IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 31, 2017, accepted February 27, 2017, date of publication March 20, 2017, date of current version April 24, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2682640

Live Data Analytics With Collaborative Edge and Cloud Processing in Wireless IoT Networks

SHREE KRISHNA SHARMA, (Member, IEEE), AND XIANBIN WANG, (Fellow, IEEE)
Department of Electrical and Computer Engineering, Western University, London, ON N6A 3K7, Canada

Corresponding author: S. K. Sharma (sshar323@uwo.ca)

ABSTRACT Recently, big data analytics has received important attention in a variety of application domains including business, finance, space science, healthcare, telecommunication and Internet of Things (IoT). Among these areas, IoT is considered as an important platform in bringing people, processes, data and things/objects together in order to enhance the quality of our everyday lives. However, the key challenges are how to effectively extract useful features from the massive amount of heterogeneous data generated by resource-constrained IoT devices in order to provide real-time information and feedback to the end-users, and how to utilize this data-aware intelligence in enhancing the performance of wireless IoT networks. Although there are parallel advances in cloud computing and edge computing for addressing some issues in data analytics, they have their own benefits and limitations. The convergence of these two computing paradigms, i.e., massive virtually shared pool of computing and storage resources from the cloud and real-time data processing by edge computing, could effectively enable live data analytics in wireless IoT networks. In this regard, we propose a novel framework for coordinated processing between edge and cloud computing/processing by integrating advantages from both the platforms. The proposed framework can exploit the network-wide knowledge and historical information available at the cloud center to guide edge computing units towards satisfying various performance requirements of heterogeneous wireless IoT networks. Starting with the main features, key enablers and the challenges of big data analytics, we provide various synergies and distinctions between cloud and edge processing. More importantly, we identify and describe the potential key enablers for the proposed edge-cloud collaborative framework, the associated key challenges and some interesting future research directions.

INDEX TERMS Big data, data analytics, internet of things (IoT), cloud computing, edge computing, fog computing.

I. INTRODUCTION

The current trend in the Internet world is to connect all the devices/objects/things to the Internet with the objective of enhancing the quality of our everyday lives, thus leading to the emergence of Internet of Things (IoT) [1], [2]. In this direction, there has been a tremendous growth in the number of Internet-enabled smart devices and connections such as smartphones, Machine to Machine (M2M) connections, smart home appliances, and smart wearable devices, and this trend is expected to continue in the future. According to CISCO, more than 50 billion devices are expected to be connected to the Internet by 2020. Recent advances in sensing, computing, wireless communications, Internet protocols, and networking technologies have made the concept of IoT feasible [1]. However, the main challenge is how to handle the real-time processing of a huge amount of data/information, called big data, generated from heterogeneous wireless IoT environment.

Big data may be generated from various environments such as e-Healthcare environment, online business/e-commerce, broadband and multimedia contents, cloud radio access networks, and distributed storage/sensing [3]. Besides the content and traffic-related data, there is continuously increasing volume of signaling data due to the rapid deployment of various wireless networks including mobile and IoT networks. The complexity of big data generated from the IoT environment depends on the computational cost required in processing the data rather than the size of data itself. Besides, this massive amount of data needs to be transferred from the edge nodes to the cloud, leading to the need of enormous communication bandwidth which is precious and expensive natural resource. In addition, this massive data needs to be stored for further processing and also to facilitate real-time delivery at the edge-side, thus leading to the storage/caching constraints.

Existing wireless networks are mainly designed by considering communication resources as the primary resources with the connection-oriented approach, and other resources such as computing and caching are considered as secondary [4]. However, the demand is towards content-oriented networks facilitated by the recent advances in Internet technologies and cloud computing, where computing and caching will be the integral parts of the network. Also, in the fifth generation (5G) and beyond wireless networks, all types of resources such as communication, computing and caching will be distributed throughout the network, and it is an important challenge to coordinate among these resources towards their effective utilization in handling the massive amount of distributed data. One of the promising approaches to tackle the issues of the big data could be to enable synergies among communications, computing and caching components of future wireless IoT networks [3]. The integration of these paradigms may lead to additional degrees of freedom in effectively optimizing the resources of communication systems. In this regard, it has become an essential requirement to take all the involved resources into account while designing future wireless IoT networks by exploiting the synergy among communications, caching and computing paradigms [4], [5].

One of the recent developments in the computing world is the Internet-based computing, called cloud computing, which provides an ubiquitous and on-demand access to a virtually shared pool of configurable computing and storage resources [6]. Cloud computing is an excellent platform to handle the enormous data generated from the IoT environment due to cheaper and large amount of virtual computing/processing power available at the cloud centre. Therefore, the current trend is towards IoT-cloud convergence with most of the IoT platforms supported with cloud computing. However, it is not suitable for the applications demanding low-latency, real-time operation and high Quality of Service (QoS) [7]. In addition to low latency and location awareness requirements, the emerging IoT platform requires the support for seamless mobility and ubiquitous coverage which can not be fully supported by cloud computing solutions.

On the other hand, the concept of edge computing, also called fog computing,[1] is receiving important attention in order to address some of the drawbacks of cloud computing [8]–[10]. The main goal of edge computing is to extend the cloud computing functions to the edges of the network. Due to proximity to the end-users and geographically distributed deployment, it can support the applications/services demanding the requirements of low-latency, location-awareness, high mobility and high QoS [9]. However, edge computing units usually do not have enough storage and computing resources in handing the massive amount of IoT data. In addition, due to several involved constraints such as low-power, heterogeneity and weak capability of devices, IoT environment is more vulnerable to the information security. Therefore, there is a clear need to investigate suitable network architecture and control mechanisms to handle the processing of massive IoT data in a secured manner.

Although there are ongoing parallel advances in the fields of cloud computing and edge computing, interactions between these platforms in handling live data analytics from the communication perspective have not been investigated in the literature. A few works have recently highlighted the need of coordination between edge computing and cloud computing [8], [11], [12], however, they do not consider various practical aspects of live data analytics in wireless IoT networks. In this regard, this paper proposes a novel framework of collaborative edge-cloud processing in order to handle live data analytics in wireless IoT networks by combining advantages from both the cloud and edge computing paradigms. Due to huge amount of storage and computing resources available at the cloud-end, it is beneficial to offload much of the computational tasks to the cloud. However, it becomes highly advantageous to handle delay-sensitive tasks at the edge-side in order to ensure real-time processing and feedback to the end-users. More importantly, in the proposed framework, cloud processing can utilize its network-wide global knowledge and historical/delayed information and can act as a monitoring or guidance platform to guide edge processing units towards the effective-utilization of available resources. Starting with the basic features and key enablers of big data analytics, we discuss various challenges in performing live data analytics in wireless IoT networks. Subsequently, we provide synergies and differences between cloud and edge computing platforms. Then, we propose a novel framework for collaborative processing between edge and cloud computing along with some interesting applications. Subsequently, we discuss various key enablers, associated challenges and future research directions with the objective of stimulating future research activities in this emerging domain.

The remainder of this paper is organized as follows: Section II provides the basic features, the key technology enablers for big data analytics and the associated challenges in wireless IoT networks. Section III compares edge processing and cloud processing platforms from the perspective of live data analytics. Section IV proposes a novel collaborative edge-cloud processing framework for handling live data analytics in wireless IoT networks while Section V discusses various technology enablers, issues and future research directions. Finally, Section VI concludes the paper.

II. DATA ANALYTICS IN WIRELESS IoT NETWORKS

In this section, we provide the basic features, key enablers and the challenges for big data analytics in wireless IoT networks.

[1] In this paper, the terms "edge computing" and "fog computing" are used interchangeably, and they refer to the computing/processing at the IoT gateway/aggregator nodes/edge servers.

A. BASIC FEATURES

The term "Big data" usually refers to extremely large, heterogeneous and complex (semi-structured and unstructured) data-sets, which cannot be handled by the conventional data processing and storage tools/applications such as Relational Database Management System (RDBMS) [13]. The importance of big data lies on how meaningful information can be extracted from it for a particular application rather than the size of the data, and this extraction process requires novel data analysis methods and huge processing power. In wireless IoT environments, big data may be generated from a variety of application scenarios ranging from smart home scenario to e-Healthcare applications. In addition to the importance of content and control signaling data in wireless networks, location-based data from various sensors such as GPS sensors and embedded sensors in mobile devices can provide significant inputs to the government bodies in developing specific strategies for public facilities, transportation system, emergency responses and crime/risk warnings. Moreover, by analyzing the habits and interests of customers, industries may plan their future products in order to address their customers' personalized as well as group needs [3].

