to save energy by offloading tasks to other devices. However, as with vehicles, it is necessary to know the right portion of tasks to be transferred. This is because transferring tasks to other devices also consumes energy [106]. Some works made it their objective to reduce the energy consumption of UEs that are inside vehicles [106], [47].

c: Decrease Financial Cost
Another objective stated by some papers is to reduce the financial cost when offloading. For example, vehicles may have to pay a fee for computation and communication services. In the case of cellular communications, the vehicle may have to pay for the transmission and reception of data. In the case of computing, it may be that the server device (vehicle or edge server) also charges a fee to process the client’s tasks. From the server’s point of view, it can benefit economically by providing computing services. However, it may have to pay for electricity from grid operator and for renting wireless bandwidth from network operator [96]. Some of these costs may vary depending on hardware, technology, location and time and can depend on energy consumption [145], [146].

Several papers on VANETs have proposed algorithms to reduce the financial cost of computation offloading [96], [145], [147].

d: Decrease Overload
Some offloading works on vehicular systems also focus on reducing computational loads on overloaded devices. Such devices end up needing offloading to alleviate their loads, meet the QoS requirements of the running applications, and maintain connectivity and system stability. These devices can be both vehicles and edge servers. This overload can be caused by factors such as multiple clients choosing the same server to process their tasks, low computational capacity, or many computationally intensive application tasks being performed simultaneously. Some papers have the specific objective of reducing the overhead of devices such as [131] and [148]. Decreased overload can also refer to the reduction of network overload, in some cases [148].

e: Increase System Utility
In computation offloading in vehicular networks, it is very important that there is a good balance in the use of system resources. That is, the ideal scenario is when there are no overloaded devices and not many idle devices. Thus, the objective of many works is to increase the system utility by better balancing workloads between devices that can act as servers. Thus, tasks will be better distributed, resulting in better QoS for applications and less overloaded devices. Some works have proposed solutions to maximize the system utility of offloading in vehicular networks, such as [135] and [118].

f: Increase Incentive
Some previous work assumes that all devices will share their resources unconditionally. However this assumption is very optimistic for practical implementations. Due to increased delays and processing overload, selfish vehicles may be reluctant to act as servers unless they are rewarded in some way. Without proper encouragement, device owners are not motivated to share their computing resources. In addition, issues such as preference of shared resources, amount of available resources and transparency of information are not the same for all devices on the network. Thus, some works also note the need for more incentive mechanisms to have more devices acting as servers in the network and optimizing the economic benefits of those involved (e.g. the reward being paid in money) [119]. Thus, incentive mechanisms encourage voluntary devices to contribute their computational resources and to mitigate eventual overloads or lack of resources. However, care must be taken that these devices do not become greedy to maximize their profits even if they are already overloaded [149]. Therefore, some studies have...
proposed algorithms to increase incentives, with restrictions, in vehicular networks such as [119] and [149].

g: Other
Other objectives found in the works include: reduce the loss of quality of service experienced by the user [136], increase offloading reliability [150], and enrich the user experience [151].

2) Strategy
Following the taxonomy, in this section we deal with the algorithmic strategies used for solving the problem at hand. The following strategies were found: stochastic, game theory-related, mathematical programming, heuristics/metaheuristics, and machine learning methods.

a: Stochastic Methods
Stochastic processes have been used for problem modelling in several works. The Markov decision process and semi-Markov decision processes are the ones that were highlighted in [111], [120], [122], [123], [152]–[155].

b: Game Theory
Game theory is the science that studies the interaction between cooperating or competing individuals, and it is usually used in computer science for modelling applications such as optimization problems [156]. The game theory approaches that were employed the most were Stackelberg game [130], [144], [149], [157], [158], contract theory [119], [159], [160], and matching theory [161], [162].

c: Mathematical Programming
In our taxonomy, we consider mathematical programming techniques for solving optimization problems. In short, an optimization problem is a problem of finding the maximum (minimum) value of a function called objective function, subject to different restrictions, and can be formulated as follows [163].

\[
\begin{align*}
\min \ F(x), \ x \in \mathbb{R}^n \\
\text{Subject to:} \\
h_i(x) & \geq 0, \ i = 1, \ldots, m \\
j_i(x) & = 0, \ i = 1, \ldots, p,
\end{align*}
\]

where \(x\) are the decision variables, which can take either discrete or continuous values.

There are several classes of optimization problems, and for each class different algorithmic techniques are employed. For instance:

- In integer programming, the decision variables must be integers. There is also a variant of the integer programming for which some, but not all, variables must present integer values, called mixed-integer programming. Some works also formulate the problem as integer and mixed-integer programming [103], [120], [136], [148], [166], [167] and well known algorithmic approaches, such as branch-and-bound [168], cutting planes [169], and dynamic programming [150] can be seen.
- Some authors also formulate the problem as a convex optimization problem [47], [106], and mathematical programming techniques, such as alternating direction method of multipliers (ADM) are among the most commonly employed ones.

d: Metaheuristic
This subsection of our taxonomy is a direct consequence of the last one, as the decision version of diverse optimization problems is NP-complete. Therefore, it might be costly to solve bigger instances of such problems to optimality. In this sense, metaheuristics are algorithms that do not guarantee to deliver a proved optimal solution to a given instance of a problem, but usually return a good solution in a feasible time [170]. More specifically, metaheuristics are general and higher-level algorithms that incorporate operators designed to avoid getting stuck in a local optimum, called intensification and diversification operators [171].

