WILEY | Hindawi

Review Article

A Survey on Mobile Edge Computing: Focusing on Service Adoption and Provision

Kai Peng [iD],[1,2,3] **Victor C. M. Leung** [iD],[2] **Xiaolong Xu** [iD],[4] **Lixin Zheng,**[1,3]
Jiabin Wang,[1,3] **and Qingjia Huang**[5]

[1]*College of Engineering, Huaqiao University, Quanzhou, Fujian 362021, China*
[2]*Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada V6T 1Z4*
[3]*Fujian Provincial Academic Engineering Research Centre in Industrial Intellectual Techniques and Systems,*
  Quanzhou, Fujian 362021, China
[4]*School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China*
[5]*Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China*

Correspondence should be addressed to Kai Peng; pkbupt@gmail.com

Received 18 April 2018; Revised 29 August 2018; Accepted 23 September 2018; Published 10 October 2018

Academic Editor: Pierre-Martin Tardif

Mobile cloud computing (MCC) integrates cloud computing (CC) into mobile networks, prolonging the battery life of the mobile users (MUs). However, this mode may cause significant execution delay. To address the delay issue, a new mode known as mobile edge computing (MEC) has been proposed. MEC provides computing and storage service for the edge of network, which enables MUs to execute applications efficiently and meet the delay requirements. In this paper, we present a comprehensive survey of the MEC research from the perspective of service adoption and provision. We first describe the overview of MEC, including the definition, architecture, and service of MEC. After that we review the existing MUs-oriented service adoption of MEC, i.e., offloading. More specifically, the study on offloading is divided into two key taxonomies: computation offloading and data offloading. In addition, each of them is further divided into single MU offloading scheme and multi-MU offloading scheme. Then we survey edge server- (ES-) oriented service provision, including technical indicators, ES placement, and resource allocation. In addition, other issues like applications on MEC and open issues are investigated. Finally, we conclude the paper.

1. Introduction

In recent years, with the continuous development of cloud computing (CC), big data, mobile network, software defined network (SDN), and upgrading of intelligent mobile terminals [1–11], the number of mobile users (MUs) has exploded rapidly. According to Cisco's latest forecast, global mobile data traffic (MDT) in 2020 will be 8 times of that in 2015, and the number of MUs will reach 2.63 billion [12]. However, compared to a traditional device such as PC, a MU has certain limitations in terms of computing power, storage capacity, and especially battery capacity, which greatly restricts the further promotion of the service. The emergence of mobile cloud computing (MCC) provides a good opportunity to relieve such limitations. MCC brings new kinds of services and facilities MUs to make the best use of CC [13].

However, cloud is usually located far away from the MUs, which leads to high network delay for data transmission between MUs and cloud. In order to solve the issue of network delay, a new paradigm named mobile edge computing (MEC) has been proposed. The MEC can be seen as a specific case of the MCC.

A number of surveys on MEC have been published recently [14–17]. A survey on MEC focusing on the general overview was presented in [14]. A comprehensive survey of the MEC research from the perspective of communication was provided in [15]. Wang et al. presented a comprehensive a survey on service migration in MEC in [16]. Two conceptions similar to service migration, such as live migration for data centers and handover in cellular networks, were investigated. Wang et al. [17] made an exhaustive review on the current research efforts on MEC. They firstly gave an overview

of MEC which included the definition, architecture, and advantages. Additionally, the issues on computing, caching, and communication techniques are also presented.

Different from the existing surveys, we provide a comprehensive survey of the state-of-the-art MEC research, focusing on service adoption and provision. More specifically, MEC is mainly composed of two parts, including MU and edge server (ES). On the one hand, we surveyed MUs-oriented service adoption, such as computation offloading and data offloading. On the other hand, ES-oriented service provision, including ES deployment and resource allocation, is investigated.

The summary of abbreviations in this paper is shown in Table 1. The rest of the paper is organized as follows. In Section 2, we give an overview of MEC. We describe MUs-oriented service adoption in Section 3. Followed by Section 4, ES-oriented service provision is mainly investigated. And then applications of MEC are introduced in Section 5. Section 6 elaborates open issues. Finally, we conclude the paper.

2. MEC Overview

In this section, we introduce the definition and architecture of MEC in Section 2.1, and then show the comparison of several modes in Section 2.2. A summary of literatures on different modes is shown in Table 2.

2.1. Definition, Architecture, and Service of MEC

2.1.1. Definition. MEC is a new network paradigm that provides information technology services and CC capabilities within the mobile access network of MUs and has become a technology. European Telecommunications Standards Institute (ETSI) proposed standardization of MEC in 2014 [18] and pointed out that MEC provided a new ecosystem and value chain that can use MEC to migrate intensive computing tasks of MUs to nearby ESs ([19, 20]). Since the MEC is located within the radio access network (RAN) and is close to MUs, it can achieve higher bandwidth with lower latency to improve quality of service (QoS) and quality of experience (QoE). MEC is also a key technology for the development of 5G [21], which helps to meet the high standards of 5G from delay, programmability, and scalability. By deploying services and caching at the edge of network, MEC can not only reduce congestion, but also efficiently respond to user requests. The most similar concept to MEC is EC. The EC refers to a new computing mode that performs computation at the edge of the network. The edge data in the EC represents the cloud service, and the upstream data represents the interconnection service. The edge of the EC refers to any computation and edge of the network from the data source to the CC center. The MEC emphasizes that the ES between the CC center and the edge device performs the task of computing the MU data on the ESs, but the MU does not basically have the computing capability. On the contrary, the MU in the EC model has strong computing capability. Therefore, the MEC can be seen as a part of the EC model [22]. In [23], Mach et al. firstly described the main use cases and reference scenarios for MEC. And they reviewed the existing notions of integrating

MEC functionalities into mobile networks and discussed the progress of MEC standardization. They mainly reviewed user-oriented cases in the MEC, i.e., computation offloading. More specifically, the study of computation offloading was dived into three subissues: decision making, resources allocation, and mobility management. Liang [24] mainly focuses on MEC in 5G system and beyond.

In this paper, we mainly focus on the MEC and present a comprehensive survey from the perspective of service of MEC. More specially, we mainly review the service adoption of MU's and service provision of ES'. MUs in this paper can be mobile phones, tablets, and other intelligent devices. In fact, it is not very critical whether MUs have computing capability or not. The main issue of MUs is offloading to ESs. Meanwhile, the ES placement and resource allocation of ES is very important.

2.1.2. Architecture. As shown in Figure 1 [25], we argue that the MEC architecture can be seen as the middle layer between MUs and cloud which is closer to MUs and provide service for them and can communicate directly with cloud. In addition, it can also be seen as an independent two-tier architecture consisting of MUs and ESs. MU generally refers to mobile phones, tablets, laptops, and so on. In addition, ES generally refers to base station combined with server or cloudlet. MUs offload computing tasks to ES, which not only enables MUs to execute applications efficiently, but also reduces latency due to access to public cloud.

2.1.3. Service. In order to better understand of MEC, the authors in [26] begin with a discussion of the potential service scenarios and identified the design challenges of a MEC enabled network. MEC services mainly contain many aspects, such as service provision and service deployment. More specifically, as MUs and ESs are the core parts of MEC, and thus we mainly focus on MUs-oriented service adoption and ES-oriented service provision. For MUs, how to obtain service they need (i.e., offloading) to meet application requirements is more critical. For ESs, how to manage resources (i.e., ESs) effectively is critical.

2.2. Comparison of Several Modes. In this section, we introduce the existing surveys on conceptual comparison in Section 2.2.1. After that the similar terms are described in Section 2.2.2.

2.2.1. Surveys of Conceptual Comparison. Comparison of EC implementations: fog computing, cloudlet, and MEC are shown in [27]. In [28], the basic issues of the distribution and active caching computing tasks in fog environment are studied under the constraints of delay and reliability. We describe our comparison in Section 2.2.2.

2.2.2. Similar Terms. In this section, similar terms, such as mobile cloud computing (MCC), fog computing (FC), and cloudlets, are introduced and compared.

(1) Mobile Cloud Computing (MCC). The definition of MCC was shown in [29, 30]. The main idea is as follows. MCC refers

Table 1: Summary of abbreviations.

Access Point(s)	AP(s)
Base Stations	BSs
Cloud Computing	CC
Deep Supervised Learning	DSL
Dynamic Voltage Scaling	DVS
Edge Computing	EC
Edge Server(s)	ES(s)
Efficient Computation Offloading	ECO
End-to-End	E2E
Energy-Efficient Computation Offloading	EECO
European Telecommunications Standards Institute	ETSI
Fiber-Wireless	FiWi
Fog Computing	FC
Full Duplex	FD
Green-energy Aware Avatar Placement	GAP
Incentive-compatible Auction Mechanism	ICAM
Integer Linear Programming	ILP
Internet of Things	IoT
Intrusion Detection System	IDS
Iterative Improvement	II
Markov Decision Process	MDP
Mixed Integer Linear Programming	MILP
Mobile Cloud Computing	MCC
Mobile Data Traffic	MDT
Mobile Data Traffic Offloading	MDTO
Mobile Edge Computing	MEC
Mobile Edge Internet of Things	MEIoT
Mobile Edge Computing-wireless Power Transfer	MEC-WPT
Multiple Knapsack Problem	MKP
Multi-objective Optimization Problem	MOOP
Multi-user Computation Offloading	MUCO
Multi-user Data Offloading	MUDO
Mobile User(s)	MU(s)
Physical Resource Block	PRB
Profit Maximization Avatar Placement	PRIMAL
Quality of Experience	QoE
Quality of Service	QoS
Radio Access Networks	RANs
Radio Access Points	RAPs
Software Defined Networking	SDN
Stochastic Reward Net	SRN
Two-Phase Optimization	TPO
Vehicular Ad Hoc Networks	VANET
Unmanned Aerial Vehicles	UAVs
Vehicular Cyber-physical Systems	VCPSs
Virtual Machine	VM
Wireless Access Point(s)	WAP(s)

TABLE 2: Summary of literatures on different modes.