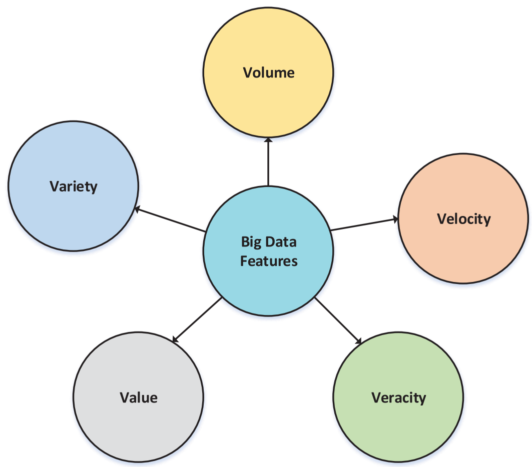

FIGURE 1. Main attributes of big data.

As depicted in Fig. 1, the commonly discussed attributes of big data are [13]: (i) volume, (ii) variety, (iii) veracity, (iv) velocity, and (v) value. The first two attributes, i.e., volume and variety, reflect to the hardware and software requirements in handling massive heterogeneous data-sets while the features veracity and velocity translate into the real-time processing ability with sufficient trustworthiness. On the other hand, acquisition of the highest useful value from the complex big data-sets in wireless IoT networks requires interdisciplinary cooperation among academia, enterprises and wireless industries [13].

B. CHALLENGES

In contrast to the traditional data, big data mainly differs in the following way [14]: (i) data rate is more rapid and data volume is constantly updated, (ii) data is of semi-structured or unstructured nature, (iii) data source is fully distributed, (iv) data access is in batch mode or real time instead of more interactive feature in the traditional data, and (v) integration of heterogeneous data from different sources becomes complicated. The heterogeneous data generated from IoT devices may have certain statistical and strong correlative features across several dimensions such as time and location, and also the devices may have social relations among themselves [15]. In addition, in hierarchial IoT networks, the aggregated features of the data traffic can be exploited in order to regulate the peak content demand, for example, cluster planning based on data distribution, peak load shifting and cache provisioning. Besides, IoT devices with similar interests may share the contents from their nearby devices and this content sharing may be enabled using either infrastructure-based communication or some infrastructure-free communication. Such a peer to peer nature of resource sharing, called as crowd computing [15], can exploit the spatial correlation as well as the mobility of IoT devices.

Big data analytics at the cloud centre can easily integrate the data collected using distributed sensors and aggregator nodes while exploiting correlation among the data-sets [16]. More importantly, this has the ability to analyze the massive data with ever-increasing scale and complexity, and can provide a global point of view across the whole network. Moreover, this cloud-based approach will lead to lower-error, higher-precision, and more dynamic treatment of data than the conventional data analytic approaches [16]. However, handling IoT data in the cloud platform in a traditional way creates several issues due to specific features of IoT data described in the following [17].

1) **Distributed and heterogeneous data structure**: IoT data is generated from distributed heterogeneous nodes which is largely diverse and may range from integer to character, and can be semi-structured and unstructured such as audio, images, and video. In addition, big data in wireless networks is usually distributed across several domains such as frequency, space, time, codes, and antennas. Also, the involved data sources may have distinct characteristics in terms of data rate, mobility, power levels and transmission schemes.
2) **Real-time requirements**: The dynamic environment of wireless IoT systems creates the need to handle a large volume of real-time and high-speed continuous data streams.
3) **Weak data semantics**: The data acquired from IoT sensors are mainly of low-level having weak semantics and are gathered with the help of resource-constrained sensors/devices/objects. In order to extract meaningful information from the collected data, they need to go through effective processing by exploiting various aspects such as spatial-temporal correlation and event-driven knowledge.
4) **Data inaccuracy**: Since the information gathered by the employed sensing system may not be accurate

enough due to various practical constraints, suitable multi-dimensional sensing, data analysis and processing techniques need to be investigated for wireless IoT applications.

The data analytics life-cycle to handle IoT data starts with the collection of the raw data collection at the IoT sensors, and then the acquired raw information is aggregated at the aggregator/gateway, and is subsequently sent to the cloud for further processing [17]. The raw information gathered from the IoT sensors may represent various parameters such as vibration, pressure, temperature, motion, heart rate etc., and they need to be converted into recognizable formats for further analysis. The acquired data may be transmitted to the cloud centre either in a single hop or multiple hops based on the employed network topology. The completion of this life-cycle requires effective coordination of various activities among IoT sensors, IoT aggregator nodes/gateways, in-transit network devices such as routers/relays and software/hardware resources in the cloud center [18]. Furthermore, it is crucial to have reliable and effective wireless communication links among these entities in order to ensure proper operation of wireless IoT networks.

Another key issue for handling big IoT data is how to turn the burden of handling big data to the benefits in improving the performance of wireless networks [15]. The big data usually exhibits highly useful features such as user activity/mobility patterns and temporal, spatial and social correlations of data contents. By properly extracting and effectively utilizing these features, the performance of various wireless networks could be significantly enhanced. For example, by analyzing social ties and common interests of people in a certain region, the network can fetch the popular contents to the corresponding edge gateway in order to reduce latency as well as the communication bandwidth overhead.

In wireless IoT networks, it is crucial to develop techniques to convert a massive amount of raw data into the meaningful information. The main challenge during this data interpretation and knowledge formation is to develop suitable data inference techniques to deal with the noisy and real-world data. Furthermore, dealing with the uncertainty in the interpreted data is another challenge due to dynamics of time-varying wireless environment. In order to make a reliable and trustworthy decision, reliable transport protocols and suitable in-field sensor calibration techniques need to be investigated [19]. Otherwise, it may lead to wrong or incomplete data, resulting in false conclusions, which if used in practical scenarios may lead to serious problems. In order to tackle this, the inferred information can be supported by a probability-based confidence level which can be used to ensure the safe operation of the devices [19]. Moreover, in several scenarios, it may be necessary to combine the current sensor data with the historical data in order to derive an effective conclusion.

C. ENABLERS FOR BIG DATA ANALYTICS

The conventional data management tools such as RDBMS are mainly designed to handle structured data and are not suitable for handling semi-structured or unstructured data. In addition, for handling the massive IoT data, RDBMSs can scale up with the costly hardware but not with the commodity hardware [14]. To address the aforementioned issues, some ad-hoc solutions have been recently proposed by the research community. In terms of infrastructure level, cloud computing has been considered important for handling big data due to its several features such as scalability, elasticity and cost-effectiveness. In order to handle massive unstructured data-sets, NoSQL databases and distributed file systems have been suggested. Furthermore, programming frameworks like MapReduce have been proposed in order to handle group-aggregation tasks like website ranking. Moreover, in terms of system-level solution, open-source software frameworks like Hadoop have been proposed to integrate data storage, data processing and other modules [14]. However, these solutions are not mature enough in order to handle huge data collection and transmission, and to provide real-time processing and feedback to the end-users. Also, high power consumption, large network overhead, latency, privacy and security are of important issues to be addressed.

Figure 2 illustrates the key technology enablers for big data analytics. In the following, we provide brief description of these techniques from the data analytics perspective. Stochastic models are probabilistic models and are usually used to capture the explicit features and dynamics of the data traffic. The commonly used stochastic models are Markov models, time series, geometric models, and Kalman filters [15]. On the other hand, data mining approach tries to extract implicit information from the data-sets and transform this information to a known structure for further usage by employing suitable anomaly detection, classification, clustering, and regression analysis methods.

Machine learning techniques aim to create a functional relationship between input data-set and output actions, and are capable of performing predictions and decisions based on the input data without requiring the need of following static program instructions. These techniques can be broadly categorized into unsupervised, supervised and reinforcement learning, and they may comprise of various classification techniques, regression analysis and Q-learning techniques. In addition to the conventional machine learning techniques, several advanced learning techniques such as active learning, deep learning, and online learning can be utilized to extract useful information from incomplete or complex data-sets. Active learning is mostly useful for partially labeled data-sets while deep learning is suitable for modeling complex behaviors of heterogeneous data-sets [15]. On the other hand, online learning deals with learning in real-time and is useful for applications where data arrives in a sequential order.

