Future Generation Computer Systems 97 (2019) 219–235

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Edge computing: A survey

Wazir Zada Khan [a,*], Ejaz Ahmed [b,*], Saqib Hakak [c], Ibrar Yaqoob [d], Arif Ahmed [e]

[a] Farasan Networking Research Laboratory, Faculty of CS & IS, Jazan University, Jazan, Saudi Arabia
[b] Centre for Mobile Cloud Computing Research (C4MCCR), Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia
[c] Canadian Institute for Cyber-security, Faculty of Computer Science, University of New Brunswick, Fredericton, Canada
[d] Department of Computer Science and Engineering, Kyung Hee University, Yongin-si 17104, South Korea
[e] University Rennes, Inria, CNRS, IRISA, F-35000, Rennes, France

HIGHLIGHTS

- A comprehensive survey on edge computing, i.e., Fog, Mobile-edge and Cloudlet.
- A new classification of multi-facet computing paradigms within edge computing.
- Identification of key requirements to envision edge computing domain.
- Exploration of open research challenges.

ARTICLE INFO

Article history:
Received 20 August 2018
Received in revised form 12 February 2019
Accepted 22 February 2019
Available online 26 February 2019

Keywords:
Mobile edge computing
Edge computing
Cloudlets
Fog computing
Micro clouds
Cloud computing

ABSTRACT

In recent years, the Edge computing paradigm has gained considerable popularity in academic and industrial circles. It serves as a key enabler for many future technologies like 5G, Internet of Things (IoT), augmented reality and vehicle-to-vehicle communications by connecting cloud computing facilities and services to the end users. The Edge computing paradigm provides low latency, mobility, and location awareness support to delay-sensitive applications. Significant research has been carried out in the area of Edge computing, which is reviewed in terms of latest developments such as Mobile Edge Computing, Cloudlet, and Fog computing, resulting in providing researchers with more insight into the existing solutions and future applications. This article is meant to serve as a comprehensive survey of recent advancements in Edge computing highlighting the core applications. It also discusses the importance of Edge computing in real life scenarios where response time constitutes the fundamental requirement for many applications. The article concludes with identifying the requirements and discuss open research challenges in Edge computing.

1. Introduction

Edge computing constitutes a new concept in the computing landscape. It brings the service and utilities of cloud computing closer to the end user and is characterized by fast processing and quick application response time. The currently developed internet-enabled applications such as surveillance, virtual reality, and real-time traffic monitoring require fast processing and quick response time [1,2]. End users normally run these applications on their resource-constrained mobile devices while the core service and processing are performed on cloud servers. Leveraging services of cloud by mobile devices result in high latency and mobility-related issues [3,4]. Edge computing fulfills the above-mentioned application requirements by bringing the processing to the edge of the network. The cloud computing issues can be resolved through the three Edge computing models Cloudlets [5,6], Fog computing [7] and Mobile Edge computing [8,9].

The European Telecommunications Standards Institute (ETSI) has introduced the concept of Mobile Edge computing where mobile users can utilize the computing services from the base station. The Fog computing concept has been introduced by Cisco [10], which enables the applications to run directly at the network edge through billions of smart connected devices. Satyanarayanan et al. [11] introduced the concept of Cloudlets to solve the latency problem in accessing the Cloud by using the computer resources available in the local network. Similarly, Mobile Edge computing provides the off-load processing, application services and storage close to the end users.

* Corresponding authors.
E-mail addresses: wazirzadakhan@yahoo.com (W.Z. Khan), ejazahmed@ieee.org (E. Ahmed).

https://doi.org/10.1016/j.future.2019.02.050
0167-739X/

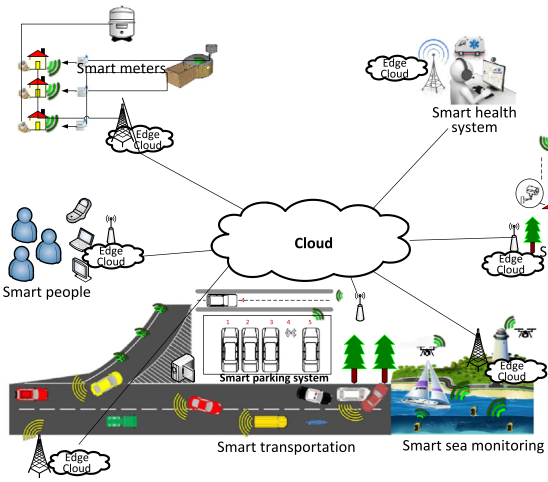

Fig. 1. Edge computing applications.

Among the promising features of Edge computing are included mobility support, location awareness, ultra-low latency, and proximity to the user [12]. These features make Edge computing suitable for different future applications like industrial automation, virtual reality, real-time traffic monitoring, smart home, smart sea monitoring and data analytics as shown in Fig. 1. Edge devices like routers, access points, and base station host different services (e.g., QoS, VPN, and Voice over IP etc.) [13]. These Edge devices act as a bridge that connects the smart mobile devices with the cloud. Several surveys (S. Yi et al. [14], L. M. Vaquero et al. [15], I. Stojmenovic [16], F. Bonomi et al. [17], T.H. Luan et al. [18], E. Ahmed et al. [19]) have studied various aspects of Edge computing like Fog computing while a few studies focus on the Edge computing domain like Mobile Edge computing that focuses on specific application domains. However, there is no comprehensive study has been completed yet that includes all aspects of Edge computing including Mobile-Edge, Fog and Cloudlet computing.

The main contribution of this article includes:

- A comprehensive survey on all aspects of Edge computing (Cloudlet, Fog and Mobile-Edge).
- A new classification of multi-facet computing paradigms within Edge computing.
- The identification of key requirements to envision the Edge computing domain.
- The exploration of open research challenges.

The rest of the manuscript is structured as follows: Section 2 presents the core concepts of cloud computing, edge computing and similar concepts along-with its architecture and differences. Section 3 discusses the characteristics of Edge computing. Section 4 explores the state-of-the-art in Edge computing and provides a comparative study of the research work that highlights the merits and demerits of the existing frameworks. Section 5 outlines the requirements for realizing the vision of Edge computing. Section 6 highlights the research challenges and open issues. Finally, Section 7 concludes the paper. Table 1 provide a list of acronyms used in this paper.

2. Background

This section presents the fundamental concepts of Cloud computing and its main difference with the Edge computing. The basic concepts of Cloudlet, Fog and Mobile Edge computing are also presented in this section. The aim of this section is to provide the reader with a solid foundation of the research subject.

Table 1
List of acronyms and their descriptions.

Acronym	Description
ETSI	European Telecommunications Standards Institute
IAAS	Infrastructure as a Service
PAAS	Platform as a Service
SAAS	Software as a Service
VPN	Virtual Private Networking
RAN	Radio Access Network
IoT	Internet of Things
LISP	Locator ID Separation Protocol
A-BCI	Augmented Brain–Computer Interfaces
M2M	Machine-to-Machine
E-HAMC	Emergency Help Alert Mobile Cloud
HEM	Home Energy Management
IOV	Internet of Vehicles
RSU	Road Side Units
D2D	Data to Decision
VM	Virtual Machine
FC-MCPS	Fog Computing-enabled Medical Cyber–Physical System
VMD	Virtual Medical Device
MINLP	Mixed Integer Non-linear Programming
PPFA	Privacy-preserving Fog-enabled Aggregation
IDS	Intrusion Detection System
VHD	Virtual Honeypot Device
WMAN	Wireless Metropolitan Area Network
BS	Base Stations
UE	User Equipment
MECC	Mobile Edge Cloud Computing
CSo	Cloud Server Operator
ESOs	End server Owners
MD	Mobile Device
QoS	Quality of Service

2.1. Cloud computing

Cloud computing is a computing paradigm that offers on-demand services to the end-users through pool of computing resources that includes storage services, computing resources and so on [20]. The key services that Cloud computing offers includes infrastructure as a service (IAAS), platform as a service (PAAS) and software as a service (SAAS) [21]. All these services provide on-demand computing services such as storing and data processing [22]. Besides offering the above-mentioned services, the Cloud computing also focuses on dynamic optimization of shared resources among multiple users. For instance, a Cloud computing resource (such as email) is allotted to western user based on his/her time zone. Through Cloud computing, the same resource is allotted to Asian user as well based on his/her time zone.

2.2. Edge computing

Edge computing directs computational data, applications, and services away from Cloud servers to the edge of a network. The content providers and application developers can use the Edge computing systems by offering the users services closer to them. Edge computing is characterized in terms of high bandwidth, ultra-low latency, and real-time access to the network information that can be used by several applications [23–25].

The service provider can make available the radio access network (RAN) to the Edge users by opening access to new applications and services. Edge computing enables several new services for enterprises and consumers [26]. The use cases of Edge computing are location services, augmented reality, video analytics, and data caching. Thus, these new Edge computing standards and deployment of Edge platforms become key enablers for new revenue streams to vendors, third-parties, and operators.

2.3. Main differences between Cloud and Edge computing

Edge computing is an advanced version of Cloud computing that reduces latency by bringing the services close to the end-users. Edge computing minimizes the load of a cloud by providing resources and services in the Edge network. However, Edge computing complements Cloud computing by enhancing the end user service for delay-sensitive applications [27]. Similar to Cloud, Edge service providers furnish application, data computation, and storage services to the end users.

Despite these service similarities, several considerable differences exist between these two emerging computing paradigms. The main difference between Edge computing and Cloud computing lies in the location of the servers. The services in Edge computing are located in the Edge network whereas the services in the Cloud are located within the Internet.

The availability of the Cloud services within the Internet depends on the distance of multi-hop between the end user and the Cloud servers. The significantly high distance between the mobile device and the Cloud server induces high latency in Cloud computing as compared to the low latency found in Edge computing. Similarly, Cloud computing has high jitter whereas Edge computing has very low jitter. Unlike the Cloud computing, Edge computing is location-aware and high support mobility. Edge computing uses a distributed model for server distribution as compared to Cloud computing that uses a centralized model. The probability of data en-route attacks is higher in Cloud computing than Edge computing caused by the longer path to the server. The targeted users for Cloud computing are general Internet users whereas the targeted service subscribers for Edge computing are the Edge users. Unlike the global scope of Cloud computing, the scope of Edge computing is limited. Lastly, Edge hardware possesses limited capabilities that make it less scalable than the Cloud.

2.4. Edge computing and similar concepts

Edge computing is the extension of Cloud computing where the computing services are brought closer to the end-users at the edge of the network. The Edge vision has been developed to address the issue of high latency in delay-sensitive services and applications that are not properly handled within the Cloud computing paradigm. These applications have the following requirements: (a) very low and predictable latency, (b) location awareness, and (c) mobility support. Although Edge computing provides several advantages over Cloud computing, the research on the emerging domain is still in its infancy. This survey article comprehensively covers all the three aspects of Edge computing, which are Cloudlets, Fog computing and Mobile Edge computing.

According to [15], Edge computing is an autonomous computing model comprising of numerous distributed heterogeneous devices that communicate with the network and perform computing tasks such as storage and processing. These tasks can also support in offering lease based services where a user lease a device and get incentives in return.

According to Cisco, Fog computing constitutes an extension of the Cloud computing paradigm that brings resources and services from the core network in the edge network [10]. It is a virtualized platform that furnishes storage, computation, and networking services in the edge network. Cloudlet [11] and Mobile-Edge computing [28] constitute some of the concepts that are similar to Fog computing paradigm. Cloudlet and Mobile-Edge computing are designed to provide the services only to mobile users with the flexibility of using the locally available resources. However, Fog specifically relies on the hardware designed by Cisco that possesses computational capabilities along with the normal functionality of the device such as router and switches.

3. Edge computing characteristics

Edge computing possesses several characteristics similar to the Cloud computing. However, the distinguishing characteristics of Edge computing that makes it unique are as follows:

3.1. Dense geographical distribution

Edge computing brings the Cloud services closer to the user by deploying numerous computing platforms in the edge networks [29]. The dense geographical distribution of the infrastructure assists in the following ways: (a) The network administrators can facilitate location-based mobility services without traversing the entire WAN; (b) Big data analytics can be performed rapidly with better accuracy [30,31]; (c) The Edge systems enable the real-time analytics on a large scale [32,33]. Examples include sensor networks to monitor the environment and pipeline monitoring.

3.2. Mobility support

As the number of mobile devices is rapidly growing, Edge computing also supports mobility, such as the Locator ID Separation Protocol (LISP), to communicate directly with mobile devices. The LISP protocol decouples the location identity from the host identity and implements a distributed directory system. The decoupling of the host identity from the location identity constitutes the key principle that enables the mobility support in Edge computing.