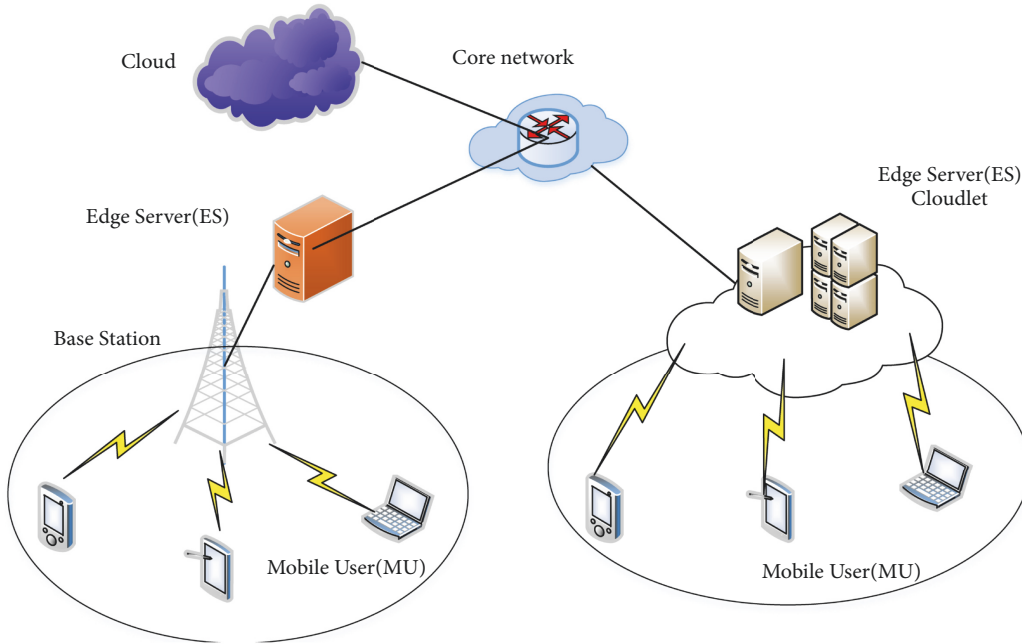| Item	Related work	Key points
MEC	Definition [22–24]	[22] The difference between EC and MEC [23] The main use cases and reference scenarios standardization [24] MEC on 5G
	Architecture [25]	[25] MEC architecture
	Different network types [26]	[26] Potential service scenarios
Comparison of several modes	[27, 28]	[27] A set of parameters, decision tree [28] The basic problems of the distribution and active caching computing tasks
MCC	[29–35]	[29, 30] MCC definition [31] Review of MCC [32] Public cloud [33] Cloudlet [34] Ad hoc mobile cloud [35] MCC Integrating cloudlets and MEC
FC	[41, 42]	[41] The definition and similar concepts [42] A more comprehensive definition
Cloudlet	[46, 47]	[46] Cloudlet's computer cluster as a complement to computing offloading for cloud [47] Cloudlet can only be used for wireless access environments

FIGURE 1: Mobile edge computing (MEC) architecture.

to the infrastructure that data storage and data processing happens outside of MU. Three kinds of MCC architectures are surveyed in [31], including the public cloud [32], the cloudlets [33], and the ad hoc mobile cloud [34]. The authors in [35] propose that the future of MCC will be integrated cloudlets and MEC. We think that the MCC and the MEC are different if the destination of the offloading in the MCC is cloud,

but their concept is the same if the offloading destination is cloudlet.

As shown in Figure 2, a basic MCC system architecture is presented. MEC mainly consists of two parts, such as MUs and cloud servers. In general, it is assumed that the resource of cloud server is infinite and faraway from MU, while the MEC is closer to the MU. The MEC can reduce the MU's
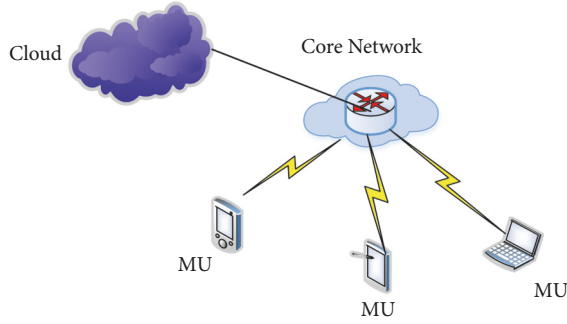
FIGURE 2: Mobile cloud computing (MCC) architecture.

access delay and improve QoS of MU. However, the ES of MEC has limited resource types and computing capabilities. When the request of MUs cannot be executed by ESs which needs to be further decided whether to execute the MU's request locally by MU itself or forward the request to the cloud.

(2) Fog Computing (FC). Although CC is widely used, due to the inherent problems (i.e., unreliable delay, lack of mobility support, and location awareness), there are still some problems needed to be figured out. FC can figure out the above issues by providing flexible resources and services for the MUs. FC was proposed by Cisco in 2012 [36] and was defined as a highly virtualized computing platform that previously performed tasks from CC centers to MUs. A comprehensive definition of FC was given by Vaquero in [37]. FC [38] extended the cloud-based network architecture by introducing an intermediate layer between cloud and MUs, and the middle layer is essentially a fog layer composed of fog servers deployed at the network edge [39]. The bandwidth load and energy consumption of the backbone link can be significantly reduced by fog servers [40]. In addition, the fog servers can be interconnected with cloud and use the computing power and rich applications and services of cloud. In [41], Yi et al. discussed the definition and similar concepts of FC, introduced representative application scenarios, and pointed out various problems that may be encountered when designing and implementing FC system. Yi et al. [42] discussed the current definitions of FC and similar concepts and proposed a more comprehensive definition. They also implement and evaluate a prototype FC platform. The authors in [22, 43] argue that MEC is substitutable with FC, and MEC mainly focuses more on the things side, whereas the latter one focuses more on the infrastructure side. In this paper, we think that there is no essential difference between MEC and FC. Both of them are closer to MUs and can be seen as an effective complement to CC and an important foundation for the realization of the Internet of Things (IoT).

(3) Cloudlets. Cloudlet was firstly proposed by authors in [33] and widely used in pervasive computing [44]. In the existing surveys, for one thing, this computing mode appeared relatively early. For another, the MUs are severed by a single cloudlet. And thus this mode is considered to be an independent service mode. Different from them, we hold the opinion that cloudlet can provide service for MUs by multiple cloudlet collaborations, which is the important way of service provision. In addition, the authors in [45, 46] also supported this view. In [47], the authors believe that cloudlet can only be used for wireless access environments. However, we believe that the MEC is a more general concept that can support Wi-Fi and is an effective supplement to 5G. Thereby, we hold the opinion that cloudlets play an important part in MEC as ESs for service provision. More specifically, cloudlets play an important role in WLAN and WMAN. More details will be described in Section 4.

3. Mobile Users- (MUs-) Oriented Service Adoption

In this section, we review the research on offloading. A summary of literatures on MUs-oriented service adoption are shown in Table 3, where key points in each paper are concluded. We firstly give a brief introduction of decision making on offloading.

In [48], Zhang et al. present a survey on decision making for task migrations which included decision factors and related algorithms. They point out that more attention should be paid to this issue. In [23], the authors review the decision making offloading issue of MEC in terms of local execution, full offloading, and partial offloading. A survey of computation offloading of mobile systems was proposed in [49]. They describe the two purposes (i.e., improving performance and saving energy) for offloading and provided a review from this perspective. Different from them, we mainly focus on the offloading issue from the perspective of offloading taxonomy. More specifically, we divide the studies on offloading into two key taxonomies: computation offloading and data offloading.

3.1. Offloading Taxonomy. We review computation offloading in Section 3.1.1 and data offloading in Section 3.1.2. The main concern of computation offloading and data offloading is different. Computation offloading refers to the MUs which send the heavy computation task to ESs and receive the results from them [49], while data offloading refers to use novel network techniques to transmit mobile data originally planned for transmission through cellular networks [50].

3.1.1. Computation Offloading. Computation offloading was originally studied in MCC. The offloading problem model in MEC is similar to MCC, but the main difference is the destination of offloading. It is commonly assumed that the implementation of computation offloading relies on a network architecture with the public cloud, while the offloading destination of MEC is ES. In addition, the offloading goals are basically the same which is to minimize total energy consumption or overall task execution time, or both of them. Furthermore, MEC computation offloading is divided into single mobile user computation offloading (single MUCO) and multiple mobile user computation offloading (Multi-MUCO).

(1) Single MUCO. There are many studies on single MUCO. The more representative ones are as follows ([51–58]). We

TABLE 3: Summary of literatures on MUs-oriented service adoption.