In terms of computing platforms, edge computing and cloud computing are considered as key solutions for handling big data analytics, and they will be detailed later in Section III. In addition, crowd computing is another emerging paradigm in which nearby devices with social ties and similar interests can share resources in order to maximize the overall

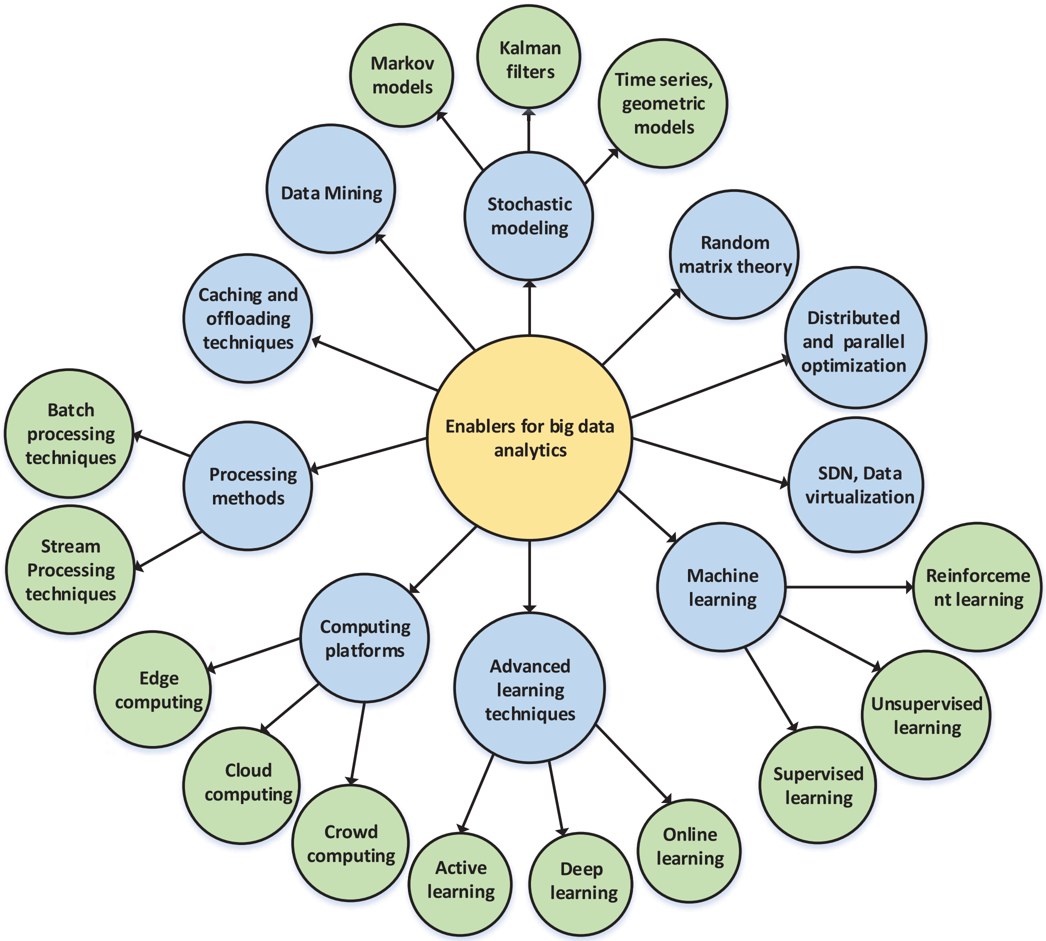

FIGURE 2. Main technology enablers for big data analytics.

system performance. By implementing this paradigm, devices having higher computing capability and power resource can help other devices with lower computing capability and battery level in achieving their performance targets. Moreover, crowd computing-based solutions can exploit temporal, spatial correlations as well as mobility of IoT devices for information/resource sharing and can subsequently reduce the backhaul/fronthaul bandwidth in infrastructure-based wireless IoT networks [15].

Based on the processing time required for big data analytics, the potential techniques can be broadly classified into stream processing and batch processing [14]. In the first approach, the data arrives to the system continuously in the form of a stream and the data size is infinite or unknown in advance. Whereas in the batch processing method, data size is finite and known in advance, and they are stored and analyzed. In this approach, the data are divided into small chunks and processed in parallel in a distributed manner.

The widely used linear algebra methods such as Cholesky decomposition and matrix multiplications, and convex optimization algorithms can not be straightforwardly applied for big data analytics because of extremely large sizes of data as well as the involved parameters [20]. This requires the need of adapting the existing algorithms and investigating new optimization techniques for big data analytics problems. Optimization algorithms suitable for big data can be categorized into the following types [20]: (i) first order methods, (ii) randomization, and (iii) parallel and distributed computation. Out of these, distributed optimization algorithms play a key role in solving resource allocation and management problems in heterogeneous wireless IoT networks where information and resources are often distributed across various entities of the networks. Some examples of distributed optimization algorithms are alternating direction method of multipliers and primal/dual decomposition, which can decompose complex optimization problems into sub-problems in such a way that they can be simultaneously computed [15]. This parallel computation capability significantly helps in reducing computational burden in a single computing unit as well as the communication overhead over the fronthaul/backhaul.

One possible way of dealing with the high dimensionality of big data is to use random matrix theory [21], [22] by representing big data in the form of large random matrices. This analysis is based on some of large dimensional matrix

TABLE 1. Key differences between edge computing and cloud computing.

Features	Cloud computing	Edge computing
Computational capacity	High	Medium to low
Size and Operating mode	Server very large in size and centralized	Edge servers smaller in size and placed over many locations
Applications	Suitable for delay-tolerant and computationally-intensive applications	Suitable for applications demanding low latency, real-time operation and high QoS
Fronthaul/backhaul communication overhead	High since devices need to be connected to Internet throughout entire duration	Low since devices can get cached contents directly from edge gateway
Deployment	Requires complicated deployment planning	Possibility of ad-hoc deployment with no or minimal planning

analysis tools such as high-dimensional statistics, matrix analysis, and convex optimization [16]. For example, a massive Multiple Input Multiple Output (MIMO) system can be regarded as a big data system considering storage and processing at the IoT gateway, and the principles of large random matrices can be applied to the architecture of large antenna array of massive MIMO system [23]. In addition, several dimension reduction approaches such as Principal Component Analysis (PCA) and tensor decomposition can be employed in order to reduce the data volume without changing the main features of the data [15]. By reducing the dimension of data, a significant gain can be achieved in saving the system cost for storage, processing and communication resources. As illustrated in Fig. 2, other enablers of big data are caching techniques [24], offloading techniques [25], Software Defined Networking (SDN) [26] and virtualization technologies [27], which will be described later in Section V.

III. EDGE COMPUTING VERSUS CLOUD COMPUTING

Existing wireless networks and Cloud/Centralized Radio Access Networks (C-RAN) under investigation are mainly designed to deliver contents without having the capability of analyzing and making use of the data-specific features in optimizing the system performance [15]. In this regard, it is important to make existing wireless networks scalable in order to handle massive data contents. Also, it is crucial to investigate suitable network architectures/mechanisms to incorporate and utilize big data awareness in wireless networks in order to enhance the system performance. As the amount of data generated by a large number of distributed sensors is highly unstructured and heterogeneous in nature, it becomes extremely complex to handle them with the conventional approaches. Mainly, how to acquire, integrate, store, process and utilize the data in highly distributed environments has been an important challenge for researchers, engineers and data scientists.

To address these issues, CISCO has recently proposed the concept of fog computing which aims to support cloud computing platform in handling a part of the workload locally at the edge devices such as switches, routers, and IP-enabled video cameras instead of transmitting the whole work load to the cloud [28]. The fog/edge computing can be enabled in the existing cloud-based networks by introducing an intermediate layer, called fog/edge layer, which may comprise of several edge servers distributed over various places such as shopping centres, parking areas and bus stations. The edge server can be regarded as a low-capacity version of the cloud server and has communication, computing and data storage capabilities.

Edge computing becomes highly advantageous for mobility support, geo-distribution, and location/context awareness. The geo-distributed nature of edge computing helps to provide rich contextual information such as event status, local network conditions, and the end-user's status. These information can be subsequently used for context-aware optimization of edge/fog applications. More specifically, edge computing can support the existing cloud computing platform in handling the following different types of applications, whose requirements can not be met with the cloud processing [28].

1) Applications demanding very low and predictable latency such as video conferencing, online gaming, and e-Healthcare
2) Real-time mobile applications such as smart connected vehicles
3) Geographically distributed applications such as wireless sensor networks for environmental monitoring
4) Large-scale distributed control systems such as smart traffic lights, smart grid and smart energy distribution

As mentioned earlier, the main benefits of cloud computing platform in wireless IoT networks are massive storage, very high computational efficiency, wide-area coverage while the main advantages of edge computing are real-time data handling, edge resource pooling, user-centric process, the support for high mobility and high QoS [7], [12]. In Table 1, we provide the key differences between edge computing and cloud computing platforms from the perspective of handling big-data in wireless IoT networks [12].

Due to ever-increasing demand for data content, the transmission of massive amount of data to the cloud creates a huge burden on the communication bandwidth of wireless networks. Furthermore, this results in intolerable latency and degraded service to the end-users. Moreover, the support of geo-distribution and mobility is an essential requirement, which can not be fulfilled by the cloud computing platform due to its centralized nature of storage and computing functions [28]. Therefore, it is crucial to explore the collaborative processing of edge and cloud computing platforms, which will be described in the following section.

IV. PROPOSED COLLABORATIVE EDGE-CLOUD PROCESSING

As highlighted in Section III, edge computing and cloud computing solutions have their own distinct advantages and disadvantages from the perspective of live data analytics in wireless IoT networks. The integration of centralized feature of the cloud and the real-time advantage of edge computing can address various issues in dealing with real-time data analytics in wireless IoT networks. Motivated by this aspect, in this section, we propose a novel framework for collaborative edge-cloud processing in wireless IoT networks.

