IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 5, 2018, accepted March 12, 2018, date of publication March 15, 2018, date of current version April 18, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2815989

Fog Based Intelligent Transportation Big Data Analytics in The Internet of Vehicles Environment: Motivations, Architecture, Challenges, and Critical Issues

TASNEEM S. J. DARWISH, (Member, IEEE), AND KAMALRULNIZAM ABU BAKAR
The authors are with the Pervasive Computing Research Group, Faculty of Computing, Universiti Teknologi Malaysia, Skudai 81310, Malaysia

Corresponding author: Tasneem S. J. Darwish (tasneem83darwish@gmail.com)

This work was supported by the Post-doctoral Fellowship Scheme of Universiti Teknologi Malaysia for the project through Big data analysis on vehicular networks for intelligent transportation systems.

ABSTRACT The intelligent transportation system (ITS) concept was introduced to increase road safety, manage traffic efficiently, and preserve our green environment. Nowadays, ITS applications are becoming more data-intensive and their data are described using the "5Vs of Big Data". Thus, to fully utilize such data, big data analytics need to be applied. The Internet of vehicles (IoV) connects the ITS devices to cloud computing centres, where data processing is performed. However, transferring huge amount of data from geographically distributed devices creates network overhead and bottlenecks, and it consumes the network resources. In addition, following the centralized approach to process the ITS big data results in high latency which cannot be tolerated by the delay-sensitive ITS applications. Fog computing is considered a promising technology for real-time big data analytics. Basically, the fog technology complements the role of cloud computing and distributes the data processing at the edge of the network, which provides faster responses to ITS application queries and saves the network resources. However, implementing fog computing and the lambda architecture for real-time big data processing is challenging in the IoV dynamic environment. In this regard, a novel architecture for real-time ITS big data analytics in the IoV environment is proposed in this paper. The proposed architecture merges three dimensions including intelligent computing (i.e. cloud and fog computing) dimension, real-time big data analytics dimension, and IoV dimension. Moreover, this paper gives a comprehensive description of the IoV environment, the ITS big data characteristics, the lambda architecture for real-time big data analytics, several intelligent computing technologies. More importantly, this paper discusses the opportunities and challenges that face the implementation of fog computing and real-time big data analytics in the IoV environment. Finally, the critical issues and future research directions section discusses some issues that should be considered in order to efficiently implement the proposed architecture.

INDEX TERMS Vehicular and wireless technologies, intelligent transportation systems, data preprocessing, real time systems, ubiquitous computing.

I. INTRODUCTION

Every year, about 8 million traffic accidents occur, resulting in injuring 7 million people and ending the lives of 1.3 million people, approximately. Around 90 billion hours of our time is wasted because of traffic problems (accidents, traffic jams), which causes a drop of 2% in the global total domestic productivity. Moreover, 220 million metric tons of carbon equivalents are generated by vehicular transportations [1], [2]. Every year, in the United States the cost of personal cars transportation (excluding commercial and public transports) is around 3 trillion USD, where 40% of this cost is due to parking, crashes, pollution, and traffic services [2], [3]. To improve the performance of transportation systems, enhance road safety, and preserve the environment, the concept of the intelligent transportation system (ITS) was introduced [4]. The emergence of ITS was highly supported by the advancing in sensing and communication technologies and the evolution in the effective integration of networked information systems, decision making, and physical infrastructure [5], [6].

Over a decade ago, the conceptual idea of Vehicular Ad Hoc Networks (VANETs) was introduced, where vehicles equipped with wireless communication devices can form networks [7]. Basically, in a VANET the communication can be among vehicles (V2V) or between vehicles and infrastructure networks (V2I). The main goal of VANETs is to enhance road safety. However, to serve the expanding and new requirements of ITSs applications the concept of VANETs evolved into the Internet of Vehicles (IoV) [8].

In IoV environment, connected vehicles behave as platforms for sensing and monitoring the traffic congestions status, the conditions of roads, and the environment pollution levels [9]. Gartner forecasts that over 250 million connected vehicles, will be in use worldwide by 2020.[1] According to CISCO, the number of connected devices to the Internet is expected to exceed 50 billion by 2020 and it will increase to 1 trillion in 2030 [10]. Advanced driver assistance systems and ultimately self-driving capabilities need a lot of computing and communication capacity to excel in their compute-intensive and latency-sensitive tasks. According to Intel, a car needs to analyze and fuse a massive amount of sensor data (approximately 1gb/s) in order to make safe decisions. Meanwhile, the infotainment services for passengers are evolving from Internet services like Facebook and Email, to video streaming and multiplayer online gaming, and to augmented and virtual reality applications [11]. Thus, the generated ITS data in IoV environment is voluminous and exploiting such data requires big data analytics [12].

Data analytics, sensors power, high-speed networks, and cloud computing are utilized by more cities recently [5]. The report of the International Data Corporation predicted that by 2019 the market of big data will exceed US$125 billion [13]. Big data analytics has the potential to enable the desired ITSs real-time control [5]. The ITS big data analytics can be used to reveal trends, hidden patterns, unseen correlations, and achieve automated decision making [14], [15]. Nevertheless, big data analytics created not only many opportunities but also several challenges, including capturing, analysis, data processing, searching, sharing, storage, transfer, visualization, querying, updating [16]. In particular, employing big data analytics in the dynamic environment of IoV is still in an infancy phase and demands new solutions [17].

The ITS big data processing can be centralized or distributed. The centralized approach utilizes the power of cloud computing. However, the requirements of real-time computation are difficult to be met as the centralized approach fail to provide feedback responses quickly. In fact, the chance to prevent damages from hazardous events might be missed while waiting to receive a control decision from a centralized center [15]. In addition, the centralized processing requires that the ITS big data, which are normally geographically distributed, be transferred to the cloud center. Such big data collection at a centralized data center generates high network overhead, consumes the network resources and creates bottlenecks. As a result, extra delays will be experienced while waiting for a feedback from the data center, which contradicts the requirements of real-time or latency-sensitive ITS applications (e.g. safety applications).

One promising technology that serves the distributed processing of big data is fog computing. In contrast to cloud computing, intelligent data processing with faster responses and higher quality can be provided by fog computing [18]. Basically, data processing is parallelized at the network edge by fog computing technologies, which fulfil the low latency requirements and reduces the network overhead [15]. However, implementing fog computing in the dynamic environment of IoV to provide real-time ITS big data analytics is facing many challenges that are not covered by previous research studies.

A. MOTIVATIONS, GOALS AND PAPER STRUCTURE

This paper points out the opportunities and challenges that are rising due to employing fog computing for the purpose of real-time ITS big data analytics in IoV environment. After identifying the limitations of the previous related work, a three dimension system architecture is proposed, which includes the dimensions of IoV, intelligent computing and real-time big data analytics. The proposed architecture aims to serve the real implementation of real-time ITS big data applications.

Using Google, a statistical search was carried out for the occurrences of four combinations of searching keywords, which are: (1) IoV + big data + fog; (2) VANET + big data + fog; (3) Vehicular ad hoc networks + big data + fog; and (4) Intelligent transportation systems + big data + fog. The search considered the published research contributions of the years 2012-2017. Surprisingly, when the search considered only the occurrences of searching keywords in the published papers titles, the obtained result was zero. This means that uptodate there is no published research paper that focus on integrating the three dimensions of IoV, fog computing, and ITS big data. On the other hand, Figure 1 demonstrates the occurrences number of each of the aforementioned searching keywords combinations in the published papers text (paper body). By analysing Figure 1, it is obvious that there is a strong relation between IoV, fog computing and Big data in the literature and the number of papers related to these topics is increasing recently. However, as on the integration issues of IoV, fog computing and real-time big data analytics there are no published technical contributions and/or research, this consideration motivates the current work.

The reminder of the paper is organized as follows:

- Section II gives a brief overview about the Internet of vehicles concept, characteristics and advantages. Then the existing IoV system architectures are discussed.

[1]http://www.gartner.com/newsroom/id/3165317

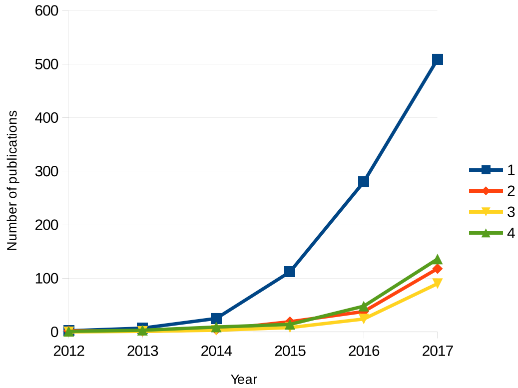

FIGURE 1. **Research trends of the four search strings over the years 2012-2017.**

- Section III explains the characteristics of ITS big data and its life-cycle. Afterwards, the opportunities and challenges of employing ITS big data analytics in IoV environment are discussed.
- Section IV presents the lambda architecture for real-time big data analytics.
- Section V discusses the concept of cloud computing and its weaknesses in handling real-time big data analytics in IoV environment. Then, several edge computing technologies are introduced with an emphasis on fog computing. Afterwards, the vehicular fog computing concept is introduced and the motivations to use fog computing for real-time big data analytics are explained. This section was concluded by a comparison between the discussed edge computing technologies.
- Section VI describes a general fog node and fog system architecture, then the available architectures in the existing studies are discussed.
- Section VII presents the proposed real-time ITS big data analytics architecture and its three dimensions.
- Section VIII discusses the critical issues and future research directions, which are related to the implementation of fog computing in the IoV environment, performing big data analytics in IoV fog environment, and security and privacy issues.
- Section IX concludes this paper.

II. OVERVIEW OF INTERNET OF VEHICLES

Worldwide, the number of vehicles for both private and commercial use was one billion in 2010 and is anticipated to be 2 billion by 2030 [19]. The conceptual idea of Vehicular Ad Hoc Networks (VANETs) emerged over a decade ago, and since then it has been a highly active area of research [7], [20]. The basic idea of VANETs considers vehicles as mobile nodes that can communicate to create a network [2]. Basically, due to mobility constraints, VANETs are considered as conditional networks, where their performance is affected by the vehicular density and distributions [21], [22], and various other factors such as bad drivers behaviours and tall buildings [2]. In addition, the vehicles are considered as unstable, temporary and random nodes. Thus, VANETs cannot guarantee the sustainability of applications/services for customers on large scale areas [23]. Therefore, VANETs are more suitable for limited scale applications that require ad hoc services such as preventing collisions or notifying drivers of hazards on roads. However, due to the Internet of Things (IoT) technology development and the increase in the number of Internet-connected vehicles new VANETs communication requirements are emerging. One more weakness of VANETs is their limited capabilities to process all the information that is captured by themselves and surrounding actors (such as mobile devices and sensors) [2]. To serve the new requirements of ITSs, vehicles must work as a smart platform of multiple sensors with IP-based Internet connectivity, several communication technologies, powerful computational components, and the ability to communicate with other vehicles and ITS devices [24]. In this context, the evolution of the conceptual idea of VANETs resulted in the introduction of the Internet of Vehicle (IoV) concept [8]. Thus, as a special case of IoT, IoV has distinctive characteristics and special requirements to serve the intelligent transportation systems.

An IoV is defined as a platform that realizes in-depth the integration and the information exchange between humans, vehicles, things, and the environment [25]. The main goal of IoV is to enhance the safety and efficiency of transportation, improve the service level of cities, save the environment, and ensure that humans are satisfied with the transportation systems services [23]. In contrast to VANETs, IoV integrates vehicles intelligence with vehicles networking, which results in intelligent networks with communication and computing capabilities that provide transportation services on large scale areas [23].

In IoV environment, as vehicles have permanent Internet connections, they can provide information for the various ITS applications categories (i.e. road safety, management and control of traffic, and infotainment). Consequently, information exchange is enabled among sensors and electric actuators, road infrastructures, and vehicles as well as drivers and passengers [2]. IoV collects large volume of data with various structures from a large scale area, which conforms with the big data heterogeneity concept [26].

With the significant advantages that IoV has over VANETs many new opportunities are opened. IoV offers various benefits to drivers, societies and economies. Cisco IBSG Automotive and Economics practices anticipated that every year the benefits of utilizing the IoV technology may reach $1,400 US dollars for each vehicle (summarized in Figure 2) [27], [28]. Moreover, traffic congestion reductions and road safety improvements can yield to major financial savings in public health sector. Furthermore, utilizing real-time traffic solutions through connected vehicles will lead to spending less time in traffic jams and increase productivity. More importantly, through IoV deployment, service providers will find opportunities to introduce new transportation services such as real-time traffic reporting, locating parking lots, and location-based customer service.

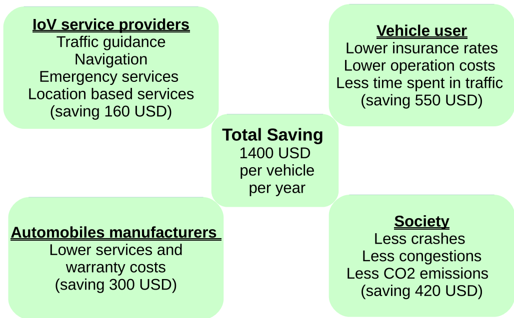

FIGURE 2. Some financial benefits of employing IoV.

Such services have high value not only for users but also for businesses [2]. The European Union estimated that by 2020 the global market value for IoV technologies and services will reach 115.26 billion Euros [29].

A. EXISTING IoV SYSTEM ARCHITECTURES

A network model of IoV was introduced in [23], which integrates humans, vehicles, things and the environment. A three-layer architecture was identified in [30], which describes the different IoV environment technologies interactions. The first layer consists of all the vehicle's sensors that collect environmental data and detect certain important events such as vehicles situations, driving patterns, and the conditions of surrounding environment. The second layer is for communications which supports different modes of wireless communications such as Vehicle-to-Pedestrian (V2P), Vehicle-to-Infrastructure (V2I), Vehicle-to-Vehicle (V2V), and Vehicle-to-Sensor (V2S). Through the communication layer, seamless connectivity is ensured to several networks such as IEEE 802.15.4, IEEE 802.11p, GSM, LTE, Wi-Fi, Bluetooth. The third layer has the IoV intelligence resources and is responsible for making decisions in risky situations (e.g. dangerous road conditions and accidents). This layer contains statistics tools as well as the collected big data storage and processing resources.