3.3. Location awareness

The location-awareness attribute of Edge computing allows the mobile users to access services from that Edge server closest to their physical location. Users can employ various technologies such as cell phone infrastructure, GPS, or wireless access points to find the location of electronic devices. This location awareness can be used by several Edge computing applications such as Fog-based vehicular safety applications and Edge-based disaster management.

3.4. Proximity

In Edge computing, computation resources and services are available in the proximity of the users that can improve their experience. The availability of the computational resources and services in the local vicinity allows the users to leverage the network context information for making offloading decisions and service usage decisions. Similarly, the service provider can leverage the mobile user's information by extracting device information and analyzing the user's behavior in order to improve their services and resource allocation.

3.5. Low latency

Edge Computing paradigms bring the computation resources and services closer to the users, which reduces the latency in accessing the services. The low latency of Edge computing enables the users to execute their resource-intensive and delay-sensitive applications on the resource-rich Edge devices (e.g. router, access point, base station, or dedicated server).

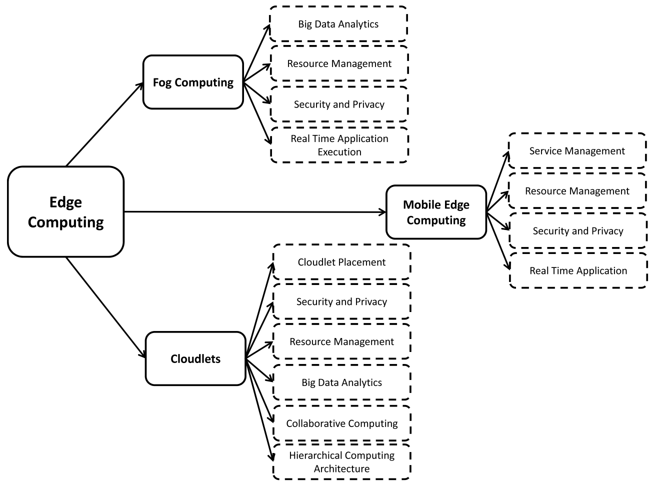

Fig. 2. Classifications of Edge computing.

3.6. Context-awareness

Context-awareness is the characteristic of mobile devices and can be defined interdependently to location awareness. Context information of the mobile device in Edge computing can be used to take offloading decisions and access the Edge services [34]. The real-time network information, such as network load and user location, can be used to offer the context-aware services to the Edge users. Further, the service provider can use the context information to ameliorate the user satisfaction and quality of experience.

3.7. Heterogeneity

Heterogeneity in Edge computing refers to the existence of varied platforms, architectures, infrastructures, computing, and communication technologies used by the Edge computing elements (end devices, Edge servers, and networks). In end device heterogeneity, software, hardware and technology variations constitute the main factors of the heterogeneity. Edge server side-heterogeneity is mainly due to APIs, custom-built policies, and platforms. Such existed differences result in interoperability issues and render it a main challenge in the successful deployment of Edge computing. The network heterogeneity refers to the diversity of communication technologies that impact the Edge service delivery.

4. State-of-the-art on Edge computing

This section aims to critically analyze the literature available on Edge computing paradigms, including Fog computing, Cloudlets, and Mobile Edge computing. Fig. 2 gives a picture of classification of Edge computing and similar concepts.

4.1. Fog computing

In this subsection, we classify the literature produced on Fog computing into real-time application execution, big data analytics, resource management, and security and privacy. Table 2 provides a comparison of Fog computing solutions based on their objectives.

4.1.1. Real-time application execution

J.K. Zao et al. [35] proposed an architecture of augmented brain–computer interfaces (A-BCI) system. The key component of the architecture is the online data server known as the Fog server, which can be installed on television set-top boxes, personal computers, or gaming consoles. Each Fog server works as a signal processor and data hub. Besides that, the server supports preprocessing of the signal, identification of a source, and fitting of an autoregressive model. The availability of the Fog server in the proximity of the user and wireless sensors form a triangular connection among the nodes. The triangular association can be leveraged to augment the flexibility and performance of the system. The proposed system also support machine-to-machine (M2M) publish/subscribe protocols. System security constitutes a paramount issue of the pervasive computing system and is implemented by employing transport layer security protocol.

M. Aazam et al. [36] proposed an emergency alert and management service architecture called emergency help alert mobile cloud (E-HAMC), that leverages on Fog and Cloud platforms. The proposed system handles different types of emergency situations efficiently. In case of an emergency, data is transmitted to the Fog server that sends alert messages to family members and appropriate emergency handling departments. Further, the data is pre-processed and uploaded to the Cloud that analyzes it and also creates an extended services portfolio. The E-HAMC system keeps a list of close family members that eliminates the need to search for family member contact details in time of emergency. The user clicks on the event type, and the remaining task is performed by the application in cooperation with the Fog server. The initial location is exchanged with the Fog server through global positioning.

M.A. Al Faruque et al. [37] presented energy management as a service over the Fog computing platform. The proposed platform is scalable, adaptive and open source that enables the end users to implement energy management as a service with low implementation cost and time-to-market. The hardware of the platform consists of multiple devices based on its operating domain. The devices can be gateways, sensors, actuator, computing and connecting devices. In the home energy management (HEM) platform, multiple devices are being monitored to minimize energy consumption. The main computing device that manages the monitoring and controlling tasks is known as the HEM control panel responsible for discovering and monitoring the devices, handling load requests, and instructing the devices depending upon the algorithms implemented. The service-oriented architecture abstracts the hardware and communication heterogeneity.

4.1.2. Big Data analytics

The proliferation of smart technologies and the deployment of intelligent cyber–physical systems have transformed traditional industries into modern industries. In practice, most factory systems are based on a centralized Cloud infrastructure, which no longer affords heavy computing workload from thousands of smart devices deployed in the factory. In order to resolve this issue, an approach of deploying Fog computing resources located near to the factory devices was introduced in [38]. The proposed approach helps to provide real-time feedback on sites. The study also focused on the deployment of an intelligent computing system comprised of data centers, gateways, Fog/Edge devices in a logistics center. The study also established an integer programming model that minimizes the total installed cost of the Fog-enabled devices. Furthermore, it solved the NP-hard facility location problem through a genetic algorithm. Simulation results verify high performance when deploying intelligent computing systems in a logistics center. Despite the many advantages of the proposed approach, there are some indeterminate factors

Table 2
Comparison of Fog computing solutions based on objectives.

	Minimizing latency	Obtaining real-time data insights	Optimizing resource utilization	Maximizing privacy	Strengthening security
J.K. Zao et al. [35]	✓	✗	✗	✗	✗
M. Aazam et al. [36]	✓	✗	✗	✗	✗
M.A. Al Faruque et al. [37]	✓	✗	✗	✗	✗
C. Lin et al. [38]	✗	✓	✗	✗	✗
R. Iqbal et al. [39]	✗	✓	✗	✗	✗
J. He et al. [40]	✗	✓	✗	✗	✗
S. K. Datta et al. [41]	✗	✓	✗	✗	✗
M. Aazam et al. [42]	✗	✓	✗	✗	✗
J. Preden et al. [43]	✗	✓	✗	✗	✗
L. Yin et al. [44]	✗	✗	✓	✗	✗
B. Jia et al. [45]	✗	✗	✓	✗	✗
L. Gu et al. [46]	✗	✗	✓	✗	✗
M.A. Hassan et al. [47]	✗	✗	✓	✗	✗
K. Hong et al. [48]	✗	✗	✓	✗	✗
F. Bonomi et al. [49]	✗	✗	✓	✗	✗
M. Aazam et al. [50]	✗	✗	✓	✗	✗
C. T. Do et al. [51]	✗	✗	✓	✗	✗
T. Wanget al. [52]	✗	✗	✗	✓	✗
L. Zhang et al. [53]	✗	✗	✗	✓	✗
L. Lyu et al. [54]	✗	✗	✗	✓	✗
A. S. Sohal et al. [55]	✗	✗	✗	✗	✓
S.J. Stolfo [56]	✗	✗	✗	✗	✓

concerning deployment, such as data traffic, latency, energy consumption, load balance, heterogeneity, fairness, and QoS, which were not considered.

R. Iqbal *et al.*[39] proposed a Fog-based data analytics framework to overcome the challenges involved in offering context-aware services in an Internet of vehicles (IoV) environment. The framework consists of the five layers of the data collection layer, the pre-processing layer, the reconstruction layer, the service layer, and the application layer. The first layer is responsible for collecting IoV systems data from distinct sources such as vehicles, road sensors, autonomous cars, pedestrians, and roadside units. The pre-processing layer helps in data extraction, trimming, reconstruction and modeling and provides a baseline to the layers above that extract meaningful information from the given datasets. The analytics layer constitutes the prime layer of the framework as it helps bring data intelligence near to the user network and is based on the Fog computing concept. The fourth layer is responsible for managing and exposing the results of analytics while the application layer is responsible for dealing with different IoV applications based on their QoS requirements. Although the appropriateness of the framework was evaluated through two indispensable use cases, the benefits of the framework were not verified through implementation.

A multitier Fog computing model with the objective of enabling large-scale data analytics for smart city applications was proposed in [40]. The model is comprised of ad hoc and dedicated fogs that mitigate the indispensable problems associated with dedicated computing infrastructure and Cloud computing in terms of latency. The experiments were performed on Raspberry Pi computers based on a distributed computing engine, which helps measure the performance of the analytics tasks and creates easy-to-use workload models. In order to maximize the analytics service utilities, quality-ware admission control, offloading, and resource allocation approaches were recommended. The experimental results revealed that the proposed multitier Fog-based model significantly improves analytics services for smart city applications. However, it lacked a rigorous evaluation and validation of the model.

S. K. Datta et al. [41] presented consumer centric IoT based applications and services that support delay-sensitive operations. They also proposed an architecture comprised of connected vehicles with road side units (RSUs) and M2M gateways. The architecture facilitates the provisioning of several consumer centric services, such as data analytics using semantic web technologies, IoT services discovery, and connected vehicles management. The data generated by smartphones, sensors, and vehicles constitute raw data that must be processed using semantic web technologies. The processed data contains actionable intelligence that can be used by connected vehicles to take smart decisions. In order to support this functionality, the M2M measurement framework is installed on RSUs and M2M gateways that collect the sensor metadata and produce inferred data. The services (including applications) that is offered by the RSUs and M2M gateways is accessed by the IoT services discovery module. The connected vehicles management framework is based on OMA lightweight M2M technical specifications that assists in high mobility of the devices. The framework enables the consumers to update, read and write sensor configurations.

M. Aazam et al. [42] proposed a smart gateway (SG) architecture for Fog computing. SG is responsible for performing different tasks in IoT, such as data collection and preprocessing, data filtering, uploading only necessary data, monitoring IoT objects activities and their energy level, and security and privacy of data. The data collected from IoT can be transmitted either directly or through the base station(s) to the SG. The SG based communication can be classified into two types namely single-hop connectivity and multiple hop-connectivity. In single-hop connectivity, smart objects are directly connected to the gateway that collects data and forwards it to the Fog. This type of communication is suitable for smart sensing devices with restricted roles that are directly connected to the Gateway. When multiple networks of sensor and IoT are deployed, the direct communication with the SG is not possible. Each sensor network has its own sink node, and Gateway collects the data from multiple sink nodes and base stations. These types of scenarios are appropriate for large-scale IoT and mobile objects.

J. Preden et al. [43] presented a Fog computing-based approach for combining data with the decision mechanism while computing on the edge of the user network. This approach enables the use of the sensor-generated big data while considering the bandwidth constraints. The data-to-decision (D2D) concept is used to find and provide actionable data to decision-makers considering their data requirements. The D2D further discusses the data collection and integration of actionable data to extract the suitable situational information for examining threats and decision consequences. The authors suggested a system-of-systems approach for performing the tasks where each single system has autonomy and computational resources to execute the necessary tasks. Although the D2D concept is applicable to all levels of organizations, the concept can also be applied to individuals on the edge who have access to modern communication platforms.

4.1.3. Resource management

Fog computing is capable of providing a variety of services in smart manufacturing as it enables real-time analysis able to diagnose device faults. The devices involved in manufacturing upload their voluminous amounts of data to fog nodes. However, fog nodes have scarce resources in terms of computation and storage, which requires efficient resource utilization mechanisms. Applying virtualization technology to Fog computing can help improve resource utilization and avoid the effects of resource competition. Virtual machine (VM) is regarded as one of the widely adopted virtual technologies, despite its high boot-time and monitoring cost. In order to overcome these issues, a lightweight virtualization technology called container has been introduced. In [44], the researchers investigated task scheduling problems by using the container concept in Fog computing and proposed container-based task scheduling algorithms. The results revealed that the proposed algorithms help improve resource utilization by minimizing latency. However, the study ignored the important factor of finiteness of Cloud resources usually taken into account in the real-scenarios. Additionally, the image placement of containers constitutes a crucial problem that needs to be addressed.