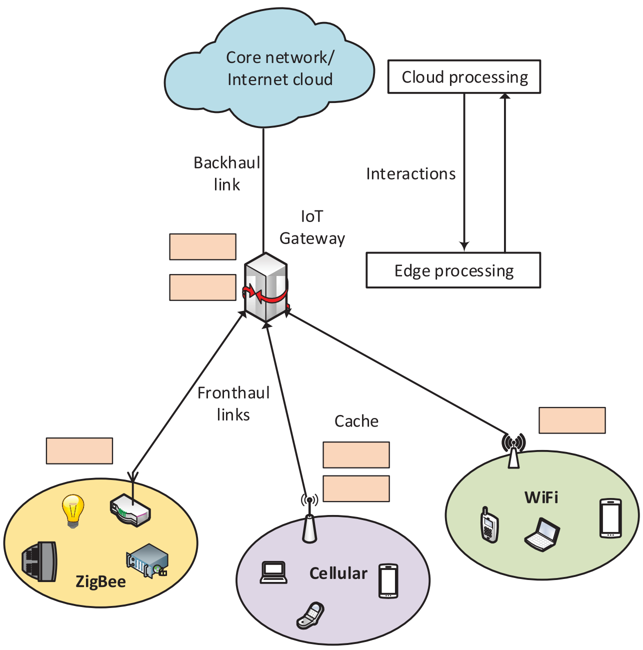

FIGURE 3. **Proposed generalized system model for collaborative edge-cloud processing in heterogeneous IoT networks.**

Figure 3 presents a generalized system model for collaborative edge-cloud processing in heterogeneous wireless IoT networks. In the proposed model, IoT edge gateways are equipped with cache memory and are capable of performing edge-caching in order to deliver the popular contents locally. The edge computing nodes may be any devices having the capability of computing, storage and network connectivity such as routers, switches, and video surveillance cameras. Depending on the application scenarios, IoT networks may comprise of various networks having distinct characteristics. For example, in the smart home scenario, wireless IoT networks may consist of a WiFi network, a bluetooth network, a Zigbee network and a cellular network. The raw-data coming from different domains/sensors is largely diverse and need to be collected over time. In addition, data dimensions and sizes may be different depending on the considered IoT application scenario. Besides the real-time processing of massive IoT data, this collaborative framework an enable new wireless IoT applications which may require collaborations among different edge computing units, and between edge computing units and the cloud centre.

The proposed system will benefit from the advantages of both the cloud computing and edge computing. In addition to this, we envision cloud centre as a monitoring and guidance platform to have effective real-time data processing at the edge-side of wireless IoT networks. In practical scenarios, IoT devices/sensors are heterogeneous in nature in terms of their computing capabilities, intelligence as well as the computing/processing power. In this regard, it becomes highly beneficial to guide the operation/processing of edge-nodes in order to utilize the available communication and computing resources in an effective manner. In the considered framework, edge computing helps to gather information from the surrounding radio environment while the cloud computing assists by providing suitable instructions to the edge-side nodes for their operations. For example, the operations at the edge-side such as data compression, filtering, sampling rate, power control, and making decisions on the type of data to be sensed/acquired can be supported by the cloud centre by providing suitable control signals over the feedback links.

Since the cloud centre can have a global view of information collected from a large number of sensors deployed over a large geographical region, the control of edge processing from the cloud-side can provide significant improvements in future wireless IoT networks. Due to huge amount of computing resources available at the cloud end, it is beneficial to offload much of the computational tasks to the cloud. On the other hand, it is advantageous to handle delay-sensitive tasks at the edge-side. Depending on various levels of information such as traffic types, location information, processing delay and transmission overhead, the decision on whether to offload data to the cloud or not can be made. It can also be considered that all the edge nodes are operated in a coordinated fashion in order to help each other in terms of communication, computing and storage/caching resources. Another important aspect which can be exploited in the proposed framework is that cloud processing can utilize the history/delayed information available at the cloud-centre in order to infer certain decisions for the edge processing without the need of waiting for the instantaneous data collected from IoT nodes.

The introduction of edge-computing in the conventional centralized cloud computing set-up brings up new opportunities to balance the trade-off between centralized and distributed network architectures [12]. In this regard, the proposed collaborative edge-cloud processing will be significantly useful to decide on what actions to perform locally and what actions to be sent to the cloud. Various constraints in the considered system model include computational rate, computing power, processor speed, cache size, communication bandwidth, latency and transmit power. The performance of the considered framework in Fig. 3 can be evaluated in terms of various metrics such as energy efficiency, spectral efficiency, throughput, operational efficiency, cache hit ratio, computational efficiency, end to end latency and offloading efficiency.

Since wireless environment is time varying in nature, it is crucial to adapt the proposed collaborative processing dynamically. Depending on different network conditions, physical channel conditions and data features such as temporal and spatial correlations, the proposed collaborative platform may adaptively choose among central processing at the cloud centre, local processing at the edge gateway and parallel processing at both the ends. The historical data available at the cloud side and the instantaneous data collected by the edge nodes at the current instance can be used to predict various important parameters such as future traffic flow, energy usage, weather forecasting, community activity, and geo-location. In this context, it is important to investigate dynamic prediction algorithms which can be updated based on time-varying situations such as source node failure and security threat.

Due to massive amount of IoT data and continuously increasing number of service requests, power consumption for operating servers in the cloud centre is rapidly increasing [29]. Therefore, it is crucial to investigate suitable strategies to reduce energy consumption in the edge-cloud coordinated platform. On the other hand, it is important to guarantee the latency requirements while delivering services to the end-users. In this regard, it is essential to investigate the fundamental tradeoff between energy consumption and latency in the considered cloud-edge platform of wireless IoT networks [28]. Moreover, the massive data-sets gathered from various locations have heterogeneous formats and are usually semi-structured and unstructured in nature. In most cases, the data-sets are in a raw form with inconsistency, high redundancy and much useless information. Without performing pre-processing operations on the data-sets, not only a huge storage is required but also the data may not fit into the predefined database structure. Therefore, it becomes highly beneficial to carry out data pre-processing such as data integration, redundancy elimination and data cleansing at the edge-side in order to avoid the unnecessary storage space, and also to enhance the computational efficiency.

Furthermore, it may not be necessary for all IoT sensor nodes to upload the collected/generated data to the cloud. In this regard, the edge computing unit can inform IoT devices when to stop sending the data to the cloud in order to effectively utilize the network resources. In addition, a certain portion of the data may not be required for a certain time period. In this case, the IoT gateway can inform the IoT sensor/device when it needs to stop uploading the data. This collaborative procedure will help to save the cloud as well as the network resources for that idle period. Besides, in the existing scenarios, many nodes are directly connected to a sink node/access point/coordinator, which leads to significant increase in the complexity for node scheduling, and subsequently it will lead to the intolerable system delay. In this context, the proposed framework follows a hierarchical network structure in which the nodes can be grouped locally into different clusters and they can be connected to the sink via their cluster heads [30]. By enabling collaborations among the devices within a cluster using their social relations, devices can share various communication, storage and computing resources, and subsequently the network performance can be enhanced in terms of different performance metrics such as latency and energy efficiency, content delivery efficiency and security.

The proposed collaborative edge-cloud processing framework can be applied to handle real-time data analytics in wireless IoT networks with different performance objectives. Some of the potential applications are listed below.

- Cloud-assisted adaptive optimization of computing, communications and caching resources
- Cloud-assisted energy-efficient caching and task/data offloading
- Spectrum monitoring and dynamic spectrum management using collaborative edge-cloud processing
- Event-driven resource allocation and network management using collaborative edge-cloud processing
- Cloud-assisted security and privacy enhancement

V. POTENTIAL ENABLERS, CHALLENGES AND FUTURE RESEARCH DIRECTIONS

In this section, we provide various potential enablers, key challenges and potential future directions for the proposed collaborative edge-cloud processing by categorizing them into the following topics.

A. COORDINATION MECHANISMS BETWEEN EDGE COMPUTING AND CLOUD COMPUTING

In the proposed collaborative edge-cloud architecture, communication, computing and storage functionalities need to be dynamically allocated among the edge-side units, cloud and the things (devices/sensors) in order to handle the massive IoT data in the real-time. Besides, there is a strong need to have the coordinated management of cloud computing and edge computing units in handling massive IoT connections in a reliable and secured manner [8]. In other words, effective coordination mechanisms between edge processing units and cloud processing units are crucial in order to realize this architecture effectively. In order to enable these interactions, suitable interfaces between edge processing unit and the cloud centre, among different edge processing units, and between edge processing units and IoT devices/objects need to be defined.