A four layers IoV architecture was proposed by CISCO [31]. The end users layer includes IEEE 802.11p based V2V communications, vehicles, and required software. All technologies that are necessary for communications between IoV actors are defined in the infrastructure layer. Afterwards, for the flow-based management and to monitor the policy enforcement, the operation layer was introduced. Finally, the services offered to drivers through cloud computing are specified through a service layer. However, the aforementioned IoV architectures suffer the following weaknesses: 1) network congestions may occur due to transmitting collected data without preprocessing, especially in high vehicular density situations, 2) limited interaction with car users that uses car devices to provide notifications only, 3) they do not provide a clear integration between communication and intelligence.

In [8], a layered IoV protocol stack and architecture were introduced. The architecture consists of five layers. The first layer is the perception layer which is represented by the different types of personal devices, RSUs, actuators, sensors, and vehicles. The second layer is the coordination layer which provides a virtual network that involves heterogeneous network technologies such as 4G/LTE, Wi-Fi, WAVE, and satellite networks. The artificial intelligence is the third layer which represents the virtual cloud infrastructure, where storage, processing and analysis the received information is carried out. The fourth layer is the application layer which involves the smart ITS applications. The last layer is the business layer and it represents the operational management module of IoV. In addition, the author designed a protocol stack to organize the existing protocols based on to the proposed five layers architecture. The designed protocol stack has three planes including management, operation and security.

In [2], a seven-layered model architecture for IoV is introduced. The seven-layers are: 1) user interface, 2) data acquisition, 3) filtering and preprocessing, 4) communication, 5) control and management, 6) processing, and 7) security. A user-vehicle interface is supported by the seven-layer architecture to manage the interactions between the driver and the vehicle. Also a communication interface was introduced for optimal transmission network selection.

By analysing the existing IoV architectures, it is clear that there was no consideration for the real-time Big data processing requirements. In addition, all the previous architectures assume that all the collected information must be sent to the data centres (i.e. cloud computing centres) for processing and analysis. Therefore, such architectures are not suitable for many of the ITS applications that require real-time big data analytics. In particular, high latency and communication network overloading are the expected results of deploying any of the previous architectures.

III. OVERVIEW OF INTELLIGENT TRANSPORTATION SYSTEM BIG DATA ANALYTICS

The evolution of sensing and communications technologies and the advances in intelligent data processing are the driving forces for realizing the intelligent transportation systems concept, which is a main component of smart cities [32]. Similar to many modern life aspects, transportation management and control is now becoming more data-driven [33]. The applications of ITSs are data-intensive, complex, and the "5Vs of Big Data" can describe their characteristics precisely [34]:

- The first "V" is the volume of ITS data, which has exponential growth. For example, in 2013 each automotive manufacturer collected around 480 TB of data and an increment to reach 11.1 PB/year is expected by 2020 [34].
- The second "V" of ITS data is the variety. This characteristic describes the various ways of collecting data in different formats such as numerical data gathering through sensors on both infrastructure and vehicles, multimedia and text data capturing from social media, and GIS and image data loading for digital maps. The organization level of such data varies from

semi-structured to structured data. The variety of this data creates highly heterogeneous data sets that impose serious challenges in the ingestion, integration and processing stages of a data analytic system.

- The third "V" is the velocity of ITS data, which varies widely. Data generation and collection rate can be continuous real-time collection and in certain applications data are collected at regular intervals. Similarly, the requirements of processing vary greatly from real-time event processing to batch processing. However, real-time data collection and processing induce high requirements on networks and data processing centres.
- The fourth "V" is for veracity which describes the ITS data trustworthiness level. In fact, the ITS community is facing significant challenges in providing timely and reliable transportation related data collection.
- The fifth "V" is for the ITS data value, which depends on the data age, their sampling frequency, and their usage purpose. For instance, for a collision avoidance application, few minutes old data may have no value. On the other hand, route planning applications can benefit from non real-time data. The value is a characteristic to measure the ability to extract from data meaningful and actionable business insights [35].

One of the distinguished characteristic of ITS big data is the geographically distributed data sources and consumers [6]. In fact, the relevant ITS data have many sources, which are classified into four main classes [34]: (1) data obtained from roadways, (2) data obtained through vehicles, (3) data obtained from travellers, and (4) data obtained from wide geographical area. Roadway data can be obtained through sensors, loop detectors, microwave radar, infra-red sensors, ultrasonic sensors, and CCTV camera. An increasing number of vehicles equipped with different types of sensors. Such sensors provide a huge amount of data related to vehicles' status, driving pattern, and traffic conditions. In nature, This data is spatio-temporal as it depends upon location and time, and it is collected through electronic tolls tags, global positioning systems (GPS), cellular networks, Wi-Fi access points, and bluetooth radios. While roadway data can cover specific location, vehicle-based data can cover different areas due to vehicles mobility. With the development of V2V and V2I communications [36], [37], vehicle-based data is becoming one of the main sources of ITS data. A rich source of ITS data is the cell phones applications of passengers and drivers. For example, the Waze cell phone application utilizes travellers' location information to collect traffic flow status. Wide area data are collected through technologies such as space-based radar and unmanned aircrafts that provide photogrammetry and video recording.

Big data analytics provide the option to distribute queries processing across multiple datasets using commodity computing units and return timely results [38]. In [39], a survey of Big data analytics methods was carried out, where the methods are categorized to classification, clustering, association rule mining and prediction. Basically, scientists and data miners may employ big data analytics to analyze large amounts of data and elicit knowledgeable information that can be used in predicting or identifying trends, making decisions, and finding hidden information [14], [40].

A. BIG DATA ANALYTICS LIFE-CYCLE

Figure 3 shows the data analytics life-cycle to handle the ITS big data. Obviously, the intelligent transportation big data analytics system consists of five main components: (1) data producers, (2) data consumers, (3) data storage system, (4) Intelligent data processing and analysis system, and (5) network and communication components to connect the previous components through reliable communications. Thus, data produced by billions of sources are collected and delivered to storage and processing centres through the provided communication medium [34]. Afterwards, intelligent analysis is carried out and then the consumers can use the outcome to make wiser and more accurate decisions. The life-cycle completion necessitates efficient coordination of activities among the producers, consumers and network devices such as relays/routers and the cloud centre hardware/software resources [41]. To ensure efficient operation of wireless IoV networks, it is significant to have effective and reliable wireless communication among these entities.

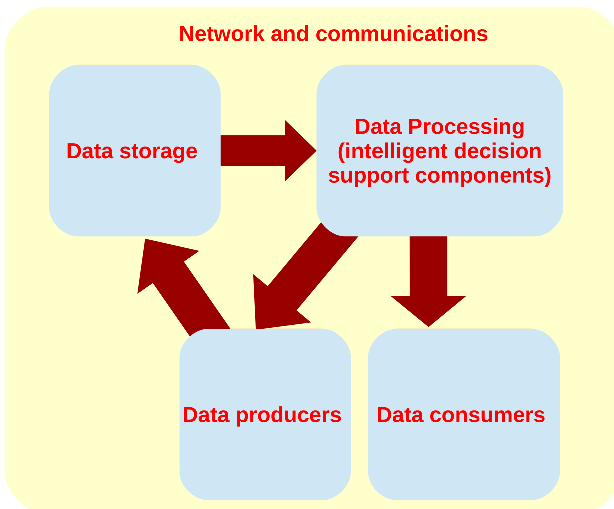

FIGURE 3. ITS big data analytics life-cycle.

B. EMPLOYING ITS BIG DATA ANALYTICS IN IoV ENVIRONMENT: OPPORTUNITIES AND CHALLENGES

Extracting meaningful information from the oceans of ITS big data through the application of big data analytics has many benefits and challenges [32]. Big Data in transportation arena is often collected continuously from different sources over vast geographic scale. Although this data are huge in size, rich in information, and highly disorganized, they can considerably enhance ITS understanding and evaluation [42]. Big data analysis can provide descriptive, diagnostic, predictive, and perspective analytics for ITS applications. In addition, the interactive data visualization provided by the Big Data analysis tools may help in describing the characteristics of transportation data [34]. On the other hand, in ITSs environment, the large amount of shared information about traffic contributes towards the optimization of smart cities management. Moreover, proactive traffic management for better system performance is made possible due to the real-time nature of the Big Data in transportation [42]. In particular, in comparison to high cost traditional infrastructure-based roads development, deploying ITS applications has higher returns on investment [43].

ITS Big Data supports many ITS applications to improve roads safety, mitigate traffic congestions, optimize energy usage, alleviate negative environmental impacts, and improve surface transportation productivity [34]. For example, transportations systems can be improved by employing traffic data to clear traffic congestions and provide alternative routes. In addition, accidents can be reduced by analysing drivers behaviour and accidents factors. While the collected big data from ITSs can assist in consolidating shipments and optimizing the shipping routes to reduce the wastage of supply chain [32]. In addition, ITS big data analytics can help in enhancing collaborations and communications among the smart city's various objects [44]. However, the mining of ITS Big Data gold cannot be achieved without the support and participation of government and private sectors. In addition, users, especially drivers, should be given motivations and incentives to be fully involved in the ITS Big Data system as they are the main owners and consumers of data.

As IoV integrates vehicle-mounted mobile Internet, vehicular ad hoc networks, and in-vehicle networks, it is considered the core of ITSs and the main source and medium to collect the ITS big data [26]. In comparison to conventional systems big data, the ITS big data collected through IoV have different characteristics. This is due to the various objects (e.g. vehicles and sensors) that participate in data collection, which produce noisy data and lead to data heterogeneity with fast growth [39]. Therefore, handling ITS big data in IoV environment faces many challenges and critical issues due to the high QoS requirements of many ITS applications and the IoV unique characteristics which are: 1) the highly dynamic topology, 2) the large scale and distributed network, and 3) heterogeneity of resources, consumers and communications.

In fact, one of the main objectives of introducing the ITS was to increase road safety. Therefore, to efficiently utilize big data analytics in improving ITS applications, the real-time collection, processing and analysis of data is crucial [32]. For instance, accident avoidance applications require timely warnings messages that are sent to drivers based on the analysis of collected data. Any delays in collecting, processing or analysing the related data or delivering such messages may result in a catastrophe [42]. Moreover, autonomous vehicles need a special reliable and high-speed transmission due to mobility in real time environment [17]. Thus, to harness the power of ITS big data, it is vital to take full advantage of the on-the-fly and real-time processing of data, whereas traditional store-then-process approaches may no longer be appropriate for many ITS applications [32], [42].

In an ITS, the data storage system is divided into two parts distributed data storage and cloud based storage [17]. In particular, every car has its own mobile database of information (distributed) [45]. However, to apply the centralized big data mining and analysis techniques, the mobile databases and the other ITS devices' data need to be collected at a central server where batch processing can be applied. The main disadvantage of centralized processing is the high latency in obtaining the data analytics outcome. Rather than the high volume of the ITS big data, the data sources are geographically distributed and transferring this massive amount of data from the nodes at the network edge to the cloud leads to consuming the available communication bandwidth [6]. As a matter of fact, the centralized approach is not suitable for many emerging ITS big data applications as they require low latency (i.e. real-time or near real-time responses), location awareness, and mobility support [6]. In such scenario, the complexity of big data processing is more related to the required computational costs and time.

Big data can be processed using distributed and parallel processing techniques. However, utilizing these techniques in IoV environment is not straightforward. In addition, such distributed techniques need to be applied at the network edge, which produces analytics based on the local view and not the global view of the network. Though there are several advances in edge and cloud computing to address some data analytics issues, they have their own pros and cons. The merging of these two computing paradigms, i.e., edge computing based real-time data processing and cloud based massive resources of computing and storage, may enable effective real-time data analytics in IoV environment [6]. Therefore, providing a hybrid big data management and processing system that merges the distributed and central approaches is more efficient for processing an ITS big data. In summary, the serious challenge is how to address the real-time collection, processing and analysis of the big data sets, generated from heterogeneous devices in several formats, in order to serve end-users with real-time information and feedbacks.

IV. LAMBDA ARCHITECTURE FOR REAL-TIME BIG DATA PROCESSING

A generic real-time big data processing architecture called lambda architecture was introduced in [46]. The premise behind this architecture is that ad-hoc queries can be employed against all the data to get results, however, such queries are very expensive in terms of resources. To solve this problem the results are pre-computed as a group of views, and then the query is done on the views. The abstract description of lambda architecture is shown in Figure 4. The architecture have three layers. The batch layer is the first layer, where views are computed on the collected data, and the computation is repeated when it is necessary. The second layer is a speed layer for parallel processing, where the newly

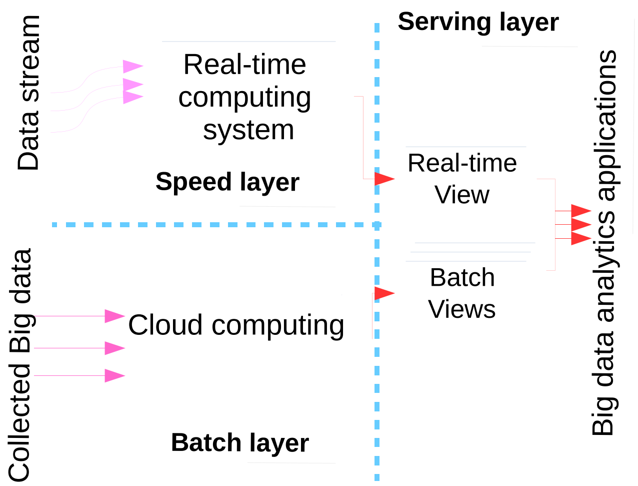

FIGURE 4. The generic lambda architecture for real-time big data processing.

received data is continuously processed in near real-time manner. The serving layer answers the any queries against the data. Therefore, both the batch and the speed layers' views are used by the serving layer to answer the queries. The batch and speed layers both are forgiving and can recover from errors by being recomputed, rolled back (batch layer), or simply flushed (speed layer) [47].

