A Comprehensive Review of Computing Paradigms, Enabling Computation Offloading and Task Execution in Vehicular Networks

ABDUL WAHEED, MUNAM ALI SHAH, SYED MUHAMMAD MOHSIN, ABID KHAN, CARSTEN MAPLE, (Member, IEEE), SHERAZ ASLAM, (Member, IEEE), AND SHAHAB SHAMSHIRBAND, (Member, IEEE)

1Department of Computer Science, COMSATS University Islamabad, Islamabad 45550, Pakistan
2Department of Computer Science, Aberystwyth University, Aberystwyth SY23 3FL, U.K.
3Cyber Security Centre, University of Warwick, Coventry CV4 7AL, U.K.
4Department of Electrical Engineering, Computer Engineering and Informatics, Cyprus University of Technology, 3036 Limassol, Cyprus
5Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Douliu 64002, Taiwan

Corresponding authors: Abdul Waheed (gallian92@gmail.com), Carsten Maple (cm@warwick.ac.uk), and Shahab Shamshirband (shamshirbands@yuntech.edu.tw)

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) through the Academic Center of Excellence in Cyber Security Research-University of Warwick under Grant EP/R007195/1, The Alan Turing Institute under Grant EP/N510129/1, Autotrust under Grant EP/R029563/1, and the National Centre of Excellence for the IoT Systems Cybersecurity, PETRAS under Grant EP/S035362/1.

ABSTRACT Road safety, optimized traffic management, and passenger comfort have always been the primary goals of the vehicle networking research community. Advances in computer and communication technologies have made the dream of modern intelligent vehicles a reality through the use of smart sensors, cameras, networking devices, and storage capabilities. Autonomous operation of modern intelligent vehicles requires massive computations where tasks are outsourced. In recent years, various computing paradigms, e.g., mobile cloud computing (MCC), vehicular cloud computing (VCC), multi-access or mobile edge computing (MEC), vehicular edge computing (VEC), vehicular fog computing (VFC), and volunteer computing based VANET (VCBV), have been developed to move computational resources close to the user and handle the delay-sensitive applications of modern intelligent vehicles. Therefore, in this study, we provide a comprehensive overview of all computing paradigms related to vehicular networks. We also present the architectural details, similarities, differences, and key features of each computing paradigm. Finally, we conclude the study with open research challenges in vehicular networks along with future research directions.

INDEX TERMS Multi-access edge computing, vehicular cloud computing, vehicular edge computing, vehicular fog computing, vehicular network, volunteer computing based VANET.

I. INTRODUCTION

With the rapid development of intelligent transportation systems (ITS), vehicular networks have attracted great interest from academia and industry. ITS plays a very important role in traffic management and safety enhancement [1], [2]. Various basic applications of vehicular networks are used to send important messages about incidents, road conditions, environment and vehicle alerts to enable ITS. Low latency is the basic requirement for effective and efficient dissemination of this critical information [3], [4].

Hardware and software advances have improved processing and communication capabilities allowing the emergence of versatile networks. These developments have paved the way for many new complex applications such as autonomous driving, augmented reality, video streaming in vehicles, automatic navigation, natural language processing, and online gaming [5]. These delay-sensitive complex applications pose a challenge to the computational capabilities of vehicles [6].

Cloud computing (CC) has served as the first solution to the international community to overcome the computing and storage challenges [7], [8]. The concept of cloud computing for customers arose from the expectation of moving computing infrastructure to a paid computing and storage service offered over the Internet [9]. An application that...
cannot be run on a local device can be moved to the cloud over the Internet. This type of computing service is called “on-demand computing” or “pay-as-you-go model”. The cloud computing service model consists of many services. Three general services, software-as-a-service (SaaS), platform-as-a-service (PaaS), and infrastructure-as-a-service (IaaS) [10]–[12], are shown in Figure 1.

Earlier, vehicular networks and cloud computing were merged to meet task execution needs and take advantage of the additional resources in vehicular networks [13]–[15]. ITS applications such as video surveillance and traffic management, etc. [16] are the best known examples of such applications. Computation offloading [17] is a process in which a device, which has fewer resources, offloads the tasks to other resource-rich servers to overcome the limitations of its limited resources. The integration of mobile devices with cloud computing technology to offload computing or storage tasks to the cloud is called mobile cloud computing (MCC) [18]. Basic goal of MCC is to make the devices more resourceful in terms of computation, storage and energy [19]. Using wireless communication, MCC can leverage vehicular networks by providing intensive computation and storage in a remote cloud [20]–[30].

Despite the tremendous benefits of using clouds for storage and computation of resource-constrained devices, the distance from remote cloud servers, high dependency on the internet, and backhaul network congestion are the major drawbacks of the cloud computing paradigm [31]. The greater distance of cloud servers from users and the ever-increasing data generation rate at the user level affect efficiency by increasing the amount of time involved in offloading data to the cloud. Therefore, to increase the efficiency, the data could be processed at the edge of the network. The type of computation where data processing is moved to the edge of the network is called edge computing [32], [33] and the use of these edge computing servers for mobile computing devices, including mobile phones and vehicles in their vicinity, is called multi-access edge computing (MEC) [34], [35].

Even before the advent of edge computing, the concept of Fog computing was introduced by Cisco [36], [37]. Both fog and edge computing have the same goal, which is to move the remote cloud functions close to the users [38]. These computing paradigms are not a complete replacement but a complement to cloud computing technology [39]. Fog computing and edge computing are used alternatively by many researchers [40], [41] ignoring minor variations depending on focus, task reliability, type of dedicated servers etc. [42].

Computing paradigms such as MCC and vehicular cloud computing (VCC) provide centralized services using cloud servers or cloudlets with the limitation that they incur a higher makespan. New promising computing paradigms called vehicular edge computing (VEC) and vehicular fog computing (VFC) have been developed to move processing to the edge of the vehicular network. In VEC, the road side units (RSUs) have computing servers to meet the computing needs of the vehicles in their communication range [43], [44]. To reduce the load on the dedicated servers and utilize the excess resources, a new computing paradigm called volunteer computing based VANET (VCBV) has been recently developed. It uses volunteer computing to utilize the excess resources of VANETs [45].

