Journal of Network and Computer Applications 128 (2019) 90–104

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Fog data analytics: A taxonomy and process model

Aparna Kumari [a], Sudeep Tanwar [a], Sudhanshu Tyagi [b], Neeraj Kumar [b], Reza M. Parizi [c], Kim-Kwang Raymond Choo [d,*]

[a] Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, Gujarat, India
[b] Thapar Institute of Engineering and Technology, Deemed to be University, Patiala, Punjab, India
[c] Department of Software Engineering and Game Development, Kennesaw State University, GA, USA
[d] Department of Information Systems and Cyber Security and Department of Electrical and Computer Engineering, University of Texas at San Antonio, San Antonio, TX 78249, USA

ARTICLE INFO

Keywords:
Big data
Fog data analytics
Data collection
Data storage
Data reduction
Data security and privacy

ABSTRACT

Through the exponential growth of sensors and smart gadgets (collectively referred to as smart devices or Internet of Things (IoT) devices), significant amount of heterogeneous and multi-modal data, termed as Big Data (BD), is being generated. To deal with such BD, we require efficient and effective solutions such as data mining, analytics, and reduction to be deployed at the edge of fog devices on a cloud. Existing research and development efforts generally focus on performing BD analytics overlook the difficulty of facilitating fog data analytics (FDA). In this paper, we discuss the unique nature and complexity of fog data analytics. A detailed taxonomy for FDA is abstracted into a process model. The proposed model addresses various research challenges, such as accessibility, scalability, fog nodes communication, nodal collaboration, heterogeneity, reliability, and quality of service (QoS) requirements. To demonstrate the proposed process model, we present two case studies.

1. Introduction

With the increasing demand of Internet of Things (IoT) and the Internet of Everything (IoE) technologies, big data analytics has attracted the attention of researchers and practitioners. The sensors, physical devices, etc, are referred to as things in IoT, which can be sensing the physical environment are the building blocks. These devices are connected through various protocols without minimal human interaction and involvement. The snowballing use of sensors and IoT devices in cloud-based applications resulted in various big data (BD) challenges. Cloud computing (CC) can be leveraged to help overcome some of the challenges and issues associated to BD storage and analysis. This growth can be categorized by high-velocity, high-volume, and high-variety data that demand efficient and intelligent processing for decision making and analysis (Tang et al., 2015). Currently, CC is widely used in industry with pay-as-you-go model. It addresses BD challenges, such as high latency and cost, due to its scalability and data distribution management methodology. Nevertheless, cloud data centers constantly face challenges due to explosion of BD and additional requirements of low latency and locality awareness required at the edge of cloud network. Fog computing (FC), a concept coined by Cisco, runs on geo-distributed applications over the cloud (Bonomi et al., 2012) to handle the aforementioned issues. Specifically, FC handles latency-sensitive applications at the edge of a network (Roman et al., 2018). Moreover, it provides latency-tolerant tasks very efficiently with the help of influential fog nodes (FNs) at the intermediary of the cloud infrastructure. A cloud data center is still used for in-depth data analytics in the network.

A significant amount of data is generated by sensors, gadgets, temperature sensors, online networking, and healthcare applications. These devices persistently produce a huge volume of data, which generates BD (Kambatla et al., 2014), which can be structured, semi-structured, and/or unstructured. Traditional database systems are not sufficiently efficient to store, process, and analyze such rapidly developing data. According to McKinsey global institute, BD requires better tools for storing, processing and analyzing such voluminous amount of data (Manyika et al., 2011). The term BD has been used in IT sectors, academia and the research industry (Keller, 2011). Digital Universe has labelled BD technologies as “new generation-technologies”, which aim to extract value from a huge volume of different types of data. Business intelligence (BI) is a type of analytics required when the data size is

* Corresponding author.
E-mail addresses: 17ftphde22@nirmauni.ac.in (A. Kumari), sudeep149@rediffmail.com (S. Tanwar), s.tyagi@thapar.edu (S. Tyagi), neeraj.kumar@thapar.edu (N. Kumar), rparizi1@kennesaw.edu (R.M. Parizi), raymond.choo@fulbrightmail.org (K.-K.R. Choo).

https://doi.org/10.1016/j.jnca.2018.12.013
Received 2 September 2018; Received in revised form 18 November 2018; Accepted 25 December 2018
Available online 29 December 2018
1084-8045/

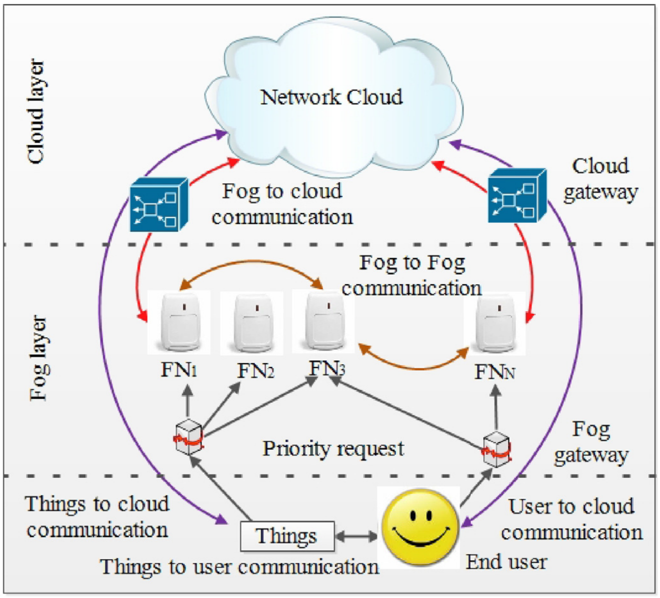

Fig. 1. Basic architecture of FDA communication.

more than the available memory. In such scenario, data can be moved from data generation sources to a BI analytics environment (Chen and Zhang, 2014). These analytics currently support tera bytes of data level (Amin et al., 2017), and additionally helps to determine strategic business prospects from the torrent of data. A typical FDA life cycle includes a collection of BD, storage, distribution and visualization as presented in Fig. 1. Here, data collected from various IoT devices is stored at the edge devices on a cloud. After performing the required analytics on the collected data, relevant information is sent to the cloud. In Fig. 1, the bottom (physical) layer contains IoT/sensor devices and users that are linked to the cloud through the communication channel. Data extracted from these devices are forwarded to the next layer, i.e., fog layer (FL). IoT devices communicate to FNs using fog gateway. FL combine various FNs and handle the priority requests. Then, fog to cloud communication takes place via a cloud gateway where the upper most layer is known as a cloud layer (CL).

Fog computing technology is used for applications that require low latency, and works based on geo-distribution, which is quick and portable, and has an extensive scale distributed control system. These requirements could not get adequate results from the standalone CC paradigm. Hence, for such applications a CC model is used along with fog computing to get a satisfactory outcome. The primary standpoint of fog was to get the benefits of fast service and computation to enhance the QoS, decrease latency and to place the information near the user. The favorable position and utilization of fog would not completely be disregarded by a cloud, but rather, they would work together to achieve better results, especially for delay sensitive applications. The gadgets which are geographically distributed would generate exchange information with the fog devices, while hierarchical organization of the computation, network, and storage are preserved (Bonomi et al., 2014). One of the key applications of FC is in the healthcare domain, where wearable sensors allow 24/7 collection of data to monitor human health conditions. These wearable sensors are especially designed medical sensors placed on the human body to permit non invasive observations (Cerina et al., 2017; Farahani et al., 2018). Another major application is the development of smart cities, the construction of which poses many challenges, such as the development of accurate, real-time, and geo-spatially scattered large-scale sensing networks to monitor the robustness of infrastructure components, which can include gas pipelines, bridges, oil pipelines, water pipelines, subways, and roads (Farahani et al., 2018; He et al., 2018).

1.1. Motivation

Currently, several review research proposals exist in the literature (Mohammadi et al., 2018; Mahmud et al., 2018; Yi et al., 2015; Chen and Zhang, 2014), which mainly address the big data analytics issues within fog computing paradigm. However, data collection, data storage, processing and analysis, security and privacy, and QoE/QoS aspects of the fog data analytics have not been fully discussed or explored to the full extent in prior study. Thus, in this paper, we will survey the existing literature pertaining to FDA and their foci as follows:

- Develop a taxonomy abstracted into an extended process model that covers all possible aspects of FDA.
- Provide a comparative analysis of several existing approaches on FDA with respect to different performance evaluation parameters.
- Develop an FDA process model and test it with case studies.

1.2. Organization

The remainder of the paper is organized as follows: First, FDA computing and different characteristics of FDA are presented in the next section. A novel taxonomy and process model for FDA is discussed in Section 3. In Section 4, two case studies will be used to test the process model. Research challenges in FDA computing are presented in Section 5, and finally Section 6 presents the conclusions of the research.

2. Background

Uniqueness of the process design is described here, as well as research challenges for the advanced process model for fog data analytics. The flow of computational load in terms of data analytics from cloud to fog is found to be similar to cloud computing, the only variation of which is inclusion of edge devices. This approach defines different characteristics of FDA and enables the development of wide variety of IoT-based applications and services as listed below:

Heterogeneity: Heterogeneity in FDA comprises of hierarchical building blocks in a distributed fashion. FC structure provides the data storage, computation and networking services between core cloud and end-devices in a virtualized way.

Cognition: Client-centric objectives are defined as cognition in this case. Data access and analytics of fog provides better understanding of user requirements, and also provides the best position to store, control, and transmit the required data throughout cloud to end-IoT-devices. Fog-based applications provide better responses and due to their closeness to end-devices, reproduces the customers' requirements more efficiently (Chiang and Zhang, 2016).

Geo-graphical Environment Distribution: Fog computing provides QoS for both dynamic and static IoT-devices. Its network contains geographically distributed nodes and sensors in a different environment such as healthcare monitoring, temperature monitoring in smart homes, and weather monitoring sensors.

Real-Time Interaction: Fog data analytics has the capability to work in real-time and perform the appropriate interaction to provide better QoS, such as real-time transmission of traffic in ITS, electricity distribution and monitoring system applications in smart grid.

Edge Location with Low Latency: Recent development of applications for smart devices has been found to be unsatisfactory due to the absence of proximity among the devices. To provide QoS at the edge of the network, applications with low latency services have been required, such as live gaming applications and video streaming TV services (More, 2015).

Table 1
Features of FDA with merits and demerits.

Features of FDA	Merits	Demerits
Network bandwidth	Improved	WAN security issues
Cost-to-Cloud	Minimizes	Additional cost for fog hardware
Latency at cloud	Minimizes	Cost/Availability of hardwares
Data flow-to-cloud	Sent only relevant data	Data benefit of any time and anywhere concept of cloud reduces
System response time	Improved	Authentication and trust issue arises
Mobility support of smart devices	Systematic	security issues with IP bluffing
Security	Improved	Privacy issue arises due to distributed data-processing
Privacy	Improved	New privacy issue arises due to distributed processing
Time-to-access data	Improved	–
Context	User, data, application context improved	–

Interoperability: Fog devices work in inter-operating environment to support wide range of services such as real-time analytics, predictive decisions, and data streaming.

These highlighted characteristics enable business models & services to increase revenues, reduce cost and speed up the product roll-outs in the industry. On the other hand, few characteristics of FDA supporting better communication and services as the support for mobility, large-scale sensor network and wide spread wireless access (Anawar et al., 2018). In personal data, sharing policies for FDA was given and claimed that the methodology was best suited for real-world scenarios (Gosman et al., 2018). A hybrid model for accurate disease prediction was proposed by Liu et al. (2018). Suggested protocol was used to calculate the non-linear activation function in a neural network. In addition to this, the protocol can also support a computation overflow problem. Smart contracts and blockchain are the most prevalent research areas (Parizi and DehghantanhaAmritraj, 2018; Parizi et al., 2018) in which FDA can play a key role. Choo et al. (2018) have proposed a novel kernel function with intrinsic mappings aimed for BD analysis. Table 1 provides the detailed description of features of FDA with respective advantages and disadvantages. To clarify the main deference between the work performed in this paper and other FDA-specific surveys, a comprehensive comparison including objectives, merits, and demerits of peer surveys are given in Table 2.