Among the selected works, a great majority employ metaheuristics for solving optimization problems. The highlights are particle swarm optimization (PSO) [114], [115], [128], ant colony optimization (ACO) [75], [92], and genetic algorithm (GA) [77], [102], [149], [151]. Other examples of metaheuristics are bat algorithm [83] and iterated local search (ILS) [154].

e: Machine Learning
According to Mitchell [172], machine learning is a field of artificial intelligence (AI) that aims at improving a given algorithm automatically though experience. Among the collected works, support vector machine (SVM) [129], adaptive learning [100], reinforcement and deep reinforcement learning [14], [104], [108], [117], [146], [173], [174] are the machine learning techniques that stand out.

f: Network
Network-related strategies are approaches to network management through configurations and arrangements adopted to improve performance or achieve specific objectives. In this sense, two strategies were frequently adopted by offloading works on VANETs: Software-Defined Networks (SDN) and clustering. SDN is an architectural approach that simplifies and optimizes network operations by bringing interactions between applications and devices (real or virtual) and network services closer together, making them programmable. This is achieved by employing a central logical control
point, also called an SDN controller, which orchestrates and mediates interactions between applications and network elements. For this, the controller makes use of interfaces and the separation of data and control plans [175], [176]. The clustering strategy is a management technique that organizes nodes into a set of groups called clusters based on pre-defined criteria such as network load balancing, affinity, etc. Each cluster has one or more leaders called Cluster Heads that collect data from other nodes in the cluster and send the data (usually merged with various other data) to other devices on the network. This can decrease interference, cost, energy consumption and inefficiency [177].

Some works have used SDN as a strategy to improve offloading performance such as [153], [148], and [178]. Others have used clustering such as [46], [83], and [179].

g: Other
Few works have also employed graph theory algorithms [113] and fuzzy logic [127] to solve the problems. In addition, many works proposed problem-specific algorithms [120], [131], [142], [180].

3) Discussion
The formulation of the problem and the way to solve it is also of fundamental importance for computation offloading in vehicular networks. In general, the most common objective in problem formulations is to decrease the response time of applications. In addition, other objectives are also widely used, such as reducing energy consumption and financial cost. However, if trying to achieve a objective in a dynamic environment as vehicle networks is already challenging, trying to achieve more than one objective becomes even more complicated. Therefore, considerable efforts are still needed to obtain solutions that are suitable for vehicular environments and that achieve multiple objectives.

In this sense, several proposals have been published in the literature, such as: algorithms of mathematical programming, stochastic models, metaheuristics, game theory, machine learning, etc. However, producing valid and feasible complete solutions that minimize the objective function while achieving the objectives formulated in the offloading problems still requires major research efforts. One of the approaches that has stood out is machine learning. In this respect, as they are algorithms that need a lot of computational resources, there is also an important trade-off. For example, using an algorithm with deep learning can result in better choices for the offloading process, but it can take a lot of time and computational resources to reach the best result. On the other hand, using weak learning algorithms can give the result in less time and spend less computational resources, but it may not generate a good decision or result for the offloading process. In addition, it is important to analyze the use of the mixture of learned features offline/statically along with learned features online/dynamically, transfer learning and domain adaptation.

C. EXPERIMENT
To validate new proposals (algorithms, schemes, frameworks, systems, etc.) in the area of computation offloading in vehicular environments, it is very important to make good choices regarding test tools, scenarios and applications. In this section, we review the details of the experiments that previous works have used in their proposals related to offloading in vehicular networks. For more details of the works used, see Table 4 at the end of the paper.

1) Network
New systems, applications, protocols, etc., appear with increasing frequency as technologies evolve. This leads to the need to be able to quickly test these research and development proposals, so that they can be validated as quickly as possible [181]. To carry out and validate these experiments at the network and computation level, the works use different approaches (7(a)). Below, we detail the main approaches.

a: Real
Real network experiment is a technique that employs an experimental setup that consists entirely of real network systems, equipments, protocols, and applications so that various parameters can be quantified and the performance of the system can be assessed [181]. Real experiments can be performed by using the developed prototypes in the vehicular driving environments so that implemented services can be better evaluated. Performance metrics obtained in real scenarios provide more reliable values for analysis due to constraints such as power consumption, volume data processing, high-latency links, among others. Real experiments, however, have disadvantages such as a high financial cost incurred, and less control over the assessed environment. Few works have used real network experiments to evaluate their proposals of offloading in vehicular networks. Among these are [180], [76], and [182].

b: Simulation
In order to reduce the high costs in vehicular networks, the use of simulators becomes a suitable alternative. Network simulation is an experimentation technique that employs a setup that consists entirely of computer models of network systems, applications and protocols. Although this type of experiment allows controlled and reproducible tests, the lack of real equipment and components in a simulation leads to a lack of realism in the results. Even so, simulation is the most widely used technique for carrying out network experiments during the development and research stage [181], as seen in Figure 7(a).

Simulation tools provide multi-access control, resource management, and measurements of packet delivery ratio and delay, as well as, different models to evaluate networks such as ad hoc, sensors, and optical networks. Some simulators used to do computation offloading in VANets were: OMNeT++ [183], ns-3 [184], and MATLAB-based environments [185]. For example, in [97], [75], and [131] OMNeT++ was
used; in [83], [94], and [128] ns-3 was used; and MATLAB-based environments were used in [96], [157], and [186].