Work area	Related Work	Key Points
Single MUCO	Cloudlet based single MUCO [51–55] Base station based single MU CO [56–58]	[51] Application-aware cloudlet selection in multi-cloudlet environment [52] Cloudlet selection mechanism in a decentralized environment [53] A power and latency aware cloudlet selection in multi-cloudlet environment [54] MILP in the Multi-cloudlet environment [55] A deep learning approach [56] Offloading in task allocation and computational frequency scaling [57] Tradeoff between the latency and reliability in task offloading [58] Partial computation offloading using DVS
Multi-MUCO	Cloudlet based multi-MUCO [59] Base station based multi-MUCO [60–70]	[59] A Game-theoretic machine learning approach [60] MUCO game [61] Power-delay tradeoff in MU MEC systems [62] Multi-objective optimization [63] Sum energy consumption minimization [64] Joint computation offloading and interference management [65] Offloading in 5G heterogeneous networks [66] Joint offloading and computing optimization [67] Multiple knapsack problem for 5G mobile edge computing [68] Joint radio and computational resource management [69] Dynamic offloading approach [70] Joint task offloading and resource allocation
Single MUDO	[50, 71–73]	[71] An adaptive data offloading model. [72] Combination optimization in emerging VCPSs [50] Data offloading technologies (four taxonomies) [73] The UAV trajectory to offload traffic for BSs
Multi MUDO	[74]	[74] A joint coalition-pricing based on coalitional game theory and pricing mechanism

review the literature from the perspective of cloudlet based single MUCO and base station based single MUCO.

Cloudlet Based Single MUCO. To offload applications to the most appropriate cloudlet is very important. In [51], Roy et al. proposed an application-aware cloudlet selection strategy for multiple cloudlets scenario. They assumed that different cloudlets could deal with different types of applications. The application type was first verified when the request came from the MU. And then, the most suitable cloudlet is chosen from multiple cloudlets near the MU on the basis of the application type. By using the proposed strategy, both the energy consumption and the latency of application execution of the MU can be reduced. In [52], a mechanism to identify a cloudlet for computation offloading in a distributed manner was proposed. The mechanism mainly is composed of two phases. In the first phase, cloudlets within Wi-Fi range of the MU were identified. In the second one, the ideal offloading cloudlet was selected. Mukherjee et al. [53] proposed a power and latency aware optimum cloudlet selection method for multiple cloudlets scenario by introducing a proxy server. Theoretical analysis showed that the power and the latency consumption were reduced by about 29%-32% and 33%-36%, respectively, compared with offloading to cloud. A two-stage optimization strategy is proposed in [54]. Firstly, a cloudlet selection model based on mixed integer linear

programming (MILP) is proposed to obtain the cloudlet for MUs by optimizing latency and the mean reward. Secondly, a resource allocation model based on MILP is presented to allocate the resources in the selected cloudlet by optimizing reward and the mean resource usage.

In [55], the authors develop a dynamic offloading framework for MUs, taking into account the local overhead of the MU, as well as the finite communication resource and computation resource of ES. They develop the offloading decision problem as a multilabel classification problem and a deep supervised learning (DSL) method is proposed to minimize the computation and offloading overhead. It is shown that system cost can be reduced compared to the existing four schemes, including no offloading scheme, random offloading scheme, total offloading scheme, and multilabel linear classifier based offloading scheme.

Base Station Based Single MUCO. An optimization framework of offloading from a single MU to multiple ESs was proposed in [56]. The authors considered two cases, such as fixed CPU frequency and elastic CPU frequency for the MU. In addition, two different methods were proposed to solve these two cases respectively. It was shown that the proposed algorithms achieved near optimal performance. In [57], the tradeoff between the latency and reliability in task offloading to MEC is studied. A framework is provided, where MU

partitions a task into subtask and offloads them to multiple ESs. In this framework, an optimization problem is formulated to jointly minimize the latency and offloading failure probability. Compared to the previous work, it is shown that the proposed algorithm achieves a good balance between these two objects, namely, latency and reliability. In [58], the authors focus on jointly optimizing communication resource and computation resource for partial computation offloading based on dynamic voltage scaling (DVS) technology. For one thing, the MU severed by a single ES is considered; for another, they investigate the energy consumption of minimization problem and latency of application execution minimization problem in a multiple ESs' scenario, where the MU could offload computation to a group of ESs.

(2) Multi-MUCO. Compared to single MU issue, multi-MUCO is more complex. Considering that resources are finite and MUs are mutually restricted, multi-MUCO is hard to solve. The more representative ones are as follows ([59–70]). We review the literature from the perspective of cloudlet based single MUCO and base station based single MUCO.

Cloudlet Based Multi-MUCO. Cao et al. [59] propose the problem of MUCO for cloudlet in MCC in multichannel wireless contention environment. They formulated this MUCO problem as a noncooperative game. And then a completely distributed computation offloading algorithm was developed so as to achieve the Nash Equilibrium Point (NEP). Finally, the effectiveness of their proposed algorithm was proved.

Base Station Based Multi-MUCO. In [60], the authors firstly studied the MUCO problem for MEC in a multichannel wireless interference environment. They showed that computing a centralized optimal solution is NP-hard, so a game theory was used to implement ECO in a distributed manner. The problem was modeled as MUCO game. They analyzed the structure of the game and showed that the game recognized Nash equilibrium. And then they designed a distributed computing offloading algorithm for the solution. Furthermore, the study is extended to MUCO scenario. It was demonstrated that their proposed algorithm could achieve higher performance. In [61], Mao et al. investigated the tradeoff between two key but contradictory goals in MU MEC systems, such as the power consumption of MUs and the execution delay of computation tasks. Based on the Lyapunov optimization method, an online algorithm for local execution and computation offloading is developed. The analysis showed that power consumption and execution delay followed the $[V (1/V), O (V)]$ (V is the control parameter). The simulation results confirmed the theoretical analysis. In [62], Liu et al. used queuing theory to study the energy consumption, execution delay, and price of offloading issue in MEC. A multiobjective optimization problem (MOOP) was formulated to minimize these three goals. The effectiveness of the proposed method by extensive simulations is demonstrated. In [63], the authors consider energy-efficient resource allocation for a multi-MU MEC system. The resource consumption for downloading the computation results back to MUs is taken into account. Meanwhile, the authors establish on two

computation-efficient models with negligible and nonnegligible BS executing durations, respectively. Then, under each model, they develop a total weighting and energy consumption minimization problem by optimally allocating communication resources and computing resources. In [64], Wang et al. propose joint consideration of computation offloading and interference management so as to improve the performance of wireless cellular networks with MEC. In addition, the computation offloading decision, physical resource block (PRB) allocation, and MEC computation resource allocation are formulated as optimization problems. Simulation results show the effectiveness of the proposed scheme. In [65], Zhang et al. formulated an optimization issue to minimize energy consumption. Based on the multiple access characteristics of 5G heterogeneous networks, they have designed an energy-efficient computation offloading (EECO) method to minimize energy consumption under time delay constraints by integrating optimized offloading and radio resource allocation.

In [66], the authors propose a unified mobile edge computing-wireless power transfer (MEC-WPT) design by considering a wireless powered multi-MU MEC system, where a multiantenna access point (AP) (integrated with an MEC server) broadcasts wireless power to charge multi-MU and each MU relies on the harvested energy to execute computation tasks. They propose an optimal resource allocation scheme which minimizes the total energy consumption of AP's meeting the MUs' individual computation latency constraints. The authors [67] provide a general model of the system that takes account of the E2E computational latency of MEC applications. An evaluation of a multi-MU MEC offloading model reducible to the Multiple Knapsack Problem (MKP) is formulated. In [68], an online joint radio and computational resource management algorithm for multi-MU MEC systems is developed to minimize the long-term average weighted sum power consumption of the MUs and the ES, which are constrained by the stability of the task buffer. In [69], the authors focus on how to dynamically offload the computation intensive tasks of newly executed mobile applications from MUs to the ESs so as to minimize the average application completion time. They consider a system model in which a group of MUs are connected to the ES. Furthermore, they consider possible transmission collisions on a shared network when more than one MU try to transmit data simultaneously. It is shown that the proposed approach outperforms significantly the previous ones. In [70], a MEC enabled multicell wireless network is considered where each BS is equipped with an ES that can be used to assist the MU in performing computationally intensive tasks by offloading. The problem of joint task offloading and resource allocation is formulated to maximize the MUs' computation offloading profit, which is measured by reducing completion time and energy consumption of task.

3.1.2. Data Offloading (DO). Similarly, DO also can be divvied into single MU data offloading (Single MUDO) and multi-MU data offloading (Multi-MUDO).

(1) Single MUDO. Wang et al. proposed multiple effective data offloading methods and obtained better results for different

scenarios [71, 72]. In [71], they proposed a mobile data traffic offloading (MDTO) model which could determine whether to adopt opportunistic communications or communicate by cellular networks adaptively. In addition, Wang et al. [72] proposed a pioneering effort to promote MDTO in the emerging vehicular cyber-physical systems (VCPSs), aiming to reduce the MDT for the QoS aware services. They investigated MDTO models for Wi-Fi and Vehicular ad hoc networks (VANET). In particular, they formulated MDTO as a MOOP to simultaneously minimize MDT and QoS aware service provision. The optimal solutions were obtained by using mixed integer programming (MIP). The simulation results showed that the proposed scheme can offload MDT by up to 84.3% and meet the requirements of the global QoS guaranteed. Zhou et al. [50] discussed the current development of mobile data offloading techniques. In view of the diversity of data offloading sponsor, they divided the current mobile data offloading technologies into four taxonomies, thus, by small cell networks, by Wi-Fi networks, by opportunistic mobile networks, and by heterogeneous networks. In [73], the authors focused on the Unmanned Aerial Vehicles (UAV) trajectory at the edges to offload traffic for the base stations (BSs). An iterative algorithm was developed to solve the optimization issue which was a mixed integer nonconvex problem. The effectiveness of proposed scheme was shown.