The lambda architecture is independent from the different technologies that can be used to implement the three layers. Actually, there are many technologies that are useful in creating a processing system for real-time big data. To compensate for the batch layer high latency updates, the speed layer processes only the new data. This layer can exploit stream processing systems, such as Spark [48], S4 [49], and Storm [50]. The batch layer is only updated through batch processing, and it requires horizontal scalability with a support for random reads. The technologies such as Hadoop [51] with Cascading [52], Hive [53], Pig [54], Python streaming [55], and Scalding [56], match the batch layer requirements. A system that is capable of performing fast random read/write processes is required for the serving layer. The serving layer data is replaced by the views computed by the batch layer and the processed real-time data from speed layer. Therefore, the data size is considered small in this layer. Since replacing the views is done in one operation, no complex locking mechanisms are required for the serving layer system. An in-memory data store such as Memchached or Redis, is sufficient for the serving layer. If it is required to store large volume of data in the serving stores, and to guarantee low latency with high availability, then systems as DynamoDB, MongoDB, ElephantDB, Cassandra, or HBase can be a good choice [47].

Implementing the Lambda architecture for real-time big data analytics in IoV environment faces many challenges. For example, implementing this architecture in the cloud center is not going to serve the real-time applications due to the high latency of cloud-based processing. Thus, the implementation need to be at the network edge and it should be distributed to match the geodistributed nature of ITS big data. However, several issues still need to be considered such as where to place each layer (batch, speed, and serving layers) and how to mange the network communication between the three layers.

V. INTELLIGENT COMPUTING PLATFORMS FOR ITS BIG DATA ANALYTICS

This section compares the definitions, functionalities, advantages, and disadvantages of cloud and edge computing. Moreover, it highlights the important role that fog computing can play in real-time big data analytics.

A. CLOUD COMPUTING

As one of the most significant improvements in modern data storage and computation technologies, cloud computing provides a powerful platform to perform complex and large-scale computing. The definition of cloud computing is "a model for allowing ubiquitous, convenient, and on-demand network access to a number of configured computing resources (e.g., networks, server, storage, application, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [57]. Cloud computing main advantages are security, parallel processing, virtualized resources, and the integration of data services with data storage scalability. Thus, not only cloud computing reduces the cost and constraints for computerization and automation by enterprises and individuals but also cuts the cost of infrastructure maintenance and provides ubiquitous users access with efficient management [58]. Cloud computing with its rich set of storage, infrastructures, and computations services provides a very interesting environment for scientific experiments, data analysis and research [59].

The need to store, process, and analyse large datasets has led to the adoption of cloud computing by many organizations and individuals. To support the users needs and demands in the big data era, the cloud services providers are doing efforts to integrate in their services parallel data processing frameworks. Big data and cloud computing are highly correlated fields. Basically, cloud computing provides the required processing infrastructure through the use of Hadoop, which is a class of distributed data-processing platforms. In particular, the cloud technology provides a distributed fault-tolerant large database for data storage, where data can be processed with parallel distributed algorithms [38]. Therefore, the big data systems evolution is prompt by the cloud-based distributed processing and storage technologies utilization.

At cloud centres, performing big data analytics and exploiting the data-sets correlation result in excellent integration of collected data. In particular, this gives the ability to analyse huge amount of data while considering its growing complexity and scale, and can provide for the whole network a comprehensive viewpoint. In addition, the cloud-based approach will result in improved precision and reduced error rate, and more dynamic data treatment as compared to the traditional data analytics approach [60]. Nevertheless, processing IoV generated data at cloud centres faces several challenges due to the unique characteristics of IoV data that are discussed in the following subsection.

B. IS CLOUD COMPUTING SUFFICIENT FOR ITS BIG DATA ANALYTICS?

It is predicted that IoV components will produce big data in high speeds and video data such as surveillance video are inferred as the 'largest' big data, which can easily make the IoV data grow to TB/PB level in seconds [61]. In addition, many IoV applications are highly sensitive to latency and require fast big data processing with fast feedbacks. For instance, in popular projects on self-driving cars, e.g. Googles cars, the data generated from complicated surroundings via various sensors and cameras is massive (approximately over 1Gb/s). The processors must give accurate orders to the steering system in milliseconds by computing these data. Until the technology is mature, an onboard computer alone is insufficient and cloud participation becomes indispensable [62]. However, if the data is delayed in a cloud server due to queueing or networking failure, a "smart" car may lose its intelligence and cause accidents.

Basically, cloud based big data system architecture cannot guarantee the latency sensitive requirements for many ITS applications for the following reasons:

1) The ITS data produced and collected by IoV components are distributed and heterogeneous in structure [63]. Thus transferring this data to the cloud consumes the network precious resources, which should be saved to serve the real-time ITS applications.
2) The networks near the cloud data center may suffer traffic overload and bottlenecks due to data floods. Consequently, network congestions and failures occurrence results in critical delays [62].
3) The IoV wireless environment dynamic nature created the need to deal with large volumes of high-speed and real-time continuous streams of data [63]. However, transferring such data to the cloud centres for processing results in high latency, which contradicts the requirements and constraints of real-time and safety-related ITS applications.
4) Many types of ITS big data is described as spatio-temporal and various ITS applications have location awareness requirements. Therefore, to save the data characteristics and to fulfil the requirements of location awareness, such data should be locally processed instead of sending it to the cloud centres.
5) The emerging ITS applications require the support for ubiquitous coverage and seamless mobility that cannot be offered adequately by cloud based platforms [6].
6) Transferring the collected data through multi-layer network hierarchy and multi-hops increases the risk of data loss and alteration. In addition, it raise up many security and privacy issues.

As the era of latency sensitive big data is coming, sending all the data to the cloud has many disadvantages. Recent research efforts are investigating how to better exploit capabilities at the edge of the network to support the needs of data-intensive applications. The concept of edge computing is very attractive to solve some of the cloud computing critical problems [64]–[66]. The main objective of edge computing is extending the cloud computing functionality to the edge of networks.

In edge computing, instead of transmitting the large amount of data generated by different kinds of IoV devices to the cloud center such data can be processed at the network edge, which will save network bandwidth and reduce communication delays [65]. In the deployed cloud-based networks, edge computing can be enabled by creating an intermediate layer, which may include many distributed edge servers in several areas such as bus stations and parking areas. The edge server is considered as a cloud server with low-capacity, and has data storage, computing, and communication capabilities [6]. Edge computing can support the services/applications that require low-latency processing, location-awareness, fast mobility management and specific QoS. This is due to the geographically distributed nature of edge computing and its proximity to users [65]. In fact, the geographically distributed nature of edge computing assists in providing valuable contextual information such as the end-user status, local networks conditions, and events status. Subsequently, such information can be utilized in edge applications context-aware optimizations [6].

Although edge computing inclusion in the centralized cloud computing brought up new opportunities, balancing the trade-off between distributed and centralized network architectures requires intelligent decisions. Basically, the network edge component cannot handle all the IoV generated data and still there is a need to offload data to the cloud centres. In addition, the network edge resources such as computing power, bandwidth and communication capabilities need to be efficiently used to serve the real-time applications. Large-scale and latency tolerant tasks can be efficiently processed at cloud centres while the processing of delay-sensitive tasks is required to be at the network edge [6]. The previous studies introduced three different edge computing concepts namely, fog Computing, Mobile Edge Computing, and cloudlet. The following sections provide a brief discussion of these edge computing technologies concepts and Table 1 presents a comprehensive comparison.

C. CLOUDLET

To support low-latency requirements for resource-intensive applications, a new architectural element called cloudlets, has been proposed by [67]. A cloudlet is a mobility-enhanced small-scale cloud data center that is located at the edge of the Internet. It represents a trusted, resource-rich computer or cluster of computers that is well-connected to the Internet and available for use by nearby mobile devices [68]. The main goal of the cloudlet is to support resource-intensive and interactive mobile applications by providing powerful computing resources to mobile devices with lower

TABLE 1. Edge computing platforms comparison.

Comparison parameter	Fog computing	Mobile edge computing	Cloudlet computing
Node devices	Gateways, Access Points, Switches, Routers, Vehicles, ITS smart devices, personal devices	Servers installed in base stations	Data Center in a box
Node location	Ranging from End Devices to cloud	Macro Base Station/Radio Network Controller	Outdoor/Local installation
Architecture	One or more layers	One layer	One layer
Software Architecture	Fog Abstraction Layer based	Mobile Orchestrator based	cloudlet Agent based
Flexibility	High	Low	Low
Computational capabilities	multiple levels	High	High
Context awareness	Medium	High	Low
Proximity	One or Multiple Hops	One Hop	One Hop
Access Mechanisms	Mobile Networks, Wi-Fi, Bluetooth, IEEE 802.11p (DSRC)	Mobile Networks	Wi-Fi
Supports non-IP based communications	Yes	No	No
Internode Communication	Supported	Partial	Partial
Latency	Low	Medium	Medium
Fault tolerance	High	Low	Low
Cost	Low (uses legacy or commodity devices)	High (requires special devices)	High (requires special devices)
Deployment	Possibility of ad hoc deployment with no or minimal planning	Planned deployment	Planned deployment
Mobility support	High	Medium	Medium

latency [18]. Through the provisioning of real-time data analysis at the edge of Internet, cloudlets can reduce the amount of data traffic migrating to the cloud [69].

Compared to cloud data centres, a cloudlet needs to be much more agile as users mobility requires highly dynamic association. In addition, to support user mobility, virtual machine (VM) handoff technology needs to be used to seamlessly migrate the offloaded services on the first cloudlet to the second cloudlet as a user moves away from the currently associated cloudlet. As cloudlets are geographically distributed small data centres, a user first has to discover, select, and associate with the appropriate cloudlet among multiple candidates [18]. Due to the fast mobility of transportation entities, utilizing cloudlets in the IoV environments faces many challenges. Moreover, providing cloudlets services over a large geographical area requires high deployment and maintenance costs and efforts. In addition, the demand for cloudlet services might vary based on vehicular traffic variations, which leads to under utilized cloudlet facilities in certain areas and times.

D. MOBILE EDGE COMPUTING

The growing popularity of wireless networks and mobile devices has taken cloud computing to new directions due to the limited battery lifetime, processing capability, and storage capacity of such devices [70]. Mobile edge computing (MEC) was defined by ETSI as a new technology that ''provides an IT service environment and cloud-computing capabilities at the edge of the mobile network, within the Radio Access Network (RAN) and in close proximity to mobile subscribers'' [71]. Compared to cloud computing, MEC is characterized by advantages of low latency, proximity, high bandwidth, and real-time insight into radio network information and location awareness. Therefore, MEC is deemed to be an enabling technology for a large number of new applications and services for multiple sectors, such as consumer, enterprise, health, etc. In particular, MEC is considered a promising solution for handling video streaming applications [18]. However, MEC was introduced to work within the radio access networks only.

E. FOG COMPUTING

Another technology of edge computing is known as fog computing, which was initiated by Cisco in 2012 [72]. Fog computing is defined as ''a system-level horizontal architecture that distributes resources and services of computing, storage, control and networking anywhere along the continuum from cloud to Things'' [73].

Fog nodes are heterogeneous in nature. They may consist of set-top boxes, access points, edge routers, high-end servers and even end devices such as mobile phones, vehicles etc. The different hardware platforms have diverse levels of secondary storage, RAM, and real estates that can be supporting to the new functionalities. Different types of software applications and OSes run on the platforms, which provide high diversity in software and hardware capabilities. The communications infrastructure in the fog layers is normally heterogeneous, ranging from multiple wireless networks technologies (e.g. WiFi,LTE, 3G/4G) at the edge to high-speed links to connect the core and the enterprise data centres [74].

Fog computing is different from other edge computing technologies as it provides tools for distributing, orchestrating, managing and securing resources and services across networks and between devices that reside at the edge. Unlike cloud computing, which is characterized with a centralized deployment and management to its physical resources that are mostly homogeneous, fog extends and complements the cloud through distributing heterogeneous resources at the network edge and managing such resources in a decentralized way [74]. By employing fog computing, not only latency-sensitive application can be supported at the network edge, but also latency-tolerant tasks can be performed efficiently at the intermediate network nodes as they have more powerful computing capabilities. Cloud computing data centres can still be used for deep analytics at the top of fog layers. In addition, mobility and location awareness are supported by fog computing even though they are lacking in current cloud computing commercial models.

1) MOTIVATIONS TO UTILIZE FOG COMPUTING IN IoV FOR REAL-TIME BIG DATA ANALYTICS

In highly distributed environments, acquiring, integrating, storing, processing and utilizing big data have created serious challenges for researchers, data scientists and engineers. Therefore, recently CISCO has introduced the fog computing concept with a goal of supporting cloud computing platforms by locally processing part of the computational tasks at the edge devices such as IP-enabled video cameras, routers, and switches, which reduces the transmitted workload to the cloud. More specifically, fog computing can support the cloud computing platforms in handling the following applications types, which the cloud computing concept cannot fulfil their requirements [74], [75]:

- Applications that require predictable or very low latency such as traffic safety applications.
- Real-time fast mobility-based applications (e.g. smart connected vehicles).
- Applications that are geographically distributed such as environmental monitoring and traffic management.
- Distributed large-scale control systems (e.g. smart energy distribution, smart grid, and smart traffic lights).