2) Mobility
An important feature in VANETs is the mobility of the nodes. The nodes can move at high speeds because they are the vehicles themselves. These nodes follow the driver’s behavior, being able to bend on a street, change direction or perform other sudden maneuvers. This can cause frequent changes in the network topology and a short connection time between the nodes, since they have a limited communication range. In fact, in VANETs, very few nodes remain connected for a long time [187]. However, there is a certain predictability of the movement of the nodes because they have to follow the patterns of the traffic routes, such as: direction of the road, traffic lights, speed limits, physical limits of road width etc. [188] [189]. Experimenting to replicate these characteristics is a challenge. Below, in this sense, we present what the computation offloading works have used, as can be seen in Figure 7(b).

a: Real
Some configuration parameters of road traffic simulator and mobility generator are often hard to set in simulated environments and may not provide adequate realism. In contrast to this, mobility studies in real scenarios provide more accurate results. Real experiments using 4G network in Aalborg, Denmark can be found at [190]. Some real world environments were also created for testing at Aldenhoven Testing Center (ATC) and Mcity (a test facility by the University of Michigan, USA) [191]. Some works used real mobility for offloading, such as [180], [182], and [84].

Although we have much more reliable data in real experiments, the main disadvantages of the mobility experiments in real scenarios are the high cost and the risks involved. For example, the costs of an autonomous vehicle can exceed US$ 300,000 without taking into account the additional expenses with scenery and other edge devices [192]. Furthermore, real experiments demand multiple individuals, equipment and vehicles and the situation depends on the other vehicles. Therefore, another alternative to have a real mobility of the vehicular offloading environment is to collect GPS data from previously chosen vehicles [174], [155] (e.g., taxis in a city) [136], [133] or traces of known vehicle routes (e.g., buses) [125], [151].

b: Simulation
Due to economic issues, logistic difficulties, and technology limitations, simulation tools are a widely adopted choice for validation of experiments in VANETs. A critical aspect is to approximate the data generated by simulators with data which reflects the real world behaviour of vehicular traffic as closely as possible. This behavior must be both macro-mobility (with macroscopic aspects such as road topology, per-road speed limits, number of lanes, safety rules, traffic signs, etc.) and micro-mobility (with drivers’ individual behavior as interacting with other drivers, speed and acceleration in different traffic conditions, overtaking criteria, etc.) [193]. Thus, simulators are becoming increasingly robust to overcome this gap with realistic mobility models for the feasibility and validity of the research [194]. Another challenge related to the use of simulators is the compatibility with network simulators, since mobility simulators cannot be used to validate network experiments in isolation [195]. However, realistic mobility simulators as SUMO [196], VanetMobiSim [193], and PTV Vissim [197] already have ways to integrate with different network simulators.

In order to provide vehicular mobility and help in the generation of scenarios to validate the offloading experiments, the mobility simulator most used by the works was SUMO [160], [107], [92]. VanetMobiSim was also used [128]. Some works used real and simulated mobility [136], [133], [182]. However, they were not simultaneous experiments as if it were an emulation.

3) Scenario
Realistic vehicular scenarios are of utmost importance for providing reliable metrics in experiments. Some scenarios are more frequent in computation offloading for VANETS, such as urban and highway. Others are less frequent such as parking lots and university campuses. Defining a vehicular scenario is also important to evaluate models that can be used in specific situations. In this way, computation offloading algorithms can be designed to adapt more efficiently
to different scenarios. The following subsections present a description of the vehicular scenarios, as shown in Figure 7(c).

a: Urban
Urban areas are regions in which relevant traffic is evidenced. The presence of many segments with intersections makes communication and routing decisions more complex to deal with. Besides that, obstacles such as trees and buildings, high density of vehicles, 2D vehicular mobility, presence of pedestrians, viaducts and tunnels are difficult factors in this regard [198]. A scenario widely used as an urban environment is the Manhattan mobility model (Figure 8), which has streets organized in the form of a grid [199].

![Figure 8. Urban scenario in the Manhattan model (adapted from [200]).](Image)

The urban scenario was used in several works of computation offloading, such as: [153], [90], and [106].

b: Highway
Highway scenarios (Figure 9), also called freeway scenarios, are generally characterized by a single road (with one or more lanes in each direction), 1D vehicular mobility, high speed vehicles, few obstacles, stable connection if the vehicles travel in the same lane or in the same direction (e.g., platooning) and unstable connection if the vehicles travel in opposite directions [201]. The provision of continuous connectivity or coverage is a major challenge in highway scenarios. High speeds and long distances require RSUs to be deployed efficiently for cost reduction [202]. Constant modifications of network topology in highways with no fixed structure creates a challenging technical issue. Frequent interruptions are also an obstacle in V2V communications due to diverse velocities of vehicles and short connection times for vehicles in opposite directions [203].

![Figure 9. Highway scenario (adapted from [204]).](Image)

According to Figure 7(c) and Table 4, the scenario most used by the works to carry out offloading experiments was the highway [100], [46], [149]. Some works also used the highway and urban scenarios in the experiments such as [160] and [155].

c: Other
Other scenarios used less commonly are parking lots, university campuses, and industrial parks. In the case of parking lots, parked vehicles can be used to process tasks or to share communication resources [42]. Parked vehicles can be seen as static infrastructure, as RSUs, and help other nodes in the network. Since RSU infrastructure can significantly increase costs, demanding high maintenance overhead, parked vehicles offer a possibility to mitigate this problem [205]. These parked vehicles generally do not change location for long periods of time. With the help of power supplies, such as rechargeable vehicle batteries built into vehicles, parked vehicles can process tasks when their engines are turned off. In fact, the energy consumption for processing tasks can be very small, compared to other moving vehicle activities [66]. Some computation offloading works use scenarios of vehicles parked in the experiments [80], [66]. In the case of university campuses and industrial parks, these are scenarios characterized by low vehicle density, low vehicle speed and traffic generally limited to authorized vehicles. Some of these scenarios were used in [180] and [182].