(2) Multi-MUDO. The authors in [74] engaged in the issues of scheduling multiple MUs' offloading and choosing the ESs for offloading. They proposed a joint coalition-pricing based data offloading approach based on coalitional game theory and price mechanism. The numerical results showed the effectiveness of the proposed approach.

4. Edge Server- (ES-) Oriented Service Provision

In this section, we mainly review the ES-oriented service provision. In Section 4.1, the technical indicators are introduced, including expenditure in Section 4.1.1 and load balancing in Section 4.1.2. Followed by Section 4.2, ES placement for different scenarios (WLAN (Section 4.2.1) and WMAN (Section 4.2.2)) are reviewed. And then we summarize the resource allocation in Section 4.3, such as resource allocation in cloudlet based MEC system in Section 4.3.1; resource allocation in base station based MEC system in Section 4.3.2 is investigated. A summary of literatures on ES-oriented service provision is shown in Table 4.

4.1. Technical Indicators. In this section, expenditure is introduced in Section 4.1.1, and the load balancing is described in Section 4.1.2.

4.1.1. Expenditure. Different from cloud, the resources of ES are limited. It is critical to minimize the cost of ES and meet the requirements of user tasks. And therefore, the expenditure indictor is highlighted. In [75], an incentive-compatible auction mechanism (ICAM) was proposed for the

resource trading between MUs and cloudlets. ICAM could effectively allocate cloudlets meeting MU's service needs and determine pricing. Based on Lyapunov optimization techniques combined with weighted perturbation techniques, Fang et al. [76] proposed a new random control algorithm that makes online decisions for computational request admission and scheduling, computational service purchase, and computational resource allocation. In particular, it was highly effective in practice. It was proved that the proposed algorithm achieved profit optimization and system stability. In [77], Sun et al. proposed a cloudlet network architecture. Each MU can communicate with its Avatar. Live Avatar migration was enabled to maintain the low E2E delay between each MU and its Avatar. They proposed a Profit Maximization Avatar placement (PRIMAL) policy to optimize the tradeoff between the migration profit and the migration cost. The effective of PRIMAL was demonstrated.

4.1.2. Load Balancing. Load balancing is an important indicator for evaluating the MEC system. Jia et al. [78] investigated how to balance the workload between multiple cloudlets. A fast and scalable algorithm was developed to solve the problem. Moreover, the performance was evaluated by experimental simulations which demonstrated the significant potential to reduce the response times of tasks. Xu et al. proposed [79] a dynamic resource allocation for load balancing in FC environment. In [80], the issue of load balancing in multi-MU of FC scenario was solved and the authors proposed a low-complexity algorithm for fog clustering.

4.2. ES Placement. In this section, we mainly review ES placement issue. In base station based MEC system, the ES are assumed to have been placed and are placed in the same location as the base station. So we mainly study the placement of edge servers (i.e., cloudlets) in the cloudlet based MEC system. We mainly review the case of cloudlet placement in VLAN and WLAN. We review the cloudlet placement in WLAN in Section 4.2.1 and the cloudlet placement in WMAN in Section 4.2.2.

4.2.1. Cloudlet Placement in WLAN. In [81], the authors presented a survey on the current mobile cloudlet architectures. They also classified the existing cloudlet solutions by presenting a hierarchical classification. The demands and challenges for deploying the cloudlet in a wireless local area networks (WLAN) were highlighted. In summary, the size of the MUs in the WLAN is relatively small, and the network coverage is relatively small. Therefore, the method cannot be directly used in WMAN. In the next section we will discuss the issue of cloudlet placement in WMAN.

4.2.2. Cloudlet Placement in WMAN. Figure 3 [82] illustrates a WMAN. It is assumed that there are K (K=3) cloudlets (cloudlet A, cloudlet B, and cloudlet C) to be placed to K different locations. For simplicity, it is assumed that the cloudlets will be colocated with some APs. Given the K placed cloudlets, MUs can offload the computation tasks to the cloudlets

TABLE 4: Summary of literatures on ES-oriented service provision.

Work Area	Related work	Key Points
Technical indicators	Expenditure [75–77] Load Balancing [78]	[75] An incentive-compatible auction mechanism [76] A new stochastic control algorithm [77] PRIMAL strategy [78] Balance the workload optimize mobile application performance
Cloudlet placement	WLAN [81]	[81] Cloudlet placement in WLAN
	WMAN [82–84, 86–88]	[83] The cache placement [84] The server placement [82] Minimize the average access delay [86] Cloudlet placement and mobile user allocation [87] A location-aware services deployment algorithm [88] AP ranking in the cloudlet placement
Resource scheduling	ES placement	[51] Offloading an application to the most appropriate cloudlet [52] Identify a cloudlet in a decentralized manner [53] Power and latency aware cloudlet selection
	Cloudlets deployments[46, 89, 90] Deploy the servers[91, 92]	[46] A MILP optimization model [89] Optimize the power consumption for large scale cloudlets deployment [90] TPO algorithm and an Π algorithm [91] Deploy the servers in a cost effective manner [92] A novel MCC architecture low latency and low energy consumption
	VM migration [99–106],	[99] A dedicated VM is provisioned on a PM [100] Combined performance and availability model using the Stochastic Reward Net [101] GAP minimize the total power consumption [102] Minimize services delay in a scenario with two cloudlet servers [103] A one-on-one contract game model and developed a learning-based price control mechanism [104] Dynamic proxy VM migration minimize the E2E delay [105] Dynamic services migration problem [106] A virtual FD-enabled small cell network with cache

via the local APs. If a cloudlet is colocated with an AP, the MUs at that AP will obtain the minimum cloudlet access delay of the MUs (i.e., MU (B) and MU(C) in Figure 3); otherwise, the MU requests at that AP must be relayed to nearby cloudlets for processing which leads to a cloudlet access delay due to the cumulative delay of multiple hop relays (i.e., MU (A) in Figure 3).

Cloudlets are particularly suited for wireless metropolitan area network (WMAN). And the cloudlet placement problem in WMAN consisting of many WAPs is very promising. There are two classical optimization problems which are closely related to this placement problem, namely, the cache placement [83] and the server placement [84]. Actually, both of these two issues can be solved by a direct reduction to the capacitated K-median problem [85]. However, this placement problem is essentially different from the above two problems; namely, it is assumed that either there is no capacity limitation on caches (or servers), or all caches (or servers) have identical capacities; however, the capacity of each cloudlet may be different and different MU requests may also have different computing resource needs. Xu et al. [82] firstly described the problem as a new capacity cloudlet placement problem.
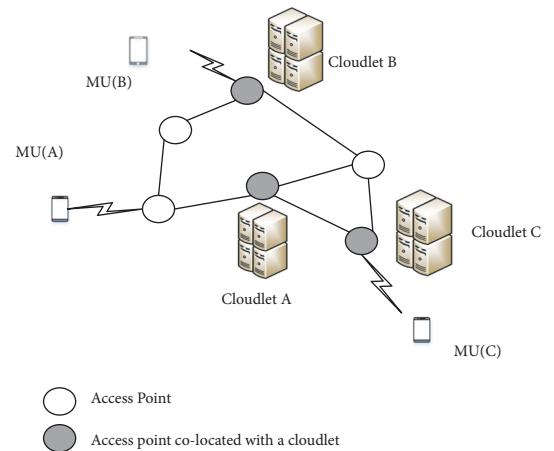

FIGURE 3: Cloudlet placements in WMAN.

And the object of this issue is to put K cloudlets in some strategic locations so as to minimize the average access delay between MUs and the cloudlets. In addition, they proposed an effective heuristic solution for this issue. It was

shown that the proposed algorithm had good scalability. In [86], Jia et al. formulated cloudlet placement and MUs allocation to the cloudlets issue in WMAN. They developed an algorithm, which enabled the placement of the cloudlets in WMAN and allocated MUs to the placed cloudlets while their workloads were balanced. They conducted experiments through simulation which demonstrate the effectiveness of the algorithm. As known to us, the MU in WMAN is moving, which made necessary to deploy and switch services anytime and anywhere to achieve the minimum network delay of MU services requests. However, the cost of this solution is usually too high for services providers and is invalid for resource utilization. A location-aware services deployment algorithm was proposed by Liang et al. [87] based on K-means to solve this problem. Generally speaking, the proposed algorithm divided the MUs into multiple MU clusters according to the geographic location of the MUs and then deployed the services instances to the ESs nearest to the centers of MU clusters. The performance evaluation showed that the algorithm could not only effectively lower the network latency, but also reduce the number of services instances while meeting the tolerable network latency. In [88], Yang et al. solved the problem of AP ranking in the cloudlet placement of the EC environment. AP ranking is an important step in the cloudlet placement. They proposed an adaptive integrated AP ordering method by analyzing the connection characteristics of APs. The results verified the effectiveness of their proposed approach. Above all, cloudlets deployment in WMAN has great challenges and great application prospects, which is an important part of MEC.