In the IoV environment fog computing can play an essential role in handling the ITS big data. Due to the heterogeneity of storage, computing and communication resources of fog technology, it is considered as a perfect match for the heterogeneous and dynamic environment of IoV. In comparison with cloud computing, fog computing has many advantages which are summarized in the following points:

- The transmission of huge amount of data to the cloud adds a enormous burdens on the wireless networks communication bandwidth, which results in high response delays and degraded services [75]. Fog computing has a decentralized architecture that brings computing resources and application services to the edge where the data is being generated and consumed [11]. Instead of direct raw data transmission to the cloud, only high-level data representations can be uploaded while utilizing fog nodes to perform associated processing tasks, which highly reduces the transmitted data size [15].
- At the network edge, the big data analysis workload is parallelized on large number of computing units. Each unit must perform only light-weight computing tasks, however, their massive aggregated computational capabilities provide computing power of high-performance [15]. In addition, the throughput and load among all computing nodes can be easily balanced to avoid potential computing bottlenecks.
- Fog computing provides the essential requirements of mobility support, which can not be satisfied by the cloud computing platforms due to their centralized nature of computing and storage functions [75].
- The distributed nodes for fog storage and computing are perfect for supporting the large number of sensors distributed in the IoV environment. If only cloud computing is employed to handle this task, high power consumption is wasted in the process of transmitting such data to the data centres [15].
- By "moving the processing to the data", fog computing distributes the intelligence computing at the network edge near the data consumers, which suits the natural characteristic of geo-distribution of IoV generated big data [15]. Thus, data can be processed at the edge, which provides quick feedback for the data consumers and fulfil the QoS requirements of ITS applications.
- Fog computing is well positioned for real time Big Data analytics. It supports real-time interactions and delay sensitive applications [15], [76]. For instance, largely-deployed camera sensors in cities are important to support traffic management, surveillance etc. With fog computing sufficient resources of storage and computation can be provided to save and process video streams for tasks such as object tracking, object recognition, and data mining. Thus, through exploiting fog services, real-time processing and feedbacks of large volume video streaming can be achieved without wasting the network bandwidth [77].
- Fog supports densely distributed data-collection points and with the utilization of multiple fog computing

layers, the processed data scope narrows in time and space at the edge, and widens at the top [74]. As a result, different levels of response speed, coverage area, and data scope can be provided based on applications requirements.

- With fog computing, flexible storage, computing and communication resources, and services can be shared in network, which reduces the need of special servers and data centres deployment [76]. Consequently, high reductions in deployment, operation, management and maintenance cost can be achieved, which contributes towards the fast penetration and implementation of ITS big data analytics applications.
- In [78], a quantitative analysis of energy consumption in a scenario where 25% of the IoT applications demand real-time and low-latency services is presented, and it is shown that the mean energy expenditure in fog computing is 40.48% less than the conventional cloud computing model. In fact, fog computing is considered as an eco-friendly computing platform that can support green technology [79], [80].
- At the fog side, privacy-preserving mechanisms can be employed to address the concerns of personal privacy leakages in undesired ways [77].

The computing and communication workload varies over time and between places. This poses challenges to capacity planning of fog nodes. Without enough capacity at the edge, users service demands cannot be satisfied during rush hours. On the other hand, more capacity requires more investment and then most of the capacity will be wasted for the rest of the day [11]. To solve this issue, vehicular fog computing (VFC) concept is proposed, which is explained in the following subsection.

2) VEHICULAR FOG COMPUTING CONCEPT

Fog computing, which focuses on moving computing resources to the edge of networks, complements cloud computing by solving the latency constraints and reducing ingress traffic to the cloud [11]. However, in the IoV environment vehicles are considered as huge underutilized resources. Vehicular fog computing (VFC), which employs vehicles as part of the fog computing resources, makes the best exploitation of the vehicular computational and communications resources. More precisely, VFC is an architecture that is based on collaboration among near-user edge devices and end-user clients to perform an essential amount of computations and communications. VFC is distinguished from other computing technologies for its high mobility support, dense geographical distribution, and high proximity to end-users. More precisely, fog computing architectures consider smart devices (i.e. ITS smart devices and personal devices) and vehicles as end-users, and in between the cloud and these end-users the fog layer exists. On the other hand, the VFC architecture consider the smart devices and vehicles as part of the "fog" [81].

Considering vehicles as a fog infrastructure is a practical idea from a technical perspective. For example, Malandrino *et al.* [82] viewed the parking vehicles as infrastructure network static nodes. To serve as service infrastructure and static backbone, parked vehicles are providing a very good option. This is due to their characteristics such as long-time static location and large numbers. Hou *et al.* [81], considered that vehicles in traffic congestion can form a cluster of vehicles that serve the VFC layer, where V2V communications are utilized to connect such vehicles. The deployment of fog nodes on certain connected vehicles like buses and taxis, and to move these fog nodes along with the traffic was proposed by [11]. Thus, VFC can harness a huge computational power through the resource aggregation of individual vehicles and other ITS smart devices. With the evolution of vehicular communications technologies and equipments, clusters of moving vehicles can act as fog infrastructure and they become perfect candidates for providing "on wheels data centres". As a result, both computational and communications capacity can be highly improved in urban environment when vehicles share tasks processing and contribute to the fog computing system.

As an extension of fog concept, the VFC and fog computing have many similar characteristics, such as low-latency communications and the wide geographical distribution. However, VFC concept highlights the new features of exploiting the resources of the collaborating near-located vehicles. Actually, many of the ITS applications are distributed and location-based. Thus, they do not require the collection of information at remote servers from a wide geographical scope, this significantly reduces deployment cost and time delay [81]. However, the main challenge in VFC is the ad hoc nature of fog nodes (vehicles), which makes the VFC layer highly dynamic and requires efficient management of collaboration, computation, storage, and communication among vehicles and moving devices.

F. COMPARISON OF EDGE COMPUTING PLATFORMS

This section introduces a comprehensive comparison of the discussed edge computing platforms (i.e. fog computing, Mobile edge computing and cloudlet). Table 1 summarizes the comparison that is constructed based on the previous discussion and the information presented in [6], [62], and [83]. Obviously, fog computing provides high flexibility in terms of architecture, resources, computational capabilities, communication technologies and deployment. Moreover, fog computing has low latency and high mobility support which make it ideal choice for ITS big data analytics and ITS applications in general. In addition, fog computing has high fault tolerance for two main reasons. First, it does not depend only on fixed deployment and it can allocate resources in ad hoc manner. Second, fog computing can be in multiple layer architecture, which allows the deployment of higher specification servers at higher layers.

VI. FOG COMPUTING BASED SYSTEM ARCHITECTURES

In fog computing, the main component is the fog nod which can be a facility or infrastructure that provides resources for

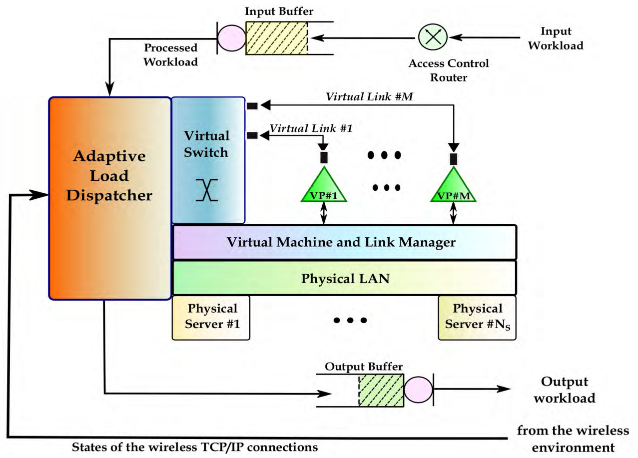

FIGURE 5. Virtualized fog node [84].

network edge services. A fog node might be a resource-poor device such as an end device, a base station, a switch, a router, and an access point, or machines with rich resources such as IOx, which is a fog device produced by Cisco [77]. Figure 5 describes a virtualized fog node essential blocks, which includes [84]:

- The input and output buffers: received data is queued in input buffer, whereas processed data is pushed to the output buffer.
- The physical resources: they comprise storage, computational, control, and communication resources such as routers, servers, and physical channels that are associated to the fog node.
- The bank of Virtual Processors (VPs): where the assigned workload is processed on behalf of the served devices.
- The virtualization layer: manages the available physical resources among the running VPs in real-time.
- The Virtual Switch: sustains an end-to-end based inter-VP TCP/IP transport connections.
- The Adaptive Load Dispatcher: over the available group of VPs it dispatches the input workload in a balanced way, in order to meet the QoS requirements of serviced devices and minimize the consumed energy.

In general, a fog node has several functions, including networking, computing, storing and control [18]. Fog nodes can communicate with each other through wireless or wired transmit. Fog nodes can form a fog cluster to provide load balancing, resilience, fault tolerance, data sharing, and minimization of cloud communication.

A general hierarchical system architecture based on fog computing is shown in Figure 6. There are often three tiers in a fog computing system, but more tiers can be allowed for special application scenarios. At the network edge, fog nodes are typically focused on sensor data acquisition/collection, data normalization, and command/control of sensors and actuators. This layer is designed for Machine-to-Machine (M2M) interactions through Internet interconnections, which

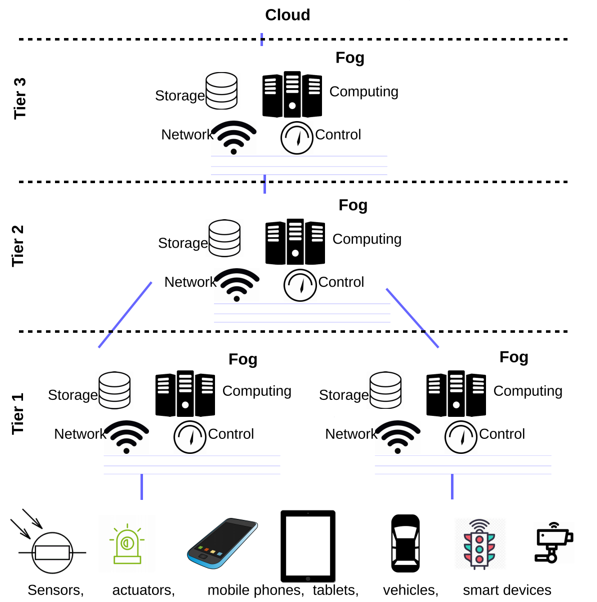

FIGURE 6. General hierarchical structure of fog based system.

gives the ability to react and make decisions in real-time. At second tier, fog nodes are focused on data filtering, compression, and transformation. At the higher tiers or near to the back-end cloud, fog nodes are focused on aggregating data to create knowledge out of data. Basically, the second and third tiers deal with visualization and reporting (e.g. Human to machine interactions), as well as systems and processes. The time scales of these interactions over the fog platform range from milliseconds to minutes (real-time analytics), up to days (transactional analytics). Therefore, fog must support multiple types of storages, from transient at the lowest tier to semi-permanent at the highest tier [76]. Architecturally, fog nodes at the edge may have less processing, communications, and storage capabilities than nodes at high levels. With the increase in the number of tiers, each tier would be sifting and extracting meaningful data to create more intelligence [18].

A. FOG COMPUTING BASED SYSTEM ARCHITECTURES IN EXITING STUDIES

This section discusses some existing fog computing based system architectures that are proposed to fulfil the requirements of certain environments such as IoT, smart cities, and IoV.

Through two use cases, [74] identified some high-level architectural requirements such as the mobility management of end users, heterogeneity of fog resources, multi-tenancy support and the geographical distributional of data and processing. The highlighted requirements are essential for utilizing fog computing in the IoT environment. The author introduced the fog software architecture which presents the required technology components. fog data mining is an important strategy for IoT in order to reduce the cloud storage requirement, the energy consumption, and package transformation across the wireless network. For efficient knowledge extraction, homoscedasticity and statistical features extraction technique was introduced in [85]. The aim of such technique is to extract the important events in sensor data in real time when used with neural classifiers, which reduces the amount of data transfer and helps in simplifying the process of knowledge management. Nevertheless, this study focused on utilizing fog computing for data analytics without discussing the requirements and consequences of introducing fog computing in such system. In [62], a framework for resource allocation of fog computing in IoT was introduced, which considers latency reduction combined with reliability, fault tolerance and privacy. The author formulated the resource optimization problem and used a genetic algorithm to solve this problem. The framework was based on four layer architecture with big data awareness. The four layers include application layer, data processing layer, network layer and perception layer. The author introduced a simple architecture for the fog node, which was elaborated in two parts: computing and networking. For the computing part, four layers are considered namely hardware platform, software and virtualization platform, functional components, and fog computing applications interface. The networking side was proposed in three layers of wireless technology, single hop/ ad hoc communications, and the software defined network concept. On the other hand, a three tier Internet-of-Everything-Fog-Cloud ecosystem was introduced in [86]. Basically, the first tier, the Internet-of-Everything (IoE) tier, encompasses a model for peer-to-peer communication between the proximate devices that offer their storage and computing capabilities. The fog layer is on top of IoE layer and it is the middle layer between IoE layer and cloud layer. The fog layer communicates with both layers and it facilitates two communication models with the cloud layer namely offloading and aggregation. In [39] an architecture, which integrates IoT and big data analytics was proposed. It has of four layers including IoT devices, network devices, cloud computing, and big data analytics layers. In this architecture, the IoT devices layer consists of all the objects (e.g. sensors) that use wireless networks. The network devices layer manages the wireless network communication that can be Bluetooth, ZigBee, ultra-wideband, WiFi, and RFID. The author introduced an IoT gateway which manges the communication between IoT devices and the cloud computing layer. However, there was no clear description for the fog computing role in this architecture. For wireless heterogeneous IoT networks, a general system model for collaborative fog-cloud processing was proposed in [6]. The presented model exploits the historical information and network-wide knowledge available at the cloud center, which assists fog computing units to satisfy varied requirements of live data analytics in IoT wireless heterogeneous networks. In this model, an IoT network edge gateway is supported with cache memory to perform fog-caching, which aims at locally delivering the contents that has high demands. The author considered a fog computing node as any device that has the capabilities of network connectivity, storage, and computing such as video surveillance cameras, switches, and routers.