B. Jia et al. [45] formulated the resource allocation problem in Fog computing networks as a double two-sided matching optimization problem and proposed a double-matching strategy based on deferred acceptance algorithm. The numerical analysis result demonstrated that high-cost proficiency can be achieved in terms of resource allocation using the proposed scheme. Furthermore, under the proposed approach, three participants within the Fog networks could achieve stable results. In the future, the work further can be extended to unmanned aerial vehicles.

L. Gu et al. [46] proposed the Fog computing-enabled medical cyber–physical system (FC-MCPS), which provides hosting services to virtual medical device (VMD) applications. In particular, the study focused on efficient resource management problems with the objective of guaranteeing QoS in FC-MCPS. The problem was formulated in a form of mixed-integer non-linear programming (MINLP) involving high computational complexity. Furthermore, a low-complexity two-phase linear programming based heuristic algorithm was proposed. Extensive experimental results show that the algorithm outperforms a greedy approach.

M.A. Hassan et al. [47] proposed an adaptive mobile application offloading and mobile storage augmentation mechanisms in context of Fog computing to assist the mobile application execution. The proposed offloading decision requires the active monitoring of the resources. The decision incorporates the resource dynamics and availability to predict the performance of various tasks execution on different computing platforms. The proposed storage augmentation mechanism combines the available storage space of a user's multiple devices to form a distributed storage service of Fog computing. The active monitoring of resources consumes the resources of mobile devices, and combining the storage space of devices requires the uniform storage management and access mechanism.

K. Hong et al. [48] presented a high level PaaS programming model, called Mobile Fog for emerging mobile applications that are latency-sensitive and geo-spatially distributed. The development of Fog computing-based applications is challenging due to orchestration of highly dynamic heterogeneous resources at various levels of hierarchy. The proposed PaaS programming model mitigates the complexity involved in the Fog-based application development. Mobile Fog supports on-demand allocation of computing instance to support the dynamic workloads of real applications. This enables the feature of dynamic scalability that depends upon user-defined scaling policy. The scaling policy incorporates monitoring metrics including network bandwidth, CPU utilization, and scaling conditions, such as CPU utilization more than 70%. The proposed high level PaaS programming model makes the programming, maintenance, and debugging easier and thus ensures faster prototyping. However, the high-level PaaS program runs slower than the low level programming model and requires more computation time.

F. Bonomi et al. [49] presented a high-level Fog software architecture and various enabling technologies that are vital for realizing the vision of Fog computing. Fog architecture enables the co-existence of applications from different users where each user receives the perception of dedicated resources. The distributed architecture of the Fog relies on the virtualization of different components such as computing, storage, and networking. For automatic management of resources, Fog supports a mechanism which is similar to Cloud based on policy-based provisioning mechanism. The authors also presented the indispensable use cases to highlight the need for Fog within Big Data and IoT space.

M. Aazam et al. [50] formulated a Fog-based service oriented resource management model for an IoT environment. The model aims to fairly and efficiently manage the resources. The authors classified the IoT devices into three groups considering the mobility factor and the nature of the device and manages the resources accordingly. The work mainly focused on resource prediction, estimation, reservation and pricing of resources considering device, and customer type. The resources estimation model is formulated as:

$$\Re = \sum_{i=0}^{n}((U_i * ((1 - \bar{x}(P_i(L|H)_s)) - \sigma^2)) * (1 - \omega_i) * \phi) \tag{1}$$

where $\Re$, U_i, $\bar{x}(P_i(L|H)_s)$, σ^2, ω_i and ϕ represent required resources, basic price of requested service i, average of service oriented relinquish probabilities of particular customer, variance of service oriented relinquish probabilities, constant decision variable value, and the type of accessing device, respectively. Although the proposed model involves resource estimation, it possesses high complexity due to the incorporation of multiple parameters.

C. T. Do et al. [51] focused on convex optimization problem where costs and location diversities are modeled. The authors proposed a scheme based on joint-resource allocation for video streaming service in Fog computing. Besides, the issue of minimizing carbon footprint problem was also analyzed. The joint resource allocation and carbon footprint minimizing problem constitutes a large-scale convex optimization problem as it involves a large number of Fog computing nodes. The authors proposed a distributed solution for attaining the scalability and improving the performance based on a proximal algorithm to divide the large-scale global problem into several sub-problems. Contrary to conventional methods, proximal algorithms converge rapidly and possess robustness without strong assumptions. The social welfare maximization problem in terms of utility and cost function can be written as:

$$\overset{max}{x \geq 0, y \geq 0} \sum_{n=1}^{N} U_n(x_n) - C(y) \quad s.t. \sum_{n=1}^{N} x_n = y \leq \zeta / \kappa \tag{2}$$

where the constraint represents the capacity constraint on data center. $U_n(x_n)$ is the utility function for using x_n units of resources and $C(y)$ is the total cost.

Table 3
Comparison of Cloudlet solutions based on objectives.

Cloudlet solution	Maximizing useful computations	Optimizing Cloudlet placement/selection	Reducing latency	Minimizing cost	Minimizing average response time	Minimizing energy consumption
K. Habak et al. [57]	✓	✗	✗	✗	✗	✗
L. Liu et al. [58]	✗	✓	✗	✗	✗	✗
M. Satyanarayanan et al. [59]	✗	✗	✓	✗	✗	✗
L. Zhao et al. [60]	✗	✓	✓	✗	✗	✗
M. Jia et al. [61]	✗	✓	✗	✗	✗	✗
F. Hao et al.'s [62]	✗	✓	✗	✓	✗	✓
Q. Fan et al. [63]	✗	✗	✗	✗	✓	✗
Z. Xu et al. [64]	✗	✗	✓	✓	✓	✓

4.1.4. Security and privacy

Recent years have seen an explosive growth of unstructured data mostly stored on cloud-based technology. In case of cloud storage, users completely lose control over their data, and thus face privacy leakage risks. Typically, privacy protection solutions are based on encryption technology. However, these schemes are no longer effective enough to resist against attacks launched inside of cloud server. To resolve this issue, a Fog computing-based three-layer privacy-preserving storage scheme was proposed in [52]. The scheme reaps the benefit of cloud storage while ensuring user data protection. Furthermore, a Hash-Solomon algorithm was designed to compute the distribution proportion stored in the cloud, Fog and local machine. The theoretical analysis results validated the feasibility of the proposed scheme. Resource allocation schemes are widely adopted in Fog computing as they ensure optimal utilization of the resources. However, these schemes are vulnerable to cyber and physical attacks. In order to overcome this problem, a privacy-preserving scheme was proposed in [53]. The scheme is based on constant message expansion. Formal analysis results show that the scheme helps to mitigate against eavesdropping and smart gateway attacks as they are mostly launched on resource allocation algorithms. Additionally, the proposed scheme is robust, as it fulfills users' privacy needs by having full key compromise resistance.

L. Lyu et al. [54] proposed a smart metering aggregation framework called privacy-preserving Fog-enabled aggregation (PPFA). The framework is based on Fog computing architecture and additive homographic encryption. PPFA enables smart meters to encrypt the noisy results by adding concentrated Gaussian noise to their data. PPFA guarantees aggregator obliviousness as intermediate nodes are involved. Robustness was ensured by using a two-layer encryption scheme. The first layer helps in achieving aggregator obliviousness, whereas, the second layer is used for authenticating. The results demonstrated that PPFA ensures privacy by saving a significant amount of energy, alleviating bandwidth bottleneck problem, and minimizing network latency. However, the framework still needs to be tested for distributed applications by considering heterogeneous queries.

A. S. Sohal et al. [55] proposed a cybersecurity framework for identifying malicious edge devices in a Fog environment by using the Markov model, intrusion detection system (IDS), and virtual honeypot device (VHD). A two-stage hidden Markov model is used for categorizing four different edge devices (legitimate device (LD), sensitive device (SD), under-attack device (UD) and hacked device (HD)) into four different levels. A secure load balancer (SLB) is used to provide services to all edge devices so that a load of Fog layer resources is balanced. The IDS is used for continuous surveillance of the services which are granted to LDs and prediction of hacked devices. The proposed cybersecurity framework works as follows: First, on the identification of any attack by IDS, attack alarm is generated, and the detection phase is triggered for that edge device. The identification and category of the attacked edge device are sent to Markov 1 which is responsible for calculating the attack probability of the edge device and its category. In the second step, Markov 2 utilizes the information of Markov 1 to predict whether the edge device should be shifted to VHD or not. If any LD is detected on VHD that is shifted mistakenly, LDP of that edge device is calculated using Markov 3, and its value is sent to Markov 4 which is further responsible for predicting whether to shift edge device back to LD status or not. VHD is responsible for storing and maintaining the log repository of all the devices identified as malicious by IDS as well as continuously monitoring the shifted edge devices to VHD. The experimental results show a better response time for achieving high accuracy. However, the important parameters of cost of response and cost of damage which affect the response time are not considered. While achieving high accuracy with a lower cost of damage, the cost of response for the proposed algorithm is high. The cost of response will increase due to the increase and severity of total number of attacks with time.

S.J. Stolfo [56] proposed an approach to secure the Cloud by utilizing decoy information technology. The authors introduced disinformation attacks against malicious insiders using the decoy information technology. Such attacks prevent the malicious insiders from differentiating the real sensitive data from customer fake data. The security of the Cloud services is enhanced by implementing the two security features of user behavior profiling and decoys. User profiling constitutes a common technique used to detect when, how, and how much a user accesses the data in the Cloud. The normal behavior of a user can be regularly monitored to identify abnormal access to a user's data. The decoy consists of worthless data generated to mislead and confuse an intruder into believing that they have accessed the valuable data. The decoy information can be generated in several forms, such as decoy documents, honeypots, and honeyfiles. The decoy approach may be used combined with user behavior profiling technology for ensuring the information security in the Cloud. When abnormal access to the service is observed, fake information in the form of decoy is delivered to the illegitimate user in such a way that it seems completely normal and legitimate. The legitimate user immediately recognizes when decoy data is returned by the Cloud. The inaccurate detection of unauthorized access is informed by varying the Cloud responses using different means such as challenge response to the Cloud security system. The proposed security mechanism for Fog computing still requires further study to analyze false alarms generated by the system and miss detections.

4.2. Cloudlets

In this subsection the literature related to Cloudlets is classified into resource management, Big Data analytics, service management, Cloudlet placement, collaborative computing, and hierarchical computing architecture. Table 3 shows the comparative summary of Cloudlets based schemes.

4.2.1. Resource management

FemtoClouds is a dynamically self-configuring Cloudlet system that coordinates multiple mobile devices to provide computational offloading services [57]. Contrary to traditional Cloudlet phenomena, FemtoClouds utilizes the surrounding mobile devices to perform computational offloading tasks, thereby reducing the network latency. In the proposed system, a device (laptop) acting as a Cloudlet creates a Wi-Fi access point and controls the mobile devices (compute cluster) willing to share computation. Mobile devices can offload a task by giving proper information about the task (such as input and output data size and computational code) to the Cloudlet. Cloudlet first inspects the available computation resources in the compute cluster and the time required to execute the task. Then, the task is scheduled to the available devices using the greedy heuristic approached optimization model. The presented system creates a community cloud with minimal dependent on corporate equipment. The performance of the presented system depends on the number of mobile device participation in the compute cluster.

L. Liu et al. [58] proposed a two-level optimization mechanism to improve resource allocation while fulfilling the user demands in a multi-Cloudlet environment. In the first level, MILP is used to model the optimal Cloudlet selection among multiple available considering the user demand. If the available Cloudlets are unable to meet the user requirements, the request is forwarded to the remote cloud. Then, in second level, MILP is used to develop a resource allocation model for allocating resources from the selected Cloudlet. The Cloudlet selection is based on objective function of optimizing latency and mean reward whereas selection of resource allocation is based on optimizing reward and mean resource usage. Although the authors claim that the proposed system has improved latency and maximized system resources usage and reward, there is a need to study the overhead and complexity before the practical deployment of the system.

4.2.2. Big Data analytics

In order to enable the analytics at the edge of the Internet for high data rate sensors used in IoT, the GigaSight architecture was proposed in [59]. GigaSight is a repository (of crowd-sourced video content) used to enforce measures like access control and privacy preference. The proposed architecture constitutes an amalgamation of VM-based Cloudlets that also offers a scalable approach for data analytics. The aim of the study was to demonstrate that how Edge computing can help improve the performance for IoT-based high-data rate applications.