3. Taxonomy of fog data analytics communication

A master taxonomy of FDA communication is summarized in Fig. 2, that includes existing approaches available for data collection and storage, data processing and analytics, connectivity at FN, data distribution, QoE/QoS, and security and privacy. Fig. 3 represents a novel process model for FDA communication, in which multi-modal heterogeneous data are collected from various IoT devices. Pre-processed data are forwarded for real-time analysis, and are ultimately processed or stored as per their requirements. Subsequently, the data intelligence on processed data takes place. Once intelligence and useful information are retrieved, data will be transferred to the cloud or other FNs for further consolidated analytics. Subsections which follow present the detailed discussion on individual components of master taxonomy.

3.1. Data collection and data storage

This section has included the data collection and storage mechanisms for the FDA system.

3.1.1. Data collection

The advent of the Internet, IoT, and social networks has provided a big rise in collection of massive real-world data. Previously there was a social sensing system, where users acted as a source and provided the information at run-time (Ali et al., 2011). These social sensors provided the redundant data and needed a faithful preprocessing. Following are the popular worldwide components for the data collection purpose:

Internet of Things: This is a network of computing devices, home appliances, digital apparatus, software, sensors, actuators and items embedded with bio chips. The connectivity between these devices help to communicate and exchange information. The internet of things has a unique internet protocol (IP) address to establish a communication link with other devices, and does not require interaction as used in human-to-human (H2H), human-to-computer (H2C) or computer-to-human communication (C2H). The IoT has opened a new door in the field of communication with a combination of the physical ecosphere with a computer, and enhances accurateness and proficiency as it decreases human intervention. Smart grids, smart cities, smart homes, virtual power plants, and intelligent transportation also incorporate different applications (Darwish and Bakar, 2018) as every element is uniquely identifiable within network and IoT delivers the global connectivity of everything. Alvi et al. (2015) have proposed a new terminology for IoT communication known as the internet of multimedia things (IoMT) for multimedia objects. Additionally, the emergent technologies for interrelated things of IoT and sensors has played a major role in the healthcare domain and provides improved patient care service.

Sensor devices: The development of social media has made possible for the users to act as a sensor in the real world. The social media websites are generating 24 by 7 users data using an online data stream. Some of the examples of these social media are YouTube, Facebook, twitter etc. There is an approach proposed by Ali et al. (2011) for social-sensing, where users act as monitors and help to provide real-time data. Data generated from sensor devices can be classified into weather, vibration, voice, current, pressure, vehicles, temperature etc. These data are mostly transferred through wireless networks or LAN for collection, storage, and processing. An application such as video surveillance system is easily installed and managed over a LAN network to gather real-time data. Sometimes efficient data transmission between sensors is possible using wireless communication with the limitation of communication capabilities and energy. Raafat et al. (2017) have proposed a knowledge extraction approach which helps to reduce the data transfer with the Fog to Cloud. They used statistical and homoscedasticity features extraction approach which helps to extract the essential events in real-time sensor data. The next section will focus on FDA data storage.

3.1.2. Data storage

Data storage is categorized into three categories, named as clustering, indexing and replication. In clustering approach, a bunch of the data are collected and stored in fog storage devices (Bonomi et al., 2012). In indexing approach, the collected data have indexation for fast retrieval and access. Lastly, replication is an approach where the same data are duplicated to the other machine for fault tolerance. A real-time indexing approach for complex data was proposed by Chmelar et al. (2015). This approach deals with the novel indexing framework termed as ReTIn (Real-Time Indexing). In their work, the objective was to allow indexing of complex data coming as a stream of soft real-time restrictions. Idea of ReTIn was essentially to combine sequential access to the utmost recent data and an index-based access to old database.

Table 2
Comparison of the existing surveys with the proposed survey.

Year	Authors	Objective(s)	Merits	Demerits
2017	Hu et al. (2017)	To explore the hierarchical architecture and the characteristics of Fog computing	Detailed comparison between Fog and Cloud computing	Role of data analytics on fog layer was not covered
2017	Subbu and Vasilakos (2017)	To investigate the context-aware computing and big data applications	Detailed taxonomy of context aware computing	Security and privacy was completely missing
2018	Dias et al. (de AssunÃ§Ão et al., 2018)	To explore the stream processing solutions for deployment of data stream processing on cloud and edge computing	Data stream processing was explored in details	Issues like, QoE, QoS, security and privacy were not explored
2018	Zhang et al. (2018)	To investigate security and trust issues for fog computing	Fog architecture, it's implementation, and layer-wise gateway was covered in details	Discussion was limited to the security and trust, data analytics was completely missing
2018	Luiz et al. (Bittencourt et al., 2018)	To explore different aspects for the integration of fog and cloud computing to be suitable for several applications using IoT	Layered common architecture for fog-IoT-cloud was given	Only data collection was covered at the edge of a network
2018	Kumari et al. (2018a)	To explore the different components on multimedia big data	Detailed taxonomy and case study was given on multimedia big data	Analytics on fog layer was not given in details
2018	Naha et al. (2018)	To explore the resource management schemes of fog computing	Authors have discussed several definitions of fog computing and suggested their definition too	Data analytics at the edge was not covered
2018	Yeow et al. (2018)	To present state-of-the-art for decentralized consensus for edge-centric IoT	Detailed taxonomy	–
2019	Fei et al. (2019)	To investigate the data streams analytics for cyber physical system using machine learning approach	Machine learning based applications were covered	Most of the data analytics oriented parameters were missing for fog computing
–	Proposed review	To explore the data analytics for fog computing	Various parameters and components of FDA are covered	–

For provisioning and maintaining of resources like hardware virtualization, storage space, utility computing and high capacity networks etc, various research works have been proposed. Verma et al. (2016) have presented a Cloud-fog duo approach for load balancing and data replication for FC infrastructure. This approach helps to manage BD over closer FNs.

Yuan et al. (2012) have suggested a solution for identifying and storing frequently required intermediate datasets through data dependency. In this solution, the storage of frequently approached datasets is located using intermediate data dependency graph (IDG), which is meant to be cost-effective. Given the IDG of four datasets shown in Fig. 4, nodes are interconnected like if n_1 is pointing to n_2, indicating n_2 is derived from n_1, and n_2 is pointing to n_3 and n_4 means n_2 was used to generate either n_3 or n_4 based on the operations. IDG was specifically used by their system to find predecessor datasets of the data in demand to know how to regenerate any intermediate dataset without processing the entire dataset. In this case, the total cost of computation is computed for every dataset, which is the sum of storage cost and computation cost for regeneration, where the ultimate objective was to minimize this quantity. Intermediate data storage strategy based on the dependency graph is an algorithm that decides whether a set must be stored or not. This algorithm guarantees that the intermediate dataset it chooses to store only necessary data, while deleting the rest that would upsurge the total system's cost. Since the storage status of any set changes dynamically, the algorithm must be dynamic in nature. The algorithm also considers a user's tolerance and takes the decision of data storage into the dataset. Similarly, Hosseinpour et al. (2016) developed a data storage concept known as smart data for hierarchical FC system. Their proposed work restructures the raw BD to self-manage and intelligent data cells to keep only the meaningful information with a compact size. After collecting and storing data, there will be a requirement of processing data for analysis purposes. Next section focuses on the data processing and analysis.

3.2. Data processing and analysis

This is one of the major phases where data are communicated in FDA network. The data that have been collected from sensor devices, get filtered on the basis of specific data processing and analysis on FL, which helps to forward only relevant information to CL.

3.2.1. Data processing

Fogging is a distributed infrastructure of communication where couple of services are managed remotely and other are managed at the edge of a network. FC has aimed to achieve high efficiency and reduces the volume of data, which need to be sent to CC for storage, processing and further analysis. In addition, FC can also be used to achieve higher data security and privacy. In a FC environment, data processing takes place in a smart device, using smart gateway or router devices at the edge of the network.

In earlier approaches (those without FC), sensors and devices typically generate gigantic volumes of data and send them all to the cloud for analysis and processing. This is not an efficient approach due to the requirement of enormous bandwidth that affects the cloud performance, and increases the latency and maintenance cost. Within FC, data processing, analysis and decision-making-action all occur simultaneously through IoT smart devices. Once data are processed and analyzed at FL, only relevant data are pushed to the cloud for further analytics. IoT verticals or smart devices, the orchestration layer, and the abstraction layer are the three basic components used in any FC environment (Storey and Song, 2017):

IoT verticals comprise of occupant applications or products which are lent for use. They support multi-tenancy to enable multiple clients to host the application on a single FC server, which provides flexibility and interoperability in FC.

Orchestration layer is used for data aggregation, data migration, data

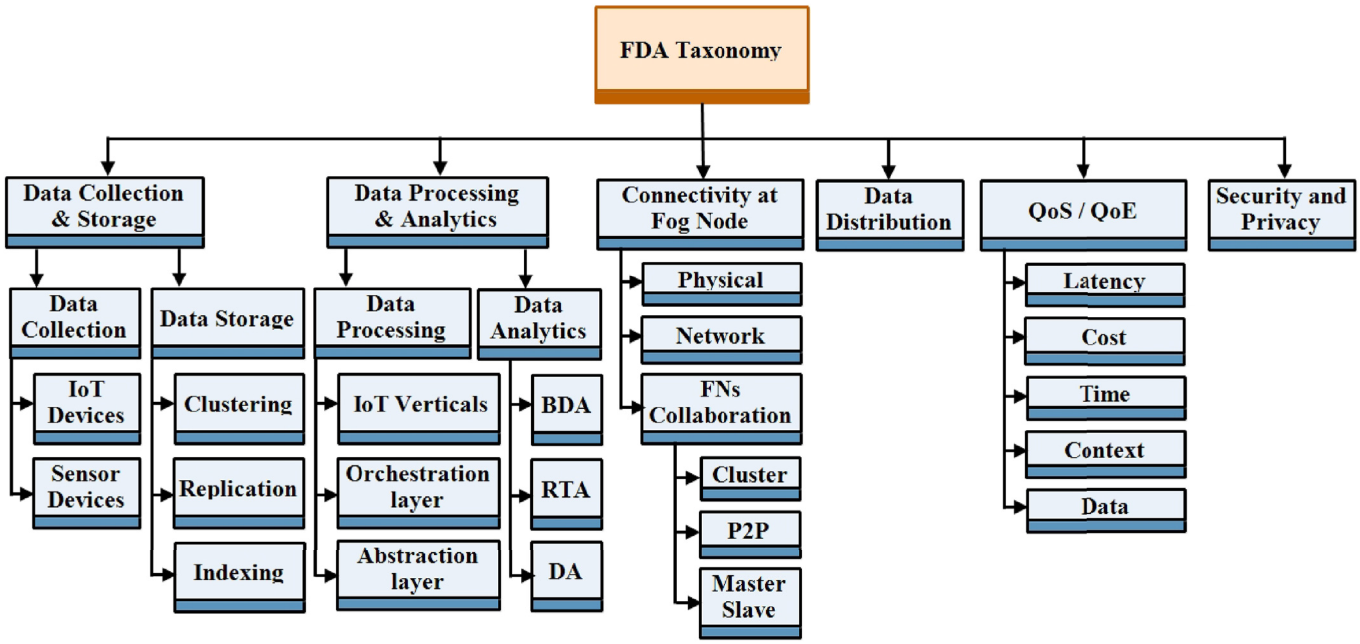

Fig. 2. Main Taxonomy of FDA communication.