Moreover, in the proposed coordinated platform, the cloud center is assumed to be capable of dictating the edge computing units about which parameters are important to monitor, how frequent a parameter needs to be monitored and which task should be processed to meet the real-time requirements of end-users. In this regard, various aspects such as coordination framework definition, parameters to be coordinated, control signaling design from the cloud center, and load balancing at the edge and cloud sides need to be investigated.

Some of the research challenges and future research directions under this research topic are provided below.

- **Definition of edge-cloud interaction mechanisms**: Various aspects such as what should be the suitable interfaces between cloud and edge, between cloud and things, between edge and things/devices and between different edge computing units, which computing and communication parameters will be involved in establishing relationship between edge and cloud units, and how to design common control signaling and management schemes in order to control edge units from the cloud, need to be investigated under the proposed framework.
- **Resource mapping between edge and cloud computing**: In order to enable effective collaborations between edge and cloud units, it is important to investigate suitable techniques to map edge-side computing/communication resources with the cloud-side resources, and also suitable strategies to share resources among multiple edge units for handling live data anlytics.
- **Load balancing among edge, cloud and things**: It is crucial to investigate which tasks can be handled at the edge gateway and which tasks should be sent to the cloud, which tasks can be handled at the things/devices, in order to effectively optimize the available edge and cloud resources in the proposed platform.
- **QoS enhancement schemes**: In the considered cloud-edge coordinated platform, it is important to identify the performance bottlenecks (communication bandwidth, cache size, computing power, etc.) and then to develop suitable techniques to minimize service response time and to improve reliability in the case of network/link failures.

B. BIG-DATA AWARE EDGE-CLOUD COLLABORATIVE PROCESSING

One promising way of dealing with the big data in wireless IoT networks is to understand the features of big data and to incorporate this awareness in order to enhance the system performance [15]. In contrast to multimedia contents, IoT data has several peculiar features such as bursty nature, small data-size and transiency, i.e., they expire in short time [31]. Besides, in the proposed framework, only enriched data with the meaningful information can be forwarded to the cloud instead of sending all the raw data to the cloud. In this direction, it is interesting to explore model-based data processing techniques such as data compression, device cooperation, and distributed coding/encoding by extracting certain features of IoT data such as temporal and spatial correlations, and social relations. For example, the data collected from IoT sensor nodes can be compressed at the aggregator node before forwarding to the cloud, and also only certain features of the raw data can be extracted and sent to the cloud for some specific applications.

One of the main tasks of IoT sensors is to sense various parameters of the environment under the practical constraints such as low-cost, low-power and weak processing capability. In the proposed architecture, it is crucial to optimize edge-computing resources at the IoT gateway along with the communication resources such as transmit power, bandwidth and antennas in order to meet the desired performance metrics of the proposed system such as latency, data rate and energy efficiency. Also, in various IoT applications such as smart health and smart car, it is important to acquire and process contextual information such as location and speed to provide meaningful information. In this regard, the cloud centre can facilitate the optimization of edge-side processing under the proposed framework. For example, sampling rate to be employed at the edge-node depends on the characteristics of information (such as radio spectrum usage) to be acquired from the environment, and higher sampling rate causes more power consumption and also requires costly equipment. In this example, the cloud centre can provide guidance to the edge-nodes on the use of suitable sampling rate based on its network-wide intelligence as well as history information in order to reduce the power consumption. Besides, other transmission and operating parameters at the edge-nodes such as transmit power, modulation and coding can be adapted based on the feedback provided by the cloud-centre.

Another key challenge in the proposed framework is how to minimize the closed-loop (end-to-end) latency in order to process IoT big data in the real-time. In practice, the closed-loop latency of the considered system should be within the data coherence time, i.e., lifetime of the data. By employing suitable data prioritization techniques, the processing of the prioritized data can be handled at the edge computing units to provide faster response to the end-users. In addition, by employing caching techniques at the edge computing units in the proposed platform, content delivery efficiency can be maximized and fronthaul/backhaul bandwidth overhead can be significantly reduced as in 5G wireless networks [24]. During the off-peak hours, popular contents can be pre-fetched to the edge-side and during the peak periods, the delivery phase has to only deal with the transmission of additional contents requested from the users. The content popularity can be predicted using the available historical big data at the cloud side, and then a suitable cache replacement policy can be employed on the basis of the predicted content popularity distribution.

Based on the above discussion, we highlight some of the research directions under this topic below.

- **Energy-efficient caching strategies**: Various aspects such as time-varying popularity estimation, popularity modeling/prediction based on historical data, cache placement, delivery and replacement strategies, different performance tradeoffs such as data rate versus cache size (memory), latency versus memory and energy versus delay, cooperative and coded caching can be investigated in the proposed framework.
- **Edge-side data acquisition and processing techniques**: At the edge-side of the proposed platform, various cloud-assisted data processing techniques such

as data filtering, compression and feature extraction techniques, cooperative sensing/monitoring/acquisition, distributed encoding/decoding, combining and decision making schemes can be investigated.

- **Closed-loop latency minimization**: In order to minimize closed-loop latency in the proposed framework, suitable device/node/task/data prioritization techniques can be investigated under various practical constraints such as transmit power and backhaul/fronthaul bandwidth based on different criteria such as residual energy, requirement, emergency level, and expected delay.
- **Adaptive learning/prediction algorithms**: Several issues such as how to make the best use of available prediction algorithms (Neural network, artificial intelligence, support vector machine, clustering, regression analysis, etc.) in different application scenarios and how to adapt algorithms based on time-varying situations are promising future directions.

C. TASK/DATA/COMPUTATION/PROGRAM OFFLOADING

It is understood that huge amount of IoT data cannot be handled at the edge-side and need to be offloaded to the cloud-side. Besides, resources at the edge side such as transmit power, bandwidth and computing power need to allocated efficiently to handle real-time applications at the edge-side. Latency tolerant and large-scale tasks can be processed efficiently at the cloud centre while it is advantageous to process latency-sensitive tasks at the edge-side. The transfer of computational load/data/task from the edge-side to the cloud in the proposed collaborative edge-cloud framework can be enabled by employing suitable task/data/computation/program offloading techniques at the IoT edge gateway/aggregator. In this regard, it is crucial to take effective decisions about which data/task to offload to cloud and which data/task to be processed at the edge-side based on the guidance from the cloud center.

By using suitable data prioritization techniques, critical data (which needs real-time treatment) can be identified and processed at the edge-side to provide faster response to the end-users. The prioritized data can be later sent to the cloud for further processing and storage for future usage. In this regard, suitable data classification, data prioritization, data/task partitioning/scheduling, task/data offloading with backhaul constraints (limited bandwidth, delay, power consumption) and delay-based node prioritization need to be investigated under the proposed framework. Besides, since there is a huge amount of cost involved in renting cloudlets from the cloud providers, it is essential to investigate the tradeoff between service execution time and the cost required to use the cloud resources [32].

Based on the above discussion, we provide some research topics under this theme below.

- **Task scheduling techniques**: In the proposed framework, suitable task scheduling techniques need to be investigated for multiple dependent and interdependent tasks under the constraints of cloud/edge computing resource cost and service/workflow execution time deadline.
- **Data offloading techniques**: The investigation of suitable data/task offloading techniques under backhaul and edge-cloud resource constraints and decision making strategies on which data to offload to the cloud and which to process at the edge-side are important future research directions.
- **Code partitioning techniques**: In order to maximize the overall computational efficiency of the proposed framework, it is crucial to investigate which parts of the code to be run locally at the edge-side, and which parts should be offloaded to the cloud and at what state of the program considering practical constraints such as battery level, delay constraint and channel state.
- **Performance tradeoffs**: Various performance tradeoffs such as offloading gain versus energy, and cost of cloud resources versus service execution time under the proposed framework are interesting aspects to be investigated in future works.

D. ADAPTIVE OPTIMIZATION OF COMPUTING, COMMUNICATIONS AND CACHING RESOURCES

In contrast to the traditional way of managing computing and communication resources in a separate manner, future wireless IoT networks require novel solutions towards the adaptive optimization of computing, storage and communication resources at both the edge and cloud sides in order to deal with the massive amount of heterogeneous data. Besides, in the considered cloud-based IoT environment, the overall system dynamics vary due to different causes such as device movements, system parameter variations, and wireless channel variations, thus making the designed system unstable over the time. In this context, it is important to adapt the system model and decisions/control actions to be taken based on the global knowledge available at the cloud side and instantaneous information collected at the edge-side. Moreover, objective functions can be adapted based on the varying state of the system as well as instantaneous requirements from the end-users.