4) Vehicular Density
VANETs rely heavily on having vehicles nearby so that they are able to exchange information and messages, especially on networks without infrastructure. Therefore, vehicular density is an important factor in these networks [206]. Thus, in the next subsections we describe the types of vehicular density and we show some works that used each type, as shown in Figure 7(d).

a: Low
Low vehicular density scenarios, also called sparse scenarios, have few vehicles to exchange information with each other and maintain good network connectivity. Thus, low densities can cause loss of messages and network packets
due to reduced communication capabilities [206]. Vehicular densities are considered low when the vehicular density is approximately 11 vehicles/km in a highway scenario and 25 vehicles/km² in an urban scenario [207], [206]. As shown in Figure 7(d), low was the vehicular density most used in the offloading experiments [129], [96], [13].

b: Medium

Medium vehicular density scenarios are intermediate scenarios between low and high density scenarios. They have a larger number of network nodes, with better connectivity, and generally without traffic jams. Vehicular densities are considered medium when the vehicular density is approximately 55 vehicles/km in a highway scenario and 120 vehicles/km² in an urban scenario [207], [206]. This approximate density has been used in experiments in some offloading computation work [103], [173], [115].

c: High

High vehicular density scenarios have better connectivity because there are more vehicles in the network that are more likely to be within the communication range of others. However, these scenarios suffer with traffic jams, mainly during peak hours, and can provoke reduced message delivery due to packet collisions, redundancy, and contention at MAC and physical layers, caused by simultaneous forwarding, also known as broadcast storm [206]. Vehicular densities are considered high when the vehicular density is approximately 120 vehicles/km in a highway scenario and 250 vehicles/km² in an urban scenario [207], [206]. According to Figure 7(d), the high density was the least used in the offloading experiments [153], [165], [112].

Some works have also used more than one type of vehicular density in offloading experiments, such as: [135], [116], and [178].

5) Application

Applications in VANETs are most commonly divided into safety applications and comfort applications. Safety applications are responsible for preventing accidents, improving road safety, saving people’s lives, and increasing the driver’s ability to react in various situations during the trip. They are the most important applications and, therefore, a priority. Comfort applications, or non-safety applications, aim to make travel more pleasant for drivers and passengers, through information and entertainment, opening up the possibility of commercial activity to VANETs applications and increasing traffic efficiency [9], [208]. Although this classification is important and systems differentiate applications by their priority, the type of data that will be processed has a greater impact on application performance and offloading systems. For example, the size of the data affects the offloading transfer time and the complexity of operations on that data affects the offloading processing time. Thus, we classify the applications according to the type of data that will be processed: video, image, audio and others [209], [210]. In addition, most of these applications consider tasks independent of each other [100], [164], [92], although some consider the dependency between tasks [102], [91].

Next, we will see a description of these applications, as shown in Figure 7(e).

a: Video

Video-related applications in vehicular networks can be used for both safety and comfort. For example, they can increase safety through applications such as 3D scene reconstruction, augmented reality, emergency video call, real-time navigation, overtaking assistance and surveillance systems, in addition to a wide range of possibilities. This type of application can also be used in comfort applications such as online games, video streaming, and tourist information. Although video file sizes vary (depending on quality, amount of time, etc.), they are usually larger than files of other data types [209]. In addition, video applications can cause a long processing delay and consume a lot of energy (e.g., augmented reality) [211]. For this reason, offloading this information between two or more devices in VANETs, with the desired quality, delay and resolution, is a big challenge for this type of application [212], [213].

Some works have tested application tasks related to video in computation offloading systems on VANETs. For example, [46] tested a 3D scene reconstruction video application; [102] used a video navigation application in the experiments; and [114] used a video streaming application.

b: Image

Image-related applications are used for safety as identification of stolen vehicles through the license plate, systems that warn of hazards, searching for drivers on the road, and recognition of traffic light, gestures, faces, and objects through cameras installed in the vehicle [214], [215], [216], [217], [218], [219]. They are also used for comfort as social networks and contextual images of interest [201]. Since images are generally smaller in size and require less processing time when compared to videos, using images instead of videos can allow a drastic reduction in the data load circulating the network and the processing time of the applications [209]. Even so, some applications related to images have great challenges in offloading schemes in vehicular environments. Therefore, these applications were used in computation offloading experiments. For example, [93] used an application for one vehicle to recognize license plates from other vehicles; [125] used a facial recognition application; and [182] used an object recognition application.

c: Audio

Audio-related applications are also used for safety (e.g., emergency calls, theft detection, etc.) and comfort (e.g., voice chat, guided tour, etc.) [201] and can provide a large amount of information to drivers in a short time and with minimal deviation from attention to driving [220]. Audio applications are based on sound, typically of the human voice. The size of
the audio files depends on the quality and amount of time that was recorded. In general, these files are also smaller than video files and require less processing time, depending on the application [209]. In offloading systems, this type of application also has transmission/reception and processing challenges. Some works have used audio applications in their computation offloading experiments on vehicular networks. The most used application was voice recognition [221], [134], [149].

d: Other

The development of other types of applications is also interesting in vehicular environments. Besides that, such applications can benefit from computational offloading for performance improvement. For example, text messaging or character-oriented data offers a lightweight and low operating cost alternative for both safety and comfort applications. It is possible to send critical information to vehicles with less delay and less chance of a bottleneck. Warnings about floods, accidents, and emergency vehicles passing by can be passed quickly and to a large number of vehicles, especially when the network can handle the sending of this information by other mediums. Also, there is a possibility of commercial exploitation of this type of application, with tourist information, blockchain [35], [222], [223], [224], [225], [226], business suggestions, and events nearby [227]. In general, this type of character-oriented applications or other data types has smaller files, depending on the application, than other types of files [209]. Some types of applications used in computation offloading were: recommendation based on location [137], traffic information, online chatting [149], and real-time financial trading [165].