For smart city environment, [15], [87] introduced a hierarchical distributed fog Computing architecture for big data analysis in smart cities. Based on the use case of smart pipelines monitoring, a fog-based four-layer computing prototype was implemented as a demonstration of the feasibility and effectiveness of the future city-wide systems implementations. The fog computing architecture consists of four layers. As a sensing network Layer 4 consists of an enormous number of sensors. On top of Layer 4, many high performance and low-power computing units form Layer 3. Then, the intermediate computing devices form Layer 2, and the cloud computing centre is the first layer.

In the context of IoV, a fog computing architecture was presented in [88], where IoV knowledge is semantically represented, published and subscribed. The author focused on how to share information between fog nodes consistently and explicitly. The interaction between the fog nodes is based on a topic-based publish/subscribe model. A publishing fog node defines the class of an event pattern to which a subscribing fog node can subscribe. This means that even though an event can be propagated to all fog nodes, only the ones that are subscribing the event can accept and receive information. In a different approach, the idea of exploiting vehicles as resources infrastructure for computations and communications was introduced in [81]. The architecture was named vehicular fog computing (VFC), which involves vehicles and near-user edge devices to carry out communication and computation. The VFC framework has a major role in increasing computing speeds and decreasing the delays for the applications that need intensive computations. For example, rescheduling traffic lights is necessary when an accident happens, in order to clear the massive backlogged traffic efficiently. Through employing VFC, a virtualized powerful supercomputer can created using the computational resources pool of vehicles. Xiao and Zhu [11] presented a visionary concept on vehicular fog computing that turns connected vehicles into mobile fog nodes and utilises mobility of vehicles for providing cost-effective and on-demand fog computing for vehicular applications. Basically, the fog nodes are placed on certain connected vehicles (vehicular fog nodes) such as self-driving buses and taxis. These vehicular fog nodes form a hybrid fog computing platform together with the static fog nodes located at the edge of networks. Vehicular fog nodes take care of local processing of aggregated sensor data (e.g. real time video processing), and host vehicular applications that are latency/privacy-critical. In case the vehicular fog nodes are overloaded or if the vehicles are moving out of the V2V communications range, the workload can be offloaded to nearby cellular fog nodes or vehicular fog nodes.

Based on the previous discussion of existing fog computing system architectures, it is obvious that many architectures and framework models exists for utilizing fog computing in

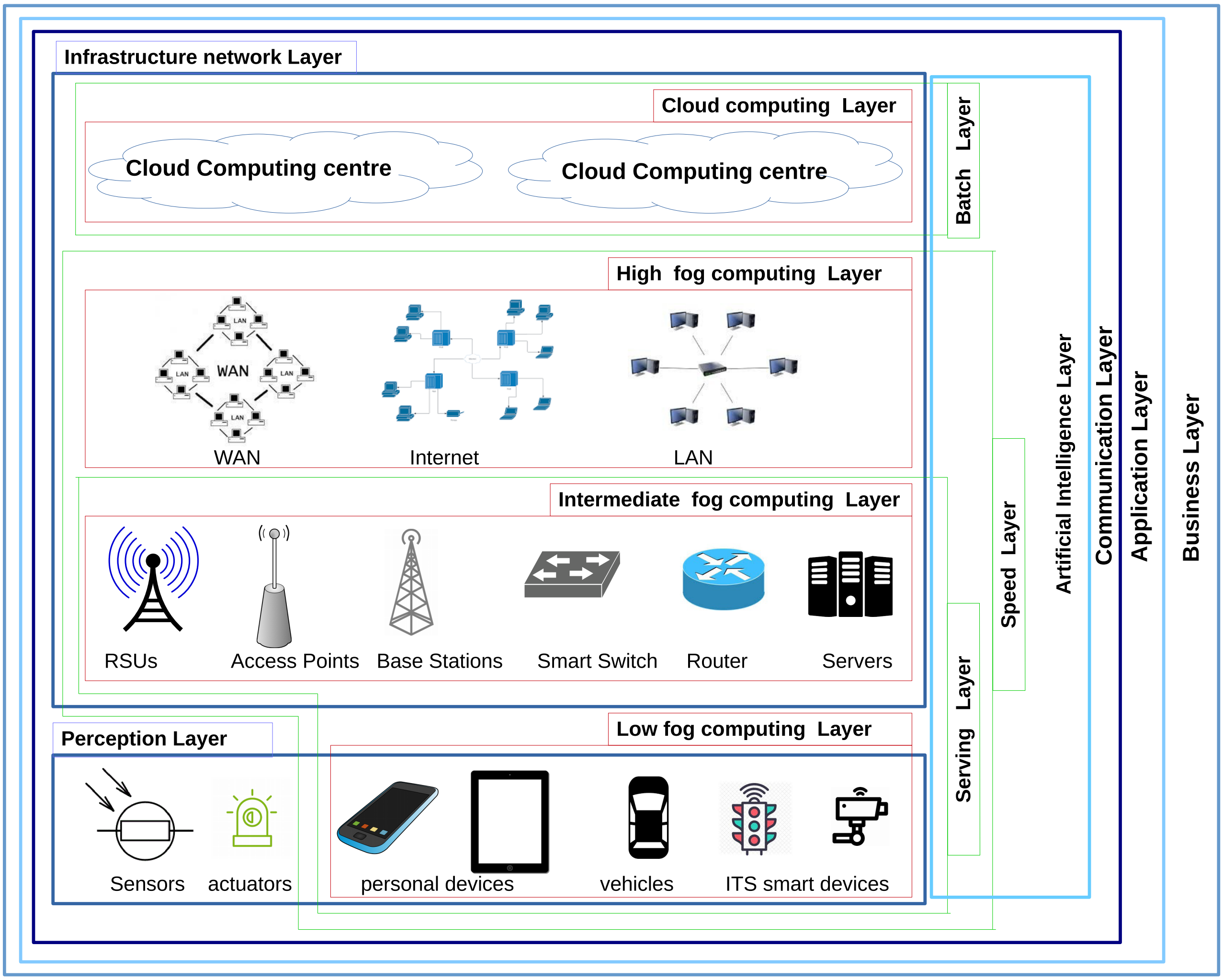

FIGURE 7. Real-time intelligent transportation system big data analytics (RITS- BDA) architecture.

different environments. However, all the available architectures do not focus on the integration of fog technology in the IoV environment for serving the ITS big data analytic, especially the analytics in real-time. Therefore, there is a real need to introduce a detailed architecture design that integrates the three dimensions of IoV, fog computing, and real-time big data analytics.

VII. PROPOSED REAL-TIME INTELLIGENT TRANSPORTATION SYSTEM BIG DATA ANALYTICS (RITS-BDA) ARCHITECTURE

Designing a layered architecture in the IoV domain which comprises heterogeneous devices and networks is a significant task. Optimizing the number of layers, identifying the different functionalities of each layer and defining the relations among layers are serious challenges of the layered architecture design. In addition, various network characteristics such as interoperability and scalability must be considered. The proposed real-time intelligent transportation system big data analytics (RITS-BDA) architecture is to efficiently utilize the fog computing facilities in order to provide real-time ITS big data analytics in the context of IoV. The design of RITS-BDA architecture should consider three different perspectives which are: the real-time big data analytics, the IoV, and the intelligent computing (i.e. fog and cloud computing).

Due to the fact that the research in the field of real-time big data processing in IoV is in initial stage, especially in research projects and industries [8], [89], to the best of our knowledge, this effort is the first towards designing an architecture for real-time ITS big data processing based on fog computing in the IoV environment.

A multi-dimensional layered architecture is designed including the intelligent computing dimension, the real-time big data analytics dimension, and the IoV dimension. The layers, components and organizations of each dimension are described below in details. Figure 7 shows an abstract view of the proposed architecture, where the blue color blocks represent the IoV dimension, the green color blocks represent the real-time big data analytics dimension, and the red color blocks represent the intelligent computing dimension (i.e. cloud and fog computing).

A. RITS-BDA INTELLIGENT COMPUTING DIMENSION

This dimension is responsible for providing the required facilities for the big data processing in the IoV environment. In this dimension cloud and fog computing are the two main components. The intelligent computing dimension is divided into four layers, which are described in the following points:

- **Low fog computing layer:** it consists of electronic devices and machines that can perform some processing for the collected data locally such as vehicles, personal devices and ITS smart devices (e.g. sign boards, traffic lights, tolls). In this layer the components play multiple roles as they can generate, collect and process the data. In addition, limited resources IoV components (e.g. sensors) can utilize the resources of this layer for fast and simple data processing. However, the main challenge in this layer is managing the computational resources while considering the ad hoc communication nature of some components (i.e. vehicles).
- **Intermediate fog computing layer:** this layer provides a more resource rich fog nodes which are located at the edge of the fixed network such as RSUs, AP, BS, and Routers. This layer receives the collected data from the lower layer. In comparison to lower layer, more complex data processing and analysis can be carried out utilizing a wider scope ITS data in this layer. As this layer is close to data consumers, it can provide fast responses to real-time ITS applications. In addition, this layer can provide better management for the low fog layer components.
- **High fog computing layer:** the components of the Internet WAN (e.g. ITS data servers and mini data centres) are used to provide the fog computing facilities in this layer. With plenty of resources, this layer performs more complicated data processing in comparison to the intermediate layer. Although processing data in this layer might experience some slight latency due to the data size and communication delays, still this layer can accelerate the stream-based data processing.
- **Cloud computing layer:** this layer is where the cloud-based big data centres are located. With the cloud computing and the complex artificial intelligence capabilities, massive big data analytics processes can be accomplished in this layer.

In the fog layers, fog nodes can form clusters, where computational and storage resources are shared among the cluster members. Basically, clusters might be formed horizontally (i.e. among fog nodes from same layer) or vertically (i.e. among fog nodes from multiple layers). In order to mange clusters resources, virtualization mechanisms need to be applied in horizontal and vertical clusters. Moreover, clusters formation can be dynamic to provide better quality of service in peak hours or high services demand areas. However, managing clusters with ad hoc resources is a critical issue in IoV environment. In this dimension, the raw data flow from the lower fog layer towards the cloud computing layer, whereas the processed data flow in the opposite direction. Nevertheless, deciding which data to send to the upper layers is a critical decision. Therefore, data size, applications requirements, and the resources of each layer need to be considered to make such decision. For example, a data set of 400TB is very difficult to be processed in the lower fog layer and it might take many days. On the other hand, processing 400TB in the cloud computing center might take few hours or less.

B. RITS-BDA REAL-TIME BIG DATA ANALYTICS DIMENSION

The real-time big data analytics dimension is focusing on providing the required layers for real-time big data processing while considering the intelligent computing and the IoV environment characteristics. Based on the general lambda architecture for real-time big data processing, this dimension is divided into three layers which are explained as follows:

- **Batch layer:** this layer utilizes the cloud computing facilities to perform the big data batch processing, and it provides virtually unlimited storage and computational capabilities. Although this layer can perform deep analysis on huge data sets, the batch processing speed is considered slow and its not sufficient to serve real-time applications. However, this layer generates batch views which are used in the serving layer to answer the queries of real-time applications. The level of data analytics granularity of each view is related to the requirements of real-time applications. Thus, different real-time applications may require different views with variable view updating frequency.
- **Speed layer:** the three fog computing layers resources (low, intermediate, high) are utilized in this layer which performs constant processing for data streams in near real time fashion. In particular, the procedure of data processing is divided into small tasks that can be processed in parallel and distributed manner. For this purpose the layers, clusters and nodes of fog computing facilities (e.g. WAN servers, AP, routers, vehicles, ITS smart devices and personal devices) are utilized. Each fog layer provides different granularity and speed of data processing, where higher fog layers generate wider scope data analytics through using larger sets of data. In comparison to batch layer, this layer processes smaller size data sets and it focuses on the recent data with less consideration for historical data. To speed up the processing, the components of this layer mostly employ the concept of in-memory processing. The outcome from this layer is represented in the real-time views which are delivered to the serving layer to be used with the batch layers views in answering the real-time applications queries.
- **Serving layer:** the intermediate and low fog computing layers resources are exploited in this layer. This layer

answers the queries of real-time applications through utilizing both the batch and the speed layers views. Obviously, the component of low and intermediate fog layers are serving in the stream data processing and answering the real-time applications queries. The data flow in this dimension is from the speed and batch layers towards the serving layer. As the size of data saved in this layer is relatively small, in-memory data storage systems such as Redis or Memchached, are adequate for this layer.

C. RITS-BDA INTERNET OF VEHICLES DIMENSION

The Internet of Vehicles dimension represents the components and activities of the the IoV environment. The structure of this dimension is inspired by the IoV architecture presented by [8]. This dimension consists of six layers which are described below:

- **Perception layer:** includes different types of sensors, actuators, vehicles, personal devices (e.g. smart phones) and ITS smart devices (e.g. tolls, traffic lights, sign boards). The primary responsibility of this layer is to collect the ITS big data which is related to users and vehicles as well as traffic environment and ITS devices. In addition, some components of this layer are involved in the low level data processing, as explained in the previous dimensions. Moreover, this layer includes many of the big data analytics outcome consumers such as vehicles and ITS smart devices, which benefit from the resulting intelligent decisions and predictions. The main issue in this layer is handling the energy constrains of some components, and the mobility of vehicles and mobile devices.
- **Infrastructure network layer:** the components of this layer includes the network edge nodes (i.e. RSUs, AP, BS, Routers), the Internet wide area network, and the data and cloud computing centres. This layer is the backbone of the IoV as it provides the essential infrastructure for communications as well as data storage and processing. Managing the heterogeneous network components' resources under high and unbalanced network loads is one of the critical issues related to this layer.
- **Artificial intelligence layer:** this layer utilizes the available cloud and fog computing resources to store, process and analyse the data received from both perception and infrastructure network layers. In fact, this layer is the IoV brain for making smart decisions based on the big data intelligent analysis. The critical issue in this layer is the efficient usage of the cloud and fog computing resources to fulfil the different analysis requirements of various ITS applications.
- **Communication layer:** it provides the required communication infrastructure in the form of a virtualized universal network that consists of different types of networks involving satellite networks, 5G, 4G/LTE, Wi-Fi, and WAVE. This layer is responsible for securely and efficiently transferring the perceived information from the lower layers to the artificial intelligence layer for processing, and then carrying back the outcome of the data analysis to the end users and devices. In addition, this layer must handle the different delay constrains imposed by various applications requirements. The insufficient standards, cooperation and interoperability among networks various types is one of the main obstacles faced in this layer.
- **Application layer:** this layer represents the ITS applications in IoV environment, ranging from traffic safety and management to web based utility and infotainment applications. The goal of this layer is to provide smart services to end users and devices using the intelligent big data analysis produced by the artificial intelligence layer. The main issues in this layer is providing efficient service discovery mechanisms to suite the user requirements.
- **Business layer:** this is the driving force behind the development of ITS applications. This layer main responsibility is to prevision future strategies for ITS business models development utilizing the big data analytics. Moreover, this layer produces variety of big data analytics representations (e.g. graphs, comparison tables and use cases), which supports decision makers in fields of economic investment and resource utilization, pricing different applications usages, allocating budgets for managerial and operational tasks.