MCC, MEC, VCC, VEC, VFC and VCBV are different computing paradigms of today. The communication architecture of different computing paradigms is shown in Figure 2. The importance of computing paradigms has attracted the attention of the research community and much research has been done in these areas. A detailed discussion and the current state of the literature on the various computing paradigms can be found in section II of this paper.

In this study, we have focused on execution of vehicular network tasks in various computing paradigms. We have tried to present our research in an effective and systematic manner to cover all computing paradigms. Figure 3 depicts the timeline, showing the evolution of computing paradigms whereas, Table 1 shows list of all abbreviations used in the manuscript.

Study of state-of-the-art literature has shown that a lot of research is being done on different computing paradigms, namely MCC, MEC, VCC, VEC, VFC, and VCBV. However, as far as we know, the existing review studies deal with only one of the computing paradigms. In this study, in order to provide a good and sufficient knowledge to the aspirants, we have discussed, summarized, critically analyzed and compared all the above computing paradigms along with the research challenges. We believe that this manuscript is useful for both novices and experts in this field. Our three-fold contributions are as follows:

1) Task execution and resource utilization concepts related to all computing paradigms of vehicular networks.
2) Similarities and dissimilarities of all vehicular networks related computing paradigms.
3) Application areas, use cases and future research challenges of all computing paradigms are comprehensively summarized.

The manuscript is organized as follows. Section II discusses the primary computing concepts and Section III gives a comparison of related computing paradigms. Section IV describes some use cases of computation offloading in vehicular networks whereas, research challenges and future research directions are stated in Section V. Finally, Section VI concludes this study.

II. PRELIMINARIES
In the following section we will briefly introduce major terms related to computing paradigms of vehicular networks.

A. COMPUTATION OFFLOADING
Computation offloading is a method of shifting all or some parts of a job to other processing device(s), usually with higher resources [46]. Decision regarding offloading and selection of appropriate server for task offloading are the main factors involved in computation offloading [47].
Network bandwidth and latency affects the network access that, in turn, affects the cloud connectivity hence, the whole computation offloading process is also affected. Figure 4 shows the process of offloading services from mobile device to cloud and from mobile device to cloud and cloudlet.

In task offloading, the transmission delay is a major contributing factor to the total delay. Further, the resource availability, heterogeneity, and computation delay are also the factors which affect the task offloading process. Wu et al. [48] proposed a model consisting of the components required for task offloading that uses an iterative algorithm.

High range sensors and high accuracy are mandatory for vehicles with the capability of automated driving. To cope with outcomes of limited computational resources within a vehicle, Du et. al., formulated a neural network based algorithm, light gated recurrent unit (Li-GRU) [49]. With the help of experimentation, it is shown that coverage and sensing are improved for a single vehicle. In another study at [50], the authors have considered fair access problem among high velocity and low velocity vehicles. Simulations of the proposed velocity adaptive fair access scheme prove that all vehicles can fairly access the road side units. Distributed computing is another effective method to complete the tasks well in time. Resource-aware parallel offloading [51] has been presented to discover appropriate nodes to offload the tasks. Simulations are used to verify the efficiency of the proposed method.

### B. VEHICULAR CLOUD COMPUTING

Recent advances in technology have lead to the emergence of advanced networking capabilities in various environments. One such network is a vehicular network formed between moving vehicles having computing and communication capabilities. With increase of vehicles on the roads, the
requirements of efficient traffic management and road safety are also increasing [52], [53]. A vehicle is considered as a computer-on-wheels due to the computing power of an on-board computer mounted on it. At the same time, great advances in cloud computing made it possible to merge with the vehicular network for common interests [54]. There is a new emergence of VCC [55] to meet the challenges of traffic management and safety. VCC utilizes the online resources of vehicles and provides the services on pay-as-you-go model. It played a major role in addressing the challenges of ITS [56]–[58].

Traffic monitoring and management are some of the major issues of ITS. A sudden reduction in speed may lead to accidents and human injuries. Proper traffic monitoring and management can reduce the number of accidents [59]. VCC has a positive impact on traffic management by utilizing the underutilized resources of other vehicles [60]. Traffic management consists of traffic flow management and road safety.
management using static and dynamic vehicular clouds. For traffic management, basic services such as processing, storage, and communication are provided by the VCs formed near the vehicles [61] or by the cooperation of the permanent and temporary VANET clouds [62]. For efficient traffic flow, timely computing services are very effective. The completion of these tasks really depends on minimizing the deadline violation of the task assigned to the VCs. Task replication is used to increase the probability of success by distributing a single task among multiple vehicles in VCs [63]. ITS depends on data collection and analysis for traffic safety. VCC enables the traffic control authority to collect and analyze the data for better traffic management and user safety [64], [65].

Advances in computation and communication technologies through sensors, cameras, computing capabilities, network devices and storage facilities have made the modern vehicles intelligent and improved the passenger comfort and road safety. Furthermore, modern intelligent vehicles have enabled autonomous operation of vehicles which involves massive computation. Consequently, the need
of task offloading arises to ensure safe autonomous operations. De Souza et al., in [66] have studied the computation offloading in vehicular environments. They have classified the related studies, presented the tools and strategies involved in computation offloading, and discussed offloading scenarios presented in different review studies. Main research challenges and future research directions conclude their review study. Silva et al., have recently reviewed various computing paradigms in emerging vehicular environments in [67].

The main purpose of VCC is to effectively utilize the surplus resources of vehicles. VCC works in two aspects, i.e., utilization of surplus resources and fulfillment of users’ computational requirements according to the pay-as-you-go model [68]. Vehicles with surplus resources form the vehicular clouds (VCs) for effective and efficient resource utilization. The fundamental difference between conventional cloud computing and VCs is the ownership of the data center. Unlike conventional cloud computing, VCs have multiple owners. When vehicles use their resources cooperatively, two types of VCs are formed, namely static VC and dynamic VC [69], [70]. Architecture of VCC showing static and dynamic vehicular clouds is shown in Figure 5.