6) Discussion

The experiments are of great importance for the reliability of new proposals and solutions for computation offloading processes in vehicular scenarios. In fact, if an experiment is badly done or carried out without due scientific rigor, it can compromise the credibility of the results and the proposal may not be valid. Accordingly, the topics presented in this section should be analyzed carefully. For example, consider the question of choosing the testing platform related to networks and mobility. The ideal approach would be to experiment in real environments so that the results are reliable and realistic. However, it may require a large amount of financial resources and the environment may need careful verification and configuration. On the other hand, using network and mobility simulators may be cheaper and better to control and replicate, but the results may be less realistic. In terms of scenarios, the type is also very important. Using an urban type scenario implies different street and avenue layouts, usually close to the grid model. A highway scenario usually consists of just one road with lanes with different directions and other scenarios have other configurations. Thus, the scenario impacts the mobility of vehicles, and, in turn, the performance of applications and offloading systems. Vehicular density is another factor that impacts offloading processes. The center of a large city at peak times can have many vehicles traveling and generating and exchanging data. This can cause many network packet collisions, make the network congested and offloading more difficult. In turn, a rural road with very little vehicular traffic may not even have adequate connectivity between the few network devices. Finally, the execution of different applications also affects the computation offloading processes. Generally, video applications have larger files to be transferred and require powerful computational resources. Audio, image and text applications may involve smaller files to be transferred. Thus, it is important that offloading systems adapt to these issues and maintain good performance regardless of which configuration/application is used.

IV. CHALLENGES AND FUTURE DIRECTIONS

Despite significant and recent advances in the field of computation offloading in vehicular networks, there are still important research challenges and open issues. Next, we present the challenges that we have intensively examined in the literature. Such topics can guide future research in the area. These open challenges and issues need further investigation and have been organized in relation to six key topics: network, mobility, security and privacy, incentive, experiment, and algorithm.

A. NETWORK

Network-related challenges greatly affect computation offloading in vehicular scenarios. For example, if the network is very congested, with a lot of contention, collision, noise and interference, the offloading may not happen or happen without success. This can happen due to factors such as a lot of offloading happening at the same time, data being exchanged, broadcast storm, vehicles within the same geographic region, control data or signaling messages (e.g., beacon messages from IEEE 802.11p protocol), transmission of large data (e.g., movies), etc. Therefore, it is very important to periodically monitor the status of the network (e.g., situation and whether the queues at the network interfaces are empty or full). In addition, many of these messages exchanged can be redundant or repeated, which could be resolved using techniques such as multicast or geocast (based on the location of the nodes).

Other network-related challenges are signal attenuation problems, lack of a central coordination point, hidden terminal, obstacles hindering communication (e.g., buildings and trucks) or even little connectivity due to a low number of nodes in the network. If there is multi-hop transmission, a problem that can also hinder is the increased delay as there will be more transmissions, receipts and processing of the packets.

Finally, the heterogeneity of technologies used can increase the network bandwidth but may need better management. For example, devices using WAVE may have different ranges than cellular technologies and not all devices on the network can use the same technologies. Furthermore, newer
technologies need to be better studied to be fully adapted into vehicular scenarios such as 5G/mmWave [57], [85], [86] and IEEE 802.11bd (considered the evolution of IEEE 802.11p) [138]. We believe that these network-related challenges need to be better studied in future works to obtain better performance in the computation offloading processes in VANETs. In addition, the new network technologies mentioned offer good research opportunities.

B. MOBILITY

Vehicular mobility also creates some challenges that need to be solved for the offloading to be successful. For example, the fast speed of the nodes causes wireless links and paths to be constantly broken or fragmented, causing vehicles to move out of each other’s communication range and connections to be short-lived. Also, since the network may have few nodes and in a matter of seconds or minutes it may already have a large number of nodes, there may be problems with scalability and it will be difficult to maintain a good performance regardless of the number of nodes in the network. In addition, these rapid and frequent changes in vehicle topology and variable node density can lead to a high packet loss rate when density is high due to a greater contention over wireless channels and when density is smaller because connectivity is low. If there is a multi-hop transmission, the probability of success in the transmission ends up being lower since the probability of all the nodes in the path remaining connected is also less.

One solution to these challenges is to use mechanisms for predicting the mobility of network nodes. Some works use an estimated lifetime of links to predict how long they will be active or nodes will be within range of each other. Even so, these estimates are not always correct. This is because drivers’ behavior is not always predictable and, generally, nothing prevents them from choosing to turn around the next corner or make a return on a highway. Other approaches use predefined public route information (e.g., buses). However, this information is only available for a very small percentage of the vehicles in the network and even then it is not completely reliable as the vehicle can make extra stops or have other unexpected behavior. Thus, although some papers have investigated this topic, future works can still be directed to better adapt the computation offloading processes to mobility of vehicles.

C. SECURITY AND PRIVACY

Security and privacy is a great concern for any network, and many recent works have pointed out various security issues and solutions that have been come up in the recent past [228], [229], [230]. Security and privacy are also very important topics in computing offloading in VANETs. In general, the principles of cloud computing for security and privacy also apply to devices at the edges of the network. However, the edge deals with much more private information. For example, vehicles may have data on their daily routes and the exposure of that information can be dangerous for personal security. In addition, computation offloading can send sensitive data to untrusted servers (from any person or company and can be a malicious node). This, coupled with the absence of a centralized control point, makes it difficult to create integrated security and privacy policies. A possible solution for this is to offload only non-confidential tasks. However, it can happen that most tasks are confidential and there is no improvement in performance in offloading a few tasks. Another solution is to add encryption or authorization/authentication certificates. However, this can compromise offloading efficiency, Quality of Experience (QoE), and system scalability because it will require more computational power or more time to process tasks [35].