4.2.3. Cloudlet placement

L. Zhao et al. [60] proposed a network architecture for SDN-based IoT that focuses on optimal placement of Cloudlets to minimize the network access delay. The proposed architecture consists of the two components of control plane and data plane. The control plane is responsible for signaling and network management whereas data plane forwards the data packets. The SDN controller keeps the complete view of the network that helps manage the data transmission. Each access point (AP) forwards the data as guided by the SDN controller. The authors proposed enumeration-based optimal placement algorithm (EOPA) and Ranking-based Near-Optimal Placement algorithm (RNOPA) to solve the problem. In EOPA, all potential cases of k Cloudlets placement are enumerated, and the mean Cloudlet access delay of all cases are compared to obtain the optimal one with minimal average access delay. In RNOPA, those k APs collocated with Cloudlets and most fit are identified before evaluating the entropy weight of each connection feature. Although RNOPA can achieve near-optimal performance in terms of network access delay and network reliability, there is a need to investigate the stability of the Cloudlet system.

Another solution to address the problem of Cloudlet placement was proposed by M. Jia et al. [61] in the form of two heuristic algorithms for optimal Cloudlet placement in a user-dense environment of wireless metropolitan area network (WMAN). A simple heaviest-ap first (HAF) algorithm places a Cloudlet at those APs with the heaviest workloads. However, the drawback of the HAF approach is that the APs with highest workload are not constantly nearest to their users. A density-based clustering (DBC) algorithm is proposed to overcome the shortcoming of the HAF algorithm by deploying Cloudlets in dense users' places. The algorithm also assigns mobile users to the deployed Cloudlets considering their workload. Despite the optimal Cloudlet placement, the proposed solutions do not provide support for user mobility considered as a promising feature of mobile networks.

4.2.4. Collaborative computing

F. Hao et al. [62] proposed a two-layer multi-community framework ($2L$-MC^3) for MEC. The framework allocates tasks in community Clouds/Cloudlet-based environment by incorporating costs (such as monetary, access), energy consumption, security level and average trust level. The problem is optimized in two ways: horizontal optimization and vertical optimization. The horizontal optimization maximizes the Green Security-based Trust-enabled Performance Price ratio whereas vertical optimization minimizes the communication cost between selected community Clouds and Cloudlets. Although the proposed framework minimizes the accessing cost, optimizes the degree of trust and saves the energy, the use of bi-level programming makes the solution complex.

4.2.5. Hierarchical computing architecture

Q. Fan et al. [63] proposed a hierarchical Cloudlet network architecture that allocates incoming user requests to an appropriate Cloudlet. The user's offloading request is forwarded to the tier-1 Cloudlet. If the tier-1 Cloudlet is unable to handle the request; the request is forwarded to the upper tier Cloudlet. If all tier Cloudlets are unable to handle, the request is forwarded to the cloud. A Workload ALLocation (WALL) algorithm is also proposed for the hierarchical Cloudlet network. The proposed scheme facilitates in distributing the users' workload among various levels of Cloudlets and allotting optimal resources to each user. The problem is formulated as an optimization problem that aims at minimizing the average response time by incorporating network delay and computing delay.

4.2.6. Security and privacy

Z. Xu et al. [64] proposed privacy-preserving based offloading and data transmission mechanism for IoT enabled WMANs. For fast processing, offloading and transmission of IoT data, the APs with the shortest path are selected using the Dijkstra Algorithm. The IoT data transmission and task offloading are modeled as a multi-objective problem and solved using non-dominated sorting differential evolution (NSDE) technique. In order to preserve the privacy, the processing of data is offloaded to different Cloudlets based on the attributes of IoT data. The experimental results show that the proposed method has achieved better performance with lower energy consumption and transmission. However, selecting the data attributes for privacy preserving to different Cloudlets results into extra processing overhead which is not considered in evaluations of the proposed mechanism.

4.3. Mobile Edge computing

In this subsection we have classified the literature related to Mobile Edge computing into Resource Management, Service Management, Real-Time Applications and Security Privacy. Table 4 shows the comparative summary of Mobile Edge computing based approaches.

Table 4
Comparison of Mobile Edge computing solutions based on objectives.

Mobile Edge computing solution	Optimizing execution cost	Maximizing throughput	Optimizing deployment	Reducing network latency	Transparent resource augmentation	Minimizing energy consumption
Z. Zhao et al. [65]	✗	✓	✓	✓	✗	✗
M. Chen et al. [66]	✓	✓	✗	✓	✗	✗
S. Wu et al. [67]	✓	✗	✗	✓	✓	✓
Q. Fan et al. [68]	✓	✓	✗	✓	✗	✗
Y. Liu et al. [69]	✓	✗	✗	✓	✓	✗
W. Chen et al. [70]	✗	✗	✗	✗	✓	✓
R. Wang et al. [71]	✓	✓	✗	✓	✓	✗
X. Chen et al. [72]	✓	✗	✓	✗	✓	✓
J. Xing et al. [73]	✓	✗	✗	✗	✗	✗
G. Jia et al. [74]	✓	✗	✗	✗	✗	✓
J. Yang et al. [75]	✓	✓	✗	✓	✗	✗
S. Wang et al. [76]	✗	✗	✓	✗	✗	✗
L. Ma et al. [77]	✗	✗	✗	✗	✗	✓
M. Du et al. [78]	✓	✗	✗	✗	✗	✗
R. Kozik et al. [79]	✓	✗	✗	✗	✓	✗
J. Yuan et al. [80]	✓	✓	✗	✓	✓	✗
R. Rapuzzi et al. [81]	✗	✗	✗	✗	✗	✗
J. Cui et al. [82]	✓	✗	✓	✗	✗	✓

4.3.1. Resource management

S. Wu et al. [67] argued that although Mobile Cloud computing and Mobile Fog computing reduce transmission latency, the containers/virtual machines (VM) are too heavy for Edge/Fog servers due to the resource-constrained environment involving WiFi access points and cellular base stations. In order to address this problem, the authors implemented lightweight run-time offloaded codes for efficient mobile code offloading at the edge server runtime. The proposed method integrates the existing Android libraries into OSv unikernel. The OSv is a library used to convert Android application codes to unikernel. The proposed approach was evaluated using performance metrics of boot-up delay, memory footprint, image size, and energy consumption on single application with promising results. The limitations of the proposed approach include the overhead associated in converting each application to unikernel, which means that the proposed approach is not feasible for multi-process applications.

Q. Fan et al. [68] minimized the total response time of user equipment (UE) request within the network by proposing a scheme known as Application awaRE workload Allocation (AREA). The proposed approach works by allocating different user equipment (IoT based) requests among the least work loaded Cloudlets with minimum network delay between the UE and the Cloudlet. In order to minimize the delay, the UE requests are assigned to the nearest Cloudlets. The proposed approach was simulated on a medium size cluster using the evaluation performance metrics of response time, computing delay, and network layer with significant improvement over existing approaches. However, in real scenario, the simulation set-up seems infeasible as within an area of 25 km, setting up 25 base stations (BS) is impractical. There is high probability of BS overshooting resulting in interference within neighboring BS, which would cause network delay.

Y. Liu et al. [69] studied the use of Edge computing for computation offloading. The aim is to achieve the better quality of service provisioning. The authors have formulated an incentive type of mechanism involving a set of interactions between the cloud service operator and the local edge server to maximize the cloud service operator utilities in the form of Stackelberg game. In Stackelberg game, the cloud server operator (CSO) acts as a leader that specify the payments to the end-server owners (ESOs) and ESO (acts a follower) that determines the computation offloaded for CSO. As this strategy makes ESOs profit-driven, this results in ESO serving only its own users, which ensures efficient computation offloading and enhanced cooperation between the ESOs. However, with the sufficiently large revenue, there will be less computation offloading by ESOs to CSO due to low utility.

W. Chen et al. [70] addressed the problem of computation offloading in Green Mobile Edge Cloud Computing (MECC). Since MECC is composed of wireless devices (WDs), the work load from a mobile device (MD) can be computed using WDs. A multi-user multi-task computation offloading framework for MECC was proposed based on a scheduling mechanism to map the workload from MD to multiple WDs. The Lyapunov optimization approach was used for determining the amount of energy to be harvested at WDs and incoming computation offloading requests into MECC along with WDs assigned to compute the workload for the admitted incoming offloading request. This process maximizes the overall system utility.

R. Wang et al. [71] proposed a knowledge-centric cellular network (KCE) architecture to maximize the network resource utilization in learning-based device-to-device (D2D) communication systems such as social and smart transportation D2D networking systems. The KCE architecture is composed of three layers, namely the physical layer, the knowledge layer, and the virtual management layer. The physical layer collects and manages the user data among the users, the knowledge layer manages the connections between users in D2D communication system, and the virtual management layer is responsible for resource allocation in the D2D assisted network. The proposed approach seems promising in vehicle-to-vehicle communication application (V2V) whereby information can be exchanged through Mobile Edge nodes.

X. Chen et al. [72] proposed a system model for cooperative Mobile Edge computing that combines local device computation and networked resource sharing. Trustworthy cooperation is achieved between different users (e.g. mobile and wearable devices users) through their social relationships developed using the device social graph model. Cooperation among devices helps in processing and execution of different offloaded tasks and network resource sharing using the bipartite matching based algorithm. The processing power of local mobile devices and available network resources in close proximity are utilized for a different type of task executions schemes including, local and offloaded tasks such as D2D, D2D assisted cloud and direct cloud. The experimental results show cost reduction in computation and tasks offloading. However, in simulation, the maximum number of devices is set to be 500 and for each trial the simulations are run only 100 times, which is insufficient for obtaining a reliable result. For the successful practical implementation of the proposed algorithm, the protocol must run for average 10,000 times and involve large social communities of devices ($\geq 10,000$).

J. Xing et al. [73] proposed the distributed multi-level storage (DMLS) model and multiple-factors least frequently used (mFlu) replacement algorithm to address the emerging problem of limited storage and data loss within Edge computing. The proposed model uses intelligent terminal devices (ITD's) such as sensors, actuators as storage devices within the edge up to nth level from user-end to cloud. Once, the storage capacity of the single edge node is exhausted, the mFLU replacement algorithm takes the part of some data from that node and transfers them to the upper nodes. The mFLU algorithm is based on traditional replacement algorithms. For evaluation purposes, a simulation was carried out by creating six level edge storage system (having levels E, S1, S2, S3, S4 and cloud) using a bottom–top approach. The proposed approach seems promising for small scale clusters (like 31 nodes in this study) due to stable upper nodes and fixed data block size. However, the data loss may be higher within large clusters due to the frequent instability of the upper nodes and the variable data block size.

G. Jia et al. [74] addressed the core issue of data storage and retrieval within the Edge computing paradigm. A new cache policy namely Hybrid-Least, a recently used cache (LRU) based on PDRAM memory architecture to solve the issues caused by phase change memory (PRAM-a non-volatile, high-density storage), was proposed. PRAM exhibits the severe drawback of limited service life compared to DRAM although it helps safeguard the data. Hence, a hybrid approach based on PRAM and DRAM was proposed. The approach works by identifying the cache block area and places DRAM blocks at the back of the cache list and PRAM blocks at the head of the cache list. It extends the existing LRU cache policy where the most frequently used data is kept at the top and least used data is pushed to the tail until it is replaced. The results show improvement of 4.6% on PDRAM architecture. However, there is still room to improve the utilization of PRAM that achieved only 11.8% in this study.

J. Yang et al. [75] proposed an Edge computing-based data exchange accounting system for smart IoT toys. The prototype was developed using Hyperledger Fabric-based blockchain system. The data exchange and the payments are done between data demanders and suppliers through consortium blockchain based smart contracts. The consensus and validations of data exchange records and billing among peers are also performed through smart contracts. The proposed accounting system is responsible for safe and reliable interactions among smart toys and other IoT devices. The architecture of Edge computing-based data exchange (EDEC) consists of the five steps of member registration, data products release, order generation, data transmission, and accounting and payment. All transactions between the suppliers and demanders are recorded in four different logs in a distributed fashion and stored on EDCE servers in a centralized manner. For better performance and to support a high number of requests (high throughput), the proposed prototype system requires more rigorous experimental evaluation.

4.3.2. Service management

S. Wang et al. [76] discussed the quality of service (QoS) issue for context-aware services such as service recommendation . The authors highlighted the importance of QoS prediction in the service recommendation system during user mobility. They proposed a service recommendation approach based on collaborative filtering to make QoS prediction possible using user mobility. The proposed scheme works by calculating the user/edge server similarity based on the user's changing locations. Finally, to decrease data volatility, Top-K most similar neighbors are selected for QoS prediction. The proposed approach shows improved results compared to previous approaches. It is suitable when the density of the Edge server nodes is sparse, yet for larger networks involving more Edge servers, the similarity computation might produce incorrect results.