VCC provides solutions to traffic related problems dealing with traffic hazards and congestion. The solution comes in the form of VCs that provide the resources to solve the problems. Vehicles share their scattered resources to form a VC [71]. To meet the computational requirements of smart vehicles, outsourcing of computational tasks to a third-party infrastructure consisting of centralized clouds, cloudlets, and VCs is required. All tasks are efficiently distributed among earlier mentioned three layers using resource allocation and task scheduling algorithms [72]. Other work related to resource allocation consists of reward-based resource scheduling [73], heterogeneous resources [74], energy efficiency [75], semi-Markov decision process based [76], and utility- based scheme [77].

In the earliest survey of VCC, authors of [78] presented the two-tier general architecture of VCC. Services and applications are also explained in this paper. In another survey by Whaiduzzaman et al. [79], the authors presented a very comprehensive study on VCC. The architecture, application scenarios, services, and issues related to security and privacy are also discussed in this survey. In [80] and [81], the introduction of vehicular networks, its integration with cloud computing, applications, architecture, and security issues and usage of 5G are presented. Vehicular cloud network architectures, applications, challenges, role in traffic management, and future trends are discussed in [82], [83] and [61]. Skondras et al. [84] presented a comprehensive survey on medium access control schemes using 5G in VCC. An overview of the schemes as well as the implementation detail is presented with the conclusion of using software defined network (SDN) for centralized controls and 5G for distributed controls.

Vehicles using cloud (VuC) and hybrid cloud (HC) are the new fields emerged after integration of cloud computing with vehicular networks [85]. In this survey, taxonomy, applications, architecture, and services for VCC are discussed in detail. A comprehensive explanation about the other two categories, VuC and HC, is missing in [85]. VCC services and simulation frameworks are explored by authors of [86], resource scheduling is discussed at [87], and VCC security issues are elaborated by the authors of [78], [88]. Olariu [89], has presented a state of the art survey on promising applications and new challenges of VCC.

One of the prevailing problems of vehicular clouds is mobility, which affects resource prediction. A framework [90] named “learning-based task offloading is developed that allows vehicles to learn task offloading to nearby vehicles. Load awareness is also addressed in this work. Simulations show better results of the proposed framework. VCC utilizes the resources managed by RSUs and other vehicles with excess resources. This requires offloading of tasks and then collecting results in a timely manner. Another problem is managing the replication of tasks. To deal with the complexity and minimize the replicas in task assignment, balanced task assignment policy (BTAP) has been proposed. The deadline violation probability is minimized by using the “optimal multitask replication policy” [63].

Mustafa et al. [91] proposed a novel Predictive VM Migration solution for mobility prediction in VCC that reduces the impact of mobility on performance. They used a scheme based on artificial neural network (ANN) models for resource management. The primary purpose of the scheme was to reduce the negative impact of abrupt changes in vehicle locations. In the study, a comparison with other schemes is made using simulations based on the “Nagel-Shreckenberg cellular automata discrete model”. The results show that the scheme proposed at [91] achieves better results.

C. MOBILE CLOUD COMPUTING

The rapid increase in the number and size of emerging applications has increased the need for computationally intensive processing. This increase has led to the option to use third-party resources outside of the computing device. Cloud computing, a next-generation computing infrastructure, offers its services such as SaaS, PaaS, and IaaS on the pay-as-you-go model. MCC has emerged by the conflation of cloud computing and mobile computing devices which comes up with processing and storage services to mobile users by taking advantage of cloud computing. In MCC, processing of applications is shifted to the cloud servers from the mobile computing devices. The primary features of MCC comprise of increased battery life, reduced computation time, higher reliability for the task completion, and scalability [20]–[23].

However, offloading tasks to the remote clouds can introduce supplemental overhead that compromises QoS. There is a need to develop effective computation offloading algorithms that minimize the processing delay for real-time applications. Average execution time is one of the fundamental goals that need to be optimized. Energy efficiency is another objective that needs to be optimized. MCC [24] domain
focuses on the provision of storage and computation services from cloud server(s) to save time and energy of mobile devices. Applications, which need more processing, like biometric application consisting of fingerprint, face, or iris scanning can be offloaded to the cloud servers [25]. Using the game theoretic approach, the user decides the offloading decision to the cloudlet, remote cloud or local execution of the task [26]. Wu et al. [28] optimized energy usage by proposing an algorithm for selection of computation processing type either entire local computation, or entire cloud based computation, or partial local computation and partial cloud based computation. Results proved the efficacy of proposed algorithm.

In [29], Fernando et al. presented a motivations, different issues, and security concerns of MCC. Efficient energy estimation methods are discussed in [30] whereas, mobile cloud application models and rich media applications are described in [19], and [18], respectively. Mollah et al. [92] presented a comprehensive survey on security challenges, their solutions, and future directions in MCC. Three important components of MCC i.e., offloading, privacy, and distributions are discussed in another survey [93]. However, the authors of [94], and [95] have considered smartphones as mobile nodes rather than vehicles in the study.

Attempting to achieve several optimization objectives was also studied in offloading. In a solution relevant to offloading [20], two objectives, i.e., makespan and energy were considered and algorithms were designed to attain a lower makespan and address energy issues. Authors of [21] have discussed a queuing model for the central cloud and the cloudlet to minimize the average task execution time using a hybrid genetic algorithm. In [22] a heuristic algorithm with minimum financial cost was proposed to optimize the energy with low delay. Guo et al. [96] presented an offloading strategy that addresses the offloading error problem and also reduces the average execution time with less energy consumption. Energy optimization and task execution delay are addressed with three different locations for offloading the tasks, i.e., the device itself, a centralized cloud, and the cloudlet in [23].

**D. MULTI-ACCESS EDGE COMPUTING**

MCC addressed the challenges of limited resources in handheld devices, which resulted in more time spent due to the remote locations of mobile clouds. To address this issue, MEC implemented a solution that places host servers at the edge of the network. In 2013, the MEC concept was introduced when Nokia Siemens Networks and IBM proposed a platform for running applications in base stations [97]. The basic function of MEC is to offload the computationally intensive and energy consuming tasks to servers in nearby base stations. Normally, mobile computing devices use cellular communication to connect to the MEC server. Offloading to MEC servers can be partial or complete. In partial offloading, some parts of the task are offloaded to the server, whereas all parts of the tasks are offloaded for processing in full offloading. Due to the use of nearby servers in MEC, the delay caused by a remote cloud is minimized. MEC can also be used for offloading tasks from vehicular networks [98]. Figure 6 shows the offloading decisions in MEC.