Implementing the RITS-BDA architecture requires high coordination between the different dimensions and layers. In addition, reliable and secure communications are essential to achieve the best real-time ITS big data analytics performance.

VIII. CRITICAL ISSUES TO CONSIDER AND FUTURE RESEARCH DIRECTIONS

IoV, fog Computing and ITS big data analytics technologies are still in their infancy stages. Many serious research problems have not yet been addressed. This section discusses various challenges facing the integration of these three technologies and introduces future research directions.

A. IMPLEMENTING FOG COMPUTING IN IoV ENVIRONMENT

Fog computing is a very resource heterogeneous environment. Implementing fog computing in the IoV environment and involving vehicles as computing and communication devices (i.e. VFC) make the fog a highly dynamic environment. In particular, in IoV environment, the fog platform needs to handle extra challenges as vehicles move from one fog node to another while performing computing, communication and end-user roles. In addition, the high QoS requirements of many ITS big data applications creates unique challenges while implementing fog computing. The following points call researchers attention to important issues that need to be considered in future research work:

- **Fog network performance enhancement:** With the fast development of big data mining, it is feasible to extract interesting patterns or knowledge to enhance the self-organizing capabilities in fog computing [18]. Usually, the big data presents very important features such as user mobility/activity patterns and social, spatial, and temporal correlations of data contents. Thus, using the historical data and network global view, big data analytic can be used to predict events in advance, and to make the fog Computing units aware of these events in order to utilize the networks resources more efficiently [6]. For instance, through analyzing common interests and social relations of users in a specific area, the highly demanded contents can be fetched to the nearby fog units to decrease communications overhead and latency. Therefore, there is a need to introduce fog computing resource allocation prediction algorithms based on big data analytics to dynamically pre-allocated the resources based on predicted user demands.

- **Resource management:** Dynamic allocation is necessary for communication, storage, and computing resources in fog nodes in order to handle the massive and variable rate ITS data in real-time [6]. Further investigation is required on how to manage the available vehicular computation and storage resources. In addition, an adaptive optimization mechanism to allocate computing tasks effectively is necessary. Moreover, to fully utilize vehicles computational and communication capabilities, enhanced mobility models that describe vehicular behaviours accurately are highly required. In fact, modelling vehicles mobility is essential for efficient vehicular fog resource allocation as well as task distribution and scheduling. As vehicles can form vehicular data center in the fog layer, management policies and computational capacity estimation models in vehicular data centres are open research problems.
- **Cross-layer collaboration:** Multiple interface definitions is required to create suitable interfaces between fog layers and the cloud center, among multiple fog nodes, and between fog nodes and IoV objects/devices. Such interfaces are essential to cope with the multiple communication technologies in IoV environment.
- **Constructing efficient fog nodes:** How to dynamically select the fog devices in order to guarantee the availability of fog services in a certain region or certain users. The mobility of fog devices (e.g. vehicles) and end users highly affect this choice [77]. Therefore, studying the relationship between mobility patterns and the services demand is essential. In addition, fog nodes resources' capabilities and coverage area are significant parameters to consider while forming fog nodes in IoV.
- **Reliability in fog computing:** periodical check-pointing and rescheduling might be two useful techniques to provide high reliability, however, in the dynamic environment of fog networks such techniques might increase the latency [77]. Replication is a good choice but it should be considered in the early stage of fog network resource allocations and management. In general, fault tolerance techniques which mainly uses extra resources to cover some accidents/failures are facing a serious challenge in fog Computing environment. This is because fog computing has some resource constraints and implementing fault tolerance techniques may increase latency [62]. Accordingly, there is a requirement for suitable techniques to improve the reliability of fog Computing in the dynamic environment of IoV while guaranteeing the requirements of real-time big data analytics. In addition, network interruption problems must be considered to avoid fog computing services disruption. Reliable communication among the fog component and with fog nodes and clusters are highly essential to provide the fog real-time data processing services.
- **Fog computing capacity:** storage capacity and network bandwidth are two main issues related to fog capacity. Investigating data placement in fog networks is important as computation efficiency is affected by data locality [77]. The patterns of user service request and mobility are essential parameters to store data on certain fog units to reduce the cost and latency of computation and to maximize the throughput. Investigating the design of search engines that can process searching queries of scattered contents in fog nodes is an important research topic.
- **Concurrent processing in fog computing:** One of the main characteristics of IoV environment is the large number of vehicles that needs to process their data concurrently. Therefore, this issue needs to be considered while allocating and managing resources for data processing. In addition, as fog computing depends on resource virtualization, it might be difficult to fulfil the concurrent processing requirements. This is due to processing large amounts of data concurrently using limited fog computing resources. However, utilizing the dynamic resources of VFC and efficient prioritization methods might highly contribute in solving this problem. Thus, further investigations are required to study the performance of dynamic VFC resource allocation under the condition of high concurrency in data processing.
- **Virtualization in fog Computing:** In the context of IoV, virtualization of computing, communication and storage resources is important to handle the heterogeneity of resources. However, many aspects need to be considered such as mapping between logical and physical resources, devices attachment and their traffic dynamic routing, devices/networks discovery, and interfaces definitions and networks topologies [6]. In addition, it is interesting to investigate the radio resource slicing, isolation, and abstraction, which are more challenging in IoV wireless networks due to various access technologies, mobility, broadcast nature, and time-varying channels. Moreover, managing virtualization across multiple-layer fog computing platform while taking into consideration the

real-time analytics requirement of ITS big data is a new research opportunity.

- **Dynamic and adaptive system parameters configuration:** In the proposed ITS-REC architecture, to make the three dimensions (i.e. IoV, real-time big data analytics, and fog computing) work coherently, many system parameters need to be configured dynamically such as computational power, cache size, transmission power, carrier frequency, and bandwidth. In addition, the configuration of system parameters must adapt to the changes in the highly dynamic IoV environment. Configuring system parameters is essential to maintain the fog layers virtualization and the required QoS of ITS big data applications.
- **The design of control signalling:** Another important issue is the designing of control signalling while considering the communication overhead and delay constraints. In future research many questions need to be investigated such as how to make reliable signalling for device-to-device communications in device failure or device mobility cases, how to support devices scalability, is it more efficient to provide a dedicated radio channel or to share the control signalling channel. In addition, adaptable control signalling to work with heterogeneous wireless access technologies is required.
- **Coping with the ad hoc nature in IoV environment:** To control the admissions of incoming devices/users while assuring the existing devices/users QoS, policies for adequate admission control are highly required. Also, device discovery need to be done in the shortest available path to reduce the latency and bandwidth consumption [6]. With the introduction of vehicles as computational resources, providing fast admission and discovery mechanisms has a major impact on decreasing latency and increasing resource utilization.
- **Software defined network adoption:** Fog Computing in IoV is facing issues such as the tight coupling of control and data planes, expensive and complex network management, interfaces for specific vendors, and software designed for certain hardware. In addition, the dynamic adaptation to the network variable conditions is difficult to be achieved [6]. To overcome these challenges, it is important to consider the implementation of software defined networks (SDN). However, there are many questions that need to be answered such as how to establish cooperation among various controllers types such as partially connected controller (end-user devices) or constantly connected controller (at the infrastructure network edge) and where to controllers should be placed in fog networks. Moreover, it is a critical issue to create a SDN system that is distributed over large area and meets the harsh requirements of fog computing such as scalability, latency, mobility and wireless links reliability [77]. Therefore, further investigation is required in order to utilize the advantages of SDN to optimize the fog Computing performance in IoV environment.
- **Network function virtualization (NFV):** It uses instances of virtual machines to replace the network functions. The virtual machines can be created, offloaded and destroyed on demand, and fog computing highly utilizes the concept of virtualization. Thus, fog computing can benefit from NFV in many ways by virtualizing firewalls, load balancers, switches, and gateways and allowing fog nodes to use those instances. Unfortunately, NFV usage is not explored in the ITS fog computing environment [77]. For future research, NFV in IoV fog environment need to be studied while considering the requirements of ITS Big data analytics.
- **Data and computation offloading:** it is important to introduce efficient data/task offloading techniques that considers the QoS, transmission bandwidth and fog resources constraints while making decision on which data to process at the fog node and which to offload to the cloud [6]. In this context various parameters maybe considered such as computational rate, transmit power, latency, communication bandwidth, cache size, processor speed, and computing power [6]. In addition, for multi-layer fog Computing system, deciding which layer should handle which tasks need to be investigated to utilize the fog resources efficiently. In this context, methods for efficiently classifying and priotrizing data are highly necessary.
- **Tasks scheduling algorithms:** For the proposed architecture, adequate data/tasks scheduling/partitioning algorithms need to be investigated while considering task interdependency constraints and service/workflow execution time deadline.
- **Adaptive optimization problems:** Future fog based IoV networks need creative solutions to adaptively optimize communication, storage, and computing resources at both the cloud and edge sides [6]. Some examples include the processing power optimization under the constraints of transmit power, latency, mobility, and bandwidth, transmit power optimization while considering the restrictions of mobility and computational rate, and latency minimizations while taking into account the computation rate and intermittent communications constraints. Besides, optimization solutions for multi-objectives problems such as [90] and [91] can be adopted to deal with the conflicting objectives in IoV environment.
- **Motivating users to join VFC networks:** To efficiently utilize the resources of thousands of vehicles, first the participation of these vehicles by joining the VFC network need to be guaranteed. To attract people to join the VFC, they should experience the powerful features and conveniences of VFC which outweigh the paid costs [81]. One way is to provide an application where vehicles' owners can rent out vehicles' resources. In this way vehicles owners will be encouraged to share their

vehicles resources and be part of the dynamic VFC environment. Thus, proposing incentive-based vehicular resource sharing applications is highly recommended to utilize VFC and establish a good business with high investments.
- **IoV wireless communication nature:** Due to the dynamic environment of IoV, the variety of communication technologies, the energy constraints of some IoV devices (e.g. sensors), maintaining continuous and reliable communications among devices and with fog layers is challenging. Thus, more reliable and adaptive communication and routing protocols are required to suit the IoV fog based environment.

B. PERFORMING BIG DATA ANALYTIC IN IoV FOG ENVIRONMENT

- **Data heterogeneity:** It is essential to create data representation and processing models that accommodate heterogeneous or new types of data [5]. Also, intelligent data interpretation and semantic interoperability techniques are required as well.
- **Big data analytic in parallel and dynamic computing environment:** There is a real need to introduce data mining algorithms that are compatible with the latest parallel computing in fog based architectures. Synchronization issues in parallel computing may create bottlenecks for data mining methods [39]. This is considered an open issue in real-time ITS big data analytics. Moreover, as fog nodes might consist of static and mobile devices, distributing the data analytic tasks among such devices creates many challenges. For example, how to handle the tasks that were assigned to a vehicle if the vehicle is about to leaving the current fog node before finishing the task computations; and how to recover a task computation if the connection to the computing vehicle in a fog node is suddenly lost. Therefore, optimized task distribution, resource allocation and scheduling techniques are required to handle data analytic in the VFC environment. Furthermore, data processing techniques such as distributed encoding/decoding, cooperative monitoring/acquisition/sensing, compression and feature extraction techniques, and data filtering can be investigated in the dynamic environment of VFC.
- **Data visualization:** To carry out queries concurrent execution in parallel visualization algorithms, decomposing a problem into independent manageable tasks is critical [39]. The rapid increase in ITS data requires enormous parallelization which makes visualization a challenging task. In addition, designing visualization algorithms that are compatible with data heterogeneity is difficult. Different dimensionality reduction methods have been introduced in different fields [92]. However, these methods are unsuitable for ITS big data. Moreover, visualizing fine-grained dimensions effectively, increases the probability of identifying hidden correlations and patterns.
- **Decision making under uncertainty:** There are many sources of data and conditions uncertainty which must be considered carefully in decision-making. First, the understanding of intelligent transportation systems and Internet of vehicles implementation issues is incomplete. Second, the data needed to specify the boundary conditions with sufficient accuracy is not always available [5]. Thus, Decision making under uncertainty must be improved by understanding representation and propagation of uncertainty, assessment, and through conducting real-time experiments to learn more about the ITS and IoV environments.
- **Hierarchical big data analytic:** as fog computing platform may consist of multiple layer, hierarchical data mining techniques should be investigated [18]. Such techniques are essential to cope with the different computational capabilities of fog computing layers. In addition, implementing hierarchical big data analytics technique is required to fulfil the QoS of each of the fog layer's served ITS applications. However, it is challenging to choose the suitable analytic algorithms and data granularity for each level of computing. Moreover, generating the suitable batch views to serve the real-time applications, needs further research.
- **Data prioritization:** By employing suitable data prioritization techniques, the data processing can be handled at the suitable fog computing node through using efficient data analytic methods to provide faster responses to end-users [6]. Data prioritization is essential to serve the real-time big data analytics applications. However, optimum data prioritization requires the understanding of data hidden interdependencies and correlations, which makes the conventional methods of prioritization inefficient.