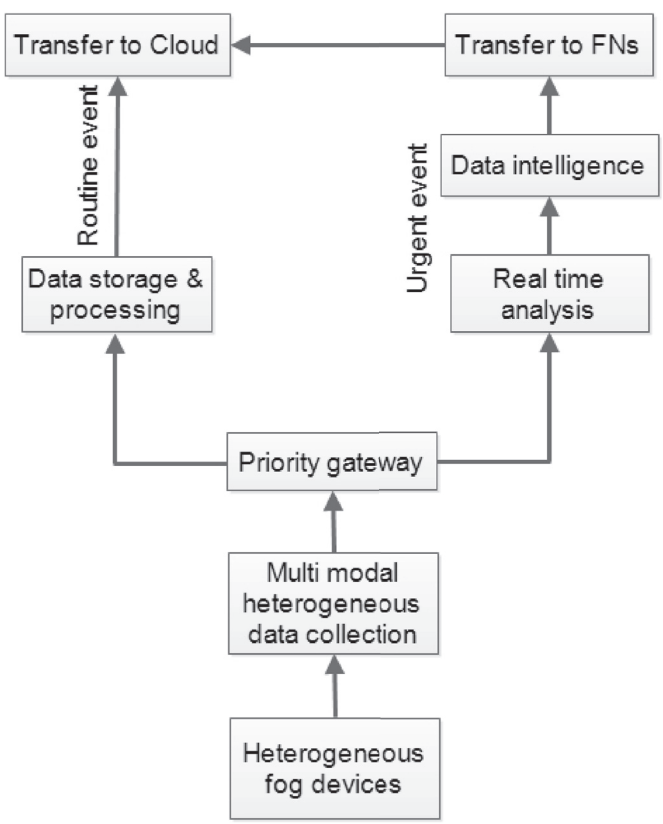

Fig. 3. Process Model for FDA communication.

sharing, decision making, and policy management for FC. It contains orchestration layer API and data API. The orchestration API provides intelligence services, policy management and analytics for FC. It also simplifies the secure communication and facilitates the sharing in a distributed network.

Abstraction layer delivers a uniform interface to the client. Similar to a CC model, it uses virtualization technologies and provides generic APIs. These APIs can be applied while creating the applications that need to be hosted on FC for clients.

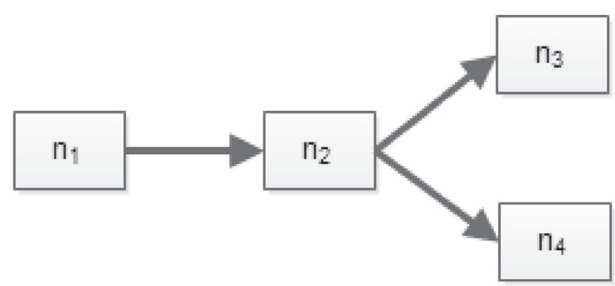

Fig. 4. IDG containing four datasets.

3.2.2. Data analysis

The services of IoT are comprised of three major functions, namely: sensing, collection, and analysis of the data. Today, device automation is required for the real time analysis of data. BD analysis has been in demands as a result of massive data generation and collection from various IoT devices, which require to be stored into a secure database for further analysis. More particularly, real-time analysis and processing are required to analyze BD with high data rates, while deep BD analysis is required for prediction of failure, power consumption, and forecasting.

IoT is a revolutionary transformation in both research and industrials settings. The emergence of the smart sensors and embedded computers can gather data from real-world with high accuracy. It also provides ubiquitous access to BD at low costs in less time. The BD analysis faces various challenges, mainly those related to data security & integrity, integrity of system, and standardization of communication protocol. These issues required a new type of analysis as suggested by Lynn et al. (2018), MTConnect data analytics system embedded with data acquisition and analysis. Table 2 provides the detailed comparison of existing approaches of data collection & storage, processing & analysis in FDA environment with reference to parameters, including response time, server load, overall cost, and pros and cons.

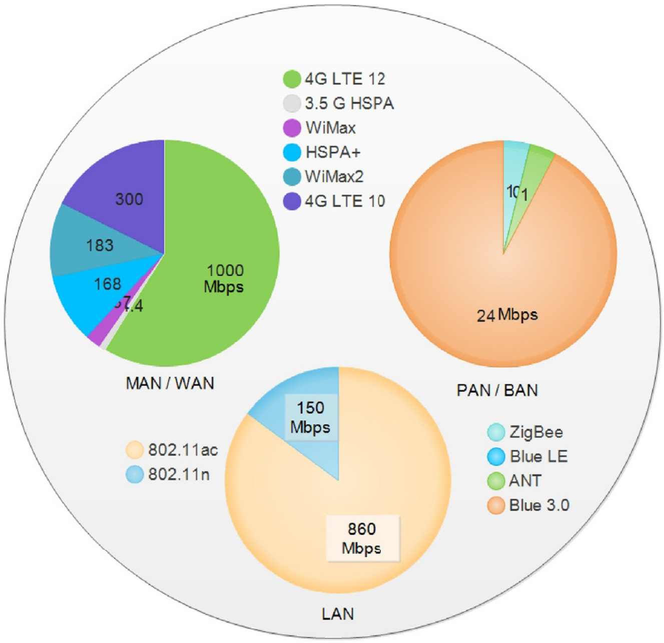

Fig. 5. Edge capacity of wireless models in FDA.

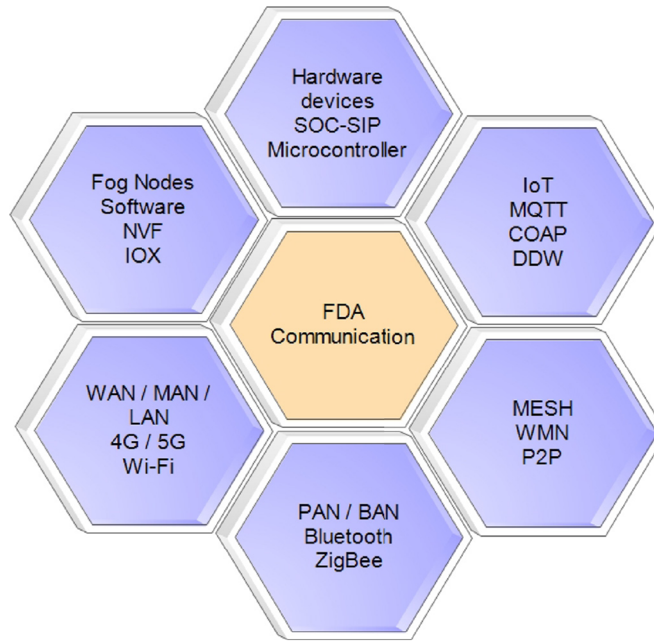

Fig. 6. Empowering network technologies for FDA communication (Vaquero and Rodero-Merino, 2014).

3.3. Connectivity at fog nodes

The nearness association of smart gadgets is a much-needed ingredient of the fog. The sheer volume of smart devices (50 billion devices in 2020), together with sensing IoT gadgets (working every minute of every day), bring attention for connectivity issues such as low bandwidth and connection failure.

3.3.1. Physical connection

Billions of smart gadgets are consuming huge energy and are generating data in relation to the end users in their edge systems, viewed as a gigantic bottleneck for such systems (Growth, 2013). System administrators have been seriously putting resources such as wireless technologies to handle the unexpected increase in devices per client. However, the LAN, WAN/MAN and PAN misses the mark in the IoT world. Greatest efforts in WAN/MAN are currently centred around LTE. LTEv12 is the first network which ful fills all the needs of the Intl.telecom.association to be marked as 4G7. As raided by Growth, 4G LTE/EPC should be completely taken by 2017 (Growth, 2013), as it will grow the accessible transmission capacity of edge systems (Astély et al., 2009). LAN innovation has enhanced to diminish blockage and increment the accessible data transmission with lower power utilization, such as cutting-edge Wi-Fi IEEE802.11ac. At long last, there have been tremendous upgrades in PNs. These short-range advances expect hubs to compose themselves, as no focal access point might be accessible like ANT+, ZigBee, Bluetooth low energy and RF4CE are the most noteworthy. Fig. 5 outlines the new wireless technologies, those which have evolved with download bandwidth capacity at the edge of the network.

3.3.2. Network connection

Apart from improvements on Wi-Fi networks, other advancements are required to provide effective communication to all end-points in FC. The network connection through LAN/WAN will not be possible due to high costs, lack of sufficient link-points for base stations. In a FDA environment, each FN should act as a router for its nearby FNs and should have mobility and resilient feature to node churn (FNs entering and leave-taking the network). To address this, it would perhaps require the mobile ad-hoc networks (MANET) (Sarkar et al., 2016) to be used. MANET is able to form dense network without having a costly and fixed infrastructure. This could be the foundation for future fog networks. ANT+, Bluetooth, LTE, RFC4CE and ZigBee each allows the creation of MANETs to the local range. One of the other solutions would be the wireless mesh networks (WMN), which is adjacent to MANETs. A WMN has little or no mobility using mesh routers at its core. FNs use these routers for connectivity or to other FNs if no direct link is able to be established with the routers.

In FDA, every FNs must have the capacity of switch/router for its neighbours and should be flexible to hub agitate (FNs entering and leaving the system). MANET has been an imperative research area since the last few decades (Sarkar et al., 2016). Hence, it could be a descent candidate solution for future FDA, as it will empower the arrangement of thickly populated systems. Here, the predefined static and costly frameworks that need to be accessible may not be required. FNs have no immediate connection to routers; thus, it has to connect via other FNs. There is still a demanding research is going on WMNs and MANET. There are protocols such as CoAP, MQTT and DDW that have been developed on the top of the WMNs/MANET for communications. They have been designed with the two major goals, which are resilience to failures and low consumption of resources. These protocols follow the subscriber model of communication (Vaquero and Rodero-Merino, 2014). Fig. 6 shows the empowering various network technologies to establish a communication channel for FDA. It includes FNs software such as NVF, IOX, hardware devices such as SOC-SIP, microcontroller, PAN, LAN, WAN and IoT devices.

Both IoT protocols and network can benefit from data locality. The choice of publisher/subscriber model will hugely improve the connectivity needs and the congestion problems at the end of communication network. Beside a better handling of the data traffic, the idea of locality would also enhance privacy and security of the data in a more effective way (Tanwar et al., 2017a).

3.3.3. FNs collaboration

The collaboration between FNs generally takes place in the forms of cluster-specific, peer-to-peer (P2P), or master-slave (Mahmud et al., 2018) approaches.

FNs which reside at the edge of the network can maintain a collaborative environment for execution by creating clusters amongst themselves. These clusters are created based on FNs' homogeneity, location, functional sub-system expansion, and computational load balancing. Collaboration through the cluster-based approach is effective in developing capabilities of various FNs simultaneously. Yet, static clusters are difficult to maintain and are not scalable at runtime, whereas dynamic clusters depend on the current load and FNs availability. In both scenarios, network overhead plays a critical role (Mahmud et al., 2018).

P2P collaboration between FNs is very common and is conducted in both hierarchical (Hong et al., 2013a) and flat order (Shi et al., 2015; Jalali et al., 2016). In P2P collaboration, the data inputs of one node are not only processed as the outputs of another node but also share virtual computing instances among the FNs. P2P collaboration is quite simple and reusable. But there exist reliability and access control issues associated with this collaboration approach.

In the master-slave approach, master FNs control functionalities, resource management, processing load, data flow and so on of underlying slave FNs. Conversely, in real-time FDA system, this approach requires high bandwidth for communication. Apart from above-mentioned collaboration approaches, there has been a hybrid collaborative network within the FDA computing environment (Nazmudeen et al., 2016).

3.4. Data distribution

Cloud computing (CC) has a centralized nature where all data are gathered and processed at one place. Unlike CC, FC is more of a decentralized platform where the data gathered from the sensor devices get processed, and run against light-weight analytics to filter the relevant data for CC. Once these necessary data are gathered and processed, they will be distributed to either cloud, users or other FNs. This distribution of data brings intelligence with it, which are listed below:

Scalability: Smart metering system includes millions of endpoints that makes it compulsory to use intelligent concentrators. For instance, smart city combines smart traffic control system (Nahri et al., 2018), smart parking management system, smart environmental monitoring system over a large population, which bring the complexity of deployment. In such systems, cloud computing alone is not enough to handle such huge volume of BD. It would be it be much safer and more efficient, if the resulting data get processed near to the edge of the network.

Network resource maintenance: The huge volume of BD generated from all actuators and sensors directly impact the network bandwidth. Remote locations are connected using wireless or wired connections with limited bandwidth. The distributed processing helps to relieve the network constraints by sending necessary information to Cloud only. At the edge, FC performs mostly the data processing, such as video analytics closer to data generation source.