4.3. Resource Allocation. In this section, we divide the issue of resource scheduling into two subsections.

4.3.1. Resource Allocation in Cloudlet Based MEC System

Service Deployments. It is critical to dispatch of MU tasks through multiple cloudlet collaborations in MEC. In [89], Al-Ayyoub et al. solved the issue of optimizing the power consumption of large-scale collaborative cloudlets deployments. The effectiveness of the proposed mode was shown. In [46], Al-Quraan et al. presented a mixed integer linear programming (MILP) optimization model for MEC systems. More specifically, two kinds of cloudlets (i.e., local cloudlets and global cloudlets) are mentioned. A MU first sends the requirements to the local cloudlet. If none of local cloudlets can provide services, then it will be transferred to a global cloudlet. They were evaluated in several practical cases to prove that it can be applied for large-scale MEC systems. The purpose of [90] is to satisfy all computing requirements of each node in network edge within a certain delay based on limited computing resources (such as cells, APs, and macro-BSs), so as to optimize the total consumption. They formulated the issue as an Integer Programming problem and then developed a Two-Phase Optimization (TPO) algorithm and an Iterative Improvement (Π) algorithm for the solution. In [91], how to deploy the servers in an economical and efficient manner without violating the predetermined QoS was investigated by Yao et al. In particular, they practically

considered that the available cloudlet servers are heterogeneous. That means the servers have different cost and resource capacities. The problem was formulated into an ILP form, and a low-complexity heuristic algorithm was proposed to address it. Extensive simulation studies validated the efficiency of the algorithm. Meng et al. [92] considered a novel MCC architecture composed of cloud server, cloudlet and MUs to ensure low latency and energy consumption. The joint optimization strategy was proposed to enhance the QoS. They formulated the wireless bandwidth and computing resource allocation model as a three stage Stackelberg game, and then used the backward method to address it. An iterative algorithm was used to achieve the Stackelberg equilibrium. The effectiveness of the method was demonstrated.

VM Migration. Some studies on virtual machine (VM) migration based on cloud have been well investigated ([31, 93–98]). However, these schemes did not take into account the relationship between MU mobility and VM migration, and thus they cannot be directly used in MEC. In [99], Raei et al. modeled a type of MCC known as the cloudlet where the MUs received services from a cloudlet as an intermediary node. A dedicated VM was provisioned on a physical machine (PM) while the PM could be located as a part of the cloudlet or a public cloud. In addition, they also proposed a combined performance and availability model based on the Stochastic Reward Net (SRN) in [100]. Sun et al. [101] proposed a Green-energy aware Avatar Placement (GAP) policy to minimize the total on-grid power consumption of the cloudlets by migrating Avatars among the cloudlets. It was shown that GAP can save 57.1% and 57.6% of on-grid power consumption. A method was proposed by Rodrigues et al. [102] for minimizing service delay in a scenario with two cloudlet servers. Differ from the previous researches, the method focused on both the computation element and communication element, controlling processing delay through VM migration. It was shown that the proposal presented the lowest service delay in all research cases. In [103], the VM migration problem was formulated as a one-on-one contract game model and a learning-based price control mechanism was developed to effectively deal with the MEC's resources. Finally, the extensive simulation results demonstrated the efficiency of the proposed approach. In [104], two dynamic proxy VM migration methods were proposed to minimize the E2E delay between proxy VMs and the IoT devices, as well as minimized the total energy consumption of the cloudlets. In [105], Wang et al. studied the dynamic service migration problem in MEC. They formulated a sequential decision making problem for service migration based on the framework of Markov Decision Process (MDP). A new algorithm and a numerical technique was developed for the solution. The effectiveness of the approach on a real-world dataset was shown. In [106], a virtual FD-enabled small cell network with cache and MEC was investigated for two heterogeneous services, namely, high-data-rate service and computation-sensitive service. Then they formulated a virtual resource allocation problem, because the original issue was a mixed combinatorial problem which was converted into a convex problem. In addition, the effectiveness of the proposed mode was verified by different system configurations.

TABLE 5: Summary of literatures on applications of MEC.

item	Related work	Key points
Applications	[104, 107–111, 113]	[104] Mobile edge Internet of Things (MEIoT)
		[107] Video streaming smart cities
		[108] A novel approach to MEC (edgeIoT)
		[109] Wearable devices and clouds in realistic setups
		[110] Special issue on the edge of the cloud
		[111] A UAV-based MCC system
		[113] OCR-like arbitrarily divisible applications

4.3.2. Resource Allocation in Base Station Based MEC System. Resource allocation and computation offloading are usually jointly considered in base station based MEC system. And thus we do not repeat introduce the literatures here. More information can be find in Section 3.

5. Applications on MEC

In this section, applications on MEC are reviewed. A summary of literatures on applications of MEC is shown in Table 5.

MEC not only remarkably reduces the cost of network operation and improving QoS of MUs by pushing computation resources closer to the network edges, but also provides a scalable IoT architecture for time-sensitive applications. In [104], Ansari et al. proposed a Mobile Edge Internet of Things (MEIoT) architecture which brought many resources (i.e., computing resource and storage resource) close to IoT devices. Taleb et al. [107] proposed an approach to enhance MUs' experience of video streaming in smart cities. The proposed approach relied on the MEC concept as a key factor to in improving QoS. It maintains QoS by guaranteeing services follow the MUs' mobility and implement the concept of "Follow Me Edge". This scheme provided an important solution to reduce core network traffic and ensure ultra-short delay. Sun et al. [108] proposed a novel approach to MEC named edgeIoT to figure out the data streams at the mobile edge. More specifically, each BS was connected to a fog node for providing computing resources locally. The SDN-based cellular core was designed to facilitate data forwarding among fog nodes. Wearable devices (i.e., smart watches, glasses, and helmets) were becoming more and more popular and were expected to become an indispensable part in our daily life. Despite the continuous upgrade of hardware, the life-time of MUs and functions (i.e., computing, storage) still need to be further improved. MCC can augment the capabilities of wearable devices by providing services. In [109], the authors presented a comprehensive analysis of computational offloading between wearable devices and cloud. The convergence of mobile computing and CC depends on reliable high bandwidth E2E network. These basic requirements are difficult to guarantee in harsh environments (i.e., military operations and disaster recovery). In [110], the authors examined how VM-based cloudlets that are located in close proximity to associated MUs can address these challenges. UAVs have been used to provide strengthened coverage or relay services for MUs in limited infrastructure wireless systems. In [111],

a MCC system based on UAV was studied, in which mobile UAVs were given computing power to provider computation offloading opportunities to MUs. The system aimed to minimize the total energy cost of MU and meet the QoS requirements of offloading applications. Offloading was achieved through uplink and downlink communications between MU and the UAV. They formulated and solved a problem of the jointly optimizing of the bit allocation in uplink and downlink communications. The numerical results showed that a large amount of energy can be saved compared with local mobile execution. In addition, a novel EC empowered radio access network architecture was proposed by dong et al. [112], where the links of fronthaul and backhaul are mounted on the UAVs for fast event response and flexible deployment. Li et al. [113] figure out the problem on how to partition and allocate divisible applications to available resources in MEC environments to minimize the completion time of the applications. A theoretical model was developed for partitioning an OCR-like arbitrarily divisible application on the basis of the load of the application and the capabilities of available resources, and the solutions were derived in closed form.

6. Open Issue

In this section, the MEC challenges are introduced. In Section 6.1 open issues for MUs-oriented service adoption and ES-oriented service provision are introduced and then we describe other open issues, including Section 6.2.1 security issue and Section 6.2.2 simulation tools.

6.1. Open Issues for MU-Oriented Service Adoption and ES-Oriented Service Provision. MEC is a very novel and promising research area. Although we have introduced many studies in this paper, there still exist open research issues. From the MU's point of view, it is necessary to further design an efficient algorithm for multiple MUs to select a cloudlet that satisfies different service requirements. From the perspective of ES, MUs and cloudlet need to be jointly optimized for cloudlet placement issues. On the one hand, considering the limited resources of cloudlet, on the other hand, there are tremendous amounts of MUs in WMAN, so it is critical to study low-complexity cloudlet placement and scheduling algorithms. In addition, considering the dynamic changes of the MU's request and the energy consumption of cloudlet, multicloudlet collaboration method and VM

migration algorithm need to be further studied. Service recommendations [114] in MEC are also a promising research issue.

6.2. Other Issues

6.2.1. Security Issue. The main purpose of [115] is to holistically analyze the security threats, challenges, and mechanisms inherent in all edge paradigms. Wang et al. propose fog-based storage technology to fight with Cyber Threat [116]. Not only the traditional security issues, but also a number of new security issues in MEC also need to be concerned about. MEC consists of a large number of mobile devices, so it is also necessary to effectively protect MUs' privacy. Moreover, because the MEC can be seen as a network of multiple nodes, the assessment of node importance [117] and the assessment of invulnerability [118] of MEC also needs to be investigated. In our previous research [119], an Intrusion Detection System (IDS) was developed based on decision tree. Firstly, we developed a preprocessing algorithm to digitize the strings in the given dataset and then normalized the whole data. Secondly, we used decision tree method for our IDS system, and then we compared this method with other two methods, i.e., Naïve Bayesian and KNN. The effectiveness and precision of our proposed IDS system are shown. Above all, security and privacy have always been a fundamental research issue. We emphasize that security issues need to attract more attention.

6.2.2. Simulation Tools. To the best of our knowledge, many tools like Matlab, JAVA, and Python can be used for simulation for CC. In addition, CloudSim [120] is a well-known tool for cloud. In this section, we mainly review the tools for EC and MEC. More especially, we also introduce the tools for FC. A new simulator tool called EdgeCloudSim streamlined for EC scenarios is proposed in [121]. EdgeCloudSim is built on CloudSim to solve the specific needs of EC research and support necessary functionality in terms of computation and networking capabilities. Gupta et al. [122] proposed a simulator called iFogSim to model IoT and FC environments. They described two case studies to demonstrate modeling of an IoT environment and the comparison of resource management strategies. In addition, the scalability of the simulation toolkit of RAM consumption and runtime was verified under different scenarios. The authors in [123] discussed resource allocation in FC in the view of MUs' mobility and introduced MyiFogSim, an extension of iFogSim, which supported mobility through VM migration among cloudlets. In sum, the development of simulation tools is very promising, which can effectively promote the development of MEC and the standardization of experimental design.