C. SECURITY AND PRIVACY ISSUES

As the intelligent transportation system is developing continuously, collecting big data is becoming more intensive through IoV various communications technologies, which leads to increasing number of security attacks and privacy violations. Nowadays, there exist some related works which focus on security of big data. Cardenas *et al.* [93] and Xu *et al.* [94] developed security and privacy mechanisms for big data applications. A key exchange scheme for secure scheduling of big data applications was proposed in [95]. For collecting big data in the large scale IoV environment, a secure mechanism is proposed in [26], where vehicles need to register in the big data center to connect into the network. Afterwards, vehicles associate with big data center via mutual authentication and single sign-on algorithm.

Guo *et al.* [26] identified some big data analytics security requirements in IoV environment which are listed in the following points:

1) Authentication to identify the big data center, sink node, and vehicle node.

2) To protect messages against destructions or modifications, integrity is required.
3) To protect the data sent to specific entity, confidentiality is highly important.
4) Nonrepudiation to prevent deny afterward.
5) To ensure that only an authorized node accesses the resources, authorization is necessary.

Mukherjee *et al.* [96]discussed the security and privacy issues of IoT fog based environment. However, the available protocols cannot be applied directly to secure the big data collection, storage and processing in large scale IoV fog based environment. This is because, the existing solutions do not consider the fog related security and privacy vulnerabilities in the IoV environment. Basically, as fog computing utilizes network edge and end-user devices to collect and process data, many security and privacy concerns may arise. Moreover, the security and privacy requirements of real-time ITS big data applications are not addressed in the existing studies. The following points summarizes important security and privacy issues that need intensive research in the area of fog computing based IoV for real-time big data analytics:

- **Security issues:** The security threats may arise in any of the data processing stages including data collection, information filtering, representation, data integration, modelling, and interpretation. In addition, an exponential growth in data rates creates difficulties in guaranteeing the security of critical data. Moreover, available security solutions are specifically designed for static data sets and not for the dynamic and continuous generation of data [39]. Nowadays, the practices of modular design for wireless communications are making the wireless IoV network vulnerable due to the poor protection mechanisms. In addition, the traditional wireless security provisioning techniques face significant challenges in the heterogeneous IoV networks environment due to the decentralization and the needs to support collaborative device to device communications [6]. Another critical issue is how to secure the connections of the huge number of heterogeneous IoV devices which have various processing capabilities levels. In a fog Computing environment, as distributed IoV devices have less protection capabilities and do not have a network global view, they are more vulnerable to attacks. In this case, fog nodes can play the role of proxies to handle the security functions of limited resources IoV devices [96]. However, sharing data and computation with other end-users (e.g. vehicles) in fog computing environment increases the vulnerability of ITS big data. Accordingly, there is an urgent need to design new security solutions that are suitable for the highly dynamic and vulnerable environment of fog based IoV, which consider the delay constraints of real-time ITS big data applications.
- **Privacy preservation techniques:** It is significant to provide efficient pseudonymizing techniques, differential privacy mechanisms, and privacy preserving data clustering algorithms, which preserve end-users privacy. Anonymity, encryptions, and temporary identification are several methods to protect data privacy. However, ethical factors, such as why to use the generated ITS big data, how to use, and what to use, have to be considered [97]. Before sending data to the cloud, fog nodes can play the role of aggregators and controllers for privacy critical data. However, the distributed storage environment of fog and the limited capabilities of some of its nodes impose some difficulties in employing the existing privacy preserving techniques. Thus, with high requirements of real-time ITS big data applications, lightweight and fast privacy protection algorithms are required, which can cope with the dynamic and heterogeneous environment of fog based IoV.
- **Trustworthiness of IoV systems:** With the growth of IoV, the collected data involve personal privacy (e.g. real-time locations of vehicles) as well as some important data like vehicles' running parameters which highly affect traffic safety [26]. Thus, to jeopardize the traffic systems or cause other traffic control problems, malicious vehicle nodes may send fraudulent messages. Some issues such as how to identify that the data source is not a malicious device, and how to evaluate the IoV sensors/devices trustworthiness, need further investigation in future research. Therefore, it is important to design a mechanism to ensure that IoV generated data is transmitted in a trusted way and not tampered with. In addition, the trustworthiness of fog nodes and the communication with them should be explored.
- **VFC concerns:** The security and privacy of the VFC network are extremely critical issues. Basically, VFC is more vulnerable to security and privacy threats because of its ad hoc nature, where vehicles which are end-users can join and leave the fog computing node at any time. In particular, vehicles can work as part of the fog node computational and storage resources, which gives them access to other devices data and this rises high security and privacy concerns. Due to the VFC environment characteristics, there are serious security challenges, such as misuse of protocols, lack of sufficient protection, and weak authentication. Thus, vehicles' operators are facing more danger from virus infection, hostile attacks, and information stealing. Therefore, it is critical to develop a suite of specifically designed security mechanisms for VFC to achieve security and privacy preservation [81].

IX. CONCLUSION

The ITS concept was introduced to increase road safety, improve transportation systems efficiency, and preserve our environment. However, as most of the ITS applications are becoming data-intensive applications, there is a need to fully utilize the power of big data analytics in ITSs. Nevertheless, employing big data analytics in the conventional way by depending on cloud computing services is not sufficient for ITS applications in the environment of IoV. This is because many ITS applications are delay-sensitive and processing the data at the cloud centers creates long delays. In addition, transferring the geo-distributed data to the cloud centres causes high network overhead and consumes network resources. Moreover, many of the ITS applications require location awareness and mobility support which are not provided through cloud based analytics.

Recently, the fog computing technology is introduced as a promising solution to support real-time big data applications. Fog computing complements the cloud computing by providing distributed, intelligent, and fast data processing at the network edge. In addition, fog computing node can consider the location awareness and mobility requirements while serving end users. However, big data applications cannot depend solely on fog computing as its computational and storage capacity is still limited in comparison to cloud platforms. Therefore both cloud and fog computing should be used to support the real-time ITS big data analytics in IoV environment.

Real-time big data analytics consists of three main stages including batch, speed, and serving. However, performing these three stages in the cloud is not going to serve the latency-sensitive applications. On the other hand, the fog platform cannot handle the batch processing stage. Therefore, big data analytics stages need to be distributed among the cloud and fog computing layers. Furthermore, the IoV environment must provide the required coordination and communication between the different layers and components.

By considering these aspects this paper proposed a novel architecture of three dimensions (intelligent computing, real-time big data analytics, and IoV) to enable the real-time ITS big data analytics in IoV environment. In addition, the opportunities and challenges that IoV and intelligent computing platforms are creating have been discussed. Moreover, a comparison between different edge computing technologies has been presented. Furthermore, critical issues and future research directions have been highlighted, which should be considered to improve the real-time big data analytics for many ITS applications.

Finally, the proposed architecture presents a good basement for future research in this field and it can be used as part of the intelligent transportation systems to enable the real-time applications such as collision avoidance, hazardous warning, advanced driver assistance systems, autonomous driving. As a result, many people lives will be saved by using more safe transportation systems. In addition, transportation systems will become more efficient and environmental friendly.

REFERENCES

[1] T. A. Litman and E. Doherty, "Transportation cost and benefits analysis II vehicle costs," Amer. Automobile Assoc., Heathrow, FL, USA, Tech. Rep. 2, 2015.

[2] J. Contreras, S. Zeadally, and J. A. Guerrero-Ibanez, "Internet of vehicles: Architecture, protocols, and security," *IEEE Internet Things J.*, Apr. 2017, doi: 10.1109/JIOT.2017.2690902.

[3] A. Mai. (2012). *The Internet of Cars, Spawning New Business Models*. [Online]. Available: https://www.slideshare.net/AndreasMai/12-1024scvgsmaciscoperspectivef

[4] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.

[5] Y. Sun, H. Song, A. J. Jara, and R. Bie, "Internet of Things and big data analytics for smart and connected communities," *IEEE Access*, vol. 4, pp. 766–773, Mar. 2016.

[6] S. K. Sharma and X. Wang, "Live data analytics with collaborative edge and cloud processing in wireless IoT networks," *IEEE Access*, vol. 5, pp. 4621–4635, Mar. 2017.

[7] C. Cooper, D. Franklin, M. Ros, F. Safaei, and M. Abolhasan, "A comparative survey of vanet clustering techniques," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 657–681, 1st Quart., 2017.

[8] O. Kaiwartya *et al.*, "Internet of vehicles: Motivation, layered architecture, network model, challenges, and future aspects," *IEEE Access*, vol. 4, pp. 5356–5373, 2016.

[9] C. Casetti and R. Frank, "Guest editorial 'Internet of vehicles,'" *Veh. Commun.*, vol. 8, p. 1, Apr. 2017.

[10] M. Chen, S. Mao, Y. Zhang, and V. C. M. Leung, *Big Data: Related Technologies, Challenges and Future Prospects*. Heidelberg, Germany: Springer, 2014.

[11] Y. Xiao and C. Zhu, "Vehicular fog computing: Vision and challenges," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, 2017, pp. 6–9.

[12] C. Casetti and R. Frank, "Guest editorial 'Internet of Vehicles,'" *Veh. Commun.*, vol. 8, p. 1, Apr. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2214209617300682

[13] J. Gantz and D. Reinsel, "Extracting value from chaos," *IDC iview*, vol. 114, no. 2011, pp. 1–12, Jan. 2011.

[14] S. F. Wamba, S. Akter, A. Edwards, G. Chopin, and D. Gnanzou, "How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study," *Int. J. Prod. Econ.*, vol. 165, pp. 234–246, Jul. 2015.

[15] B. Tang *et al.*, "Incorporating intelligence in fog computing for big data analysis in smart cities," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2140–2150, Oct. 2017.

[16] H.-H. Hsu, C.-Y. Chang, and C.-H. Hsu, *Big Data Analytics for Sensor-Network Collected Intelligence*. San Mateo, CA, USA: Morgan Kaufmann, 2017.

[17] A. Daniel, K. Subburathinam, A. Paul, N. Rajkumar, and S. Rho, "Big autonomous vehicular data classifications: Towards procuring intelligence in ITS," *Veh. Commun.*, vol. 9, pp. 306–312, Jul. 2017.

[18] Y. Ai, M. Peng, and K. Zhang, "Edge cloud computing technologies for Internet of Things: A primer," *Digital Commun. Netw.*, Jul. 2017. [Online]. Available: https://doi.org/10.1016/j.dcan.2017.07.001

[19] M. Gross, "A planet with two billion cars," *Current Biol.*, vol. 26, no. 8, pp. R307–R310, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960982216303414

[20] A. Awang, K. Husain, N. Kamel, and S. Aïssa, "Routing in vehicular ad-hoc networks: A survey on single- and cross-layer design techniques, and perspectives," *IEEE Access*, vol. 5, pp. 9497–9517, 2017.

[21] T. Darwish and K. A. Bakar, "Traffic density estimation in vehicular ad hoc networks: A review," *Ad Hoc Netw.*, vol. 24, pp. 337–351, Jan. 2015.

[22] T. Darwish and K. A. Bakar, "Traffic aware routing in vehicular ad hoc networks: Characteristics and challenges," *Telecommunication Syst.*, vol. 61, no. 3, pp. 489–513, 2016.

[23] Y. Fangchun, W. Shangguang, L. Jinglin, L. Zhihan, and S. Qibo, "An overview of Internet of Vehicles," *China Commun.*, vol. 11, no. 10, pp. 1–15, Oct. 2014.

[24] E.-K. Lee, M. Gerla, G. Pau, U. Lee, and J.-H. Lim, "Internet of vehicles: From intelligent grid to autonomous cars and vehicular fogs," *Int. J. Distrib. Sensor Netw.*, vol. 12, no. 9, p. 1550147716665500, 2016.

[25] M. Nitti, R. Girau, A. Floris, and L. Atzori, "On adding the social dimension to the Internet of Vehicles: Friendship and middleware," in *Proc. IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom)*, May 2014, pp. 134–138.

[26] L. Guo *et al.*, "A secure mechanism for big data collection in large scale Internet of Vehicle," *IEEE Internet Things J.*, vol. 4, no. 2, pp. 601–610, Feb. 2017.

[27] A. Mai and D. Schlesinger, "A business case for connecting vehicles: Executive summary," Cisco Internet Bus. Solutions Group, San Jose, CA, USA, Tech. Rep. 1, Apr. 2011.