Besides, in order to facilitate real-time collaboration between the cloud computing and edge computing units, it is crucial to optimize the involved backhaul/feedback links under the transmission bandwidth constraint without compromising the quality of the links. In this direction, suitable techniques for designing low-latency flexible backhaul links under the constraints of end-user QoS requirements can be developed by utilizing the concepts of software defined wireless networks [33]. Moreover, the interactions between various layers of protocol stacks need to be considered in devising end-end system reliability. In this direction, various cross-layer techniques such as modulation and coding adaptation at the physical layer, collision free access mechanisms at the access layer and novel routing algorithms at the network layer can be investigated for the proposed architecture.

Based on the above discussion, we highlight some of the interesting research topics below.

- **Characterization of system resources**: In the proposed framework, it is important to identify all the involved communication, caching and computing resources, and then derive link/system capacity or any other suitable metrics by considering all these resources into account in order to provide the overall characterization of the system.
- **Adaptive optimization problems**: Some examples include latency minimization under the constraints of computational rate, transmit power minimization under the constraints of computational rate, precoding matrices/beamforming vector optimization under the constraints of computational rate, and the optimization of processing power under the constraints of latency, bandwidth and transmit power.
- **Performance tradeoffs**: In the considered framework, several performance tradeoffs such as caching gain and memory size/cache content size, computing and communication delays can be investigated.
- **Joint optimization of system resources**: Suitable joint optimization solutions can be investigated under the proposed framework in order to jointly optimize the caching (cache size), computing resources (processor speed, memory size, computational rate/power) and communication resources (bandwidth, transmit power, energy, latency etc.). Besides, suitable multi-objective optimization solutions [34] can be developed in order to address conflicting objectives in designing a wireless IoT network.

E. AUTONOMOUS DEVICE COLLABORATIONS IN WIRELESS IoT NETWORKS

Device to Device (D2D) communication has got importance in various practical scenarios and plays an important role in the proposed edge-cloud coordinated framework in Fig. 3. The increasing context-aware applications require location awareness and discovery functionalities and need to communicate with other neighboring nodes. Also, D2D plays an important role in enabling the sharing of resources such as contents, applications, processing power and spectrum among spatially-closed resource-constrained nodes in order to enhance the overall system performance. In addition, D2D communication is of vital importance to form emergency communication network in disaster scenarios such as earthquake or hurricane [35]. In such disaster scenarios, D2D communications has to take care of various aspects such as unpredictability, limited resources in disaster areas and dynamically varying environment [36].

In wireless IoT networks involving different types of things/objects such as home appliances, body area sensors, smart-phones and environmental monitoring sensors, a series of communication is required to realize the process of creating, collecting and sharing information among nearby devices [37]. Due to their heterogeneous capabilities, devices can help each other with the help of suitable collaborative strategies among them (for example, in a smart home, home audio, lighting bulbs, and alarming systems can collaborate to indicate the emergency level). Similar to human social network in connecting users via Internet, one device can be connected with other devices based on its social relationships with them, thus leading to the concept of social IoT [38]. By exploring the social relations of the devices, device collaboration can be initiated without human intervention.

The heterogeneity and the diversity of the connected smart devices impose challenges in handling interoperability among IoT devices. During the operation phase, there arise several challenges such as how to maintain seamless connections among the devices, how to attach a new device to the network, how to rediscover a device in the network [39], how to form a collaborative cluster, and which device to disconnect in case of faulty conditions and security threats. By grouping correlative devices under the same collaborative cluster, efficient management strategies can be employed at the IoT gateway in order to control them, and also to enable device collaborations. In this direction, the key issues to be considered are to understand the social relations among the devices [38], to classify devices based on a suitable basis, to track the location of the devices/inhabitants [40], and to design effective device collaborative strategies based on the inferred information with the aim of enhancing the zero-configuration index of device interaction.

In the following, we highlight some of the interesting research topics and issues under this theme.

- **Definition of social relationships**: Several aspects such as how to assign social relations among IoT devices and what kind of device social relations can be extracted from human social relations are important to be investigated under the considered joint edge-cloud collaborative platform.
- **Device attachment and discovery schemes**: It is essential to develop suitable policies to admit a new device to the existing IoT network and the ways to discover a device in the shortest possible path to minimize the latency as well as energy consumption.
- **Techniques for device classification and collaborative cluster formation**: Various research questions such as what is the best classification basis, how many devices should be grouped in a network, how to support a large number of heterogeneous group/cluster of devices, how to synchronize different clusters, and how to choose a cluster head dynamically in time varying channel conditions need to be addressed in the proposed framework.
- **Performance analysis**: Balancing the tradeoff between local device interactions and maintaining stable global system state in the considered distributed system [12] is one of the important aspects to be considered. Besides, suitable incentive-based device participation approaches can be investigated to enhance the participation of IoT devices in sharing their communication, computing and storage resources.

F. DATA-DRIVEN EVENT MONITORING AND EVENT-DRIVEN SERVICE PROVISIONING IN WIRELESS IoT NETWORKS

In the considered cloud-edge collaborative framework, various features of big data can be utilized in order to monitor the events which may be periodic or random in nature. Based on the global view of the network and historical data, cloud can implement suitable prediction techniques to forecast future events, and to provide this information to the edge computing units in order to better utilize the network resources. Moreover, in order to support device interactions in IoT networks with the minimal human intervention, energy efficiency, computational efficiency and spectral efficiency are important issues to be considered. One of the promising approaches to tackle these issues is event triggering, which enables communication/computing action to take place only when a particular event or a sequence of events happens [41]. With this approach, resource utilization efficiency can be significantly improved since communication circuits can be put into the sleep mode and the computing burden is reduced during the non-occurrence of events.

The first step in data-driven event monitoring is to define a suitable behavior model which captures the normal behavior of the considered IoT system. After a behavior model is identified, a suitable even detection technique can be employed. The even detection techniques can be broadly categorized into sample-based and stream-based [41]. The first category of techniques is based on individual samples to detect events while the second category relies on the flow of data samples.

Besides, the service provisioning of IoT is significantly different than from the traditional services which are mainly targeted for human-machine interactions and the difference mainly lies in the fact that IoT services need to deal with the seamless interactions with the real-world [42]. The heterogeneous IoT environment requires distributed processing of the large-scale sensing information collected by the IoT nodes, which need to combined and shared among different entities. In this regard, it is important to dynamically adapt the services based on the instantaneous changes in the physical environment.

We highlight some of the interesting research problems and directions under this research topic below.

- **Behavior modeling techniques**: Several aspects such as definition of suitable reference models, methods to cope up with unpredictable system characteristics, exploitation of differential states in the system state, and advanced learning methods can be studied under the proposed framework.
- **Event detection techniques**: Suitable sample and stream-based mechanisms, and statistical approaches such as eigenvalue distributions and PCA can be investigated for the considered framework.
- **Event handling methods**: In order to handle events in the proposed framework, various clustering techniques, event classifiers, data-driven methods such as regression, expectation maximization and linear support vector machines, inference methods such as fuzzy logic and heuristic reasoning, and data fusion techniques can be investigated [41].
- **On-demand information sharing**: Several aspects such as resource sharing models, interactions with application layer parameters, on-demand information dissemination models, environmental awareness and information fusion techniques require further research.
- **Adaptation and validation of event triggering**: In practice, the events can be spontaneous, and may have recurrent patterns and are of dynamic nature. How to adapt the operation of the network based on various states of the event such as early sign, and intermediate state is another key issue to be considered. After acquiring information about the event, how to validate the credibility of that information is another important challenge.

G. SOFTWARE DEFINED NETWORKING AND VIRTUALIZATION IN WIRELESS IoT NETWORKS

In the existing networks, the main problems are vendor-specific interfaces and software associated with hardware, complex and expensive network operation, and the tight coupling of data and control planes [26]. Besides, the network cannot dynamically adapt based on the network conditions. To address the aforementioned issues, the emerging concept of SDN [26], [43] can be employed. Besides, virtualization technology can be used to form a virtual network on the top of the existing networks, which shields the user from the underlying hardware and becomes adaptable to diverse technologies and protocols [44].

However, in contrast to the traditional virtualized wired networks, radio resource abstraction and isolation in wireless networks is challenging due to time-varying channel, broadcast nature, mobility and heterogeneous access technologies [27]. In heterogeneous IoT environment, dynamic resource discovery and the sharing of available resources can enable the creation of effective virtual networks. However, one of the critical challenges is the efficient management of the physical resources allocated to virtual networks. How often resource discovery and allocation need to be performed is also another challenge to be addressed. Besides, mapping of the physical resources to the logical resources in order to embed a virtual network on the existing physical networks in an effective way is another key aspect to be investigated.

The creation of a virtual network requires effective interactions among the involved entities at different hierarchical levels of the wireless IoT network [26]. This requires the need of defining suitable interfaces as well as proper control signalling which can be adaptable among heterogeneous wireless access technologies. For control signalling, mechanisms such as IP-based signalling and dedicated channel assignment can be investigated considering both network overload and delay. Besides, naming the devices with logical identifiers, mapping between physical and logical addresses, slicing, device attachment and dynamic routing of the device traffic are other important operations to be considered [26], [39]. Furthermore, another key research aspect is to examine the possibility of employing various levels of slicing such as spectrum-level slicing, network-level slicing and flow-level slicing in the wireless IoT networks. In addition, investigating suitable admission control policies in order to control the admission of incoming users/devices while guaranteeing the QoS of the existing users/devices is another aspect to be considered.