There is also a risk that potential servers can deny service to clients they do not know and this may lead to offloading failure or underutilization of network resources. On the other hand, malicious nodes can upload tasks containing viruses, due to the lack of security strategies. There can also be a Distributed Denial of Service (DDoS) in which the nodes become infected and stop providing computing services. Thus, it is also necessary to ensure secure communication and avoid the spread of viruses or false information. As security and privacy is a priority, there is a need for more studies, research, solutions and mechanisms in this area [26], [231], [232], [233], [234].

D. INCENTIVE

Computation offloading requires collaboration by network devices to share resources such as CPU, storage space, battery, etc. For example, a vehicle can perform tasks for another vehicle or forward data to other nodes in the network using its own computational resources. Transmitting and receiving this information and leaving wireless network interfaces connected consumes energy. In addition, the owner of these resources could benefit more if his/her own applications used the borrowed resources. In this way, the network nodes can become selfish and refuse to share their resources without any compensation. So, the challenge is how to motivate network nodes to share their resources. Although some studies have suggested incentive mechanisms to reward nodes that share their resources (e.g., monetary payment) or analyzed the possibility of the device owners themselves voluntarily deciding, there is still a need for more incentive mechanisms that are appropriate for different scenarios and that are relatively attractive [26]. Therefore, research efforts must still be directed towards solving the challenges of this topic.

E. EXPERIMENT

To validate the computation offloading experiments in vehicular networks, reliable test environments are necessary. Realistic vehicular communication experiments have a significant impact on the credibility of the results. Although simulators have constantly improved in reproducing realistic traffic patterns and movements, including interaction between vehicles, these simulators can still improve the microscopic modeling of the individual movements of each element of the network.
Replicating through simulations the exact movements of real vehicular traffic is still a challenge. In addition, products and prototypes must necessarily be tested in real environments before they are deployed, even if simulations have already been used. This is because simulations may not provide adequate realism and real experiments provide more accurate results [181]. However, making these real experiments more accessible is a challenge due to the prohibitive cost for some research centers. In this manner, considerable future works are needed to perform and analyze computation offloading processes in vehicular environments in a reliable and realistic way.

F. ALGORITHM

A computation offloading system is a group of computational components and modules that interact with each other to improve the processing capabilities of a device by allowing the migration of task execution to other devices. Thus, some of the biggest challenges of computation offloading in VANETs are related to the algorithms of the systems. Some of these challenges are listed below.

1) Load Balancing
From the macro point of view of the network, ensuring load balance among all the devices on the network is still an open issue. For example, ensuring that all task processing requests are not addressed only to the most powerful device on the network (often an edge server attached to a BS) or that only a few nodes on the network receive most of the requests is still challenging. Thus, it is very important that the algorithms are designed to make the best possible use of computational and network resources. If not properly managed, few devices can be overloaded and many can be idle at the same time, thus impairing network and application performance.

2) Partitioning
Before performing computation offloading, an application partitioning procedure may be necessary, although not all applications can be partitioned. The purpose of this procedure is to divide the application into tasks that can be performed on different devices. This partitioning can be done through different models, techniques and levels of granularity. In addition, this partitioning can be done automatically by the system or manually by the developers of the application through code annotation or markup [236]. In this sense, an important challenge is how to partition a large and complex application workload in an optimal or near-optimal way. Thus, parameters such as the size and quantity of tasks must be taken into account in order to optimize the upload, download and processing time for the remote devices.

3) Failure Handling
As vehicle networks have dynamic topologies, there are constant disconnections between vehicles and devices. Such disconnections can affect computation offloading processes between clients and servers and result in failures. For example, as can be seen in Figure 10, orange server vehicle is moving out of range of the red client vehicle without returning processing results. Thus, mechanisms and algorithms are needed to ensure that the application is not affected by the failures. In this respect, some important challenges concern an acceptable level of fault tolerance and ways of recovering from failures. For example, some tasks may be less important and may have the results of their processing discarded. In addition, copies of tasks can be sent to run on different devices, can be kept locally in case of failure detection [93], or, even if the client and server no longer have a direct connection, it can use multi-hop transmissions to deliver processing results or tasks. Thus, despite the need for fault tolerance and failure recovery schemes, it is not a topic covered by the great majority of the papers studied in the present survey and is a key topic of research.

4) Policies
Another challenge is related to the policies adopted by the offloading systems. For example, users or applications may have different priorities, Service Level Agreements (SLAs), be classified as gold, silver, bronze, etc., and have different QoS and QoE requirements. It can be a challenge to ensure that these policy of priorities and agreements are correctly applied and managed in the offloading systems. In addition, few studies have addressed this topic in computation offloading in vehicular environments.

5) Discovery and Resource Request
An important step in the offloading processes concerns the discovery and request of computational resources from other devices. Finding the best way to do this can be challenging. For example, using periodic beacon messages to carry contextual information about devices (e.g., location, speed, CPU capacity, etc.) can be a good solution. However, this can increase network overhead. Another solution would be to use traditional request/reply messages to decrease overhead. But that could increase the delay in the offloading process. In addition, at the time of the request, the contextual situation can be in a certain way, and when the tasks are actually sent,
the situation may already be different and that information may be out of date. This can hinder the computation offloading process. Thus, further studies on this interesting topic are still lacking.