4.3.3. Real-time applications

Z. Zhao et al. [65] proposed the three-phase deployment approach for reducing the number of Edge servers to improve throughput between IoT and Edge nodes. The aim is to support real-time processing for a large-scale IoT network. The proposed approach comprises of discretization, a utility metric, and the deployment algorithm. In the discretization phase, the whole network is discretized into small sections, and the candidate node is identified from the centroid of each section. The performance gained of each candidate node along with its link quality and correlation is evaluated in utility metric phase. Finally, the best node is deployed within the network resulting in maximum throughput. The approach was evaluated using the performance metric of throughput with significant improvement in throughput for heterogeneous data. However, in case of applications requiring homogeneous amount of data, the proposed approach may not be suitable due to the accountability of traffic diversity parameter.

M. Chen et al. [66] proposed an Edge and cognitive computing (ECC) based smart healthcare system for patients in emergency situations. Cognitive computing is utilized for analyzing and monitoring the physical health conditions of patients in emergency situations. Based on the health condition of the patient, the processing resources of Edge devices are allocated. Cognitive computing processing and analysis are performed on Edge devices for low latency and fast processing. The proposed ECC system architecture consists of the two modules of data and resource cognitive engines. The role of data cognitive engine is to collect and analyze the data. Different types of internal and external data are collected by the data cognitive engine, which includes patient physical and behavioral data, type of network, data flow, and communication channel quality. The resource cognitive engine is responsible for collecting the information related to the available Edge, Cloud, and network resources before being sent to the data cognitive engine for the allocation of resources according to the requirement and risk level of the patient. However, systems related to healthcare face the challenge of user privacy. Furthermore, the proposed system lacks privacy protections of sensitive user data. The proposed system calculates the health risk level of each user based on his or her sensitive information, which is stored and processed on Edge systems that may leak.

4.3.4. Security and privacy

L. Ma et al. [77] proposed two reputation-based privacy preserving schemes called Basic Privacy Preserving Reputation Management (B-PPRM) and Advanced Privacy Preserving Reputation Management (A-PPRM) for Edge computing-based mobile crowd-sensing systems. The crowd-sensing system works in two phases. In the first phase, the central manager (CM) distribute sensing task among the participants of the system when in a particular area a sensing task is required. The participant who performs the sensing task encrypts its sense data and sends it to the CM. In the second phase, the participant receives the reputation value as a reward by the reputation manager (RM) with the help of an Edge node. In B-PPRM, the encrypted deviation of the sense data from aggregated sense data is computed by an Edge node, which sends it to the RM. After receiving the encrypted deviation results, the participant node reputation value is updated by the RM. In order to preserve the privacy of the participant, the sensed data are encrypted and accessible only for those participants of the crowd-sensing system who can also perform the sensing task. In A-PPRM, the rank of encrypted deviation results of the participant is calculated by the Edge node and the RM using DGK protocol. The reputation values are updated based on the deviation rank by the RM. The original sensed data are not available to malicious participants who can only access the deviation values. The experimental results show that the malicious participants are successfully handled by the proposed algorithms. However, preserving privacy through reputation management introduces additional computational cost as the reputation manager has to update the reputation values and calculate the rank of the deviations of the participants.

M. Du et al. [78] proposed two privacy-preserving algorithms called Output Perturbation (OPP) and Objective Perturbation (OJP). The authors have identified privacy issues in training data of machine learning algorithms processed and computed by Edge nodes. In most cases privacy can be violated through correlated datasets. The proposed algorithms preserve training data privacy by employing differential privacy through machine learning. In their proposed algorithm, they added noise in wireless big data by applying the Laplace mechanism to the smart edge. The proposed algorithms are compared with stochastic gradient descent (SGD) and private aggregation of teacher ensembles (PATE-G) using the four datasets CIFAR-10, MNIST, STL-10, and SVHN and Tensorflow. Their proposed algorithms achieved accurate privacy preserving, which also means, however, that a larger privacy budget is required and that the data utility decreases.

R. Kozik et al. [79] proposed a distributed attack detection system for Cloud and Edge assisted IoT applications. The proposed system is based on machine learning models and uses Extreme Learning Machine (ELM) classifiers for network traffic analysis and classification. The network traffic classification and inspection tasks are performed on Edge devices while those tasks that require high processing and storage are performed on Cloud servers. The ELM classifiers work in two stages for processing the data. In the first stage, the data is arranged according to its features into multi-dimensional and random order. In the second stage, the machine learning classifier is trained to arrange the data for normal and threat models. The tasks are then easily divided for processing on Cloud and Edge devices. The cloud services are mostly used for processing and storage of collected data, but the anonymized data is sent to Cloud for analysis and storage. ELM classifiers are trained using Cloud storage data and are then used by Edge devices for anomalies and attack detection. The proposed system requires more computational resources in terms of parallel processing nodes in order to reduce the training time for large data sets.

J. Yuan et al. [80] presented a hybrid trust computing approach for the assessment of IoT Edge devices. The proposed framework consists of a combination of three different procedures, namely the global trust calculation using multi-source feedback process, the trust evaluation process based on the cooperation of IoT Edge devices, and the trust factor calculation using the objective information entropy theory-based algorithm. Collaboration among IoT Edge devices is encouraged through incentive-based trust management scheme. For efficient trust calculation, the whole process is done on Edge devices. For calculating the trust and incentives of a particular IoT Edge device, the two factors considered are the interaction with other devices and the quality of service provided by the IoT Edge device. However, the proposed trust mechanism lacks any incentive mechanism, which may prevent the IoT devices from successfully adopting and collaborating with other IoT devices.

R. Rapuzzi et al. [81] discussed the network threats detected in Fog and Edge computing. This work also developed situational awareness to counter these threats during installations of Edge and Fog systems through different layers of the proposed conceptual framework. It consists of the presentation layer, the business logic layer, and the context fabric layer. The presentation layer is responsible for situational awareness, the business logic layer for analysis, processing, and correlation of data, and the context fabric layer for the collection of events from different sources and context information. The authors argued that the analysis of data from different sources and shifting the design of architectures from vertically close to horizontally open could help in identification, situational awareness, and the detection of network threats. However, the proposed framework is conceptual and in order to verify the effectiveness and adaptability of the framework, the detection protocol needs to be developed, implemented, and tested in real-world scenarios.

J. Cui et al. [82] proposed a message authentication scheme using Edge computing for VANETs. The proposed scheme consists of the RSU, trusted authority (TA), and a vehicle equipped with an onboard unit (OBU). A vehicle participating in the message authentication is called Edge computing vehicle (ECV). RSU acts as the cloud of the vehicle and is responsible for quickly verifying the feedback given by the ECVs. After obtaining the legitimacy of these messages, the information is broadcasted to the vehicle through the cuckoo filter. For message authentication, different parts of the vehicle act as Edge computing nodes in order to aid RSU. In order to validate the messages, the vehicle only needs to query the filter, and thus redundant authentication is reduced and the efficiency of the entire system is enhanced even if VANET is attacked. The experimental results show that the proposed scheme achieves better efficiency than other message authentication schemes. However, the use of the cuckoo filter results in higher communication and processing costs.

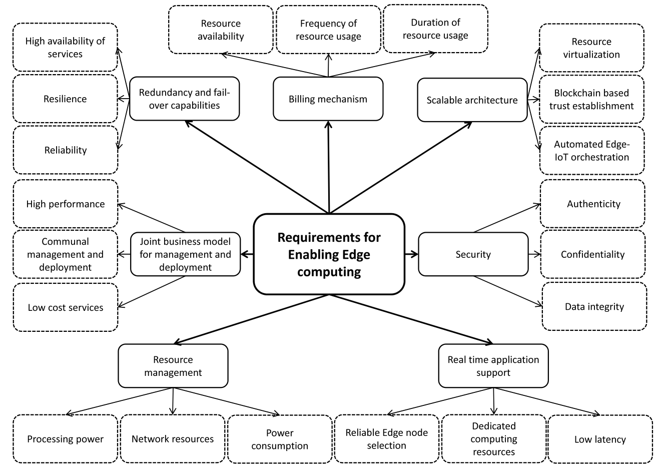

Fig. 3. Requirements for Enabling Edge computing.

5. Requirements for enabling Edge computing

This section highlights key requirements that need to be met in order to utilize Edge computing systems. The requirements for Edge computing can be classified into the Smart Billing Mechanism, Real-time applications support, Joint Business Model for Management and Deployment, Resource Management, Scalable architecture, Redundancy and fail-over capabilities and Security. Each requirement is discussed in detail. Fig. 3 summarizes the requirements for enabling Edge computing systems.

5.1. Dynamic billing mechanism

The involvement of multi-vendor systems (multi-service providers and operators), with user mobility, results in challenging user tariff mechanism. A dynamic billing mechanism is required among various service providers/operators of Edge computing systems when a mobile user is facilitated through their roaming Edge services. The main motivation behind dynamic billing mechanism is the number of different mobile users who may request specific resources (e.g., CPU and memory requirements etc.) with various network parameters (e.g., bandwidth requirement and availability for data transmission, latency, delay and level of security as per application requirements etc.) from the cloud through the Edge based systems. Hence, three factors that should be considered while developing the dynamic billing mechanism include resource availability (e.g., billing charges may vary in case of more users presence which will result in high resources demand), frequency of resource usage (e.g., how frequently a particular user use resources) and duration of resource usage (e.g., charges may differ for users with long duration resource usage as compared to users with short duration resources usage due-to management overheads).

5.2. Real-time applications support

Edge computing is meant to provide numerous services, particularly to real-time based applications. E-learning and gamification constitutes one such area where Edge computing can pay a pivotal role in enhancing the learning process in higher education. Hence, any existing Edge computing-based network must be reliable enough to handle real-time based applications like E-learning and gamification-based applications. Gamification is emerging as the latest trend for enhancing the learning process in education and is expected to replace the traditional mode of learning [83]. Gamification involves embedding game-like elements in courses to enhance learning capabilities among students. In order to gamify and share real-time based courses with individual learning institutions such as universities will incur huge computing cost. Edge computing provides a feasible solution that reduces the cost and computing resources issue for a group of universities. However, handling such type of applications requires the flexibility of Edge nodes that can adjust according to demand. In order to gamify the learning process, the neighboring Edge nodes must be able to select powerful individual Edge nodes that can dedicate their respective services to handle the cluster of universities with each Edge leader node serving per university. This approach minimizes the latency for real-time based gamified modules. The absence of priority-based Edge node selection may increase the latency due to overhead associated in passing data through all the Edge neighboring nodes.

5.3. Joint business model for management and deployment

Edge computing systems are owned by different service providers and work under different business models. Every business runs according to different business strategies and management policies and follows different rules and regulations according to the organization of its operation. Similarly, Edge devices are developed by different vendors and have their own interfaces, which affects the performance of service and incurs high cost. In order to overcome aforementioned issues, a joint management and deployment business model is required to ensure high performance and offer low cost services to end users.

5.4. Resource management

Investment in infrastructure by an individual Edge computing service provider is based on the expected service demands by users. This results in moderate resources to serve a specific number of users. On the other hand, lower service demand results in low revenue. The service providers can increase their revenue by sharing their available resources with interested service providers who have higher service demands. The increase in service demands made by users may result in increased offloading and high optimization between variable computational and network resource allocation [84]. Joint resource management for serving a large number of users among these heterogeneous Edge computing systems with varying processing, memory and network resources constitutes an important requirement. Efficient and simple joint resource management schemes are needed that offer less processing, transmission, power consumption, and balanced delay.

5.5. Scalable architecture

Due to the rapid advancements made in IoT, the number of devices in the Edge network has significantly increased and with it the demand of Edge-based services and resources. The applications running on Edge Cloud are expected to perform consistently despite the increasing load. In order to meet this performance expectation, a scalable Edge computing architecture is considered as vital as it can lower the cost. A scalable architecture for the Edge computing platform can be designed by incorporating various features including resource virtualization, blockchain-based trust establishment, and automated Edge-IoT orchestration. Resource virtualization in IoT involves sharing of IoT resources to multiple applications. These IoT resources (like software based simulators, remote access to servers via abstraction etc.) can be shared through open application package interfaces (API) using software abstraction. By incorporating resource virtualization, each IoT application is deployed as an isolated "slice" that shares the same sensing resources and enables multi-application coexistence. The blockchain and smart contracts enable IoT devices to significantly lower the cost incurred with the trust establishment. Similarly, the automated Edge-IoT orchestration eliminates the manual configuration required for selection, deployment, monitoring, and controlling resources during application execution, thereby satisfying high-level demands.