Offloading computational tasks to remote cloud servers is a very effective way to overcome the resource constraints of mobile devices. However, such offloading involves higher bandwidth and delay. MEC is proposed by the research community to bring computing and storage services in the proximity of user yao2019mobile. Many surveys have been published to explain MEC focusing on different areas such as communication perspective [99], convergence of caching [100], MEC using terminal computation and caching [101], and computation offloading in MEC [102]. In another survey, the contribution of the internet of things (IoT) is considered as a use case of MEC as a realization of future generations of networks such as 5G [103].

Minimizing average execution time and reducing energy consumption are the two fundamental goals of computation offloading in MEC. To achieve these goals in MEC, several works are proposed in the existing literature. Wang et al. [104] proposed “EdgeFlow”, a processing system to utilize the services related to computation where the entire network is used for the offloading purpose. The proposed system was a multi-layer data flow model to minimize the makespan where all the network entities like edge and cloud servers are used. To attain the lower makespan for delay-sensitive applications, use of partial offloading is very effective. A heuristic resource allocation algorithm [105] has been designed to optimize the latency in MEC. In another work [106], multiple edge users are considered for offloading the applications to cell phones and the edge servers by using non-orthogonal multiple access. In [106], layered algorithm is used to optimize the makespan delay.

With the emergence of new in-vehicle applications, a challenge of limited resources in a single vehicle can be addressed in many ways including cloud and edge. Sun et al. [107] have presented the idea of mobile MEC technology servers among the vehicles moving on the roads. A vehicle with limited resources offloads the required jobs to the nearest MEC server in its range. Simulations are carried out using NS3 and effectiveness of the algorithm regarding makespan and reliability are verified. The main problem area for most of the research work [108], [109] has been the lower connection time of users with static servers. By using scheduling algorithms in distributed environments, it is possible to effectively manage the demand on communication and compute resources. VEC assumes the placement of a server at RSU which lie between the vehicles on the roads and the distant cloud server placed thousands of miles away. Then multiple servers were installed on different RSUs to balance the load on a single server. Due to the shorter link duration between the vehicles and the static servers, servers were installed on special vehicles moving among the user vehicles.
E. VEHICULAR EDGE COMPUTING

Implementation of MEC in vehicular networks improved interactive responses during computational offloading for delay-sensitive applications due to the proximity. Because of the proximity, implementing MEC in vehicular networks improves makespan for real-time offloading. However, the extra load on MEC servers due to computation offloading from vehicles with dense traffic, jeopardizes the optimal response time [110], [111]. In VEC [112], [113], the computation and processing tasks are analyzed at the edge of the network. It is located in the middle of cloud and the vehicles. VEC is used to minimize the high latency in case of VCC. Authors of [114] and [115] have discussed reputation management and low latency caching, respectively in their studies.

The concept of VEC is shifting the computation at the edge of the network near the vehicles at one hop to fulfill the requirements of communication and computation for ever-growing vehicular applications. A detailed survey explaining the architecture, applications, and future direction is presented by Raza et al. [44]. Studies at [116] and [117] have covered computational offloading schemes and methods and offloading and content caching delivery, respectively. In another survey [118], overview, architecture, challenges, and future research directions of VEC are presented.

With advancements and growth of cloud and edge architectures, the energy requirements also increased. Therefore, the researchers working on computation offloading in VES focused on energy efficiency as a primary parameter as shown in BEGIN [119]. In the framework proposed by authors of [119], VEC and big data are integrated and the energy consumption control is demonstrated through a case study. A framework for energy efficient resource allocation to in-vehicle user devices is presented in [120]. In another work [121], a framework called Chimera has been developed to reduce energy in vehicles with crowd-sensing applications. An appropriate decision for the selection of optimal proportion of the application is made on the basis of energy states [122]. Other task offloading algorithms in VEC that consider energy efficiency as a parameter, are computation/communication bit allocation [123], dynamic application offloading [124], and heuristic resource allocation algorithm [125].

To minimize the delay for vehicle unloading, a distributed cloud-based VEC offloading algorithm [126] is designed. Numerical analysis shows the effectiveness of the proposed method. Authors have considered different strategies which are to increase the turnover of the servers and decrease the optimum for task shifting. However, the effect of mobility on delivering results to specific vehicles is not considered. Time-varying channels can affect task transfer during task offloading to vehicular edge computing. To address this issue, time-varying channels were replaced with fixed spectrum efficiency. Two algorithms, linearization-based branch-and-bound and fixed spectrum efficiency-based heuristics, were proposed in [127] and the results showed their effectiveness. In [128], the idea of fiber connection from edge to cloud server with two efficient task offloading schemes is presented to minimize the task latency. Numerical results prove the efficiency of the proposed models.

Dai et al. [129] considered both the issues of task offloading and load balancing. An algorithm named JSCO, was developed which solves the server selection and task offloading problem. Simulations proved the effectiveness of the presented solution. A task offloading algorithm [130] using a support vector algorithm has been proposed to address the increase in delay due to mobility of vehicles. In this algorithm, the task is divided into sub-tasks and then outsourced to
a particular server or processed locally. Simulations are performed to evaluate the proposed idea and the results proved the author’s claim of low response time, fast convergence and better decision accuracy. Lin et al. [131] considered the core network congestion problem and high latency problems. A social edge for VANETs was proposed to use the edge and social network services together. To increase the utility of the network, Lagrangian theory has been used to obtain the optimal resource allocation. The VEC evolution is shown in Figure 7.

F. VEHICULAR FOG COMPUTING

VFC, a new computing paradigm, emerged by merger of the fog computing and vehicular networks [132]. VCC uses cloud’s resources as well as the vehicles’ resources for task executions. However, distant cloud server and centralized decision making can lead to load on transmission capacity and delay in transmission capacity. Fog computing has resolved the issue of distant clouds by shifting computation at the edge of the network [133]. These services are provided near the end users where most of the data is generated, to meet the low latency requirement. Fog computing is not complementing cloud computing but uses cloud computation services for computation-intensive tasks. Vehicular fog node is another option for vehicular networks and fog computing to fulfill the computational needs of VANETs.