Closed loop control: For a real-time system like healthcare latency has critical criteria for data communication. Low latency is highly essential to have a stable performance in real-time systems. In various multi-hop systems and encumbered cloud server, a large delay is often found which is undesirable. This issue can be handled by a local and high performed intelligent distributed system. Utilizing such distributed system would help to minimize the timing jitter and latency. There are various time-critical applications like inflight control system, industrial automation, electrical telehealth system, internal or medical vehicle networking that require very low latency and jitter. Closed loop control combines the advanced networking approach such as deterministic networking and assured delivery of packets in a limited time period.

Resilience: Resilience is the most critical feature that works through the data communication where operation centre is not functioning. A distributed approach is a highly recommended solution by researchers to achieve a good deal of resiliency.

3.5. QoE/QoS

QoE measures the complete performance of the system with respect to customer satisfaction. On the other hand, QoS measures the complete performance of a service with respect to user's satisfaction, e.g., CC services and FC services. FDA applications and services are real-time, therefore, QoE requirements for online processing/real-time streaming, distributed/parallel processing for BD analysis, and learning are required. Transformation of IoT into reality requires better resource management. In this context, FC plays a critical role. With the emergence of Vehicular Ad hoc Networks (VANET) and healthcare industry 4.0 for remote monitoring of the health, minimal latency and quick response are required.

Though the FNs have a fluctuating behaviour due to mobility nature, FC is a localized CC at the edge of network and placed near to the essential IoT. This issue can be handled by proper estimation of resources and analyzing the behaviour of FNs. Additionally, meeting the QoS level also becomes an issue. A good QoS helps to improve the QoE parameters. There are various QoS parameters, which are used to enhance the service quality such as latency, cost, time, content and data. Users do expect cost- and time-effective services with relevant content and informative data. Aazam et al. (2016) have proposed an approach named as MEdia FOg Resource Estimation (MeFoRE), which provides an estimation of resource on the basis of relinquish rate (RR). This approach also enhances QoS by using the previous QoE and net promoter score (NPS) data. The underlying algorithm has been applied on real-time IoT data by Amazon EC2 resource pricing.

3.6. Security and privacy

FDA environment is designed on traditional networking mechanisms, which makes it vulnerable to security attacks. Implementation of security approaches can affect data centric integrity. It is hard to ensure authentication access to services and maintenance of privacy in a largely distributed FDA model. Implementing authentication, authorization and access control is really difficult. The public and private key infrastructures (referred as encryption and decryption) and trustworthy execution environments such as trusted execution environment (TEE) (Marforio et al., 2014) are one of the best solutions to provide security and privacy in FDA environment. Further access control is one of the best tool to handle the security of the fog system. This raises the concern for designing of access control for client-fog-cloud environment, to meet the resource constraints goals. Puthal et al. (2018) have proposed a secure and sustainable load balancing technique for edge data centers (EDCs). In their approach, data security and authentications are first handled for EDCs, and then it provides load balancing between EDCs. Additionally, intrusion detection can be applied to fog infrastructure to mitigate the attacks such as flooding attack, attacks on VM or hyper-visor, insider attack, and port scanning. FDA provides new opportunities to investigate intrusion detection implementation on both client side and core cloud side.

In an FDA network, privacy-preserving algorithms are required at the data collection stage, where data collection takes place from different sensors. These algorithms are resource prohibited (Lu et al., 2012). Differential privacy can be applied to safeguard the privacy of a random single entry in the dataset. Due to the ubiquity of devices and challenges remain in fog environments, some applications still do not sent data to centralized services and kept them in the hosting network for better privacy. One of the better ways to provide privacy on data is to store encrypted sensitive data in traditional clouds, even though it makes it difficult to perform processing over such data. Applying special encryption functions or crypto processors (keeping some properties as original) allows data to perform limited tasks on it with very limited applicability. Consequently, users demand an innovative way to preserve their privacy in FDA environments. Tables 3 and 4 provide a comparative summary of existing approaches of network connectivity, data distribution, QoE/QoS, security and privacy challenges in FC with reference to parameters such as-network connectivity, FNs collaboration, QoS/QoE security & privacy, Pros and Cons.

4. Case studies

In order to demonstrate the fog data analytics process model, two case studies are presented that follow the process model.

4.1. Case study 1: smart grid system

In order to validate the FDA process model for big data analytics, we present a case study on smart grid (SG) (http).

The true value of smart metering/adapter lies in analyzing the real time data to produce actionable insights for loss reduction, improved load management, and improved power quality and reliability for instances. Data analytics is not only crucial for utilities, but is equally beneficial to consumers, who can utilize actionable insights on its consumption pattern to drive down its overall bill value by optimizing their daily electricity consumption. In our case, analysis of past billing data combined with real time data from smart meters/adapters were utilized for analyzing consumers with unusual energy usage pattern, large consumption drops, near zero consumption and consistent meter failure cases in the pilot areas. Such analysis aided with field investigation of suspect cases and helped to detect cases of faulty traditional meters (Tanwar et al., 2017b).

With the increasing capabilities of communication within power grid, the SG offers justifiable services to both customers and electricity service providers and it is summarized as follows (Okay and Ozdemir, 2016): An autonomous SG system, Environment protection (referred as reduction of gas emission and climate change) (Galli et al., 2011), high quality and reliability (Vatanparvar et al., 2015), Security and Privacy, Superior asset utilization (Al Faruque and Vatanparvar, 2016), power cost minimization and empower QoS at SG (Abdelwahab et al., 2014).

Here we discuss how the proposed FDA-based process model for SG in the smart home scenario works. Some of the criteria are: The SG supplies power with a varying power tariff, the renewable energy resources generate energy locally, and various electrical appliances of smart home consume a different quantity of energy supplied by renewable energy resources. The customer can control and monitor the energy consumption information with the help of a smart meter. The various electric appliances such as vacuum cleaner, AC, washing machine, TV, and electrical car consume energy. The smart homes in a region are considerably large, the fog server of that particular region has to collect and manage a huge amount of data. Fog servers keep details of energy consumption data momentarily so that in case of customer query, fog server can respond quickly. Then, relevant data are sent to cloud to store permanently. This FDA process model empowers SG to offer detail information with a low latency to its customers. It helps customers to analyze and monitor their hourly/daily/weekly/monthly/yearly power consumption. The detailed analysis may expose significant private data of customers. For instance, electricity consumption analysis of a TV may reveal clues whether the customer is at home or not, or similar electricity consumption analysis of dishwasher can disclose eating habits of individuals. The storage of this kind of private information requires outstanding security measures by using the public and private key. In the proposed process model, data are separated as public and private at smart meters. Furthermore, security keys are not shared with the cloud, which in turn offers better privacy at FC. The detailed electricity consumption data can be accessed by customers securely using local fog servers. In addition, FDA reduces the access time and the search of data. The customer can delete their private data at local fog server. Basically, FC is a smaller version of the CC with more efficient functionality. Data storage for SG is reduced as aggregated data only stored at cloud servers (Okay and Ozdemir, 2016). Fig. 7 shows the application of the proposed FDA process model for the smart grid application.

Table 3
Comparison for Data Collection, Data Storage, and Processing and Analysis in FDA environment.

Approaches	Objective	Data Collection/ Storage	Data Processing/ Analysis	Network Used	Response time	Server Load	Overall Cost	Pros	Cons
Verma et al. (2016)	Data replication and load balancing technique for fog	Data Replication	BD Analysis	Mesh	High	Low	Low	Improve Server time, server load and overall cost	Emphasis required on confidentiality and authentication issue of security and privacy
Smart Data*et al.* (Hosseinpour et al., 2016)	Smart Management of BD in the Fog Computing Context	IoT	IoT Vertical	–	High	Low	Low	Provides better preprocessing of big data originated from IoT sensors	Smart data model need to be more realistic
IDG et al. (Yuan et al., 2012)	A data dependency graph for intermediate data storage on cloud	Sensor Device/IoT	–	–	Medium	Medium	Medium	Cost-effective computation and storage	Need to reduce data transfer among data centre
ReTIn et al. (Chmelar et al., 2015)	Real-Time Indexing of Complex Data Streams	IoT	In-stream Processing	–	High	Low	–	Indexing schema and framework to handle complex data types	System reliability need to enhance
Raafat et al. (2017)	FDA for Real-Time IoT Sensor	IoT	Real-time Analysis	LAN	High	Low	Low	Reduce transmission cost, avoid network latency, enhance security and privacy	System reliability need to enhance
Darwish and Bakar (2018)	Intelligent Transportation BD Analytics	Vehicles	Real-time BD Analysis	IoV	High	Low	Low	Improve Response time, reduce latency and overall cost	Systems need to be more efficient and environmental friendly
MTConnect (Lynn et al., 2018)	Low cost data acquisition and analysis devices	IoT	Real-time Analysis	TCP/IP	Medium	Low	Low	Provide high-speed real time sampling and analysis	Mobility of the devices need to be address

Table 4
Comparison for network connectivity, data distribution, QoE/QoS, security and privacy challenges in FC.

Approaches	Objective	Network connectivity	Fog Nodes (FNS)	FNs Collaboration	QoS/QoE	Security and Privacy	Pros	Cons
Al Faruque and Vatanparvar (2016)	An Energy Management technique in Fog environment	IoT	Network devices	Cluster	Energy Saving	–	Provide flexibility, interoperability, connectivity	Need to handle rest of the features of energy management
Aazam and Huh (2015)	Fog based dynamic resource estimation and pricing model	IoT	Network devices	–	Data-volume	Data encryption	Efficient and in time scheduling of resources	Need to handle heterogeneous services and QoS
Hassan et al. (2015)	An approach for mobile application offloading and storage expansion	Mobile	Network devices	Cluster	Low Cost	Privacy	Handle computation offloading and data storage	Need to handle service availability, the security and privacy issues
Oueis et al. (2015)	A multi-user small cell clustering optimization policy for distributed fog computing	–	Base stations	Cluster	Low Latency	–	Optimization of communication and computational resources, higher QoE	Need to handle unpredictable idle time of the computing
Deng et al. (2016)	An approach for optimal Workload Allocation	Mobile	Network devices	P2P	Data-volume	–	Handle power consumption and transmission delay	Need to handle optimization in distributed manner
Zhang et al. (2017)	Fog network optimization	LRPON	Cloudlets	P2P	Low Cost	–	Cost-effective, Delay-sensitive	Need to security issues
F-RAN (Peng et al., 2016)	An architecture for Radio Access Networks based on fog	RAN	Base stations	P2P	Data-volume	–	Radio resource management, and distributed storing capabilities	Need to handle edge caching, SDN and NFV
MobileFog (Hong et al., 2013a)	A programming model for large–scale, latency–sensitive applications in IoT	IoT	Cloudlets/MediaBroker	–	Low latency	–	Geo-spatially distributed, large–scale, and latency–sensitive	Need to improve network bandwidth utilization, load balances etc
CoAP (Shi et al., 2015)	An IoT protocol to link mobile device clouds with fog computing	IoT	Network devices	P2P	Low Latency	–	Support distributed computation and share resources among mobile devices	Need to enhance the system performance
Jalali et al. (2016)	An approach for energy saving	IoT	Nano Network devices	P2P	Energy Saving	–	Energy-efficient communication by using nano data centre	Need to enhance the system performance
VFC (Hou et al., 2016)	An architecture for better communication and computation	Vehicles	Vehicles	Cluster	Resource management	Privacy	Enhancement in the communication and computation capacity	Need to handle more advance security issues
Puthal et al. (2018)	A Secure and Sustainable Load Balancing technique	IoT	Edge Data Centres (EDCs)	P2P	Load Balancing	Data Security and authentication	Strengthens the security by authenticating the destination EDCs	Need to implement on real-time testbed