5.6. Redundancy and fail-over Capabilities

Redundancy and fail-over capabilities constitute two important requirements for the reliable functioning of Edge computing systems that support many critical business applications with low latency and uninterrupted content delivery services. In order to develop reliable and resilient Edge computing systems, redundancy and fail-over capabilities should be considered at the system development stage with a built-in option.

5.7. Security

The heterogeneous nature of Edge computing systems makes security an important requirement for network and application running. Security also constitutes a highly influential factor that motivates the consumers to embrace these new Edge computing systems. Due to the employment of Edge computing systems for applications that require intensive processing, the use of the wireless network and usage of these systems physically by different users increases the risk of intrusion, unauthorized access, and a variety of security attacks [85,86]. Strong and robust security mechanisms with advanced and efficient cryptographic schemes are required for both applications and network to ensure the success of these Edge computing systems. To enable authorizing/trusted access control requires an secure protocol for authentication of applications and mobile devices. Thus, when dealing with the security of Edge computing systems, the common security requirements of availability, authenticity, confidentiality and data integrity should be achieved before deployment, however, without compromising the performance of applications running on these Edge computing systems.

Table 5
Open challenges and guidelines.

Challenges	Causes	Guidelines
User's trust on Edge computing systems	(a) Lack of security and privacy-preserving mechanisms	The influential factors of Consumer trust e.g., security and privacy, can be adopted to deal with challenges in stimulating the consumer's trust upon Edge computing systems.
Dynamic and agile pricing models	(a) High QoS requirements, (b) Inappropriate pricing models, (c) Service provider's high cost.	Considering the three important factors e.g., resource availability, frequency and duration of resource usage by consumers can help in developing dynamic pricing models.
Service discovery, service delivery and mobility	(a) Intermitted connectivity due to mobility, (b) Non-accessibility of local resources, (c) Immature security policies.	The mobility management for wireless networks and Service discovery for peer to peer networks can be used as guidelines.
Collaborations between heterogeneous Edge computing Systems	(a) Heterogeneous architectures, (b) Interoperability problems, (c) Data privacy issues, (e) Deficiencies in terms of load balancing.	The interoperability and collaborations among ubiquitous systems can be used as guidelines.
Low-cost fault tolerant deployment models	(a) High availability, (b) Data integrity, (c) disaster recovery.	Anomaly detection and predictive maintenance through machine Learning can help in providing low-cost fault tolerance.
Security	(a) Involvement of distributed data processing, (b) Indispensable requirements of Edge computing, *e.g.*, distributed architecture, immense data processing capabilities, location-awareness, and mobility support.	Salient features of blockchain technology, *e.g.*, tamper-proof, redundant, and self-healing, can help to mitigate important security threats. Additionally, quantum cryptography based solutions can also be helpful.

6. Open challenges

The complexity of these Edge computing systems has given rise to a number of technical challenges, such as mobility management, security and privacy, scalability, heterogeneity, reliability, and resource management. There are some of the other important challenges that are still needed to be resolved. Table 5 contains the open challenges and factors along with potential guidelines to resolve the mentioned challenge.

6.1. User's trust on edge computing systems

The success of any technology is positively linked with consumer acceptance. Trust is regarded as one of the most important factors for the acceptance and adoption of these Edge systems by the users. As highlighted in the literature, security and privacy are among the basic challenges faced by Edge computing systems/ technologies. Since trust of a consumer is closely associated with the security and privacy of the technologies and thus, if security and privacy of user's data is not addressed well, it will definitely shatter the trust of a consumer, leading to non-acceptance of these edge systems/ technologies. Therefore, research efforts are required to develop consumer trust models for the adoption of Edge computing systems. A recently proposed consumer trust model [87,88] highlights the influential factors (including security and privacy requirements) that stimulate the consumer's trust for adoption and usage of IoT products that can be applied in Edge computing systems.

6.2. Dynamic and agile pricing models

The rapid growth of the Edge computing paradigm has opened up the need for dynamic pricing models that can meet the changing expectations by striking a right balance between the customer's expectations of quality of service, less delay and price, and the service provider's cost and operational efficiency. It is a challenging task to develop dynamic and agile pricing models as one pricing model may not be successful for the engagement of multiple customers. It is also challenging to provide best-fit pricing models for heterogeneous Edge computing systems that can offer mutual benefits for service providers and customers. However, the pricing model for cloud services [89] such as "pay-as-you-go" can be used for developing dynamic pricing models for Edge computing systems.

6.3. Service discovery, service delivery and mobility

Edge computing-as-service helps the individual Edge computing system operators and service providers to fulfill their customer demands with limited processing and memory resources, lower revenue and infrastructure cost. The service discovery in distributed Edge computing systems constitutes a challenging task given the increasing number of mobile devices that require services simultaneously and uninterruptedly. This task becomes more challenging when delay is involved in discovering and selecting the other available services and resources. The automatic and user-transparent discovery of appropriate Edge computing nodes according to required resources in heterogeneous Edge computing systems also poses a challenging task for service discovery mechanisms. However, service discovery solutions proposed for peer-to-peer networks [90,91] can help in the design and development of effective user-transparent solutions for Edge computing systems.

Seamless service delivery constitutes an important mechanism that ensures the uninterrupted and smooth migration of running an application between different Edge computing systems while the consumer is moving [92]. Seamless service delivery with mobility also poses a challenging task as mobility severely affects the different network parameters (latency, bandwidth, delay, and jitter) which ultimately results in application performance degradation. In addition, the non-accessibility of local resources for a mobile user from outside the network due to the implementation of different security policies and billing methods also makes seamless service delivery challenging. The seamless service delivery in the context of mobility poses a vital research problem that needs to be addressed. Effective mobility management mechanisms capable of discovering available resources in a seamless manner are required for supporting the seamless service delivery with mobility. In order to overcome this issue, seamless handoff and mobility management solutions [93–95] for wireless networks, can be used for effective design and development of seamless service delivery mechanisms in Edge computing systems.

6.4. Collaborations between heterogeneous Edge computing systems

The ecosystem of Edge computing systems consists of a collection of different heterogeneous technologies such as Edge data servers (data centers) and different cellular networks (3G, 4G and 5G). Although this heterogeneous nature of the Edge computing network allows Edge devices to access services through multiple wireless technologies such as WiFi, 3G, 4G and 5G, it makes the collaboration between such multi-vendor systems a challenging task [96]. Also, interoperability, synchronization, data privacy, load balancing, heterogeneous resource sharing and seamless service delivery are among the factors which make collaboration between heterogeneous Edge computing systems challenging. The research efforts for interoperability and collaborations in ubiquitous systems, such as reported in [97] can be used for designing and developing efficient collaboration techniques among heterogeneous Edge computing systems.

6.5. Low-cost fault tolerant deployment models

Fault tolerance ensures the continuous operation of any system in the event of failure with little or no human involvement. In Edge computing systems, fault tolerance is achieved through fail-over and redundancy techniques in order to guarantee the high availability of services, data integrity of critical business applications, and disaster recovery of the system in case of catastrophic events by using servers in different physical locations, backup power supply batteries (UPS) and equipment with harsh environmental resistance. However, it is very challenging to provide low-cost fault tolerance deployment models in Edge computing since a remote backup server requires high bandwidth and additional hardware which is very costly. Machine learning-based anomaly detection or predictive maintenance-based systems for power supply batteries/UPS system constitute a cost-effective solution. Predictive maintenance will avoid unscheduled downtime and will effectively reduce the cost required for backup/redundant batteries and UPS.

6.6. Security

Although Edge computing has revolutionized Cloud computing systems by tackling latency issues, it has brought along imperative challenges, especially in terms of security. Ensuring Edge computing security has become a major challenge due to its distributed data processing. Moreover, the distinctive characteristics of Edge computing such as location-awareness, distributed architecture, requirement of mobility support and immense data processing, hinder the traditional security mechanisms to be adopted in the Edge computing paradigm [98]. Security threats associated with Edge computing can be classified into personal point of view (*e.g.*, the user, network operator, and third-party application provider); attribute point of view, (*e.g.*, privacy, integrity, trust, attestation, verification, and measurement); and compliance, dealing with lawful access to data and local regulations.[1] In order to cope with the aforementioned security issues in Edge computing, numerous indispensable solutions with respect to identity and authentication, access control systems, intrusion detection systems, privacy, trust management, visualization, and forensics, have to be developed. The intrinsic characteristics of blockchain technology such as tamper-proof, redundant, and self-healing [99–101], can help meet certain security objectives while it engenders new challenges that have to be addressed. Additionally, quantum cryptography based solutions can also be adopted in the Edge computing paradigm.

[1] Accessed on: 17 December 2018 https://www.sdxcentral.com/edge/definitions/mec-security.

7. Conclusion

Edge computing envisions to bring services and utilities of Cloud computing closer to the end user for ensuring fast processing of data-intensive applications. In this paper, we comprehensively studied the fundamental concepts related to Cloud and Edge computing. We categorized and classified the state-of-the-art in Edge computing (Cloudlets, Fog and Mobile Edge computing) according to the application domain. The application domain areas include services such as real-time applications, security, resource management, and data analytics. We presented key requirements that need to be met in order to enable Edge computing. Furthermore, we identified and discussed several open research challenges.

This study concludes that the state-of-the-art in Edge computing paradigm suffers from several limitations due to imperative challenges remaining to be addressed. Those limitations can be compensated by proposing suitable solutions and fulfilling the requirements such as dynamic billing mechanism, real time application support, joint business model for management and deployment, resource management, scalable architecture, redundancy and fail-over capabilities, and security. This study serves as an excellent material to future researchers to comprehend the Edge computing paradigm and take the research forward to resolve the unaddressed issues. Our future research aims to explore the research trends in Multi-access Edge computing networks.

References

[1] N. Hassan, S. Gillani, E. Ahmed, I. Yaqoob, M. Imran, The role of edge computing in Internet of Things, IEEE Commun. Mag. (99) (2018) 1–6.
[2] M. Liu, F.R. Yu, Y. Teng, V.C. Leung, M. Song, Distributed resource allocation in blockchain-based video streaming systems with Mobile Edge computing, IEEE Trans. Wirel. Commun. 18 (1) (2019) 695–708.
[3] E. Ahmed, A. Akhunzada, M. Whaiduzzaman, A. Gani, S.H. Ab Hamid, R. Buyya, Network-centric performance analysis of runtime application migration in Mobile Cloud computing, Simul. Model. Pract. Theory 50 (2015) 42–56.
[4] P. Pace, G. Aloi, R. Gravina, G. Caliciuri, G. Fortino, A. Liotta, An edge-based architecture to support efficient applications for healthcare industry 4.0, IEEE Trans. Ind. Inf. 15 (1) (2019) 481–489.
[5] U. Shaukat, E. Ahmed, Z. Anwar, F. Xia, Cloudlet deployment in local wireless networks: Motivation, architectures, applications, and open challenges, J. Netw. Comput. Appl. 62 (2016) 18–40.
[6] I. Yaqoob, E. Ahmed, A. Gani, S. Mokhtar, M. Imran, S. Guizani, Mobile ad hoc cloud: A survey, Wirel. Commun. Mob. Comput. 16 (16) (2016) 2572–2589.
[7] W. Bao, D. Yuan, Z. Yang, S. Wang, W. Li, B.B. Zhou, A.Y. Zomaya, Follow me fog: Toward seamless handover timing schemes in a fog computing environment, IEEE Commun. Mag. 55 (11) (2017) 72–78, http://dx.doi.org/10.1109/MCOM.2017.1700363.
[8] E. Ahmed, M.H. Rehmani, Mobile edge computing: opportunities, solutions, and challenges, Future Gener. Comput. Syst. 70 (2017) 59–63.
[9] Y. Jararweh, A. Doulat, O. AlQudah, E. Ahmed, M. Al-Ayyoub, E. Benkhelifa, The future of mobile cloud computing: integrating cloudlets and mobile edge computing, in: Telecommunications (ICT), 2016 23rd International Conference on, IEEE, 2016, pp. 1–5.
[10] Cisco fog computing solutions: Unleash the power of the Internet of Things. Available at: https://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-solutions.pdf. (Accessed on 23 July 2018).
[11] M. Satyanarayanan, P. Bahl, R. Caceres, N. Davies, The case for VM-based cloudlets in mobile computing, IEEE Pervasive Comput. 8 (4) (2009) 14–23, http://dx.doi.org/10.1109/mprv.2009.82.
[12] A. Ahmed, E. Ahmed, A survey on mobile edge computing, in: 10th International Conference on Intelligent Systems and Control, ISCO, 2016, pp. 1–8, http://dx.doi.org/10.1109/ISCO.2016.7727082.
[13] Z. Zhang, W. Zhang, F. Tseng, Satellite mobile edge computing: Improving QoS of high-speed satellite-terrestrial networks using edge computing techniques, IEEE Netw. 33 (1) (2019) 70–76, http://dx.doi.org/10.1109/MNET.2018.1800172.
[14] S. Yi, C. Li, Q. Li, A survey of fog computing: concepts, applications and issues, in: Proceedings of the 2015 Workshop on Mobile Big Data, ACM, 2015, pp. 37–42.