7. Conclusion

With the development of mobile network and 5G, MEC has become a promising field in these years. It not only meets the user's more business needs, improves the QoS and QoE of MU, but also brings business benefits to service providers. In this paper, we present a comprehensive survey of MEC from the perspective of service adoption and provision. We firstly describe the overview of MEC. After that we review the existing MUs-oriented service adoption of MEC. And then we survey ES-oriented service provision. Moreover, other issues like applications on MEC, open issues are investigated. We highlight that more researches should focus on services of MEC.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] K. Peng, R. Lin, B. Huang, H. Zou, and F. Yang, "Link importance evaluation of data center network based on maximum flow," *Journal of Internet Technology*, vol. 18, no. 1, pp. 23–31, 2017.

[2] W. Quan, Y. Liu, H. Zhang, and S. Yu, "Enhancing crowd collaborations for software defined vehicular networks," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 80–86, 2017.

[3] L. Qi, W. Dou, Y. Zhou, J. Yu, and C. Hu, "A context-aware service evaluation approach over big data for cloud applications," *IEEE Transactions on Cloud Computing*, vol. 1, pp. 1-1, 2015.

[4] Li. W, Y. Xia, M. Zhou, X. Sun, and Q. Zhu, "Fluctuation-aware and predictive workflow scheduling in cost-effective Infrastructure-as-a-Service clouds," *IEEE Access*, 2018.

[5] X. Xu, X. Zhao, F. Ruan et al., "Data placement for privacy-aware applications over big data in hybrid clouds," *Security and Communication Networks*, vol. 2017, pp. 1–15, 2017.

[6] L. Qi, X. Xu, X. Zhang et al., "Structural balance theory-based e-commerce recommendation over big rating data," *IEEE Transactions on Big Data*, 2016.

[7] K. Peng, H. Zou, R. Lin, and F. Yang, "Small business-oriented index construction of cloud data," in *Proceedings of the International Conference on Algorithms and Architectures for Parallel Processing*, pp. 156–165, Springer, Berlin, Germany, 2012.

[8] T. Wang, M. Z. A. Bhuiyan, G. Wang, M. A. Rahman, J. Wu, and J. Cao, "Big data reduction for a smart citys critical infrastructural health monitoring," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 128–133, 2018.

[9] H. Xiang, X. Xu, H. Zheng et al., "An adaptive cloudlet placement method for mobile applications over GPS big data," in *Proceedings of the Global Communications Conference (GLOBECOM, 2016)*, pp. 1–6, IEEE, 2016.

[10] H. Zhou, V. C. M. Leung, C. Zhu, S. Xu, and J. Fan, "Predicting temporal social contact patterns for data forwarding in opportunistic mobile networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10372–10383, 2017.

[11] K. Peng, V. C. M. Leung, and Q. Huang, "Clustering Approach Based on Mini Batch Kmeans for Intrusion Detection System over Big Data," *IEEE Access*, vol. 6, pp. 11897–11906, 2018.

[12] Cisco, Cisco visual networking index: global mobile data traffic forecast update, 2016-2021 white paper (EB/OL). (2017-03-01). http://10.3.200.202/cache/10/03/cisco.com/89e8529e7886890c-828d4a976994f806/mobile-white-paper-c11-520862.pdf.

[13] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, "A survey of mobile cloud computing: architecture, applications, and approaches," *Wireless Communications and Mobile Computing*, vol. 13, no. 18, pp. 1587–1611, 2013.

[14] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile Edge Computing: A Survey," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450–465, 2018.

[15] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.

[16] S. Wang, J. Xu, N. Zhang, and Y. Liu, "A Survey on Service Migration in Mobile Edge Computing," *IEEE Access*, vol. 6, pp. 23511–23528, 2018.

[17] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, "A Survey on Mobile Edge Networks: Convergence of Computing, Caching and Communications," *IEEE Access*, vol. 5, pp. 6757–6779, 2017.

[18] European Telecommunications Standards Institute, Multi-access Edge Computing, http://www.etsi.org/technologies-clusters/technologies/multi-access-edge-computing.

[19] T. Zhao, S. Zhou, X. Guo, Y. Zhao, and Z. Niu, "A cooperative scheduling scheme of local cloud and internet cloud for delay-aware mobile cloud computing," in *Proceedings of the IEEE Globecom Workshops, GC Wkshps 2015*, USA, December 2015.

[20] R. R. Sarukkai and A. Mendhekar, Yahoo! Inc, (2004). Method and apparatus for accessing targeted, personalized voice/audio web content through wireless devices. U.S. Patent 6,728,731.

[21] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing—A key technology towards 5G," *ETSI white paper*, vol. 11, no. 11, pp. 1–16, 2015.

[22] W. S. Shi, H. Sun, and J. Cao, "Edge computing: an emerging computing model for the internet of everything era," *Journal of Computer Research and Development*, vol. 54, no. 5, pp. 907–924, 2017.

[23] P. Mach and Z. Becvar, "Mobile Edge Computing: A Survey on Architecture and Computation Offloading," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1628–1656, 2017.

[24] B. Liang, *Mobile edge computing*, Cambridge University Press, 2017.

[25] H. Tian, S.-S. Fan, X.-C. Lü, P.-T. Zhao, and S. He, "Mobile Edge Computing for 5G Requirements," *Beijing Youdian Daxue Xuebao/Journal of Beijing University of Posts and Telecommunications*, vol. 40, no. 2, pp. 1–10, 2017.

[26] B. P. Rimal, D. P. Van, and M. Maier, "Mobile Edge Computing Empowered Fiber-Wireless Access Networks in the 5G Era," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 192–200, 2017.

[27] K. Dolui and S. K. Datta, "Comparison of edge computing implementations: Fog computing, cloudlet and mobile edge computing," in *Proceedings of the 2017 Global Internet of Things Summit, GIoTS 2017*, pp. 1–6, Geneva, Switzerland, June 2017.

[28] M. S. Elbamby, M. Bennis, and W. Saad, "Proactive edge computing in latency-constrained fog networks," in *Proceedings of the 2017 European Conference on Networks and Communications, EuCNC 2017*, Finland, June 2017.

[29] M. Ali, "Green cloud on the horizon," in *Proceedings of the IEEE International Conference on Cloud Computing*, pp. 451–459, Berlin, Germany, December 2009.

[30] http://www.mobilecloudcomputingforum.com/.

[31] F. Liu, P. Shu, H. Jin et al., "Gearing resource-poor mobile devices with powerful clouds: architectures, challenges, and applications," *IEEE Wireless Communications Magazine*, vol. 20, no. 3, pp. 14–21, 2013.

[32] Z. Liu, Y. Feng, and B. Li, "Socialize spontaneously with mobile applications," in *Proceedings of the IEEE Conference on Computer Communications, INFOCOM 2012*, pp. 1942–1950, USA, March 2012.

[33] M. Satyanarayanan, P. Bahl, R. Cáceres, and N. Davies, "The case for VM-based cloudlets in mobile computing," *IEEE Pervasive Computing*, vol. 8, no. 4, pp. 14–23, 2009.

[34] D. Huang, T. Xing, and H. Wu, "Mobile cloud computing service models: a user-centric approach," *IEEE Network*, vol. 27, no. 5, pp. 6–11, 2013.

[35] Y. Jararweh, A. Doulat, O. Alqudah, E. Ahmed, M. Al-Ayyoub, and E. Benkhelifa, "The future of mobile cloud computing: Integrating cloudlets and Mobile Edge Computing," in *Proceedings of the 23rd International Conference on Telecommunications, ICT 2016*, Greece, May 2016.

[36] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," in *Proceedings of the 1st ACM Mobile Cloud Computing Workshop, MCC 2012*, pp. 13–16, Finland, August 2012.

[37] L. M. Vaquero and L. Rodero-Merino, "Finding your way in the fog: Towards a comprehensive definition of fog computing," *Computer Communication Review*, vol. 44, no. 5, pp. 27–32, 2014.

[38] T. H. Luan, L. Gao, Z. Li et al., *Fog computing: Focusing on mobile users at the edge*, https://arxiv.org/abs/1502.01815.

[39] G. I. Klas, Fog computing and mobile edge cloud gain momentum open fog consortium, etsi mec and cloudlets. Google Scholar, 2015.

[40] Blue an exclusive wifi enabled onboard entertainment system https://www.greyhound.com/en/about/media/2013/07-08-2013.

[41] S. Yi, C. Li, and Q. Li, "A survey of fog computing: concepts, applications and issues," in *Proceedings of the Workshop on Mobile Big Data (Mobidata '15)*, pp. 37–42, ACM, Hangzhou, China, June 2015.

[42] S. Yi, Z. Hao, Z. Qin, and Q. Li, "Fog computing: Platform and applications," in *Proceedings of the 3rd Workshop on Hot Topics in Web Systems and Technologies, HotWeb 2015*, pp. 73–78, USA, November 2015.

[43] M. R. Anawar, S. Wang, M. Azam Zia et al., "Fog Computing: An Overview of Big IoT Data Analytics," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 7157192, 22 pages, 2018.