Based on the above discussion, some research directions under this topic are provided below.

- **Formation of virtualized edge-cloud collaborative network**: The design of a unified logical network in the proposed framework include various aspects such as network topology and interfaces definition, networks/devices discovery/routing, mapping between physical and logical resources, naming, assignment of unique logical identifiers to the devices, and device attachment strategies/admission control policies for future devices/nodes/users.
- **Wireless virtualization technologies**: It is an interesting future research direction to investigate suitable radio resource abstraction, isolation and slicing techniques under various practical constraints such as time-varying channel, mobility and heterogeneous access technologies in the proposed edge-cloud collaborative framework.
- **Adaptive configuration of system parameters**: Some important research challenges in the proposed framework are to dynamically configure a large number of system parameters such as carrier frequency, bandwidth, computing power, cache size and transmission power, to estimate the time-varying wireless channels, i.e., loads, and to dynamically characterize the communication links in terms of stability, delay, rate loss, and collision. It is interesting to investigate the application of SDN to overcome these challenges.
- **Control signalling design**: Another key challenge is how to design control signalling under the constraints of delay and communication bandwidth overhead. Several questions such as whether it is better to provide a dedicated radio channel or to employ a shared channel for control signalling, how to support the scalability of the devices, how to make signalling reliable in the case of device mobility and device failure during D2D communication need to be investigated in future works.

H. SECURITY/PRIVACY ENHANCEMENT IN WIRELESS IoT NETWORKS

Another main challenge is how to provide secured connections to the massive number of heterogeneous IoT devices having different levels of processing capabilities. During the processing of big data generated from massive IoT environment, the security threat may arise in any of the data processing phases including data acquisition, information filtering, data integration, representation, modeling, processing and interpretation. In our proposed framework, the cloud centre can guide the processing of edge nodes by utilizing the available history information as well as the network-wide intelligence towards achieving secured flow of information in the aforementioned phases.

The existing modular design practices for wireless communications are leading to vulnerable wireless IoT networks due to the transparent air interface and weak security protection mechanisms. Also, the decentralized characteristics of wireless IoT networks, long sleeping windows of IoT sensors, and the needs for collaborative communications to support D2D communications over heterogeneous networks pose significant challenges to the traditional wireless security provisioning techniques. For example, in IoT-based healthcare environment, the delegation of both storage and computation to the untrusted party can bring serious security and privacy issues and it is crucial to maintain the security and privacy of the every patient's data [45].

Since the cloud centre has a global view of the network as well as the history information, this knowledge can be exploited in taking decision on which nodes are unreliable and need to be disconnected from the network. In the IoT environment involving heterogeneous sensors having different levels of processing capabilities and also different mobility levels, existing cloud computing security mechanisms like sophisticated access control and encryption methods may not be sufficient to prevent illegitimate and unauthorized access to the data. In this context, there is a strong need to investigate suitable physical layer and cross-layer techniques such as dynamic link adaptation and adaptive medium access scheduling to enhance the data flow security in the proposed edge-cloud coordinated platform.

In contrast to the traditional scenarios, there is a high probability of privacy loss in IoT enabled systems due to several factors such as location-based services, increased interaction with smart devices, and less awareness on the user side. In this regard, there is a strong need to investigate privacy preserving mechanisms in IoT-enabled wireless networks [46]. Moreover, in the proposed cloud-edge collaborative platform, the distributed edge units may be more vulnerable to attacks since they do not have a global view of the network and have less resources to protect themselves. In this regard, cloud can assist edge-side units in implementing security measures in practical scenarios with the help of its global intelligence in identifying attacks/threats. Furthermore, in the proposed platform, edge computing units can act as controllers and aggregators for privacy sensitive data before sending data to the cloud, and they may also act as the proxies of resource-constrained IoT devices in order to handle their security functions [12].

Based on the above discussion, the following specific research directions can be considered under this theme.

- **Security enhancement with joint edge-cloud processing**: Several research questions such as how to perform cloud-side computations on the encrypted IoT data without revealing any privacy/secrets to the cloud service

providers, how to ensure that data is not corrupted in the edge processing units while in transit to the cloud centre, and how to offload tasks/programs from the edge to cloud in a secured manner, need to be addressed.
- **Access control and cooperative data aggregation schemes**: In the proposed platform, suitable cloud-assisted encryption techniques, energy-efficient data scrambling techniques, and cooperative data aggregation schemes can be investigated under the constraints of latency and communication bandwidth. In the cooperative schemes, various aspects such as how the nodes negotiate for a shared key, and what are different ways in which cloud can assist nodes in devising these cooperative policies under the constraints of communication bandwidth and computational rate need to be investigated.
- **Privacy preservation techniques**: In the proposed cloud-assisted platform, it is important to investigate suitable privacy preserving data clustering, differential privacy mechanisms and Pseudonymizing techniques in order to preserve the privacy of the end-users.
- **Trustworthiness of IoT systems**: Various aspects such as how to measure the trustworthiness of an IoT sensor/aggregator node, how to identify that the source of the data is a desired sensor device but not a robot or a malicious device, need to be investigated under the proposed framework.

VI. CONCLUSIONS

Cloud computing and edge computing are considered as two emerging paradigms in handling the massive amount of distributed data generated by IoT devices. However, these paradigms have their own advantages and disadvantages. Cloud computing provides a centralized pool of storage and computing resources and has a global view of the network but it is not suitable for applications demanding low latency, real-time operation and high QoS. On the other hand, edge computing is suitable for the applications which need real-time treatment, mobility support, and location/context awareness but does not usually have sufficient computing and storage resources. Taking these aspects into consideration, this paper has proposed a novel framework of collaborative edge-cloud processing for enabling live data analytics in wireless IoT networks. The basic features, key enablers and the challenges of big data analytics in wireless IoT networks have been described and the main distinctions between cloud and edge processing have been presented. Furthermore, potential key enablers for the proposed collaborative edge-cloud computing framework have been identified and the associated key challenges have been presented in order to foster future research activities in this domain.

Finally, it is worthy to mention that the proposed edge-cloud collaborative framework can be exploited as an important platform for wireless networks to achieve various objectives such as dynamic spectrum management, energy-efficient caching and offloading, closed-loop latency minimization, adaptive optimization of computing, communication and caching resources, even-driven resource allocation and security/privacy enhancement.

REFERENCES

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.
[2] S. K. Sharma, T. E. Bogale, S. Chatzinotas, X. Wang, and L. B. Le, "Physical layer aspects of wireless IoT," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Sep. 2016, pp. 304–308.
[3] P. Fan, "Coping with the big data: Convergence of communications, computing and storage," *China Commun.*, vol. 13, no. 9, pp. 203–207, Sep. 2016.
[4] H. Liu, Z. Chen, and L. Qian, "The three primary colors of mobile systems," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 15–21, Sep. 2016.
[5] S. Andreev *et al.*, "Exploring synergy between communications, caching, and computing in 5G-grade deployments," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 60–69, Aug. 2016.
[6] J. Tang and T. Q. S. Quek, "The role of cloud computing in content-centric mobile networking," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 52–59, Aug. 2016.
[7] P. Corcoran and S. K. Datta, "Mobile-edge computing and the Internet of Things for consumers: Extending cloud computing and services to the edge of the network," *IEEE Consum. Electron. Mag.*, vol. 5, no. 4, pp. 73–74, Oct. 2016.
[8] X. Masip-Bruin, E. Marn-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, Oct. 2016.
[9] 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, Oct. 2016.
[10] M. Satyanarayanan, "The emergence of edge computing," *Computer*, vol. 50, no. 1, pp. 30–39, Jan. 2017.
[11] S. H. Park, O. Simeone, and S. Shamai (Shitz), "Joint optimization of cloud and edge processing for fog radio access networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7621–7632, Nov. 2016.
[12] M. Chiang and T. Zhang, "Fog and IoT: An overview of research opportunities," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 854–864, Dec. 2016.
[13] S. Yin and O. Kaynak, "Big data for modern industry: Challenges and trends [point of view]," *Proc. IEEE*, vol. 103, no. 2, pp. 143–146, Feb. 2015.
[14] H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward scalable systems for big data analytics: A technology tutorial," *IEEE Access*, vol. 2, pp. 652–687, Jul. 2014.
[15] S. Bi, R. Zhang, Z. Ding, and S. Cui, "Wireless communications in the era of big data," *IEEE Commun. Mag.*, vol. 53, no. 10, pp. 190–199, Oct. 2015.
[16] 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.
[17] 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, Feb. 2017.
[18] 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, May 2016.
[19] J. A. Stankovic, "Research directions for the Internet of Things," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 3–9, Feb. 2014.
[20] V. Cevher, S. Becker, and M. Schmidt, "Convex optimization for big data: Scalable, randomized, and parallel algorithms for big data analytics," *IEEE Trans. Signal Process.*, vol. 31, no. 5, pp. 32–43, Sep. 2014.
[21] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[22] S. K. Sharma, S. Chatzinotas, and B. Ottersten, "SNR estimation for multi-dimensional cognitive receiver under correlated channel/noise," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6392–6405, Dec. 2013.
[23] C. Zhang and R. C. Qiu, "Massive MIMO as a big data system: Random matrix models and testbed," *IEEE Access*, vol. 3, no. 4, pp. 837–851, 2015.
[24] E. Zeydan *et al.*, "Big data caching for networking: Moving from cloud to edge," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 36–42, Sep. 2016.