[15] L.M. Vaquero, L. Rodero-Merino, Finding your way in the fog: Towards a comprehensive definition of fog computing, ACM SIGCOMM Comput. Commun. Rev. 44 (5) (2014) 27–32.
[16] I. Stojmenovic, S. Wen, The fog computing paradigm: Scenarios and security issues, in: Computer Science and Information Systems, FedCSIS, 2014 Federated Conference on, IEEE, 2014, pp. 1–8.
[17] F. Bonomi, R. Milito, J. Zhu, S. Addepalli, Fog computing and its role in the Internet of Things, in: Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, ACM, 2012, pp. 13–16.
[18] T.H. Luan, L. Gao, Z. Li, Y. Xiang, G. Wei, L. Sun, Fog computing: Focusing on mobile users at the edge, arXiv:1502.01815v3.
[19] E. Ahmed, A. Ahmed, I. Yaqoob, J. Shuja, A. Gani, M. Imran, M. Shoaib, Bringing computation closer toward the user network: Is edge computing the solution? IEEE Commun. Mag. 55 (11) (2017) 138–144.
[20] P. Mell, T. Grance, The NIST definition of cloud computing. Available at: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf. (Accessed on 23 July 2018).
[21] S. Hakak, S.A. Latif, G. Amin, A review on mobile cloud computing and issues in it, Int. J. Comput. Appl. Technol. 75 (11) (2013).
[22] I. Yaqoob, E. Ahmed, A. Gani, S. Mokhtar, M. Imran, Heterogeneity-aware task allocation in mobile ad hoc cloud, IEEE Access 5 (2017) 1779–1795.
[23] P. Wang, C. Yao, Z. Zheng, G. Sun, L. Song, Joint task assignment, transmission and computing resource allocation in multi-layer mobile edge computing systems, IEEE Internet Things J. (2019) http://dx.doi.org/10.1109/JIOT.2018.2876198, in press.
[24] Y. Sahni, J. Cao, L. Yang, Data-aware task allocation for achieving low latency in collaborative edge computing, IEEE Internet Things J. (2019) http://dx.doi.org/10.1109/JIOT.2018.2886757, in press.
[25] J. Ren, Y. He, G. Huang, G. Yu, Y. Cai, Z. Zhang, An edge-computing based architecture for mobile augmented reality, IEEE Netw. (2019) 12–19, http://dx.doi.org/10.1109/MNET.2018.1800132, in press.
[26] Z. Ning, X. Kong, F. Xia, W. Hou, X. Wang, Green and sustainable cloud of things: Enabling collaborative edge computing, IEEE Commun. Mag. 57 (1) (2019) 72–78.
[27] W. Li, Z. Chen, X. Gao, W. Liu, J. Wang, Multi-model framework for indoor localization under mobile edge computing environment, IEEE Internet Things J. (2019) http://dx.doi.org/10.1109/JIOT.2018.2872133, in press.
[28] M. Patel, et al., Mobile Edge computing, 2010. Available at: https://portal.etsi.org/Portals/0/TBpages/MEC/Docs/Mobile-edge_Computing_-_Introductory_Technical_White_Paper_V1%2018-09-14.pdf. (Accessed on 23 July 2018).
[29] M. Satyanarayanan, How we created Edge computing, Nature Electron. 2 (1) (2019) 42.
[30] E. Ahmed, I. Yaqoob, I.A.T. Hashem, I. Khan, A.I.A. Ahmed, M. Imran, A.V. Vasilakos, The role of Big Data analytics in internet of things, Comput. Netw. 129 (2017) 459–471.
[31] H. Khelifi, S. Luo, B. Nour, A. Sellami, H. Moungla, S.H. Ahmed, M. Guizani, Bringing deep learning at the edge of information-centric internet of things, IEEE Commun. Lett. 23 (1) (2019) 52–55.
[32] S. Jošilo, G. Dán, Selfish decentralized computation offloading for Mobile Cloud computing in dense wireless networks, IEEE Trans. Mob. Comput. 18 (1) (2019) 207–220, http://dx.doi.org/10.1109/TMC.2018.2829874.
[33] A. Ferdowsi, U. Challita, W. Saad, Deep learning for reliable mobile edge analytics in intelligent transportation systems: An overview, IEEE Veh. Technol. Mag. 14 (1) (2019) 62–70.
[34] B. Han, S. Wong, C. Mannweiler, M.R. Crippa, H.D. Schotten, Context-awareness enhances 5G multi-access edge computing reliability, IEEE Access 7 (2019) 21290–21299.
[35] J.K. Zao, T.T. Gan, C.K. You, S.J.R. Méndez, C.E. Chung, Y. Te Wang, T. Mullen, T.P. Jung, Augmented brain computer interaction based on fog computing and linked data, in: Intelligent Environments, IE, 2014 International Conference on, IEEE, 2014, pp. 374–377.
[36] M. Aazam, E.-N. Huh, E-hamc: Leveraging fog computing for emergency alert service, in: Pervasive Computing and Communication Workshops, PerCom Workshops, 2015 IEEE International Conference on, IEEE, 2015, pp. 518–523.
[37] M.A. Al Faruque, K. Vatanparvar, Energy management-as-a-service over fog computing platform, IEEE Internet Things J. 3 (2) (2016) 161–169.
[38] C. Lin, J. Yang, Cost-efficient deployment of fog computing systems at logistics centers in industry 4.0, IEEE Trans. Ind. Inf. 14 (10) (2018) 4603–4611.
[39] R. Iqbal, T.A. Butt, M.O. Shafique, M.W.A. Talib, T. Umer, Context-aware data-driven intelligent framework for fog infrastructures in internet of vehicles, IEEE Access 6 (2018) 58182–58194.
[40] J. He, J. Wei, K. Chen, Z. Tang, Y. Zhou, Y. Zhang, Multitier fog computing with large-scale iot data analytics for smart cities, IEEE Internet Things J. 5 (2) (2018) 677–686.
[41] S.K. Datta, C. Bonnet, J. Haerri, Fog computing architecture to enable consumer centric internet of things services, in: International Symposium on Consumer Electronics, ISCE, IEEE, 2015, pp. 1–2.
[42] M. Aazam, E.-N. Huh, Fog computing and smart gateway based communication for cloud of things, in: International Conference on Future Internet of Things and Cloud, IEEE, 2014, pp. 464–470.
[43] J. Preden, J. Kaugerand, E. Suurjaak, S. Astapov, L. Motus, R. Pahtma, Data to decision: pushing situational information needs to the edge of the network, in: Cognitive Methods in Situation Awareness and Decision Support, CogSIMA, 2015, pp. 158–164.
[44] L. Yin, J. Luo, H. Luo, Tasks scheduling and resource allocation in fog computing based on containers for smart manufacturing, IEEE Trans. Ind. Inf. 14 (10) (2018) 4712–4721.
[45] B. Jia, H. Hu, Y. Zeng, T. Xu, Y. Yang, Double-matching resource allocation strategy in fog computing networks based on cost efficiency, J. Commun. Netw. 20 (3) (2018) 237–246.
[46] L. Gu, D. Zeng, S. Guo, A. Barnawi, Y. Xiang, Cost efficient resource management in fog computing supported medical cyber-physical system, IEEE Trans. Emerg. Top. Comput. 5 (1) (2017) 108–119.
[47] M.A. Hassan, M. Xiao, Q. Wei, S. Chen, Help your mobile applications with fog computing, in: 12th Annual International Conference on Sensing, Communication, and Networking - Workshops, SECON Workshops, IEEE, 2015, pp. 1–6.
[48] K. Hong, D. Lillethun, U. Ramachandran, B. Ottenwalder, B. Koldehofe, 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 - MCC 13, ACM, 2013, pp. 15–20.
[49] F. Bonomi, R. Milito, P. Natarajan, J. Zhu, Fog computing: A platform for Internet of Things and analytics, in: Big Data and Internet of Things: a Roadmap for Smart Environments, Springer, 2014, pp. 169–186.
[50] M. Aazam, E.-N. Huh, Fog computing micro datacenter based dynamic resource estimation and pricing model for IoT, in: 2015 IEEE 29th International Conference on Advanced Information Networking and Applications, IEEE, 2015, pp. 687–694, http://dx.doi.org/10.1109/aina.2015.254.
[51] C.T. Do, N.H. Tran, C. Pham, M.G.R. Alam, J.H. Son, C.S. Hong, A proximal algorithm for joint resource allocation and minimizing carbon footprint in geo-distributed fog computing, in: Information Networking, ICOIN, 2015 International Conference on, IEEE, 2015, pp. 324–329.
[52] T. Wang, J. Zhou, X. Chen, G. Wang, A. Liu, Y. Liu, A three-layer privacy preserving cloud storage scheme based on computational intelligence in fog computing, IEEE Trans. Emerg. Top. Comput. Intell. 2 (1) (2018) 3–12.
[53] L. Zhang, J. Li, Enabling robust and privacy-preserving resource allocation in fog computing, IEEE Access 6 (2018) 50384–50393.
[54] L. Lyu, K. Nandakumar, B. Rubinstein, J. Jin, J. Bedo, M. Palaniswami, PPFA: Privacy preserving fog-enabled aggregation in smart grid, IEEE Trans. Ind. Inf. 14 (8) (2018) 3733–3744, http://dx.doi.org/10.1109/TII.2018.2803782.
[55] A.S. Sohal, R. Sandhu, S.K. Sood, V. Chang, A cybersecurity framework to identify malicious edge device in fog computing and cloud-of-things environments, Comput. Secur. 74 (2018) 340–354.
[56] S.J. Stolfo, M.B. Salem, A.D. Keromytis, Fog computing: Mitigating insider data theft attacks in the cloud, in: 2012 IEEE Symposium on Security and Privacy Workshops, IEEE, 2012, pp. 125–128, http://dx.doi.org/10.1109/spw.2012.19.
[57] K. Habak, M. Ammar, K.A. Harras, E. Zegura, Femto clouds: Leveraging mobile devices to provide cloud service at the edge, in: 2015 IEEE 8th International Conference on Cloud Computing, 2015, pp. 9–16, http://dx.doi.org/10.1109/CLOUD.2015.12.
[58] L. Liu, Q. Fan, Resource allocation optimization based on mixed integer linear programming in the multi-cloudlet environment, IEEE Access 6 (2018) 24533–24542, http://dx.doi.org/10.1109/ACCESS.2018.2830639.
[59] M. Satyanarayanan, P. Simoens, Y. Xiao, P. Pillai, Z. Chen, K. Ha, W. Hu, B. Amos, Edge analytics in the Internet of Things, IEEE Pervasive Comput. 14 (2) (2015) 24–31, http://dx.doi.org/10.1109/MPRV.2015.32.
[60] L. Zhao, W. Sun, Y. Shi, J. Liu, Optimal placement of cloudlets for access delay minimization in SDN-based Internet of Things networks, IEEE Internet Things J. 5 (2) (2018) 1334–1344, http://dx.doi.org/10.1109/JIOT.2018.2811808.
[61] M. Jia, J. Cao, W. Liang, Optimal cloudlet placement and user to cloudlet allocation in wireless metropolitan area networks, IEEE Trans. Cloud Comput. 5 (4) (2017) 725–737, http://dx.doi.org/10.1109/TCC.2015.2449834.
[62] F. Hao, D. Park, J. Kang, G. Min, 2L-MC^3: A two-layer multi-community-cloud/cloudlet social collaborative paradigm for mobile edge computing, IEEE Internet Things J. (2018) http://dx.doi.org/10.1109/JIOT.2018.2867351, in press.
[63] Q. Fan, N. Ansari, Workload allocation in hierarchical Cloudlet networks, IEEE Commun. Lett. 22 (4) (2018) 820–823.