[44] S. Clinch, J. Harkes, A. Friday, N. Davies, and M. Satyanarayanan, "How close is close enough? Understanding the role of cloudlets in supporting display appropriation by mobile users," in *Proceedings of the 10th IEEE International Conference on Pervasive Computing and Communications, PerCom 2012*, pp. 122–127, Switzerland, March 2012.

[45] Y. Jararweh, M. Al-Ayyoub, M. Al-Quraan, L. A. Tawalbeh, and E. Benkhelifa, "Delay-aware power optimization model for mobile edge computing systems," *Personal and Ubiquitous Computing*, vol. 21, no. 6, pp. 1067–1077, 2017.

[46] M. Al-Quraan, M. Al-Ayyoub, Y. Jararweh, L. Tawalbeh, and E. Benkhelifa, "Power optimization of large scale mobile cloud system using cooperative cloudlets," in *Proceedings of the 4th International Conference on Future Internet of Things and Cloud Workshops, W-FiCloud 2016*, pp. 34–38, Austria, August 2016.

[47] P. Hu, S. Dhelim, H. Ning, and T. Qiu, "Survey on fog computing: architecture, key technologies, applications and open issues," *Journal of Network and Computer Applications*, vol. 98, pp. 27–42, 2017.

[48] W. Zhang, S. Tan, F. Xia et al., "A survey on decision making for task migration in mobile cloud environments," *Personal and Ubiquitous Computing*, vol. 20, no. 3, pp. 295–309, 2016.

[49] K. Kumar, J. Liu, Y.-H. Lu, and B. Bhargava, "A survey of computation offloading for mobile systems," *Mobile Networks and Applications*, vol. 18, no. 1, pp. 129–140, 2013.

[50] H. Zhou, H. Wang, X. Li, and V. C. M. Leung, "A Survey on Mobile Data Offloading Technologies," *IEEE Access*, vol. 6, pp. 5101–5111, 2018.

[51] D. G. Roy, D. De, A. Mukherjee, and R. Buyya, "Application-aware cloudlet selection for computation offloading in multi-cloudlet environment," *The Journal of Supercomputing*, vol. 73, no. 4, pp. 1672–1690, 2017.

[52] D. Parmar, A. S. Kumar, A. Nivangune, P. Joshi, and U. P. Rao, "Discovery and selection mechanism of cloudlets in a decentralized MCC environment," in *Proceedings of the IEEE/ACM International Conference on Mobile Software Engineering and Systems, MobileSoft 2016*, pp. 15-16, USA, May 2016.

[53] A. Mukherjee, D. De, and D. Roy, "A Power and Latency Aware Cloudlet Selection Strategy for Multi-Cloudlet Environment," *IEEE Transactions on Cloud Computing*, 2016.

[54] L. Liu and Q. Fan, "Resource Allocation Optimization Based on Mixed Integer Linear Programming in the Multi-Cloudlet Environment," *IEEE Access*, vol. 6, pp. 24533–24542, 2018.

[55] S. Yu, X. Wang, and R. Langar, "Computation offloading for mobile edge computing: a deep learning approach," in *Proceedings of the Personal, Indoor, and Mobile Radio Communications (PIMRC, 2017)*, pp. 1–6, IEEE, 2017.

[56] T. Q. Dinh, J. Tang, Q. D. La, and T. Q. S. Quek, "Offloading in Mobile Edge Computing: Task Allocation and Computational Frequency Scaling," *IEEE Transactions on Communications*, vol. 65, no. 8, pp. 3571–3584, 2017.

[57] J. Liu and Q. Zhang, "Offloading Schemes in Mobile Edge Computing for Ultra-Reliable Low Latency Communications," *IEEE Access*, vol. 6, pp. 12825–12837, 2018.

[58] Y. Wang, M. Sheng, X. Wang, L. Wang, and J. Li, "Mobile-Edge Computing: Partial Computation Offloading Using Dynamic Voltage Scaling," *IEEE Transactions on Communications*, vol. 64, no. 10, pp. 4268–4282, 2016.

[59] H. Cao and J. Cai, "Distributed multiuser computation offloading for cloudlet-based mobile cloud computing: A game-theoretic machine learning approach," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 1, pp. 752–764, 2018.

[60] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient Multi-User Computation Offloading for Mobile-Edge Cloud Computing," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 2795–2808, 2016.

[61] Y. Mao, J. Zhang, S. H. Song, and K. B. Letaief, "Power-delay tradeoff in multi-user mobile-edge computing systems," in *Proceedings of the In Global Communications Conference (GLOBECOM, 2016)*, pp. 1–6, 2016.

[62] L. Liu, Z. Chang, X. Guo, and T. Ristaniemi, "Multi-objective optimization for computation offloading in mobile-edge computing," in *Proceedings of the 2017 IEEE Symposium on Computers and Communications, ISCC 2017*, pp. 832–837, Greece, July 2017.

[63] J. Guo, Z. Song, Y. Cui, Z. Liu, and Y. Ji, "Energy-efficient resource allocation for multi-user mobile edge computing," in *Proceedings of the GLOBECOM 2017-2017 IEEE Global Communications Conference*, pp. 1–7, 2017.

[64] C. Wang, F. R. Yu, C. Liang, Q. Chen, and L. Tang, "Joint Computation Offloading and Interference Management in Wireless Cellular Networks with Mobile Edge Computing," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7432–7445, 2017.

[65] K. Zhang, Y. Mao, S. Leng et al., "Energy-Efficient Offloading for Mobile Edge Computing in 5G Heterogeneous Networks," *IEEE Access*, vol. 4, pp. 5896–5907, 2016.

[66] F. Wang, J. Xu, X. Wang, and S. Cui, "Joint Offloading and Computing Optimization in Wireless Powered Mobile-Edge Computing Systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1784–1797, 2018.

[67] I. Ketykó, L. Kecskés, C. Nemes et al., "Multi-user computation offloading as multiple knapsack problem for 5G mobile edge computing," in *Proceedings of the Networks and Communications (EuCNC), 2016 European Conference*, pp. 225–229, 2016.

[68] Y. Mao, J. Zhang, S. H. Song, and K. B. Letaief, "Stochastic Joint Radio and Computational Resource Management for Multi-User Mobile-Edge Computing Systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 5994–6009, 2017.

[69] K. Guo, M. Yang, Y. Zhang, and Y. Ji, "An Efficient Dynamic Offloading Approach based on Optimization Technique for Mobile Edge Computing," in *Proceedings of the Mobile Cloud Computing, Services, and Engineering (MobileCloud, 2018)*, pp. 29–36, 2018.

[70] T. X. Tran and D. Pompili, *Joint task offloading and resource allocation for multi-server mobile-edge computing networks*, https://arxiv.org/abs/1705.00704.

[71] T. Lei, S. Wang, J. Li, and F. Yang, "AOM: adaptive mobile data traffic offloading for M2M networks," *Personal and Ubiquitous Computing*, vol. 20, no. 6, pp. 863–873, 2016.

[72] S. Wang, T. Lei, L. Zhang, C.-H. Hsu, and F. Yang, "Offloading mobile data traffic for QoS-aware service provision in vehicular cyber-physical systems," *Future Generation Computer Systems*, vol. 61, pp. 118–127, 2016.

[73] F. Cheng, S. Zhang, Z. Li et al., "UAV Trajectory Optimization for Data Offloading at the Edge of Multiple Cells," *IEEE Transactions on Vehicular Technology*, 2018.

[74] T. Zhang, W. Chen, and F. Yang, *Data offloading in mobile edge computing: A coalitional game based pricing approach*, 2017, https://arxiv.org/abs/1709.04148.

[75] A.-L. Jin, W. Song, and W. Zhuang, "Auction-Based Resource Allocation for Sharing Cloudlets in Mobile Cloud Computing," *IEEE Transactions on Emerging Topics in Computing*, vol. 6, no. 1, pp. 45–57, 2018.

[76] W. Fang, X. Yao, X. Zhao, J. Yin, and N. Xiong, "A Stochastic Control Approach to Maximize Profit on Service Provisioning for Mobile Cloudlet Platforms," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 4, pp. 522–534, 2018.

[77] X. Sun and N. Ansari, "PRIMAL: PRofIt Maximization Avatar pLacement for mobile edge computing," in *Proceedings of the 2016 IEEE International Conference on Communications, ICC 2016*, Malaysia, May 2016.

[78] M. Jia, W. Liang, Z. Xu, and M. Huang, "Cloudlet load balancing in wireless metropolitan area networks," in *Proceedings of the INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, pp. 1–9, San Francisco, CA, USA, April 2016.

[79] X. Xu, S. Fu, Q. Cai et al., "Dynamic Resource Allocation for Load Balancing in Fog Environment," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 6421607, 15 pages, 2018.

[80] J. Oueis, E. C. Strinati, and S. Barbarossa, "The fog balancing: Load distribution for small cell cloud computing," in *Proceedings of the 81st IEEE Vehicular Technology Conference, VTC Spring 2015*, UK, May 2015.

[81] U. Shaukat, E. Ahmed, Z. Anwar, and F. Xia, "Cloudlet deployment in local wireless networks: Motivation, architectures, applications, and open challenges," *Journal of Network and Computer Applications*, vol. 62, pp. 18–40, 2016.

[82] Z. Xu, W. Liang, W. Xu, M. Jia, and S. Guo, "Efficient Algorithms for Capacitated Cloudlet Placements," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 10, pp. 2866–2880, 2016.

[83] L. Qiu, V. N. Padmanabhan, and G. M. Voelker, "On the placement of web server replicas," in *Proceedings of the INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 3, pp. 1587–1596, 2001.