[28] National Highway Traffic Safety Administration, "The economic and societal impact of motor vehicle crashes, 2010 (revised)," *Ann. Emergency Med.*, vol. 66, no. 2, pp. 194–196, 2015.
[29] R. Viereckl, D. Ahlemann, J. Assmann, and S. Bratzel, "Racing ahead the connected c@r," Strategy Formely Booz Company, New York, NY, USA, Tech. Rep., 2014.
[30] K. Golestan, R. Soua, F. Karray, and M. S. Kamel, "Situation awareness within the context of connected cars: A comprehensive review and recent trends," *Inf. Fusion*, vol. 29, pp. 68–83, May 2016.
[31] F. Bonomi *et al.*, "The smart and connected vehicle and the Internet of Things," in *Proc. Workshop Synchronization Telecommun. Syst. (WSTS)*, 2013, pp. 1–3.
[32] I. A. T. Hashem *et al.*, "The role of big data in smart city," *Int. J. Inf. Manage.*, vol. 36, no. 5, pp. 748–758, 2016.
[33] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.
[34] M. Chowdhury, A. Apon, and K. Dey, *Data Analytics for Intelligent Transportation Systems*. Amsterdam, The Netherlands: Elsevier, 2017.
[35] A. Luckow, K. Kennedy, F. Manhardt, E. Djerekarov, B. Vorster, and A. Apon, "Automotive big data: Applications, workloads and infrastructures," in *Proc. IEEE Int. Conf. Big Data*, Nov. 2015, pp. 1201–1210.
[36] T. Darwish and K. A. Bakar, "Lightweight intersection-based traffic aware routing in Urban vehicular networks," *Comput. Commun.*, vol. 87, pp. 60–75, Aug. 2016.
[37] T. S. J. Darwish, K. A. Bakar, and K. Haseeb, "Reliable intersection-based traffic aware routing protocol for urban areas vehicular ad hoc networks," *IEEE Intell. Transp. Syst. Mag.*, vol. 10, no. 1, pp. 60–73, Jan. 2018.
[38] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of 'big data' on cloud computing: Review and open research issues," *Inf. Systems*, vol. 47, pp. 98–115, Jan. 2015.
[39] M. Marjani *et al.*, "Big IoT Data analytics: Architecture, opportunities, and open research challenges," *IEEE Access*, vol. 5, pp. 5247–5261, 2017.
[40] R. Sowmya and K. Suneetha, "Data mining with big data," in *Proc. 11th Int. Conf. Intell. Syst. Control (ISCO)*, 2017, pp. 246–250.
[41] D. Puthal, S. Nepal, R. Ranjan, and J. Chen, "Threats to networking cloud and edge datacenters in the Internet of Things," *IEEE Cloud Comput.*, vol. 3, no. 3, pp. 64–71, Mar. 2016.
[42] Q. Shi and M. Abdel-Aty, "Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 380–394, Sep. 2015.
[43] L. Vanajakshi, G. Ramadurai, and A. Anand, "Intelligent transportation systems: Synthesis report on its including issues and challenges in India," in *Centre of Excellence in Urban Transport*. Chennai, India: Indian Institute Technology Madras, Dec. 2010.
[44] F. Cicirelli, A. Guerrieri, G. Spezzano, and A. Vinci, "An edge-based platform for dynamic Smart City applications," *Future Generat. Comput. Syst.*, vol. 76, pp. 106–118, Nov. 2017.
[45] H. Al Najada and I. Mahgoub, "Anticipation and alert system of congestion and accidents in vanet using big data analysis for intelligent transportation systems," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2016, pp. 1–8.
[46] N. Marz and J. Warren, *Big Data: Principles and Best Practices of Scalable Realtime Data Systems*. Shelter Island, NY, USA: Manning Publications, 2015.
[47] X. Liu, N. Iftikhar, and X. Xie, "Survey of real-time processing systems for big data," in *Proc. 18th Int. Database Eng. Appl. Symp.*, 2014, pp. 356–361.
[48] A. Spark, "Apache spark: Lightning-fast cluster computing," Apache Softw. Found., Wakefield, MA, USA, Tech. Rep. 2.3.0, 2016.
[49] L. Neumeyer, B. Robbins, A. Nair, and A. Kesari, "S4: Distributed stream computing platform," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Dec. 2010, pp. 170–177.
[50] A. Storm, "Storm, distributed and fault-tolerant realtime computation," Apache Softw. Found., Wakefield, MA, USA, Tech. Rep. 1.2.1, 2013.
[51] A. Hadoop, "Hadoop," Apache Softw. Found., Wakefield, MA, USA, Tech. Rep. 3.0.0, 2009.
[52] R. Sabapathi and S. Yadav, "Big data: Technical challenges towards the future and its emerging trends," *J. Manage. Technol.*, vol. 6, pp. 130–137, Mar. 2016.
[53] A. Thusoo *et al.*, "Hive: A warehousing solution over a map-reduce framework," *VLDB Endowment*, vol. 2, no. 2, pp. 1626–1629, 2009.
[54] A. F. Gates *et al.*, "Building a high-level dataflow system on top of map-reduce: The pig experience," *VLDB Endowment*, vol. 2, no. 2, pp. 1414–1425, 2009.
[55] S. Leo and G. Zanetti, "Pydoop: A python mapreduce and HDFS API for Hadoop," in *Proc. 19th ACM Int. Symp. High Perform. Distrib. Comput.*, 2010, pp. 819–825.
[56] A. Laicano, "Digital signal processing in hadoop with scalding," Cascading, USA, Tech. Rep. 1.0, 2013.
[57] P. Mell *et al.*, "The NIST definition of cloud computing," Nat. Inst. Standards Technol. Gaithersburg, Montgomery, MD, USA, Tech. Rep. 800-145, 2011.
[58] C.-W. Lu, C.-M. Hsieh, C.-H. Chang, and C.-T. Yang, "An improvement to data service in cloud computing with content sensitive transaction analysis and adaptation," in *Proc. IEEE 37th Annu. Comput. Softw. Appl. Conf. Workshops (COMPSACW)*, 2013, pp. 463–468.
[59] T. Gunarathne, B. Zhang, T.-L. Wu, and J. Qiu, "Scalable parallel computing on clouds using twister4azure iterative mapreduce," *Future Generat. Comput. Syst.*, vol. 29, no. 4, pp. 1035–1048, 2013.
[60] Y. He, F. R. Yu, N. Zhao, H. Yin, H. Yao, and R. C. Qiu, "Big data analytics in mobile cellular networks," *IEEE Access*, vol. 4, pp. 1985–1996, 2016.
[61] T. Huang, "Surveillance video: The biggest big data," *Comput. Now*, vol. 7, no. 2, pp. 82–91, 2014.
[62] Y. Liu, J. E. Fieldsend, and G. Min, "A framework of fog computing: Architecture, challenges, and optimization," *IEEE Access*, vol. 5, pp. 25445–25454, 2017.
[63] H. Cai, B. Xu, L. Jiang, and A. V. Vasilakos, "Iot-based big data storage systems in cloud computing: Perspectives and challenges," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 75–87, Jan. 2017.
[64] X. Masip-Bruin, E. Marín-Tordera, G. Tashakor, A. Jukan, and G.-J. Ren, "Foggy clouds and cloudy fogs: A real need for coordinated management of fog-to-cloud computing systems," *IEEE Wireless Commun.*, vol. 23, no. 5, pp. 120–128, May 2016.
[65] C. Vallati, A. Virdis, E. Mingozzi, and G. Stea, "Mobile-edge computing come home connecting things in future smart homes using LTE device-to-device communications," *IEEE Consum. Electron. Mag.*, vol. 5, no. 4, pp. 77–83, Apr. 2016.
[66] M. Satyanarayanan, "The emergence of edge computing," *Computer*, vol. 50, no. 1, pp. 30–39, 2017.
[67] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, "The case for VM-based cloudlets in mobile computing," *IEEE Pervasive Comput.*, vol. 8, no. 4, pp. 14–23, Dec. 2009.
[68] B. Briscoe *et al.*, "Reducing Internet latency: A survey of techniques and their merits," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 2149–2196, 3rd Quart., 2016.
[69] M. Satyanarayanan *et al.*, "Edge analytics in the Internet of Things," *IEEE Pervasive Comput.*, vol. 14, no. 2, pp. 24–31, Feb. 2015.
[70] A. N. Khan *et al.*, "BSS: Block-based sharing scheme for secure data storage services in mobile cloud environment," *J. Supercomput.*, vol. 70, no. 2, pp. 946–976, 2014.
[71] M. Patel *et al.*, "Mobile-edge computing introductory technical white paper," Eur. Telecommun. Standards Inst., Sophia Antipolis, France, White Paper 1, 2014.
[72] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," in *Proc. 1st Ed. MCC Workshop Mobile Cloud Comput.*, 2012, pp. 13–16.
[73] O. C. A. W. Group *et al.*, "OpenFog architecture overview," OpenFog Consortium Archit. Working Group, Fremont, CA, USA, White Paper 1, Feb. 2016.
[74] F. Bonomi, R. Milito, P. Natarajan, and J. Zhu, "Fog computing: A platform for Internet of Things and analytics," in *Big Data and Internet of Things: A Roadmap for Smart Environments*. Cham, Switzerland: Springer, 2014, pp. 169–186.
[75] R. Deng, R. Lu, C. Lai, T. H. Luan, and H. Liang, "Optimal workload allocation in fog-cloud computing toward balanced delay and power consumption," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1171–1181, 2016.
[76] P. G. V. Naranjo, M. Shojafar, L. Vaca-Cardenas, C. Canali, R. Lancellotti, and E. Baccarelli, "Big data over smartgrid-a fog computing perspective," in *Proc. SOFTCOM Workshop*, 2016, pp. 1–6.
[77] S. Yi, C. Li, and Q. Li, "A survey of fog computing: Concepts, applications and issues," in *Proc. Workshop Mobile Big Data*, 2015, pp. 37–42.
[78] S. Sarkar and S. Misra, "Theoretical modelling of fog computing: A green computing paradigm to support IoT applications," *IET Netw.*, vol. 5, no. 2, pp. 23–29, 2016.

[79] S. Sarkar, S. Chatterjee, and S. Misra, "Assessment of the suitability of fog computing in the context of Internet of Things," *IEEE Trans. Cloud Comput.*, vol. 6, no. 1, pp. 46–59, Jan. 2015.
[80] T. Darwish, K. A. Bakar, and A. Hashim, "Green geographical routing in vehicular ad hoc networks: Advances and challenges," *Comput. Elect. Eng.*, vol. 64, pp. 436–449, Nov. 2017.
[81] X. Hou, Y. Li, M. Chen, D. Wu, D. Jin, and S. Chen, "Vehicular fog computing: A viewpoint of vehicles as the infrastructures," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 3860–3873, Jun. 2016.
[82] F. Malandrino, C. Casetti, C.-F. Chiasserini, C. Sommer, and F. Dressler, "The role of parked cars in content downloading for vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 9, pp. 4606–4617, Nov. 2014.
[83] K. Dolui and S. K. Datta, "Comparison of edge computing implementations: Fog computing, cloudlet and mobile edge computing," in *Proc. Global Internet Things Summit (GIoTS)*, 2017, pp. 1–6.
[84] M. Shojafar, N. Cordeschi, and E. Baccarelli, "Energy-efficient adaptive resource management for real-time vehicular cloud services," *IEEE Trans. Cloud Comput.*, Apr. 2016, doi: 10.1109/TCC.2016.2551747.
[85] H. M. Raafat *et al.*, "Fog intelligence for real-time IoT sensor data analytics," *IEEE Access*, vol. 5, pp. 24062–24069, 2017.
[86] E. Baccarelli, P. G. V. Naranjo, M. Scarpiniti, M. Shojafar, and J. H. Abawajy, "Fog of everything: Energy-efficient networked computing architectures, research challenges, and a case study," *IEEE Access*, vol. 5, pp. 9882–9910, 2017.
[87] B. Tang, Z. Chen, G. Hefferman, T. Wei, H. He, and Q. Yang, "A hierarchical distributed fog computing architecture for big data analysis in smart cities," in *Proc. ASE BigData Social Inf.*, 2015, p. 28.
[88] S. Shin, S. Seo, S. Eom, J. Jung, and K.-H. Lee, "A pub/sub-based fog computing architecture for Internet-of-vehicles," in *Proc. IEEE Int. Conf. Cloud Comput. Technol. Sci. (CloudCom)*, Dec. 2016, pp. 90–93.
[89] J. Rivington and M. Swider, "Apple CarPlay: Everything you need to know about iOS in the car," Future Publishing Limited Quay House, Bath, U.K., Tech. Rep. 1.0, 2015.
[90] D. W. K. Ng, E. S. Lo, and R. Schober, "Multiobjective resource allocation for secure communication in cognitive radio networks with wireless information and power transfer," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3166–3184, May 2016.
[91] J.-H. Cho, Y. Wang, R. Chen, K. S. Chan, and A. Swami, "A survey on modeling and optimizing multi-objective systems," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1867–1901, 3rd Quart., 2017.
[92] A. T. Azar and A. E. Hassanien, "Dimensionality reduction of medical big data using neural-fuzzy classifier," *Soft Comput.*, vol. 19, no. 4, pp. 1115–1127, 2015.
[93] A. A. Cardenas, P. K. Manadhata, and S. P. Rajan, "Big data analytics for security," *IEEE Secur. Privacy*, vol. 11, no. 6, pp. 74–76, Jun. 2013.
[94] L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, "Information security in big data: Privacy and data mining," *IEEE Access*, vol. 2, pp. 1149–1176, Oct. 2014.
[95] C. Liu *et al.*, "An iterative hierarchical key exchange scheme for secure scheduling of big data applications in cloud computing," in *Proc. 12th IEEE Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Jul. 2013, pp. 9–16.
[96] M. Mukherjee *et al.*, "Security and privacy in fog computing: Challenges," *IEEE Access*, vol. 5, pp. 19293–19304, 2017.
[97] C.-W. Tsai, C.-F. Lai, H.-C. Chao, and A. V. Vasilakos, "Big data analytics: A survey," *J. Big Data*, vol. 2, no. 1, p. 21, 2015.

TASNEEM S. J. DARWISH received the B.Sc. degree in computer engineering from Islamic University of Gazah in 2005, the M.Sc. degree in electronics and electrical engineering from the University of Glasgow, Glasgow, U.K., in 2007, and the Ph.D. degree in computer science from the Universiti Teknologi Malaysia in 2017. From 2008 to 2013, she was as a Researcher and Lecturer Assistant with the Faculty of applied engineering, University of Palestine. He is currently a Researcher with the Universiti Teknologi Malaysia under the Post-Doctoral Fellowship Scheme for the project: Big data analysis on vehicular networks for intelligent transportation systems. Her current research focuses on vehicular communications, wireless ad hoc networks, intelligent transportation systems, and big data analytics. She is a member of the Pervasive Computing Research Group.

KAMALRULNIZAM ABU BAKAR received the B.Sc. degree in computer science from the Universiti Teknologi Malaysia, Malaysia, in 1996, the M.Sc. degree in computer communications and networks from Leeds Metropolitan University, U.K., in 1998, and the Ph.D. degree in computer science from Aston University, U.K., in 2004. He is currently a Professor in computer science with the Universiti Teknologi Malaysia, and a member of the Pervasive Computing Research Group. His research interests include mobile and wireless computing, ad hoc and sensor networks, information security, and grid computing. He is a member of the ACM, the Internet Society, and the International Association of Engineering. He involves in many research projects and is a referee for several scientific journals and conferences.