VFC can outsource the computational resources from fog nodes, edge devices near the users, and parked/moving vehicles. In [134], placement of number of infrastructures is reduced by replacing it with group of parked vehicles. VFC is however, based on the concept of leveraging resources from vehicles whether moving or parked, terminals, edge devices, or fog nodes. Being geographically distributed, VFC supports mobility, and is nearer to the end users [135]. In [136], VFC is presented in three layers consisting of cloud, cloudlet, and end-user layer that cooperate with each other to balance the network load. A survey on VFC [137] describes VFC taxonomy, its architecture, and the possible deployment challenges. In this study, vehicles are considered as fog-users and fog-nodes on the basis of their status. Study confirms the efficiency of VFC for vehicular resource utilization and quality of service (QoS).

Resource allocation is a significant challenge due to geographically dispersed resources of VFC. Therefore, it is necessary to allocate resources in a way that minimizes the latency. The admission control problem in [138] is tackled using a game theoretic approach for applications with varying QoS criteria whereas The goal of scalability has been met using the proposed scheduling algorithm. In another study [139], transportation vehicles are used as “fog nodes” in a semi-Markov decision method to outsource the jobs. The application-based policy is utilized for offloading to maximize the long-term benefit and attain the best resource allocation.

Zhou et al. [140] provided a concept for reducing base station load by utilizing the vehicles’ surplus resources through an efficient incentive mechanism and the use of a reliable pricing based matching algorithm. To handle the parking mechanism, VFC has been used in [141]. In this study, with the use of fog nodes and smart vehicles, a plan is devised to guide automobiles to parking places. Efficiency is achieved with the help of parked and moving vehicles with surplus resources. The vehicles that took part in task shifting were given monetary incentives. In [142], authors have discussed the scenario when occupied vehicles leave the system and their uncompleted tasks are to be offloaded to other vehicles. Simulation results proved the accuracy of proposed task offloading scheme.

G. VOLUNTEER COMPUTING BASED VANET

Recent technological developments have produced many promising applications that provide security and convenience to users in VANETs. These applications require extensive computational and storage resources that cannot be met by a single vehicle. These vehicles can offload their applications to cloud or edge servers by paying service fees. These services can be interrupted due to many reasons such as disasters or overload of certain servers. Moreover, traffic jams occur during peak hours due to heavy traffic, wasting the resources of vehicles on a daily basis. To reduce the load on the dedicated servers and to utilize the wasted resources, a new paradigm called VCBV has been recently developed. It utilizes the excess resources of the vehicles by taking over the execution of tasks for nearby vehicles. VCBV can be deployed using ad hoc, infrastructure or hybrid task coordination [45].

Use of volunteer computing in vehicular network can be done with or without the intervention of infrastructure. However, use of RSU, as a central identity, is very beneficial regarding priority based job scheduling. In a recent research [143], effectiveness of infrastructure-assisted coordination is shown where location based task replication is employed to attain fault-tolerance. To address the limitation of shorter link duration between the vehicles and centralized infrastructure, ad-hoc coordination between the host and user vehicles is inevitable. In ad-hoc task coordination, the job initiator directly interacts with volunteers and completes the task coordination process. In the absence of the infrastructure, ad-hoc coordination among nearby vehicles is very beneficial for task execution and resource utilization.

Hybrid VCBV [144] is very useful for dense traffic and might be used in the case when an RSU has several jobs to execute. In this type of computing, usually, RSU initiates the coordination and then infrastructure and ad-hoc based coordination are carried out simultaneously to utilize the more resources. In [145], Vehicular ad-hoc network (VANET) is merged into volunteer computing to accomplish the task execution using a central manager of the jobs. Different algorithms were used to evaluate the parameters like latency, throughput, and rate for job completion. A detailed description of VCBV and its types is shown in Figure 8.
III. COMPARISON OF VEHICULAR COMPUTING PARADIGMS

This section focuses on the comparison of vehicular computing paradigms in the context of remote server task execution and resource utilization. We compare the salient features of these computing paradigms to understand their significance to vehicular networks. In vehicular networks, when a task cannot be computed locally, it is outsourced to
any of the computing paradigms. This computing paradigm either uses its dedicated servers or the surplus resources of the vehicles. All the computing paradigms explained above can be categorized into two types according to their dominant characteristics over dedicated or distributed resources. Figure 9 shows the types of computing paradigms based on the type of resource usage. We compare vehicular related computing paradigms in terms of task execution and resource usage. In the Table 2 and the Figure 11, we have also shown some similarities and differences among various computing paradigms.

A. MCC VS VCC
With advancements in cloud computing, it is possible to perform data storage and data processing for mobile devices. In MCC, mobile computing devices use the facility of task offloading to remote cloud servers to save time and energy. To use this capability, a mobile device must have the ability to connect to a remote cloud server. VCC is not only used to provide data storage and processing services from the cloud, but also employed to use the excess resources of vehicles. The basic difference between these paradigms is of resource utilization. VCC utilizes the vehicles’ resources and MCC entirely relies on remote servers.

B. VCC VS VFC
VCC and VFC both belong to the category of distributed resources and satisfy task execution requirements by using the surplus resources of vehicles. VFC is an extension of VCC where dedicated servers and the central controller are kept closer to the deployed vehicles. The differences between these two domains include the physical distance to dedicated servers and the capacity of dedicated servers. VCC has servers with more resources and provides services for longer distances as compared to VFC.

C. MCC VS MEC
Although motivation/objectives behind emergence of MCC and MEC are the same, however, MEC is a better choice for the execution of delay-sensitive applications because of less distant MEC servers as compared to MCC servers. Location of the remote server is the only difference between these two domains. MCC uses internet connectivity to connect with remote servers whereas, MEC uses cellular communication for connection among participating nodes.