[25] D. Mazza, D. Tarchi, and G. E. Corazza, "A cluster based computation offloading technique for mobile cloud computing in smart cities," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.
[26] D. Kreutz, F. Ramos, P. E. Veríssimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proc. IEEE*, vol. 103, no. 1, pp. 14–76, Jan. 2015.
[27] C. Liang and F. R. Yu, "Wireless network virtualization: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 358–380, Mar. 2015.
[28] 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, Dec. 2016.
[29] S. Prathibha, B. Latha, and G. Sumathi, "Improving energy efficiency of computing servers and communication fabric in cloud data centers," in *Proc. Int. Conf. Res. Comput. Intell. Commun. Netw. (ICRCICN)*, Sep. 2016, pp. 17–21.
[30] I. Park, D. Kim, and D. Har, "MAC achieving low latency and energy efficiency in hierarchical m2m networks with clustered nodes," *IEEE Sensors J.*, vol. 15, no. 3, pp. 1657–1661, Mar. 2015.
[31] S. Vural, P. Navaratnam, N. Wang, C. Wang, L. Dong, and R. Tafazolli, "In-network caching of Internet-of-Things data," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 3185–3190.
[32] X.-Q. Pham and E.-N. Huh, "Towards task scheduling in a cloud-fog computing system," in *Proc. Asia–Pacific Netw. Oper. Manage. Symp. (APNOMS)*, Oct. 2016, pp. 1–4.
[33] J. Zhang, X. Zhang, and W. Wang, "Cache-enabled software defined heterogeneous networks for green and flexible 5G networks," *IEEE Access*, vol. 4, pp. 3591–3604, 2016.
[34] E. Bjornson, E. A. Jorswieck, M. Debbah, and B. Ottersten, "Multiobjective signal processing optimization: The way to balance conflicting metrics in 5G systems," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 14–23, Nov. 2014.
[35] M. N. Tehrani, M. Uysal, and H. Yanikomeroglu, "Device-to-device communication in 5G cellular networks: Challenges, solutions, and future directions," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 86–92, May 2014.
[36] J. Wang, Y. Wu, N. Yen, S. Guo, and Z. Cheng, "Big data analytics for emergency communication networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1758–1778, 3rd Quart., 2016.
[37] O. Bello and S. Zeadally, "Intelligent device-to-device communication in the Internet of Things," *IEEE Syst. J.*, vol. 10, no. 3, pp. 1172–1182, Sep. 2016.
[38] D. O. Kang, J. H. Choi, J. Y. Jung, K. Kang, and C. Bae, "SDIF: Social device interaction framework for encounter and play in smart home service," *IEEE Trans. Consum. Electron.*, vol. 62, no. 1, pp. 85–93, Feb. 2016.
[39] A. Papageorgiou, R. Bifulco, E. Kovacs, and H. J. Kolbe, "Dynamic M2M device attachment and redirection in virtual home gateway environments," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.
[40] M. Danancher, J. J. Lesage, and L. Litz, "Model-based location tracking of an a priori unknown number of inhabitants in smart homes," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 2, pp. 1090–1101, Apr. 2016.
[41] P. Kolios, C. Panayiotou, G. Ellinas, and M. Polycarpou, "Data-driven event triggering for IoT applications," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1146–1158, Dec. 2016.
[42] S. Zhao, L. Yu, and B. Cheng, "An event-driven service provisioning mechanism for IoT (Internet of Things) system interaction," *IEEE Access*, vol. 4, pp. 5038–5051, 2016.
[43] A. M. Akhtar, X. Wang, and L. Hanzo, "Synergistic spectrum sharing in 5G HetNets: A harmonized SDN-enabled approach," *IEEE Commun. Mag.*, vol. 54, no. 1, pp. 40–47, Jan. 2016.
[44] N. Bizanis and F. A. Kuipers, "SDN and virtualization solutions for the Internet of Things: A survey," *IEEE Access*, vol. 4, pp. 5591–5606, 2016.
[45] J. Zhou, Z. Cao, X. Dong, and X. Lin, "Security and privacy in cloud-assisted wireless wearable communications: Challenges, solutions, and future directions," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 136–144, Apr. 2015.
[46] P. Porambage, M. Ylianttila, C. Schmitt, P. Kumar, A. Gurtov, and A. V. Vasilakos, "The quest for privacy in the Internet of Things," *IEEE Cloud Comput.*, vol. 3, no. 2, pp. 36–45, Mar. 2016.

SHREE KRISHNA SHARMA (S'12–M'15) received the M.Sc. degree in information and communication engineering from the Institute of Engineering at Pulchowk, Nepal, the M.A. degree in economics from Tribhuvan University, Nepal, the M.Res. degree in computing science from Staffordshire University, Staffordshire, U.K., and the Ph.D. degree in wireless communications from the University of Luxembourg, Luxembourg, in 2014. He was a Research Associate with the Interdisciplinary Center for Security, Reliability and Trust, University of Luxembourg for two years, where he was involved in the EU FP7 CoRaSat Project, EU H2020 SANSA, ESA Project ASPIM, Luxembourgish National Projects Co2Sat, and SeMIGod. He is currently a Post-Doctoral Fellow with Western University, Canada. His research interests include 5G and beyond wireless systems, Internet of Things , adaptive optimization of distributed communication, computing and caching resources, cognitive and cooperative communications, and interference mitigation and resource allocation in heterogeneous wireless networks.

He was with Kathmandu University, Dhulikhel, Nepal, as a Teaching Assistant, and a Part-Time Lecturer with eight engineering colleges in Nepal. He was with Nepal Telecom for over four years as a Telecom Engineer in the field of information technology and telecommunication. He has authored over 70 technical papers in refereed international journals, scientific books, and conferences. He has been serving as a TPC Member of a number of international conferences including the IEEE ICC, the IEEE PIMRC, the IEEE Globecom, the IEEE ISWCS, and the CROWNCOM. He received an Indian Embassy Scholarship for the B.E. degree, an Erasmus Mundus Scholarship for the M.Res. degree, and an AFR Ph.D. Grant from the National Research Fund (FNR) of Luxembourg. He received the Best Paper Award at the CROWNCOM 2015 conference. His Ph.D. thesis received the FNR Award for Outstanding Ph.D. Thesis 2015 from FNR, Luxembourg. He has been serving as a Reviewer for several international journals and conferences.

XIANBIN WANG (S'98–M'99–SM'06–F'17) received the Ph.D. degree in electrical and computer engineering from the National University of Singapore in 2001. He is currently a Professor and the Canada Research Chair with Western University, Canada.

From 2001 to 2002, he was a System Designer with STMicroelectronics, where he was responsible for the system design of DSL and Gigabit Ethernet chipsets. He was with the Communications Research Center Canada (CRC) as a Research Scientist/Senior Research Scientist from 2002 and 2007. He has authored over 280 Peer-Reviewed journal and conference papers, 26 Granted and pending patents, and several standard contributions. His current research interests include 5G technologies, signal processing for communications, adaptive wireless systems, communications security, and locationing technologies.

Dr. Wang is an IEEE Distinguished Lecturer. He has received many awards and recognition, including the Canada Research Chair, the CRC President's Excellence Award, the Canadian Federal Government Public Service Award, the Ontario Early Researcher Award, and five IEEE Best Paper Awards. He was involved in a number of IEEE conferences including the GLOBECOM, the ICC, the VTC, the PIMRC, the WCNC, and the CWIT, in different roles such as the Symposium Chair, the Tutorial Instructor, the Track Chair, the Session Chair, and the TPC Co-Chair. He was an Associate Editor of the IEEE Transactions on Wireless Communications from 2007 to 2011, and the IEEE Wireless Communications Letters from 2011 to 2016. He currently serves as an Editor/Associate Editor of the IEEE Transactions on Communications, the IEEE Transactions on Broadcasting, and the IEEE Transactions on Vehicular Technology.