6) Task Distribution

Perhaps the most important and challenging step in the computation offloading process is the task distribution or task scheduling (the task distribution process can be seen in Figure 10). This step can have several parameters and metrics that need to be taken into account in order to obtain the best possible performance in offloading. In fact, finding the optimal way to distribute tasks in order to have the maximum reduction in application processing time and the minimum percentage of failures in the entire process has been described as a NP-hard problem [75], [94], [102], [114], [169]. Thus, distributing these tasks in an intelligent and optimal way can require considerable computational resources. Below, we list some topics that need to be analyzed to optimize task distribution:

- Dependence or independence of tasks;
- Multiple criteria, attributes, constraints, and objectives (e.g., reducing response time, energy consumption, and financial costs at the same time or reducing response time and still maintaining good data quality);
- Convergence time of distribution decision;
- Different degrees of priority and deadlines of users, applications and tasks;
- Servers with multiple and distinct characteristics;
- Tasks with different complexities, sizes and requirements;
- Reliability level so that the servers are within range of the client when returning the processing result (e.g., spreading replicas of the most important tasks for several servers to process and that at the same time does not congest the network);
- One or multi-hop scenarios;
- Specific technologies such as 5G/mmWave that need to have line of sight to transmit/receive;
- Contextual information;
- Scenarios with large amounts of data and processing (e.g., autonomous vehicles);
- Ideal number of server nodes;
- Whether offloading is worth doing or not;
- Best moment to do the offloading;
- Best number of tasks to be sent;
- Information about user behavior;
- Best place to process tasks (e.g., Local, VC, Edge, TC, or a combination of these).

Therefore, dealing with all or some of these algorithmic issues is a major challenge. Considerable research efforts are required for these topics. In addition, there are several research opportunities related to intelligent algorithms for computation offloading in VANETs, with emphasis on machine learning algorithms.

V. CONCLUSION

Computation offloading is a technique that has potential to improve application performance in vehicular networks. Although the vehicular scenarios are challenging, it is possible to apply computation offloading so as to benefit different types of clients and servers and, consequently, the end users of the applications. Furthermore, computation offloading can be applied in different paradigms such as Vehicular Cloud/Fog/Edge Computing. With the arrival of autonomous vehicles and their applications with large and complex processing requirements, applying this technique will be even more necessary. In this paper, we comprehensively covered the state-of-the-art of computation offloading in vehicular environments. We began with summarizing the various terms and paradigms used in the area. Next, we proposed a taxonomy for existing literature on computation offloading in vehicular environments. Furthermore, we classified a large number of works in this emerging area according to the proposed categories, describing the main concepts used in the works. Finally, we presented problems and challenges that can guide future research.

REFERENCES


A. B. Souza et al.: Computation Offloading for Vehicular Environments: A Survey

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3033828, IEEE Access

VOLUME 4, 2020
PAULO A. L. REGO [M’15] holds a PhD in Computer Science from the Federal University of Ceará (UFC) since 2016. He is currently an adjunct professor at Computing Department at UFC. He is an IEEE Member and works in the area of Computer Science, with emphasis on Computer Networks and Distributed Systems. His main research interests include Mobile and Cloud Computing, Vehicular Ad Hoc Networks, Edge and Fog Computing.

TIAGO CARNEIRO received the Bachelor’s (2009) and Master’s degree (2012) in Computer Science from the State University of Ceará, and the D.Sc degree in Computer Science from the Federal University of Ceará, Brazil. He is currently a Postdoctoral researcher at INRIA Lille - Nord Europe. His research interests include the use of heterogeneous architectures for solving combinatorial optimization problems.

JARDEL DAS C. RODRIGUES received his B.Sc. in Computing Science from the Federal Institute of Education, Science and Technology of Ceará (2014) where he has experienced studies in the computer architecture area. Also, he obtained an academic achievement award (first place) in 2012. In 2017, received his M.Sc in Computing Science from the Federal Institute of Education, Science and Technology of Ceará. Some topics of his research in this period were mathematical morphology and biomedical image segmentation, more precisely retinal images. Actually, is professor at Federal Institute of Education, Science and Technology of Ceará and Ph.D. student of the Post-Graduation Program in Teleinformatics Engineering from the Federal University of Ceará and his research interests include Image and Signal Processing, Pattern Recognition, Pose Estimation, and Biometric Applications.

PEDRO P. REBOUÇAS FILHO is graduated in Mechatronics Engineering from the Federal Institute of Ceará, Brazil (2008). In 2010, he obtained an M.Sc. in Teleinformatics Engineering, in the field of Biomedical Engineering, also at the Federal University of Ceará. In 2013, he earned a Ph.D. degree in Teleinformatics Engineering from the same University. He has been an Assistant Professor at the Federal Institute of Ceará since 2008. From 2015 to 2016, he was a postdoctoral researcher at the University of Porto, Portugal. He is co-author of more than 100 articles in national and international journals and conferences. He has also been involved in several research projects in the field of Computer Vision, Medical Image, and embedded systems.

ALISSON B. DE SOUZA is Assistant Professor at Federal University of Ceará, Quixadá, Brazil and is currently pursuing the Ph.D. degree in Computer Science at Federal University of Ceará, Fortaleza, Brazil. He received the B.S. and M.S. degree in Computer Science from State University of Ceará, Fortaleza, Brazil, in 2009 and 2012, respectively. His main research interests are in Vehicular and Flying Ad Hoc Networks, Wireless Sensor Networks, Cloud, Edge and Fog Computing, Internet of Things, Computer Networks, Mobile and Wireless Networks, and Distributed Systems.