D. VEC VS VFC
VEC and VFC, both domains apparently seem alike. In both computing paradigms, the main server is located at the edge of the network with the objective of decreased latency as compared to cloud computing. VEC provides the computing services to the vehicles at the edge of network and very close to RSU whereas, VFC provides the computation services to the edge vehicles through fog nodes that are the vehicles or devices having surplus resources.

Study of various computing paradigms has concluded that the physical distance of the cloud from the end-users is maximum. Therefore, the user will experience high latency, a higher number of hops, less bandwidth while it will access cloud computing for task offloading. However, complexity and severity of all the above-mentioned issues will be improved when we move from cloud to fog and edge computing, respectively. However, in other aspects, the cloud has more computational resources, and these resources decrease when we move to fog and edge, as shown in Figure 10.

IV. APPLICATION AREAS OF COMPUTATION OFFLOADING IN VEHICULAR NETWORKS
In this section, all applications or use cases for the tasks offloaded from vehicles to any of the existing computing paradigms are explained. For consistency, the term “server” is used in the following explanations to refer to all servers associated with any computing paradigm.

A. OBJECT DETECTION
The surveillance applications used for traffic monitoring and control, continuously capture real-time images/videos of interest with the vehicle-mounted high-resolution cameras [146] for object identification. This object can be a human face or a license number plate involved in an event that needs to be recognized. Object identification task involves image acquisition, recognition, pre-processing, feature extraction, and searching of database [147], [148]. The license number plate and/or human face is extracted and searched by special software on the cloud, edge, or fog servers and query results are forwarded to the appropriate authorities for consideration and necessary action(s). Since these jobs are delay-sensitive; therefore, require real-time processing and quick responses. Accurate detection/recognition of object/face is the performance objective while keeping the network capacity intact. Examples include signboard, number plate, human face and animal recognition etc.

B. TRAFFIC MANAGEMENT
Vehicular networks provide communication and sharing of data between vehicles. Increased vehicles on the roads infrastructure are increasing the threat of road safety and traffic load on the roads. Reducing the risk of accidents and avoiding human injuries is one of the basic objectives of ITS. Application used to manage the road traffic consist of headway management, lane change warning, speed management, and weather detection [149]. Best route selection for vehicles is one of the types of traffic management applications. But the selection of the same route for all vehicles may increase the congestion on that route. The analysis must be done on the pattern of the traffic and best route selection for some specific vehicle for load balancing [150]. Overcoming the main challenges like resource management, awareness for warning messages, congestion management through computing servers can lead to efficient traffic management [83].
The information of accidents and road blockage can be used for traffic management which is time-sensitive and requires accuracy.

C. AUGMENTED REALITY
Augmented reality (AR) is an advanced and emerging application which merges an additional virtual layer of data over reality and explains the scenes in more detail. Usually, visuals are used for AR, but sound and touch can also be used to further improve the user experience. Typically it adds what you see but also it can subtract it too by removing things out of the way [151]. In vehicles, the integration of synthetic information with original content is possible by using AR technology. AR warnings must be embedded in such a way that the cost of local information is minimal [152]. In vehicles, AR can be used to enhance the vision and to embed the warnings for safety and comfort using the third party remote servers like a cloud [153]–[155]. This application can also be used to provide driving assistance to those who are colorblind [156]. These applications increase the awareness of drivers about the environment to enhance safety [157]. In VANETs, AR can be used to warn drivers for hidden dangers, help drivers who are colorblind, and to provide more information about their surroundings. This application can be categorized as a safety as well as an infotainment application.

D. VIDEO ANALYTICS
Cameras are mounted in modern vehicles which can be utilized by using the pictures or videos captured for many purposes. Video analytics is the process of getting important information from the specific video. Crowdsourcing is a process in which multiple clients especially humans collect the information collectively [158]. Video crowdsourcing can be used in vehicles with dash mounted cameras.
These videos are sent to a server where these are analyzed for some common interests like traffic management or national security [159]. These videos can focus on different purposes according to situation e.g., driving assistance, road situations, weather conditions, parking spot and events detection etc.

**E. AUTONOMOUS DRIVING**

Idea of autonomous driving is taken from autopilot mechanism for aeroplanes. Nowadays, this can be used in health care, disastrous management, constructions and warfare [160]. Autonomous driving aims safe navigation from source to destination facing the changes in surroundings happened due to vehicles, pedestrians, or other obstructions. Moreover, unexpected events like road-blockage, congestion, and accidents are also needed to be dealt with using this system [161]. Three levels of autonomous driving where primary issues exist are the perception, planning, and control. Sensing and filtering are basic components of perception. Sensing collects the data around which is then sorted and filtered. In autonomous driving, the path trajectory is planned according to static or mobile obstacles’ locations. Planning level sends the signals to control level which is responsible for future trajectory and path planning. [162]. These high computational requirements are either provided by the sensors and devices installed in vehicles or from the servers of various computing paradigms [162]. These analytics are done on behalf of information collected from the sensors of the vehicle. However, sometimes the information collected from
one vehicle is not enough to have the correct decision of a path to be followed. Therefore, information from multiple heterogeneous sensors is needed to be processed for autonomous driving [163].

**F. CONTENT DISTRIBUTION**

In vehicular networks, there is a lot of content that need to be downloaded/uploaded like music, games, important information ahead like road congestion or incidents happened [164], [165]. The delivered content might be a small file having important information in the form of bits or a large video file which is captured from camera installed anywhere. The content disseminated might be required by the vehicles or uploaded to cloud or to specific department [166]. Computing servers having specific content, provide the required content services to end-users [167]. The content distribution consists of a detailed map, lane information, traffic signs, or advertisement related data. Some of the content distributions like emergency warnings might be delay-sensitive and require low latency whereas, infotainment related content distribution e.g. data distribution may not be delay-sensitive.

**G. PARKING ASSISTANCE**

With the rapid growth in the number of vehicles, limited parking spaces cause severe parking-related issues. These parking issues are also increasing traffic congestion and wastage of resources. Finding the place and parking vehicle with less effort and fuel is the major concern of the parking assistance system. There are many designs for smart parking in the literature using third party reliance. In a parking management
system [168], when a vehicle enters or leaves the parking slot its status is sent to the server for record and computation. Finding a parking slot and vehicle’s request management is done at the server side using the information from sensors indicating available spaces [169]. By taking vehicles directly to available parking places saves time, minimizes the fuel consumption, and improves traffic and parking efficiency. Real-time parking services can also be provided using remote servers for handling the requests by locating the suitable parking slots [170]. Examples of parking assistance include finding free parking spots and checking parking time etc.