[64] Z. Xu, R. Gu, T. Huang, H. Xiang, X. Zhang, L. Qi, X. Xu, An IoT-oriented offloading method with privacy preservation for Cloudlet-enabled wireless metropolitan area networks, Sensors 18 (9) (2018) 3030.
[65] Z. Zhao, G. Min, W. Gao, Y. Wu, H. Duan, Q. Ni, Deploying Edge computing nodes for large-scale IoT: A diversity aware approach, IEEE Internet Things J. 5 (5) (2018) 3606–3614, http://dx.doi.org/10.1109/JIOT.2018.2823498.
[66] M. Chen, W. Li, Y. Hao, Y. Qian, I. Humar, Edge cognitive computing based smart healthcare system, Future Gener. Comput. Syst. 86 (2018) 403–411.
[67] S. Wu, C. Mei, H. Jin, D. Wang, Android unikernel: Gearing mobile code offloading towards Edge computing, Future Gener. Comput. Syst. 86 (2018) 694–703.
[68] Q. Fan, N. Ansari, Application aware workload allocation for Edge computing-based IoT, IEEE Internet Things J. 5 (3) (2018) 2146–2153.
[69] Y. Liu, C. Xu, Y. Zhan, Z. Liu, J. Guan, H. Zhang, Incentive mechanism for computation offloading using Edge computing: a stackelberg game approach, Comput. Netw. 129 (2017) 399–409.
[70] W. Chen, D. Wang, K. Li, Multi-user multi-task computation offloading in Green Mobile Edge Cloud computing, IEEE Trans. Serv. Comput. (2019) http://dx.doi.org/10.1109/TSC.2018.2826544, in press.
[71] R. Wang, J. Yan, D. Wu, H. Wang, Q. Yang, Knowledge-centric Edge computing based on virtualized D2D communication systems, IEEE Commun. Mag. 56 (5) (2018) 32–38.
[72] X. Chen, Z. Zhou, W. Wu, D. Wu, J. Zhang, Socially-motivated cooperative mobile edge computing, IEEE Netw. 32 (6) (2018) 177–183, http://dx.doi.org/10.1109/MNET.2018.1700354.
[73] J. Xing, H. Dai, Z. Yu, A distributed multi-level model with dynamic replacement for the storage of smart Edge computing, J. Syst. Archit. 83 (2018) 1–11.
[74] G. Jia, G. Han, H. Xie, J. Du, Hybrid-LRU caching for optimizing data storage and retrieval in edge computing-based wearable sensors, IEEE Internet Things J. (2019) in press.
[75] J. Yang, Z. Lu, J. Wu, Smart-toy-edge-computing-oriented data exchange based on blockchain, J. Syst. Archit. 87 (2018) 36–48.
[76] S. Wang, Y. Zhao, L. Huang, J. Xu, C.-H. Hsu, QoS prediction for service recommendations in Mobile Edge computing, J. Parallel Distrib. Comput. (2017) 1–11, http://dx.doi.org/10.1016/j.jpdc.2017.09.014.
[77] L. Ma, X. Liu, Q. Pei, Y. Xiang, Privacy-preserving reputation management for Edge computing enhanced mobile crowdsensing, IEEE Trans. Serv. Comput. (2019) http://dx.doi.org/10.1109/TSC.2018.2825986, in press.
[78] M. Du, K. Wang, Z. Xia, Y. Zhang, Differential privacy preserving of training model in wireless big data with edge computing, IEEE Trans. Big Data (2018) http://dx.doi.org/10.1109/TBDATA.2018.2829886, in press.
[79] R. Kozik, M. Choraś, M. Ficco, F. Palmieri, A scalable distributed machine learning approach for attack detection in edge computing environments, J. Parallel Distrib. Comput. 119 (2018) 18–26.
[80] J. Yuan, X. Li, A reliable and lightweight trust computing mechanism for iot edge devices based on multi-source feedback information fusion, IEEE Access 6 (2018) 23626–23638.
[81] R. Rapuzzi, M. Repetto, Building situational awareness for network threats in fog/edge computing: Emerging paradigms beyond the security perimeter model, Future Gener. Comput. Syst. 85 (2018) 235–249.
[82] J. Cui, L. Wei, J. Zhang, Y. Xu, H. Zhong, An efficient message-authentication scheme based on edge computing for vehicular ad hoc networks, IEEE Trans. Intell. Transp. Syst. (2018) 1–12, http://dx.doi.org/10.1109/TITS.2018.2827460.
[83] S. Hakak, N.F.M. Noor, M.N. Ayub, H. Affal, N. Hussin, E. Ahmed, M. Imran, Cloud-assisted gamification for education and learning–recent advances and challenges, Comput. Electr. Eng. 74 (2019) 22–34.
[84] L. Lin, P. Li, X. Liao, H. Jin, Y. Zhang, Echo: An edge-centric code offloading system with quality of service guarantee, IEEE Access 7 (2019) 5905–5917.
[85] T. Wang, G. Zhang, A. Liu, M.Z.A. Bhuiyan, Q. Jin, A secure iot service architecture with an efficient balance dynamics based on Cloud and Edge computing, IEEE Internet Things J. (2019) http://dx.doi.org/10.1109/JIOT.2018.2870288, in press.
[86] Q. Li, S. Meng, S. Zhang, J. Hou, L. Qi, Complex attack linkage decision-making in Edge computing networks, IEEE Access 7 (2019) 12058–12072, http://dx.doi.org/10.1109/ACCESS.2019.2891505.
[87] W.Z. Khan, M.Y. Aalsalem, M.K. Khan, Q. Arshad, Data and privacy: Getting consumers to trust products enabled by the Internet of Things, IEEE Consumer Electron. Mag. 8 (2) (2019) 35–38, http://dx.doi.org/10.1109/MCE.2018.2880807.
[88] W.Z. Khan, M.Y. Aalsalem, M.K. Khan, Q. Arshad, Enabling consumer trust upon acceptance of iot technologies through security and privacy model, in: Advanced Multimedia and Ubiquitous Engineering, Springer, 2016, pp. 111–117.
[89] M. Al-Roomi, S. Al-Ebrahim, S. Buqrais, I. Ahmad, Cloud computing pricing models: a survey, Int. J. Grid Distributed Comput. 6 (5) (2013) 93–106.
[90] G. Di Modica, O. Tomarchio, L. Vita, Resource and service discovery in SOAs: A P2P oriented semantic approach, Int. J. Appl. Math. Comput. Sci. 21 (2) (2011) 285–294.
[91] C. Bertolli, D. Buono, G. Mencagli, M. Torquati, M. Vanneschi, M. Mordacchini, F.M. Nardini, Resource discovery support for time-critical adaptive applications, in: Proceedings of the 6th International Wireless Communications and Mobile Computing Conference, ACM, 2010, pp. 504–508.
[92] E. Ahmed, A. Gani, M.K. Khan, R. Buyya, S.U. Khan, Seamless application execution in Mobile Cloud computing: Motivation, taxonomy, and open challenges, J. Netw. Comput. Appl. 52 (2015) 154–172.
[93] E. Ahmed, A. Naveed, A. Gani, S.H. Ab Hamid, M. Imran, M. Guizani, Process state synchronization-based application execution management for mobile edge/cloud computing, Future Gener. Comput. Syst. 91 (2019) 579–589.
[94] S. Mitra, Seamless mobility management and QoS support for multihomed mobile node in heterogeneous wireless networks, in: Industrial and Information Systems, ICIIS, 2010 International Conference on, IEEE, 2010, pp. 145–150.
[95] E. Ahmed, A. Naveed, A. Gani, S.H. Ab Hamid, M. Imran, M. Guizani, Process state synchronization for mobility support in Mobile Cloud computing, in: International Conference on Communications, ICC, IEEE, 2017, pp. 1–6.
[96] E. Ahmed, A. Gani, M. Sookhak, S.H. Ab Hamid, F. Xia, Application optimization in Mobile Cloud computing: Motivation, taxonomies, and open challenges, J. Netw. Comput. Appl. 52 (2015) 52–68.
[97] P. Kirci, Ubiquitous and Cloud computing: Ubiquitous computing, in: Resource Management and Efficiency in Cloud Computing Environments, IGI Global, 2017, pp. 1–32.
[98] X. Li, S. Liu, F. Wu, S. Kumari, J.J.P.C. Rodrigues, Privacy preserving data aggregation scheme for Mobile Edge computing assisted IoT applications, IEEE Internet Things J. (2019) http://dx.doi.org/10.1109/JIOT.2018.2874473, in press.
[99] J. Kang, R. Yu, X. Huang, M. Wu, S. Maharjan, S. Xie, Y. Zhang, Blockchain for secure and efficient data sharing in vehicular Edge computing and networks, IEEE Internet Things J. (2019) http://dx.doi.org/10.1109/JIOT.2018.2875542, in press.
[100] R. Yang, F.R. Yu, P. Si, Z. Yang, Y. Zhang, Integrated blockchain and Edge computing systems: A survey, some research issues and challenges, IEEE Commun. Surv. Tutor. (2019) http://dx.doi.org/10.1109/COMST.2019.2894727, in press.
[101] K. Salah, M.H.U. Rehman, N. Nizamuddin, A. Al-Fuqaha, Blockchain for AI: Review and open research challenges, IEEE Access 7 (2019) 10127–10149.

Wazir Zada Khan (M'16–SM'17) received his BS and MS in Computer Science from COMSATS University Islamabad, Wah Campus, in 2004 and 2007 respectively. He received his Ph.D. from Electrical and Electronic Engineering Department, Universiti Teknologi PETRONAS, Malaysia in 2015. Dr. Khan is currently working with Farasan Networking Research Laboratory, Faculty of CS & IS, Jazan University, Saudi Arabia. His current research interests are in the area of Wireless Sensor Networks, Security and Privacy, Internet of Things. He is SMIEEE.

Ejaz Ahmed (S'12, M'17) received his Ph.D. in Computer Science from University of Malaya, Malaysia. He has worked as a research assistant at C4MCCR, University of Malaya, Malaysia, CogNet Lab, NUST, and CoReNet Lab, CUST, Pakistan. His research experience spans over more than eleven years. He is Associate Technical Editor/Editor of IEEE Communications Surveys & Tutorials, IEEE Communications Magazine, IEEE Access, Elsevier JNCA, KSII TIIS, and Elsevier FGCS. He has also served as a Lead Guest Editor for the Elsevier FGCS Journal, IEEE Access, Elsevier Computers & Electrical Engineering, IEEE Communications Magazine, Elsevier Information Systems, and Transactions on Emerging Telecommunications Technologies. He has served as Chair and Co-chair in several international conferences. His areas of research interest include Mobile Cloud Computing, Mobile Edge Computing, Internet of Things, Cognitive Radio Networks, Big Data, and Internet of Things. He has successfully published his research work in a number of international journals and conferences. He has received several performance awards during his research career.

Saqib Hakak received his Ph.D. in Computer Science with major specialization in Information security from University of Malaya, Malaysia. He worked as Research assistant in the University of Malaya from 2015–2018 and worked on numerous research projects. He is an active reviewer of numerous ISI-indexed journals and has served as program committe technical member for various international conferences held in China, Gulf, Malaysia, UK, and Canada. His research interests revolve in the respective areas of Information security including Natural Language Processing, Information retrieval, Edge Computing, Networking, Gamification and Blockchain technology.

Ibrar Yaqoob has recently been listed among top researchers based on the number of citations in six categories of Computer Science, according to Thomson Reuters from 2015–2018. He is currently working as a research professor in the Department of Computer Science and Engineering, Kyung Hee University, South Korea, from where he completed his postdoctoral fellowship under the prestigious grant of Brain Korea 21st Century Plus. Prior to that, he received his fully funded Ph.D. (Computer Sciences) from the University of Malaya, Malaysia, in 2017. He worked as a researcher and lead developer at the Centre for Mobile Cloud Computing Research (C4MCCR), University of Malaya. His numerous research articles are very famous and among the most downloaded in top journals. He has reviewed over 100 times for the top ISI-indexed journals and conferences. He has served as a technical program committee member for various international conferences, which were held in Spain, Portugal, United Kingdom, Germany, and China. He has served/serving as a guest editor in various Journals. He is also serving as an associate editor, editor, area editor, and a track chair in IEEE Access Journal, Ad Hoc & Sensor Wireless Networks Journal, EAI Transactions on Pervasive Health and Technology, and IEEE HealthCom 2017, respectively. He has won numerous awards during his stay at the University of Malaya. His research interests include big data, edge computing, mobile cloud computing, the Internet of Things, and computer networks.

Arif Ahmed received his M.Tech degree in Computer Science & Engineering from the National Institute of Technology Silchar, India in 2014, and B. Tech in Information Technology from Assam University, Silchar, India in 2012. He worked as a visiting scientist at the Centre for Development of Advanced Computing, Mumbai, India from 2014 to 2015. From 2015 onwards he is working as an assistant professor at National Institute of Techonology Silchar, India. His research interests are in the field of Cloud Computing, Fog Computing, Internet of Things etc.