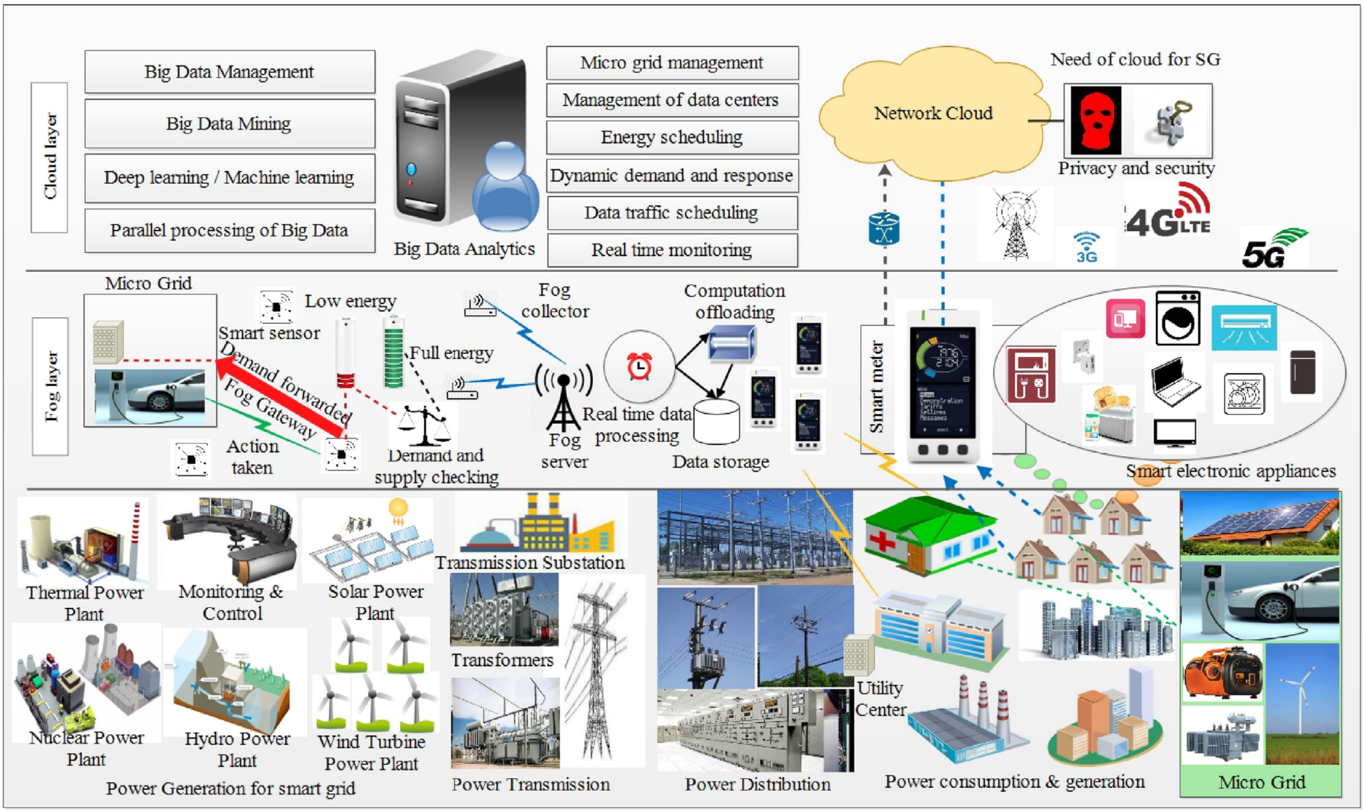

Fig. 7. FDA process model for smart grid application.

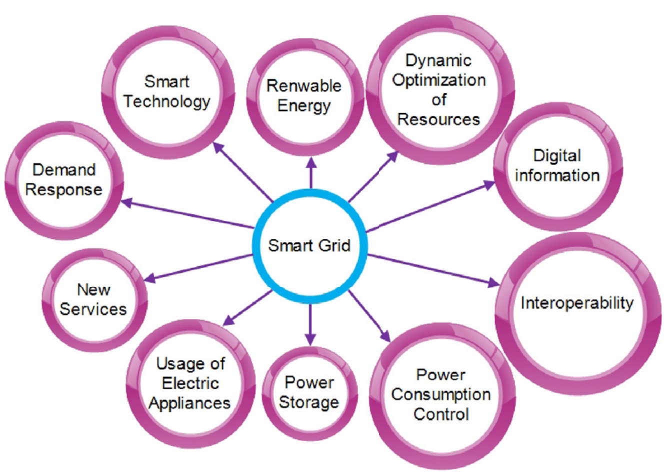

Fig. 8. Big Data Analytics services in Smart Grid systems.

Further, Fig. 8 represents the various analytics services related to BD in this smart grid system.

4.2. Case study 2: healthcare industry 4.0

This case study is related to the health observations and monitoring system. At the present time, biomedical data analytics is used to correctly predict smart healthcare decisions in industry 4.0 (Kumari et al., 2018b). This can be achieved with the deployment of omnipresent systems and transmitting data in a real-time environment through FNs in FDA environments rather than transferring to the cloud. Sensors and actuators are playing a vibrant role in most of the applications such as ITS, environmental monitoring system, smart city and healthcare applications (Vora et al., 2018a; Vora et al., 2018b). Especially, wearable medical devices are essential for gathering rich clinical information revealing of our mental and physical health. These sensors are constantly generating enormous BD. At any time, heart rate, blood pressure, respiratory rate, sugar level and body temperature surpass its standard value, the medical devices enabled at the fog send an alert message to the doctors and patients using a wireless network (Vora et al., 2017). An edge intelligence real-time and distributed analytics system was used for fall detection among stroke patients. It uses distributed data analytics between the edge device such as smartphone and server in the cloud. The fog-based fall detection system termed as U-Fall was categorized into three sections: front-end, back-end, and a communication module. The front and back-end perform fall detections independently. For the experiment, Samsung Galaxy S3 smartphone was used with Android 4.0.4, where Ice Cream Sandwich was used as a front-end, 64-bit Windows Server 2008 R2 Base was used as a back-end and 4 GB RAM for AWS server. Two subsystems were developed, one was to detect fall patterns and another one was for comparison purposes. The resulting system shows the 88% of Sensitivity and more than 74% of Specificity. This implies that U-Fall achieved a low false positive rate with a low miss rate as compared to the existing system. Consequently, a collaborative detection method maximizes the accuracy and decreases the wrong alarms rate. This health monitoring system detects the extensive fall throughout the moderate strokes. Further, the uses of the smart gateway in FDA architecture discussed in Section 3, provide complicated computing services such as embedded data mining and prioritized real-time storage (Cao et al., 2015). FAST system has been proposed for distributed data analytics system in healthcare using fog based sensors for monitoring and diminishing of strokes (Chen et al., 2016). Another framework has been proposed to monitor special diseases using body area sensor network at the nearest FNs at the edge of the network (Gia et al., 2015). Further, Dubey et al. (2015) have used pre-analyzed data collected by distributed

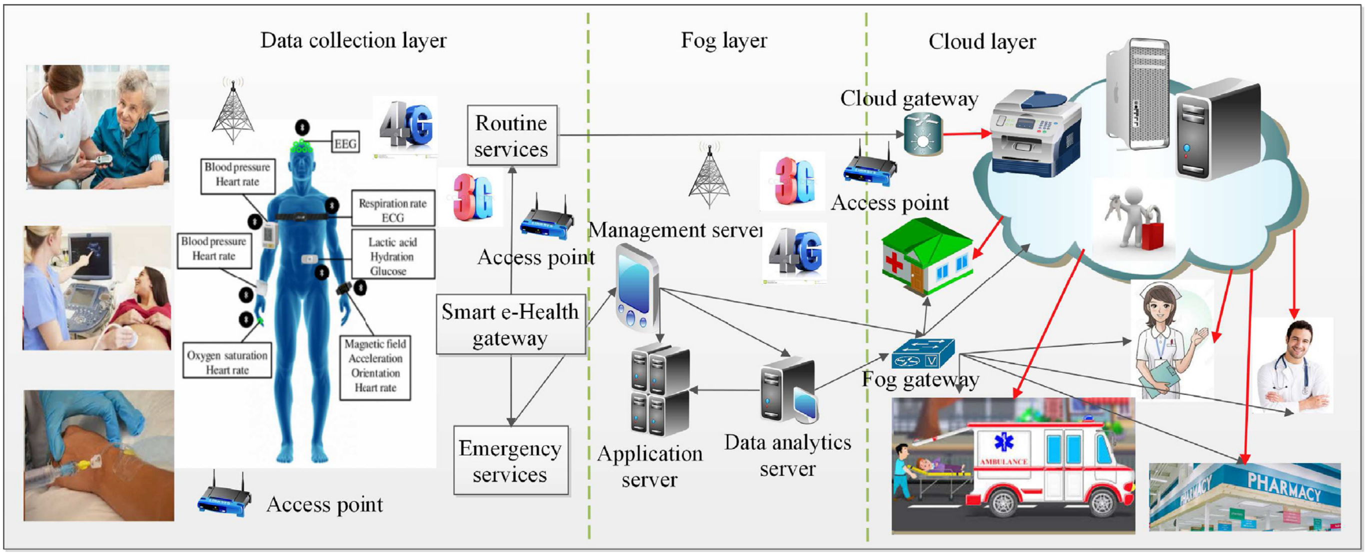

Fig. 9. Application of FDA process model in Healthcare domain.

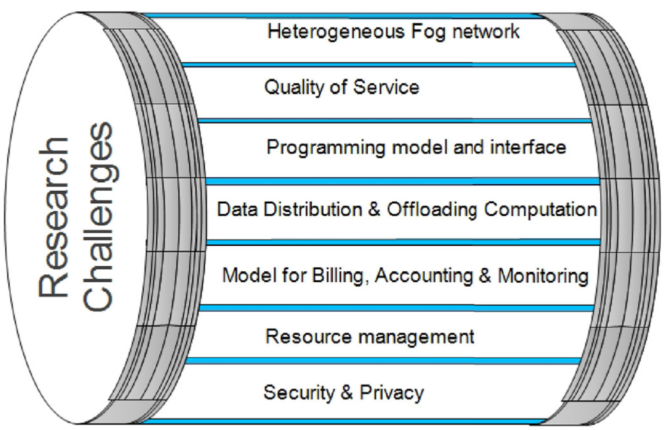

Fig. 10. Research challenges in fog data analytics.

medical sensors in smart healthcare for data mining and analytics at the edge. Similarly, Ahmad et al. (2016) presented a fog structure for the healthcare system. Hence, the proposed FDA architecture is well suited for healthcare industry 4.0. Fig. 9 represent the application of the proposed FDA process model for the healthcare Industry.

5. Research challenges in FDA

The discussion of FDA for IoT applications in the context of heterogeneity, real-time access, and particularly interoperability calls for the need of storage as well as processing of FDA. Rapid growth in technology in these and related areas generates a huge amount of BD at an exponential rate. However, the existing approaches available to process generated data are not adequate as compared to their exponential growth rate. Currently, there are very limited tools and software to address these issues related to FDA for IoT applications. Based on a review of literature, in this section, potential research issues are identified, which necessitate the development of process model for FDA life cycle. Based on the extensive literature review, challenges discussed with FDA within IoT applications, and issues delineated in Fig. 10, we have identified major research challenges related to FDA. The pursuit of these challenges will facilitate the advancement of FDA.

5.1. *Heterogeneous fog network*

Fog network has included fog gateway to manage and maintain connectivity within the network. Fog network is located at the edge of a network, which is heterogeneous in nature. To maintain the network properly, emergent technologies such as network function virtualization (NFV) and software defined networking (SDN) are the most popular solutions nowadays. These technologies help to increase the network scalability, reduce the costs of maintenance, easy execution, and management in various aspects of FC.

SDN resembles FC concepts, where fog devices act as a centralized controller (Stojmenovic, 2014). In this scenario, each FN acts as a router for neighbouring FN with node mobility and controller is available at the end node in the network. One of the major problems to integrate SDN with fog is unreliable wireless link and mobility of FN. Ku et al. (2014) proposed SDN-based mobile architecture for the cloud to handle such issues. Other challenges to deal with FN are to agitate & predicting, updating and maintaining the network graph connectivity at different granularity, connection and placement of the controller at the end of fog network (Heller et al., 2012). To accomplish minimal throughput and low latency, NVF plays a critical role in the performance of VM network applications in FDA. Throughput of VM network appliances, efficient placement, instantiations and migration of VM based applications in a dynamic network of fog (Yi et al., 2015) are the additional challenges.