V. CURRENT CHALLENGES AND FUTURE DIRECTIONS
In vehicular networks, opportunities for task execution vary from cloud to edge, and fog computing that provides alternate computing options and increases the possibility of in-time task completion. New technologies like edge and fog are introduced by research community for real-time and delay-sensitive applications. However, there are some constraints in using these paradigms to fulfill the on-time completion of tasks. Cloud computing has limitations of controller failure or disconnection to the server. Same is the case when an edge server itself or connection to the edge server is failed. A vehicle in far-flung areas with no connection to the Internet or RSU, cannot connect with a remote server for task offloading and execution. The same can happen in disaster situations when connectivity to remote servers is not possible due to infrastructure damage.

Remote and proximity vehicles provide data and storage services in the form of centralized and distributed resources. Each of the service provision technique has its pros and cons which have been discussed in previous sections of this manuscript. In this section, we discuss the prominent research challenges and future directions for task execution in vehicular networks.

A. MOBILITY
High speed vehicular networks face frequent changes in road traffic topology due to their dynamic nature. Vehicles on the road always change their neighbors and most of the time their mobility patterns are not predictable. This dynamic nature of vehicles can lower the success probability in task offloading. A vehicle requests for task execution and changes its position before completion of task. Now, the server must locate that vehicle using smart techniques and algorithms. In case of resource utilization, the change of position for vehicles joining or leaving the vehicle’s cluster can decrease the performance of the computing paradigm. Therefore, research community should focus on vehicles mobility to ensure the performance and efficiency of moving vehicles.

B. CONNECTIVITY
Connectivity refers to the connection of a vehicle with another vehicle or with a remote server. To utilize the services of a cloud server, there is a continuous need of stable internet connection from task offloading till collection of results. Facing disconnection during this interval can cause the data loss resulting in the failure of the whole process. In the case of MEC, a cellular connection is required to offload and get the result for the task execution. Even for VFC and VEC, connectivity to the central controller is always required. Having no connectivity or disconnection between the task offloading and result collection can cause failure in task execution. Therefore, connectivity issue is one of the prominent issues of all computing paradigms.

C. QoS SUPPORT
In task execution, latency is considered the major factor to meet QoS requirement of vehicular networks. However, other parameters like bandwidth, task input size, and resource utilization are also required to be achieved to have a full QoS support. Controlling the network overhead is an important method to achieve required QoS support. Therefore, managing the communication overhead, during connection setup and task offloading, is the key to QoS management.

D. VOLUNTEERS DISCOVERY AND RESOURCE REQUEST
Cloud provides services to users on pay-as-you-go model. In volunteer computing, such task computation services are offered free of cost by the volunteers having sufficient spare resources. Users are usually unaware of the availability volunteers on the way. Therefore, invention of an effective methodology to indicate or forecast the presence of volunteers willing to present their computing resources in someone’s route, is an interesting and important research area. Further, this will help promote free of cost computing.

E. FAULT TOLERANCE
In all computing paradigms, failure of the central controller always leads to task execution failure which consequently leads to re-transmission. When a submitted task is not completed and task failure happens, it is again offloaded to the same or different host which causes delay due to re-transmission. To avoid task failure, same task is offloaded to several resources for successful completion of the job which is called task replication. Fault tolerance is very critical for delay-sensitive tasks. For fault tolerance, tasks must be distributed intelligently to avoid the high transmission latency. To achieve fault tolerance, there can more resource utilization to avoid the delay or failure of some of the tasks. In such cases, establishment of a fault tolerant network, is an important issue.

F. ENERGY MANAGEMENT
Computation offloading is an important task to be executed in vehicular networks and involves energy consumption. Mobile vehicles are equipped with millions of IoT devices which use a massive amount of energy for task offloading and task execution. Future transportation system is considering electrification of vehicles as an important research area. Energy management would be the most important concern for emerging electric vehicles. Therefore, energy management of...
MEC, MCC, VEC, VCC and VFC should be focused by the researchers.

G. IMPROVED LOCAL CAPABILITIES
Mobility of devices and the vehicles is the basic feature in vehicular networks with offloading as an essential ingredient. Offloading leads to energy consumption, delays, privacy etc. issues. To overcome the offloading related issues it is pertinent to mention here that special emphasis should be paid to improve the local processing capabilities of mobile nodes (vehicles, mobiles etc.) to minimize the need of task offloading and to avoid the issues related to offloading.

H. RELIABLE LOAD FORECASTING
Vehicles on the road use RSUs for computation purpose. During busy hours, more traffic on the roads causes higher computation demand from RSUs which may lead to computation queues and consequently may add more delays in response. Increased delay in response will further complicate the task execution. Therefore, a reliable load forecasting mechanism using effective and efficient machine learning approaches is an important need of mobile nodes to manage the traffic routes, load on RSUs and minimize the possibility of processing delays caused by traffic congestion.

I. INTEGRATION OF RENEWABLE ENERGY SOURCES
Various studies have pointed out the abrupt rise in traffic load on the roads. Consequently, the mobile nodes and their processing demands have also increased a lot. Increased traffic on the roads leads to higher carbon emissions caused by higher energy demands. To deal with higher carbon emissions and thereby the environment pollution, integration of renewable energy sources with mobile nodes, RSUs and other processing servers is an important need of the hour.

J. TASK DISTRIBUTION
Task distribution is perhaps the most crucial feature of offloading computations in vehicular environment and has very critical role in application processing time, percentage of failures and overall performance of task execution. Therefore, an intelligent and optimal task distribution is very important research challenge. Some important topics covered in task distribution are given in the following.