[84] H. Yin, X. Zhang, T. Zhan, Y. Zhang, G. Min, and D. O. Wu, "NetClust: A framework for scalable and pareto-optimal media server placement," *IEEE Transactions on Multimedia*, vol. 15, no. 8, pp. 2114–2124, 2013.

[85] M. Charikar, S. Guha, E. Tardos, and D. B. Shmoys, "Constant-factor approximation algorithm for the k-median problem," in *Proceedings of the 1999 31st Annual ACM Symposium on Theory of Computing - FCRC '99*, pp. 1–10, May 1999.

[86] M. Jia, J. Cao, and W. Liang, "Optimal cloudlet placement and user to cloudlet allocation in wireless metropolitan area networks," *IEEE Transactions on Cloud Computing*, 2015.

[87] T.-Y. Liang and Y.-J. Li, "A location-aware service deployment algorithm based on k-means for cloudlets," *Mobile Information Systems*, vol. 2017, Article ID 8342859, 10 pages, 2017.

[88] G. Yang, Q. Sun, A. Zhou, S. Wang, and J. Li, "Poster abstract: Access point ranking for cloudlet placement in edge computing environment," in *Proceedings of the 1st IEEE/ACM Symposium on Edge Computing, SEC 2016*, pp. 85-86, USA, October 2016.

[89] M. Al-Ayyoub, Y. Jararweh, L. Tawalbeh, E. Benkhelifa, and A. Basalamah, "Power Optimization of Large Scale Mobile Cloud Computing Systems," in *Proceedings of the 3rd International Conference on Future Internet of Things and Cloud, FiCloud 2015*, pp. 670–674, Italy, August 2015.

[90] W. Wang and W. Zhou, "Computational offloading with delay and capacity constraints in mobile edge," in *Proceedings of the 2017 IEEE International Conference on Communications, ICC 2017*, France, May 2017.

[91] H. Yao, C. Bai, M. Xiong, D. Zeng, and Z. Fu, "Heterogeneous cloudlet deployment and user-cloudlet association toward cost effective fog computing," *Concurrency Computation*, vol. 29, no. 16, 2017.

[92] S. Meng, Y. Wang, Z. Miao, and K. Sun, "Joint optimization of wireless bandwidth and computing resource in cloudlet-based mobile cloud computing environment," *Peer-to-Peer Networking and Applications*, vol. 11, no. 3, pp. 462–472, 2018.

[93] B.-G. Chun, S. Ihm, P. Maniatis, M. Naik, and A. Patti, "CloneCloud: elastic execution between mobile device and cloud," in *Proceedings of the 6th ACM EuroSys Conference on Computer Systems (EuroSys '11)*, pp. 301–314, ACM, April 2011.

[94] F. Liu, P. Shu, and J. C. S. Lui, "AppATP: An Energy Conserving Adaptive Mobile-Cloud Transmission Protocol," *IEEE Transactions on Computers*, vol. 64, no. 11, pp. 3051–3063, 2015.

[95] F. Xu, F. Liu, L. Liu, H. Jin, B. Li, and B. Li, "iAware: making live migration of virtual machines interference-aware in the cloud," *Institute of Electrical and Electronics Engineers. Transactions on Computers*, vol. 63, no. 12, pp. 3012–3025, 2014.

[96] F. Xu, F. Liu, H. Jin, and A. V. Vasilakos, "Managing performance overhead of virtual machines in cloud computing: A survey, state of the art, and future directions," *Proceedings of the IEEE*, vol. 102, no. 1, pp. 11–31, 2014.

[97] B. Huang, R. Lin, K. Peng, H. Zou, and F. Yang, "Minimizing latency in fetching virtual machine images based on multi-point collaborative approach," in *Proceedings of the 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, GreenCom-iThings-CPSCom 2013*, pp. 262–267, China, August 2013.

[98] X. Xu, X. Zhang, M. Khan, W. Dou, S. Xue, and S. Yu, "A balanced virtual machine scheduling method for energy-performance trade-offs in cyber-physical cloud systems," *Future Generation Computer Systems*, 2017.

[99] H. Raei, N. Yazdani, and R. Shojaee, "Modeling and performance analysis of cloudlet in Mobile Cloud Computing," *Performance Evaluation*, vol. 107, pp. 34–53, 2017.

[100] H. Raei and N. Yazdani, "Performability analysis of cloudlet in mobile cloud computing," *Information Sciences*, vol. 388-389, pp. 99–117, 2017.

[101] X. Sun and N. Ansari, "Green Cloudlet Network: A Sustainable Platform for Mobile Cloud Computing," *IEEE Transactions on Cloud Computing*, 2017.

[102] T. G. Rodrigues, K. Suto, H. Nishiyama, and N. Kato, "Hybrid method for minimizing service delay in edge cloud computing through VM migration and transmission power control," *Institute of Electrical and Electronics Engineers. Transactions on Computers*, vol. 66, no. 5, pp. 810–819, 2017.

[103] S. Kim, "One-on-one contract game–based dynamic virtual machine migration scheme for Mobile Edge Computing," *Transactions on Emerging Telecommunications Technologies*, vol. 29, no. 1, 2018.

[104] N. Ansari and X. Sun, "Mobile edge computing empowers internet of things," *IEICE Transactions on Communications*, vol. E101B, no. 3, pp. 604–619, 2018.

[105] S. Wang, R. Urgaonkar, M. Zafer, T. He, K. Chan, and K. K. Leung, "Dynamic service migration in mobile edge-clouds," in *Proceedings of the 2015 14th IFIP Networking Conference, IFIP Networking 2015*, France, May 2015.

[106] Z. Tan, F. R. Yu, and X. Li, "Heterogeneous Services Provisioning in Small Cell Networks with Cache and Mobile Edge Computing," https://arxiv.org/abs/1706.09542.

[107] T. Taleb, S. Dutta, A. Ksentini, M. Iqbal, and H. Flinck, "Mobile edge computing potential in making cities smarter," *IEEE Communications Magazine*, vol. 55, no. 3, pp. 38–43, 2017.

[108] X. Sun and N. Ansari, "EdgeIoT: Mobile Edge Computing for the Internet of Things," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 22–29, 2016.

[109] C. Ragona, F. Granelli, C. Fiandrino, D. Kliazovich, and P. Bouvry, "Energy-efficient computation offloading for wearable

devices and smartphones in mobile cloud computing," in *Proceedings of the 58th IEEE Global Communications Conference, GLOBECOM 2015*, December 2015.

[110] M. Satyanarayanan, G. Lewis, E. Morris, S. Simanta, J. Boleng, and K. Ha, "The role of cloudlets in hostile environments," *IEEE Pervasive Computing*, vol. 12, no. 4, pp. 40–49, 2013.

[111] S. Jeong, O. Simeone, and J. Kang, "Mobile Edge Computing via a UAV-Mounted Cloudlet: Optimization of Bit Allocation and Path Planning," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2049–2063, 2018.

[112] Y. Dong, M. Hassan, J. Cheng et al., *An edge computing empowered radio access network with UAV-mounted FSO fronthaul and backhaul: Key challenges and approaches*, https://arxiv.org/abs/1803.06381.

[113] B. Li, M. He, W. Wu, A. K. Sangaiah, and G. Jeon, "Computation Offloading Algorithm for Arbitrarily Divisible Applications in Mobile Edge Computing Environments: An OCR Case," *Sustainability* , vol. 10, no. 5, p. 1611, 2018.

[114] S. Wang, Y. Zhao, L. Huang, J. Xu, and C. Hsu, "QoS prediction for service recommendations in mobile edge computing," *Journal of Parallel and Distributed Computing*.

[115] R. Roman, J. Lopez, M. Mambo et al., "Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges," *Future Generation Computer Systems*, vol. 78, pp. 680–698, 2018.

[116] T. Wang, J. Zhou, M. Huang et al., "Fog-based storage technology to fight with cyber threat," *Future Generation Computer Systems*, vol. 83, pp. 208–218, 2018.

[117] K. Peng, R. Lin, B. Huang, H. Zou, and F. Yang, "Node importance of data center network based on contribution matrix of information entropy," *Journal of Networks*, vol. 8, no. 6, pp. 1248–1254, 2013.

[118] K. Peng and B. Huang, "The invulnerability studies on data center network," *International Journal of Security and Its Applications*, vol. 9, no. 11, pp. 167–186, 2015.

[119] K. Peng, V. C. M. Leung, L. Zheng et al., "Intrusion Detection System Based on Decision Tree over Big Data in Fog Environment," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 4680867, 10 pages, 2018.

[120] Cloudsim. http://www.cloudbus.org/cloudsim/.

[121] C. Sonmez, A. Ozgovde, and C. Ersoy, "EdgeCloudSim: An environment for performance evaluation of Edge Computing systems," in *Proceedings of the 2nd International Conference on Fog and Mobile Edge Computing, FMEC 2017*, pp. 39–44, Spain, May 2017.

[122] H. Gupta, A. Vahid Dastjerdi, S. K. Ghosh, and R. Buyya, "iFogSim: A toolkit for modeling and simulation of resource management techniques in the Internet of Things, Edge and Fog computing environments," *Software: Practice and Experience*, vol. 47, no. 9, pp. 1275–1296, 2017.

[123] M. M. Lopes, W. A. Higashino, M. A. Capretz, and L. F. Bittencourt, "MyiFogSim: A Simulator for Virtual Machine Migration in Fog Computing," in *Proceedings of the10th International Conference on Utility and Cloud Computing*, pp. 47–52, Austin, Texas, USA, December 2017.