5.2. *Quality of service (QoS)*

QoS is one of the essential aspects in FDA as directly committed to provide good quality services to the end users. To facilitate our discussion, we divide QoS into the following categories:

Reliability: A measure of 'less number of network failures' is known as reliability. A review on reliability requirements for FC was presented by Madsen et al. (2013). Periodical check-points are used to resume the FDA network after disaster or failure. This implies a rescheduling of failed jobs and to exploit parallel execution in replication. In highly dynamic FDA network, checkpoints and rescheduling may not be a feasible solution due to latency hence, since network cannot adapt changes. Replication may be the most promising solution for the scenarios where multiple FNs are working together.

Connectivity: Network relaying, clustering and partitioning provide new prospects in a heterogeneous FDA architecture for cost reduction, data pruning, and connectivity progression. For instance, an adhoc WSN can be segregated into various clusters due to the availability of rich-resources such as sinks node, cloudlet, and powerful smartphone. Xu et al. (2010) have proposed an approach to achieve the minimal throughput and computational efficiency. Likewise, FN selection from the end-user profoundly impacts the performance. The subset of FNs can be selected as the transmit nodes to achieve optimization for a specific area or for a single user with connectivity, throughput, delay and power consumption constraints.

Delay: Fog-based applications need real-time data processing rather than batch processing. Hong et al. (2013b) have proposed a spatio temporal event processing system to handle the latency issue in the context of fog. Further, Ottenwälder et al. (2014) have proposed a system to handle overlapping interests in data, reducing resource requirements and inaccurate results to reuse computation. Despite their efforts, further research is required to improve the scalability and remunerates the delay in complex event processing systems.

Capacity: Storage capacity and network bandwidth bend capacity of a fog network. To achieve this investigation on data locality for computation is a critical task. For instance, FC needs computation on BD that is distributed in various adjacent FNs. Computation of data can start after data collection stage, which may introduce a delay in the services. To handle this issue, service request pattern and user mobility pattern need to be governing to place BD on suitable FNs. Data placement design challenge arises due to association of cloud and fog to accommodate different workloads. It needs dynamic data placement with large volume capacity in FC. This implies to redesign search engine to process a query distributed in FNs (Yi et al., 2015), which further requires redesigning of cache on FN to exploit broader coverage and temporal locality to reduce delay and save network bandwidth (Zhang et al., 2013; Wang et al., 2014).

5.3. Programming model and interface

With a specific end goal to facilitate the developers to port their applications to FC, unified programming model and interfacing are required. The reasons are: (1) the FC based application-centric computing model needs suitable optimization for a specific application; (2) it is really difficult for a developer to coordinate progressive, dynamic and heterogeneous resources to form compatible applications for diverse platforms. Hong et al. (2013c) have proposed an on-demand scaling model for Internet-based, latency sensitive and large-scale geo-spatially distributed applications. However, their approach was dedicated for tree-based network structure for fixed location of FNs only. This is to say that, a more general approach for large-scale geo-spatially distributed and diverse networks would still be required with a dynamic mobility of FNs.

5.4. Data distribution and offloading computation

Most applications support for trustworthy relationship for data flow in fog computing environment between two fog devices. In such case, some attributes need to be defined in advance to maintain the trust during data distribution. Regular measurements and monitoring of the attributes are required for smooth function of the system. Feedback of the evaluation has to be taken and applied back to the fog devices for quality service achievement. In addition to this, offloading helps with computationally intensive tasks, while giving benefits to span the battery lifetime, improve performance and save storage space. Offloading can be classified into different aspects including, granularity, objectives scheme, adaptation, distributed execution and communication (Enzai and Tang, 2014). Although a substantial amount of research has previously been conducted in computation offloading from the perspective of CC and mobile computing, it is still in the stage of infancy in FC. A computation offload method MAUI, on future invocations was proposed by Cuervo et al. (2010). This approach highlighted the dynamic network connectivity, latency, and bandwidth. Further, Clonecloud (Chun et al., 2011) automatically spot merge point with in the network using static code analyzer. Additionally, ThinkAir (Kosta et al., 2012) worked on scalability and elasticity of the CC to improve the power of cloud. Gordon et al. (2012) have proposed COMET worked on distributed shared memory and VM synchronization to enhance tablets, smartphones or computing devices available within the network.

5.5. Business model for billing, accounting and monitoring

FDA cannot be flourishing without a maintainable business model. FDA model consists of the following parties; 1) Internet service providers (ISP), to construct fog network structures; 2) Cloud service providers (CSP), concerned to expand the services to the border of the network; 3) End users interested in computation and private cloud storage to reduce the cost of tenure. Consequently, in the pay-as-you-go model various issues need to be resolved, like billing, prices for dissimilar resources need to be set, payment setting for several parties of fog. To impose the pricing policy, accounting and monitoring is required at different granularity. To maximize the revenue and utilization of fog services, pricing needs to be done alike traditional industry hotels, airline ticketing or car rental industry (Xu and Li, 2013). Another approach to improve revenue is to provide incentives to the end user. Local clouds with storage and computation capacity are needed to be deployed at the edge of network. There should be some standardization and regulation to lease additional storage and computation to fog service provider. This will help to reduce the cost and increase revenues as fog service provider has to pay the owner of private cloud for the lease.

5.6. Resource management

Within FDA, resource management can be categorized into two sections: application-aware provisioning and resource detection & sharing. In application-aware provisioning, challenges arise due to the mobility of the end nodes, storage, band-width, latency and changes in the computation dynamically. For instance, in a connected vehicular system, an on-duty ambulance can be tracked by smart traffic light and be tuned to ensure green traffic wave for the ambulance and giving alert to the neighbouring vehicles to clear the way. The resource management to provide service mobility and provisioning helps to meet the QoS requirements such as the delay. Ottenwälder et al. (2013) have proposed a MigCEP approach for migration and placement of resources for cloud and fog both. Their approach ensures end-to-end network utilization and reduces latency. Application aware provisioning plays a critical role in mobile sensing applications of fog.

Resource detection and sharing are crucial for application performance in FDA. Liu et al. (2014) have proposed a dynamic method with flooding and centralized strategies to save power in heterogeneous FDA networks. N. Takayuki et al. (Nishio et al., 2013) have proposed a resource sharing framework for FDA networks by mapping heterogeneous resources such as CPUs, storage and communication. The resource sharing optimization problems expanded to include different aspects such as energy consumption, service availability, and revenues. In FDA environments, dynamic resource management systems are required so that analytics tasks are to be pushed towards the edge of the network. This strategy helps to ensure real-time data analytics. Arkian et al. (2017) proposed a fog-based data analytics approach for resource provisioning and management of IoT based applications.

5.7. Security and privacy

Presently, there is limited research work which focuses on security and privacy issues in FDA. Designing authentication & authorization, access control and intrusion detection techniques that can work with different FNs in heterogeneous architecture with different computing capacities, is very difficult to achieve. For authentication in FC, biometric-based systems can be advantageous. FDA provides new opportunities to investigate intrusion detection on the client and core cloud sides. There are challenges associated with the implementation of intrusion detection in large-scale, geo-distributed, and high mobility FC environments. One of the potential solutions is public and private key infrastructures and reliable execution environments such as trusted execution environment (TEE) (Chen et al., 2014; Marforio et al., 2014). In any case, the end users should always have their backup plan ready for failure of the sensors, service platforms, networks, and applications.

Currently, users are highly concerned about the privacy leakage risks such as location exposure over the Internet. In the FDA networks, algorithms for privacy preserving run between the cloud and the fog layers. The storage and computation are sufficient on both sides, whereas these algorithms are resource prohibited at the end devices. FN needs the privacy preserving techniques without decryption, for example, homomorphic encryption method during data collection at the local gateways. To safeguard the privacy of a random single entry in the dataset a differential privacy approach can be applied (Lu et al., 2012). In addition to the security and privacy issues to maintain a strong working environment between end users and fog service providers, the perfect matching of trust is also required. Hence, one has to set the level of trust and ensures the regular monitoring of the services to that level for the satisfaction of end users. The challenges which needs immediate attentions before executing a trust-oriented fog computing based model include (i) How to measure the trust in fog service? (ii) What are the measuring parameters of trust? (iii) Who can decide the selection of parameters of trust measurement? (iv) What can be the frequency of trust measurement? and (v) Who decides the implementation of improvement factors for better fog services?

6. Conclusion

The BD is growing drastically in most recent couple of years with the explosion of smart devices, sensor, and IoT. The generation of massive data from these devices necessitates the storage, processing and analyzing large amounts of data at a high frequency with efficient outcomes. In this paper, we have targeted the emergent intersects of fog and cloud computing technologies with a special focus on data analytics for fog-based BD. The FDA communication has an extensive variety of applications in smart cities, healthcare, satellite imaging, intelligent transportation system, and smart grid. FDA is the key to unlocking full advantages of these applications in the modern society. This paper conducted a comprehensive survey on BD communication and analytics in FC architecture with related to the concepts of traditional cloud computing. As the results of the survey indicated, efficient and effective algorithms for data storage and the data processing are required for appropriate analytics as we move forward with this promising technology, i.e. fog computing. Within the survey, we addressed and discussed numerous challenges in the FDA data processing and analysis and developed a broad taxonomy. We further abstracted an advanced level of a process model that is an extended FDA life cycle for BD computing. This novel process model aims to expand research and practices regarding FDA computing and address the current issues in this field of research, which are growing exponentially.