- Task priority
- Task dependence or independence
- One-hop or multi-hop scenario
- Communication technology i.e. 4G, 5G, DSRC, LAN, WiFi etc.
- User behavior
- Selection of the best computing paradigm
- Granularity level of task distribution

K. ENVIRONMENT AND HUMAN HEALTH ISSUES
Biomass emissions, desert storms, wildfires and air pollution are the main contributing sources to human health problems like high blood pressure, respiratory issues, heart disease etc. As mentioned earlier, increase in road traffic is a vital contributing factor in air pollution therefore, a better traffic management policy so as to decrease its participation in air pollution is an important challenge for research community.

VI. CONCLUSION
Task offloading is an unavoidable action, especially in resource-constrained nodes, and the research community has proposed various techniques and methods for effective and successful task offloading. In this study, we reviewed the current state of literature on task execution and resource utilization in vehicular networks. We started by introducing computational paradigms for task execution in vehicular networks by describing their architectures. Similarities and dissimilarities between different computing paradigms are explored to distinguish the key features. Use cases or applications are presented to explain the need for task execution in vehicular networks. Finally, we have discussed research challenges of various computing paradigms involved in vehicular networks. This study is equally beneficial for novice, as well as, the experts of domain.

REFERENCES
A. Waheed et al.: Comprehensive Review of Computing Paradigms, Enabling Computation Offloading and Task Execution


A. Waheed et al.: Comprehensive Review of Computing Paradigms, Enabling Computation Offloading and Task Execution


[151] A. Waheed received the B.Sc. degree in double math from Punjab University, Lahore, Pakistan, the M.C.S. degree (Hons.) from Baqi University, Karachi, Pakistan, and the M.S. degree in computer science from COMSATS University Islamabad, Pakistan, where he is currently pursuing the Ph.D. degree. His research interests include provenance inference, cryptography, vehicular ad-hoc networks, flying ad-hoc networks, wireless sensor networks, underwater wireless sensor networks, volunteer computing-based VANET (VCYB), and location privacy in VANETs. He received the Gold Medal from Baqi University for his M.C.S. degree.


[165] ABDUL WAHEED received the B.Sc. and M.Sc. degrees in computer science from the University of Peshawar, Pakistan, in 2001 and 2003, respectively, the M.S. degree in security technologies and applications from the University of Surrey, U.K., in 2010, and the Ph.D. degree from the University of Bedfordshire, U.K., in 2013. His research interests include MAC protocol design, QoS and security issues in wireless communication systems, and the Internet of Things.

[166] MUNAM ALI SHAH received the B.Sc. and M.Sc. degrees in computer science from the Virtual University of Pakistan, in 2011 and 2016, respectively. He is currently pursuing the Ph.D. degree with COMSATS University Islamabad, Pakistan. He has a total of 13 years experience in system/network administration, teaching, research and development. He has presented his research work in national and international conferences. He has published many research articles in conferences and peer-reviewed journals. His research interests include smart grid, the Internet of Things, cloud computing, green energy, cyber security, and related areas. He has served as a TPC member and an invited reviewer for international conferences and journals. He was a recipient of the Higher Education Commission (HEC) Pakistan Merit Scholarship for his Ph.D. studies.

[167] SYED MUHAMMAD MOHSIN received the B.S. and M.S. degrees in computer science from the Virtual University of Pakistan, in 2011 and 2016, respectively. He is currently pursuing the Ph.D. degree with COMSATS University Islamabad, Pakistan. He was a Postdoctoral Fellow at the Politecnico di Torino, Italy, from 2009 to 2011. He has more than 14 years of teaching and research and development experience. His research interests include security and privacy issues in distributed systems, applied cryptography, and quantum cryptography.

[168] ABID KHAN received the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 2008. He is currently with the Department of Computer Science, Aberystwyth University, Aberystwyth, U.K. Before that, he was associated with COMSATS University Islamabad, Pakistan. He was a Postdoctoral Fellow at the Politecnico di Torino, Italy, from 2009 to 2011. He has more than 14 years of teaching and research and development experience. His research interests include security and privacy issues in distributed systems, applied cryptography, and quantum cryptography.
CARSTEN MAPLE (Member, IEEE) is currently a Professor of cyber systems engineering at the Cyber Security Centre, University of Warwick, where he leads the GCHQ-EPSRC Academic Centre of Excellence in Cyber Security Research. He has published over 200 peer-reviewed papers. He has provided evidence and advice to governments and organizations across the world, including being a high-level scientific advisor for cyber security to the European Commission. He is a Principal or a Co-Investigator on a number of projects in cyber security. He is a fellow of the Alan Turing Institute. He is the Immediate Past Chair of the Council of Professors and Heads of Computing in the U.K.

SHERAZ ASLAM (Member, IEEE) received the B.S. degree in computer science from Bahauddin Zakariya University (BZU), Multan, Pakistan, in 2015, and the M.S. degree in computer science with a specialization in energy optimization in the smart grid from COMSATS University Islamabad (CUI), Islamabad, Pakistan, in 2018. He is currently pursuing the Ph.D. degree with the DICL Research Laboratory, Cyprus University of Technology (CUT), Limassol, Cyprus, under the supervision of Dr. Herodotos Herodotou, where he is also a part of EU-funded research project named as STEAM. He worked as a Research Associate with Dr. Nadeem Javaid during his M.S. period at CUI. He has authored around 40 research publications in ISI-indexed international journals and conferences. His research interests include data analytics, generative adversarial networks, network security, wireless networks, smart grid, cloud computing, berth scheduling at maritime container terminal, and intelligent shipping. He served as a TPC member and an invited reviewer for international journals and conferences.

SHAHAB SHAMSHIRBAND (Member, IEEE) received the M.Sc. degree in artificial intelligence from Iran and the Ph.D. degree in computer science from the University of Malaya (UM), Malaysia, in 2014. He was an Adjunct Assistant Professor with the Department of Computer Science, Iran University of Science and Technology. He also worked as a Senior Lecturer with UM and Islamic Azad University, Iran. He has participated in many research programs within the Center of Big Data Analysis, IUST, and IAU. He has been associated with young researchers and elite club, since 2009. He has supervised or co-supervised undergraduate and postgraduate students (master’s and Ph.D.) by research and training. He has authored or coauthored articles published in IF journals and attended to high-rank A and B conferences. He is a Professional Member of ACM. He is an associate editor, a guest editor, and a reviewer of high-quality journals and conferences.