JOSÉ N. DE SOUZA [SM’13] holds a PhD degree at Pierre and Marie Curie University (PARIS VI/MASI Laboratory), France, since 1994. He is currently working as a researcher full professor at the Federal University of Ceará in the Computer Science Department, and is IEEE Senior member. Since 1999, he has been the Brazilian representative at the IFIP TC6. His main research interests are Cloud Computing and Network Management.

VINAY CHAMOLA received the B.E. degree in electrical and electronics engineering and master’s degree in communication engineering from the Birla Institute of Technology and Science, Pilani, India, in 2010 and 2013, respectively. He received his Ph.D. degree in electrical and computer engineering from the National University of Singapore, Singapore, in 2016. In 2015, he was a Visiting Researcher with the Autonomous Networks Research Group (ANRG), University of Southern California, Los Angeles, CA, USA. He also worked as a post-doctoral research fellow at the National University of Singapore, Singapore where he worked in the area of Internet of Things. He is currently Assistant Professor with the Department of Electrical and Electronics Engineering, BITS-Pilani, Pilani Campus where he heads the Internet of Things Research Group / Lab. He has over 50 publications in high ranked SCI Journals including more than 33 IEEE Transactions and Journal articles. His works have been published in Journals like IEEE Transactions on Communications, IEEE Transactions on Vehicular Technology, IEEE Journal on Selected Areas in Communications, IEEE Communications Magazine etc. Furthermore, his works have been accepted and presented in reputed conferences like IEEE INFOCOM, IEEE Globecom, IEEE ICC, IEEE PerCom to name a few. His research interests include IoT Security, Blockchain, 5G network management and addressing research issues in VANETs and UAV networks. He has served as a reviewer for several IEEE/Elsevier Journals. He is a Guest Editor in Computer Communication, Elsevier. He serves as an Associate Editor for the Ad Hoc Networks, Elsevier; IET Quantum Communications and also Frontiers in Communications and Networks journal. He is a senior member of the IEEE.

VICTOR HUGO C. DE ALBUQUERQUE [M’17, SM’19] is a full professor and senior researcher at the University of Fortaleza, UNIFOR. He has a PhD in Mechanical Engineering from the Federal University of Paraíba (2010), an MSc in Teleinformatics Engineering from the Federal University of Ceará (2007) and Bachelor in Mechatronics Engineering from the Federal Center of Technological Education of Ceará (2006). He mainly researches IoT, Machine/Deep Learning, Pattern Recognition, and Robotics. He is an Associate Editor of IEEE Access.

BIPLAB SIKDAR received the B.Tech. degree in electronics and communication engineering from North Eastern Hill University, Shillong, India, in 1996, the M.Tech. degree in electrical engineering from the Indian Institute of Technology Kanpur, Kanpur, India, in 1998, and the Ph.D. degree in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 2001. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. His current research interests include wireless network, and security for Internet of Things and cyberphysical systems. Dr. Sikdar served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS from 2007 to 2012. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON MOBILE COMPUTING.

***
**TABLE 2.** (Appendix A) Communication standard classification of works about computation offloading in vehicular networks. (✓) indicates that topic is covered.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Technology</th>
<th>Client</th>
<th>Server</th>
<th>Type</th>
<th>Wireless Hop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref. [150]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [139]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [148]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [127]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [116]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [105]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [94]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [16]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [111]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [239]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [152]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [150]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [238]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [98]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [173]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [145]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [133]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [83]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [116]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [47]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [120]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [159]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [102]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [22]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [147]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [144]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [58]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [123]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [122]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [145]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [133]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [83]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [116]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [47]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [120]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [159]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [102]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [22]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [147]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [144]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [58]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [123]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [122]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [145]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [133]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [83]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [116]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [47]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [120]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [159]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [102]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [22]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [147]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [144]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [58]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [123]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [122]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [145]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ref. [133]</td>
<td>V2V</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>
### TABLE 3. (Appendix B) Problem classification of works about computation offloading in vehicular networks. (√) indicates that topic is covered.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Objective</th>
<th>Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref. [58]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [135]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [180]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [100]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [145]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [129]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [122]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [123]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [97]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [75]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [130]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [132]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [46]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [221]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [96]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [164]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [22]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [119]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [102]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [103]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [159]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [76]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [106]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [47]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [126]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [121]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [116]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [136]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [114]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [157]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [124]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [146]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [83]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [133]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [167]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [125]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [186]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [161]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [91]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [237]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [92]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [179]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [134]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [173]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [94]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [40]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [99]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [158]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [144]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [182]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [98]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [104]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [238]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [151]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [153]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [111]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [142]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [174]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [165]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [112]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [77]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [168]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [137]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [155]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [242]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [113]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [160]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [118]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [243]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [148]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [244]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [178]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [188]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [182]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [78]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [93]</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Ref. [66]</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

#### Column Notes:
- **Response Time**: √ if considered.
- **Energy Consumption**: √ if considered.
- **Financial Cost**: √ if considered.
- **Overload**: √ if considered.
- **Incentive**: √ if considered.
- **Other**: √ if considered.
- **Stochastic**: √ if considered.
- **Game Theory**: √ if considered.
- **Mathematical Programming**: √ if considered.
- **Metaheuristic**: √ if considered.
- **Machine Learning**: √ if considered.
- **Network**: √ if considered.
- **Other**: √ if considered.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Network</th>
<th>Mobility</th>
<th>Scenarios</th>
<th>Vehicular Density</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref. [112]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ref. [129]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ref. [132]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ref. [136]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ref. [173]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ref. [179]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>