References

Aazam, M., Huh, E.-N., 2015. Fog computing micro datacenter based dynamic resource estimation and pricing model for iot. In: Advanced Information Networking and Applications (AINA), 2015 IEEE 29th International Conference on, pp. 687–694.
Aazam, M., St-Hilaire, M., Lung, C.-H., Lambadaris, I., 2016. Mefore: QoE based resource estimation at fog to enhance QoS in IoT. In: Telecommunications (ICT), 23rd International Conference on, pp. 1–5.
Abdelwahab, S., Hamdaoui, B., Guizani, M., Rayes, A., 2014. Enabling smart cloud services through remote sensing: an internet of everything enabler. IEEE Internet Things J. 1 (3), 276–288.
Ahmad, M., Amin, M.B., Hussain, S., Kang, B.H., Cheong, T., Lee, S., 2016. Health fog: a novel framework for health and wellness applications. J. Supercomput. 72 (10), 3677–3695.
Al Faruque, M.A., Vatanparvar, K., 2016. Energy management-as-a-service over fog computing platform. IEEE Internet Things J. 3 (2), 161–169.
Ali, R., Solis, C., Salehie, M., Omoronyia, I., Nuseibeh, B., Maalej, W., 2011. Social sensing: when users become monitors. In: Proceedings of the 19th ACM SIGSOFT Symposium and the 13th European Conference on Foundations of Software Engineering, ESEC/FSE '11, pp. 476–479.
Alvi, S.A., Afzal, B., Shah, G.A., Atzori, L., Mahmood, W., 2015. Internet of multimedia things: vision and challenges. Ad Hoc Netw. 33, 87–111.
Amin, A., Riyaz, S., Ali, A., Paul, Z., 2017. Review of IoT data analytics using big data, fog computing and data mining. Int. J. Comput. Sci. Mobile Comput. 6, 33–39.
Anawar, M.R., Wang, S., Azam Zia, M., Jadoon, A.K., Akram, U., Raza, S., 2018. Fog computing: an overview of big IoT data analytics. Wireless Commun. Mobile Comput. 2018, 22:1–22:5.
Arkian, H.R., Diyanat, A., Pourkhalili, A., 2017. MIST: fog-based data analytics scheme with cost-efficient resource provisioning for IoT crowdsensing applications. J. Netw. Comput. Appl. 82, 152–165.
Astély, D., Dahlman, E., Furuskär, A., Jading, Y., Lindström, M., Parkvall, S., 2009. Lte: the evolution of mobile broadband. IEEE Commun. Mag. 47 (4).
Bittencourt, L., Immich, R., Sakellariou, R., Fonseca, N., Madeira, E., Curado, M., Villas, L., DaSilva, L., Lee, C., Rana, O., 2018. The internet of things, fog and cloud continuum: integration and challenges. Internet Things 3–4, 134–155.
Bonomi, F., Milito, R., Zhu, J., Addepalli, S., 2012. Fog computing and its role in the internet of things. In: ACM Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, pp. 13–16.
Bonomi, F., Milito, R., Natarajan, P., Zhu, J., 2014. Fog computing: a platform for internet of things and analytics. In: Big Data and Internet of Things: a Roadmap for Smart Environments. Springer, pp. 169–186.
Cao, Y., Hou, P., Brown, D., Wang, J., Chen, S., 2015. Distributed analytics and edge intelligence: Pervasive health monitoring at the era of fog computing. In: ACM Proceedings of the Workshop on Mobile Big Data, pp. 43–48.
Cerina, L., Notargiacomo, S., Paccanit, M.G., Santambrogio, M.D., 2017. A fog-computing architecture for preventive healthcare and assisted living in smart ambients. In: Research and Technologies for Society and Industry (RTSI), 2017 IEEE 3rd International Forum on, pp. 1–6.
Chen, C.P., Zhang, C.-Y., 2014. Data-intensive applications, challenges, techniques and technologies: a survey on Big Data. Inf. Sci. 275, 314–347.
Chen, C., Raj, H., Saroiu, S., Wolman, A., 2014. cTPM: a cloud TPM for cross-device trusted applications. In: NSDI, pp. 187–201.
Chen, N., Chen, Y., You, Y., Ling, H., Liang, P., Zimmermann, R., 2016. Dynamic urban surveillance video stream processing using fog computing. In: Multimedia Big Data (BigMM), IEEE Second International Conference on, pp. 105–112.
Chiang, M., Zhang, T., 2016. Fog and IoT: an overview of research opportunities. IEEE Internet Things J. 3 (6), 854–864.
Chmelar, P., Drozd, M., Sebek, M., Zendulka, J., 2015. Real-time indexing of complex data streams. In: Sobh, T., Elleithy, K. (Eds.), Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering. Springer International Publishing, Cham, pp. 371–378.
Choo, K.-K.R., Lu, R., Chen, L., Yi, X., 2018. A foggy research future: advances and future opportunities in fog computing research. Future Generat. Comput. Syst. 78, 677–679.
Chun, B.-G., Ihm, S., Maniatis, P., Naik, M., Patti, A., 2011. Clonecloud: elastic execution between mobile device and cloud. In: ACM Proceedings of the Sixth Conference on Computer Systems, pp. 301–314.
Cuervo, E., Balasubramanian, A., Cho, D.-k., Wolman, A., Saroiu, S., Chandra, R., Bahl, P., 2010. MAUI: making smartphones last longer with code offload. In: ACM Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, pp. 49–62.
Darwish, T.S., Bakar, K.A., 2018. Fog based intelligent transportation big data analytics in the internet of vehicles environment: motivations, architecture, challenges, and critical issues. IEEE Access 6, 15679–15701.
de AssunÃ§Ão, M.D., da Silva Veith, A., Buyya, R., 2018. Distributed data stream processing and edge computing: a survey on resource elasticity and future directions. J. Netw. Comput. Appl. 103, 1–17.
Deng, R., Lu, R., Lai, C., Luan, T.H., Liang, H., 2016. Optimal workload allocation in fog-cloud computing toward balanced delay and power consumption. IEEE Internet Things J. 3 (6), 1171–1181.
Dubey, H., Yang, J., Constant, N., Amiri, A.M., Yang, Q., Makodiya, K., 2015. Fog data: enhancing telehealth big data through fog computing. In: ACM Proceedings of the ASE BigData & SocialInformatics 2015, p. 14.
Enzai, N.I.M., Tang, M., 2014. A taxonomy of computation offloading in mobile cloud computing. In: Mobile Cloud Computing, Services, and Engineering (MobileCloud), 2014 2nd IEEE International Conference on, pp. 19–28.
Farahani, B., Firouzi, F., Chang, V., Badaroglu, M., Constant, N., Mankodiya, K., 2018. Towards fog-driven IoT eHealth: promises and challenges of IoT in medicine and healthcare. Future Generat. Comput. Syst. 78, 659–676.
Fei, X., Shah, N., Verba, N., Chao, K.-M., Sanchez-Anguix, V., Lewandowski, J., James, A., Usman, Z., 2019. Cps data streams analytics based on machine learning for cloud and fog computing: a survey. Future Generat. Comput. Syst. 90, 435–450.
Galli, S., Scaglione, A., Wang, Z., June 2011. For the grid and through the grid: the role of power line communications in the smart grid. Proc. IEEE 99, 998–1027.

Gia, T.N., Jiang, M., Rahmani, A.-M., Westerlund, T., Mankodiya, K., Liljeberg, P., Tenhunen, H., 2015. Fog computing in body sensor networks: an energy efficient approach. In: Proc. IEEE Int. Body Sensor Netw. Conf.(BSN), pp. 1–7.
Gordon, M.S., Jamshidi, D.A., Mahlke, S.A., Mao, Z.M., Chen, X., 2012. COMET: code offload by migrating execution transparently. In: OSDI, vol. 12, pp. 93–106.
Gosman, C., Cornea, T., Dobre, C., Pop, F., Castiglione, A., 2018. Controlling and filtering users data in intelligent transportation system. Future Generat. Comput. Syst. 78, 807–816.
Growth, M.N.T., 2013. An Architecture Impact Study. Alcatel-Lucent Bell Labs Strategic White Paper.
Hassan, M.A., Xiao, M., Wei, Q., Chen, S., 2015. Help your mobile applications with fog computing. In: Sensing, Communication, and Networking-workshops (SECON Workshops), 2015 12th Annual IEEE International Conference on, pp. 1–6.
He, J., Wei, J., Chen, K., Tang, Z., Zhou, Y., Zhang, Y., 2018. Multitier fog computing with large-scale IoT data analytics for smart cities. IEEE Internet Things J. 5 (2), 677–686.
Heller, B., Sherwood, R., McKeown, N., 2012. The controller placement problem. In: ACM Proceedings of the First Workshop on Hot Topics in Software Defined Networks, pp. 7–12.
Hong, K., Lillethun, D., Ramachandran, U., Ottenwälder, B., Koldehofe, B., 2013a. Mobile fog: a programming model for large-scale applications on the internet of things. In: Proceedings of the Second ACM SIGCOMM Workshop on Mobile Cloud Computing, pp. 15–20.
Hong, K., Lillethun, D., Ramachandran, U., Ottenwälder, B., Koldehofe, B., 2013b. Opportunistic spatio-temporal event processing for mobile situation awareness. In: ACM Proceedings of the 7th ACM International Conference on Distributed Event-based Systems, pp. 195–206.
Hong, K., Lillethun, D., Ramachandran, U., Ottenwälder, B., Koldehofe, B., 2013. Mobile fog: a programming model for large-scale applications on the internet of things. In: Proceedings of the Second ACM SIGCOMM Workshop on Mobile Cloud Computing, pp. 15–20.
Hosseinpour, F., Plosila, J., Tenhunen, H., Dec 2016. An approach for smart management of big data in the fog computing context. In: 2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom), pp. 468–471.
Hou, X., Li, Y., Chen, M., Wu, D., Jin, D., Chen, S., 2016. Vehicular fog computing: a viewpoint of vehicles as the infrastructures. IEEE Trans. Veh. Technol. 65 (6), 3860–3873.
http://nsgm.gov.in/sites/default/files/Report-on-Ajmer-SGPP-Case-Study.pdf.
Hu, P., Dhelim, S., Ning, H., Qiu, T., 2017. Survey on fog computing: architecture, key technologies, applications and open issues. J. Netw. Comput. Appl. 98, 27–42.
Jalali, F., Hinton, K., Ayre, R., Alpcan, T., Tucker, R.S., 2016. Fog computing may help to save energy in cloud computing. IEEE J. Sel. Area. Commun. 34 (5), 1728–1739.
Kambatla, K., Kollias, G., Kumar, V., Grama, A., 2014. Trends in big data analytics. J. Parallel Distr. Comput. 74 (7), 2561–2573.
Keller, T., 2011. Mining the Internet of Things: Detection of False-positive RFID Tag Reads Using Low-level Reader Data. PhD thesis. University of St. Gallen, Germany.
Kosta, S., Aucinas, A., Hui, P., Mortier, R., Zhang, X., 2012. Thinkair: dynamic resource allocation and parallel execution in the cloud for mobile code offloading. In: Infocom, 2012 Proceedings IEEE, pp. 945–953.
Ku, I., Lu, Y., Gerla, M., 2014. Software-defined mobile cloud: architecture, services and use cases. In: Wireless Communications and Mobile Computing Conference (IWCMC), 2014 International, pp. 1–6.
Kumari, A., Tanwar, S., Tyagi, S., Kumar, N., Maasberg, M., Choo, K.-K.R., 2018a. Multimedia big data computing and internet of things applications: a taxonomy and process model. J. Netw. Comput. Appl. 124, 169–195.
Kumari, A., Tanwar, S., Tyagi, S., Kumar, N., 2018b. Fog computing for healthcare 4.0 environment: opportunities and challenges. Comput. Electr. Eng. 72, 1–13.
Liu, W., Nishio, T., Shinkuma, R., Takahashi, T., 2014. Adaptive resource discovery in mobile cloud computing. Comput. Commun. 50, 119–129.
Liu, J., Li, J., Zhang, L., Dai, F., Zhang, Y., Meng, X., Shen, J., 2018. Secure intelligent traffic light control using fog computing. Future Generat. Comput. Syst. 78, 817–824.
Lu, R., Liang, X., Li, X., Lin, X., Shen, X., 2012. Eppa: an efficient and privacy-preserving aggregation scheme for secure smart grid communications. IEEE Trans. Parallel Distr. Syst. 23 (9), 1621–1631.
Lynn, R., Wescoat, E., Han, D., Kurfess, T., 2018. Embedded fog computing for high-frequency MTConnect data analytics. Manuf. Lett. 15, 135–138 Industry 4.0 and Smart Manufacturing.
Madsen, H., Burtschy, B., Albeanu, G., Popentiu-Vladicescu, F., 2013. Reliability in the utility computing era: towards reliable fog computing. In: Systems, Signals and Image Processing (IWSSIP), 2013 20th International Conference on, pp. 43–46.
Mahmud, R., Kotagiri, R., Buyya, R., 2018. "Fog Computing: a Taxonomy, Survey and Future Directions," Internet of Everything: Algorithms, Methodologies, Technologies And Perspectives, pp. 103–130.
Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H., 2011. Big Data: the Next Frontier for Innovation, Competition, and Productivity. Report by McKinsey Global Institute.
Marforio, C., Karapanos, N., Soriente, C., Kostiainen, K., Capkun, S., 2014. Smartphones as practical and secure location verification tokens for payments. In: NDSS.
Mohammadi, M., Al-Fuqaha, A., Sorour, S., Guizani, M., 2018. Deep learning for iot big data and streaming analytics: a survey. IEEE Commun. Surv. Tutor. 1.
More, P., 2015. Review of implementing fog computing. Int. J. Res. Eng. Technol. 4 (06), 335–338.
Naha, R.K., Garg, S., Georgakopoulos, D., Jayaraman, P.P., Gao, L., Xiang, Y., Ranjan, R., 2018. Fog computing: survey of trends, architectures, requirements, and research directions. IEEE Access 6, 47980–48009.
Nahri, M., Boulmakoul, A., Karim, L., Lbath, A., 2018. IoV distributed architecture for real-time traffic data analytics. Proced. Comput. Sci. 130, 480–487 The 9th International Conference on Ambient Systems, Networks and Technologies (ANT 2018)/The 8th International Conference on Sustainable Energy Information Technology (SEIT-2018)/Affiliated Workshops.
Nazmudeen, M.S.H., Wan, A.T., Buhari, S.M., 2016. Improved throughput for power line communication (plc) for smart meters using fog computing based data aggregation approach. In: Smart Cities Conference (ISC2), IEEE International, pp. 1–4.
Nishio, T., Shinkuma, R., Takahashi, T., Mandayam, N.B., 2013. Service-oriented heterogeneous resource sharing for optimizing service latency in mobile cloud. In: ACM Proceedings of the First International Workshop on Mobile Cloud Computing & Networking, pp. 19–26.
Okay, F.Y., Ozdemir, S., 2016. A fog computing based smart grid model. In: Networks, Computers and Communications (ISNCC), 2016 International Symposium on, pp. 1–6.
Ottenwälder, B., Koldehofe, B., Rothermel, K., Ramachandran, U., 2013. Migcep: operator migration for mobility driven distributed complex event processing. In: Proceedings of the 7th ACM International Conference on Distributed Event-based Systems, pp. 183–194.
Ottenwälder, B., Koldehofe, B., Rothermel, K., Hong, K., Ramachandran, U., 2014. Recep: selection-based reuse for distributed complex event processing. In: Proceedings of the 8th ACM International Conference on Distributed Event-based Systems, pp. 59–70.
Oueis, J., Strinati, E.C., Sardellitti, S., Barbarossa, S., 2015. Small cell clustering for efficient distributed fog computing: a multi-user case. In: Vehicular Technology Conference (VTC Fall), 2015 IEEE 82nd, pp. 1–5.
Parizi, R.M., Dehghantanha, A., Amritraj, 2018. Smart contract programming languages on blockchains: an empirical evaluation of usability and security. In: International Conference on Blockchain (ICBC). LNCS, Springer, pp. 75–91.
Parizi, R.M., Dehghantanha, A., Choo, K.-K.R., Singh, A., 2018. Empirical vulnerability analysis of automated smart contracts security testing on blockchains. In: 28th Annual ACM International Conference on Computer Science and Software Engineering (CASCON), pp. 103–113.
Peng, M., Yan, S., Zhang, K., Wang, C., 2016. Fog-computing-based radio access networks: issues and challenges. IEEE Netw. 30 (4), 46–53.
Puthal, D., Obaidat, M.S., Nanda, P., Prasad, M., Mohanty, S.P., Zomaya, A.Y., 2018. Secure and sustainable load balancing of edge data centers in fog computing. IEEE Commun. Mag. 56 (5), 60–65.
Raafat, H.M., Hossain, M.S., Essa, E., Elmougy, S., Tolba, A.S., Muhammad, G., Ghoneim, A., 2017. Fog intelligence for real-time IoT sensor data analytics. IEEE Access 5, 24062–24069.
Roman, R., Lopez, J., Mambo, M., 2018. Mobile edge computing, fog et al.: a survey and analysis of security threats and challenges. Future Generat. Comput. Syst. 78, 680–698.
Sarkar, S.K., Basavaraju, T., Puttamadappa, C., 2016. Ad Hoc Mobile Wireless Networks: Principles, Protocols, and Applications. CRC Press.
Shi, H., Chen, N., Deters, R., 2015. Combining mobile and fog computing: using coap to link mobile device clouds with fog computing. In: Data Science and Data Intensive Systems (DSDIS), IEEE International Conference on, pp. 564–571.
Stojmenovic, I., 2014. Fog computing: a cloud to the ground support for smart things and machine-to-machine networks. In: Telecommunication Networks and Applications Conference (ATNAC), 2014 Australasian, pp. 117–122.
Storey, V.C., Song, I.-Y., 2017. Big data technologies and Management: what conceptual modeling can do. Data Knowl. Eng. 108, 50–67.
Subbu, K.P., Vasilakos, A.V., 2017. Big data for context aware computing â" perspectives and challenges. Big Data Res. 10, 33–43.
Tang, B., Chen, Z., Hefferman, G., Wei, T., He, H., Yang, Q., 2015. A hierarchical distributed fog computing architecture for big data analysis in smart cities. In: ACM Proceedings of the ASE BigData & SocialInformatics, (New York, NY, USA), pp. 28:1–28:6.
Tanwar, S., Patel, P., Patel, K., Tyagi, S., Kumar, N., Obaidat, M.S., 2017a. An advanced internet of thing based security alert system for smart home. In: International Conference on Computer, Information and Telecommunication Systems. CITS, Dalian, China, pp. 25–29.
Tanwar, S., Tyagi, S., Kumar, S., 2017b. The role of internet of things and smart grid for the development of a smart city. In: Intelligent Communication and Computational Technologies. Lecture Notes in Networks and Systems, vol. 19, pp. 23–33.
Vaquero, L.M., Rodero-Merino, L., 2014. Finding your way in the fog: towards a comprehensive definition of fog computing. Comput. Commun. Rev. 44 (5), 27–32.
Vatanparvar, K., Faruque, A., Abdullah, M., 2015. Energy management as a service over fog computing platform. In: Proceedings of the ACM/IEEE Sixth International Conference on Cyber-physical Systems, pp. 248–249.
Verma, S., Yadav, A.K., Motwani, D., Raw, R., Singh, H.K., 2016. An efficient data replication and load balancing technique for fog computing environment. In: Computing for Sustainable Global Development (INDIACom), 3rd International Conference on, pp. 2888–2895.
Vora, J., Tanwar, S., Tyagi, S., Kumar, N., Rodrigues, J.J.P.C., 2017. FAAL: fog computing-based patient monitoring system for ambient assisted living. In: 19th IEEE International Conference on e-Health Networking, Applications and Services. Healthcom, Dalian, China, pp. 1–6.
Vora, J., Italiya, P., Tanwar, S., Tyagi, S., Kumar, N., Obaidat, M.S., Hsiao, K., 2018a. Ensuring privacy and security in E- health records. In: 2018 International Conference on Computer, Information and Telecommunication Systems, CITS, Alsace, Colmar, France, pp. 1–5.

Vora, J., DevMurari, P., Tanwar, S., Tyagi, S., Kumar, N., Obaidat, M.S., 2018b. Blind signatures based secured e-healthcare system. In: 2018 International Conference on Computer, Information and Telecommunication Systems, CITS, Alsace, Colmar, France, pp. 1–5.

Wang, X., Chen, M., Taleb, T., Ksentini, A., Leung, V., 2014. Cache in the air: exploiting content caching and delivery techniques for 5g systems. IEEE Commun. Mag. 52 (2), 131–139.

Xu, H., Li, B., 2013. Dynamic cloud pricing for revenue maximization. IEEE Trans. Cloud Comput. 1 (2), 158–171.

Xu, F., Tan, C.C., Li, Q., Yan, G., Wu, J., 2010. Designing a practical access point association protocol. In: INFOCOM, 2010 Proceedings IEEE, pp. 1–9.

Yeow, K., Gani, A., Ahmad, R.W., Rodrigues, J.J.P.C., Ko, K., 2018. Decentralized consensus for edge-centric internet of things: a review, taxonomy, and research issues. IEEE Access 6, 1513–1524.

Yi, S., Li, C., Li, Q., 2015. A survey of fog computing: concepts, applications and issues. In: ACM Proceedings of the 2015 Workshop on Mobile Big Data, pp. 37–42.

Yuan, D., Yang, Y., Liu, X., Zhang, G., Chen, J., 2012. A data dependency based strategy for intermediate data storage in scientific cloud workflow systems. Concurrency Comput. Pract. Ex. 24 (9), 956–976.

Zhang, Y., Tan, C., Qun, L., 2013. Cachekeeper: a system-wide web caching service for smartphones. In: Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 265–274.

Zhang, W., Lin, B., Yin, Q., Zhao, T., 2017. Infrastructure deployment and optimization of fog network based on microdc and lrpon integration. Peer-to-Peer Network. Appl. 10 (3), 579–591.

Zhang, P., Zhou, M., Fortino, G., 2018. Security and trust issues in fog computing: a survey. Future Generat. Comput. Syst. 88, 16–27.

Aparna Kumari pursuing Ph.D. from Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, Gujarat, India. She received M.Tech from Jawaharlal Nehru University, Delhi, India in 2012. Her research interest includes big data analytics, smart grid, and cloud computing.

Sudeep Tanwar is an Associate Professor in the Computer Science and Engineering Department at Institute of Technology, Nirma University, Ahmedabad, Gujarat, India. He received B.Tech in 2002 from Kurukshetra University, India, and M.Tech (Honor's) in 2009 from Guru Gobind Singh Indraprastha University, Delhi, India. He received Ph.D. in 2016 from Mewar University, Chittorgarh, Rajasthan, India with specialization in Wireless Sensor Network. He has authored more than 50 technical research papers published in leading International Conferences and Peer-reviewed International journals from the IEEE, Elsevier, Springer, John Wiley, etc. Some of his research findings are published in top-cited journals such as the Journal of Network and Computer Application, Pervasive and Mobile Computing, International Journal of Communication System, Telecommunication System and IEEE Systems Journal. He has also published three books on Role of Heterogeneity in WSN, Big Data Analytics, and Mobile Computing with International/National Publishers. He has guided many students leading to M.E./M.Tech and guiding students leading to Ph.D. He is Associate Editor of "Security and Privacy Journal, Wiley. His current interest includes Routings Issues in WSN, Integration of sensor with the cloud, Computational aspects of Smart Grid, and Assessment of Fog Computing in BASN.

Sudhanshu Tyagi is an Assistant Professor in Department of Electronics and Communication Engineering, Thapar Institute of Engineering and Technology, Deemed University, Patiala, Pb., India. He received his Bachelor in Engineering in Electronics & Tele-Communication in 2000 from North Maharashtra University, Jalgaon, Maharashtra (India). He achieved second rank in the university at under graduation level. He received his Masters in Technology with honor's in Electronics & Communication Engineering in 2005 from National Institute of Technology, Kurukshetra, Haryana (India). He received Ph.D. in 2016 from Mewar University, Chittorgarh, Rajasthan (India). He has 30 research publications in peer reviewed journal and conferences from leading publishers like Elsevier, Springer, Wiley etc. He is reviewer for the board of international journal from leading publisher like International Journal of Ad-Hoc and Ubiquitous Computing, Inderscience and Journal of the Franklin Institute, Elsevier. His research area includes lifetime enhancement of homogeneous and /or heterogeneous WSNs. He is a member of IEEE and IAENG.

Dr. Neeraj Kumar is an Associate Professor in Computer Science & Department, Thapar Institute of Engineering & Technology, Patiala. Dr. Neeraj received his Ph.D. in CSE from SMVD University, Katra (J & K), India, and was a postdoctoral research fellow in Coventry University, Coventry, UK. Dr. Neeraj is an internationally renowned researcher in the areas of VANET, CPS Smart Grid, IoT Mobile Cloud computing, Big Data, and Cryptography. He has published more than 200 technical research papers in leading journals and conferences.

Reza M. Parizi is a Software Engineering faculty in the Department of Software Engineering and Game Development at Kennesaw State University, GA, USA. He is a consummate technologist and software engineering researcher with an entrepreneurial spirit at KSU. He is the member of Cyber Scientist- A community of cybersecurity researchers, as well as IEEE, IEEE Blockchain Community, IEEE Computer Society and ACM. Prior to joining KSU, he was an Associate Professor at New York Institute of Technology. He has applied his insights and expertise to a host of innovative and technology driven projects across start-ups, security, software, and education industries. He received a Ph.D. in Software Engineering in 2012 and M.Sc. and B.Sc. degrees in Software Engineering and Computer Science respectively in 2008 and 2005. He has more than 8 years of working experience in industrial software development and project managing. His interests are R&D in Blockchain, smart contract programming, emerging issues in software engineering, and the practice of secure software-run world applications. He has published several research papers in top reputable scientific and international conferences and also has two copyrights to his credit.

Kim-Kwang Raymond Choo received the Ph.D. in Information Security in 2006 from Queensland University of Technology, Australia. He currently holds the Cloud Technology Endowed Professorship at The University of Texas at San Antonio (UTSA), and has a courtesy appointment at the University of South Australia. In 2016, he was named the Cybersecurity Educator of the Year - APAC (Cybersecurity Excellence Awards are produced in cooperation with the Information Security Community on LinkedIn), and in 2015 he and his team won the Digital Forensics Research Challenge organized by Germany's University of Erlangen-Nuremberg. He is the recipient of the 2018 UTSA College of Business Col. Jean Piccione and Lt. Col. Philip Piccione Endowed Research Award for Tenured Faculty, IEEE TrustCom 2018 Best Paper Award, ESORICS 2015 Best Research Paper Award, 2014 Highly Commended Award by the Australia New Zealand Policing Advisory Agency, Fulbright Scholarship in 2009, 2008 Australia Day Achievement Medallion, and British Computer Society's Wilkes Award in 2008. He is also a Fellow of the Australian Computer Society, and an IEEE Senior Member.