IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 21, 2020, accepted June 6, 2020, date of publication June 10, 2020, date of current version July 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3001277

A Survey of Multi-Access Edge Computing in 5G and Beyond: Fundamentals, Technology Integration, and State-of-the-Art

QUOC-VIET PHAM[1], (Member, IEEE), FANG FANG[2,8], (Member, IEEE), VU NGUYEN HA[3], (Member, IEEE), MD. JALIL PIRAN[4], (Member, IEEE), MAI LE[5], LONG BAO LE[6], (Senior Member, IEEE), WON-JOO HWANG[7,9], (Senior Member, IEEE), AND ZHIGUO DING[2], (Fellow, IEEE)

[1]Research Institute of Computer, Information, and Communication, Pusan National University, Busan 46241, South Korea
[2]School of Electrical and Electronics Engineering, The University of Manchester, Manchester M13 9PL, U.K.
[3]École Polytechnique de Montréal, Montréal, QC H3C 3A7, Canada
[4]Department of Computer Science and Engineering, Sejong University, Seoul 05006, South Korea
[5]Department of Information and Communications System, Inje University, Gimhae 50834, South Korea
[6]Institut National de la Recherche Scientifique, Université du Québec, Montréal, QC H5A 1K6, Canada
[7]Department of Biomedical Convergence Engineering, Pusan National University, Busan 46241, South Korea
[8]Department of Engineering, Durham University, Durham DH1 3LE, U.K.
[9]Department of Information Convergence Engineering (Artificial Intelligence), Pusan National University, Busan 46241, Korea

Corresponding author: Won-Joo Hwang (wjhwang@pusan.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) funded by the Korea Government (MSIT) under Grant NRF-2019R1C1C1006143 and Grant NRF-2019R1I1A3A01060518. This work was also supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2020-0-01450, Artificial Intelligence Convergence Research Center [Pusan National University]).

ABSTRACT Driven by the emergence of new compute-intensive applications and the vision of the Internet of Things (IoT), it is foreseen that the emerging 5G network will face an unprecedented increase in traffic volume and computation demands. However, end users mostly have limited storage capacities and finite processing capabilities, thus how to run compute-intensive applications on resource-constrained users has recently become a natural concern. Mobile edge computing (MEC), a key technology in the emerging fifth generation (5G) network, can optimize mobile resources by hosting compute-intensive applications, process large data before sending to the cloud, provide the cloud-computing capabilities within the radio access network (RAN) in close proximity to mobile users, and offer context-aware services with the help of RAN information. Therefore, MEC enables a wide variety of applications, where the real-time response is strictly required, e.g., driverless vehicles, augmented reality, robotics, and immerse media. Indeed, the paradigm shift from 4G to 5G could become a reality with the advent of new technological concepts. The successful realization of MEC in the 5G network is still in its infancy and demands for constant efforts from both academic and industry communities. In this survey, we first provide a holistic overview of MEC technology and its potential use cases and applications. Then, we outline up-to-date researches on the integration of MEC with the new technologies that will be deployed in 5G and beyond. We also summarize testbeds and experimental evaluations, and open source activities, for edge computing. We further summarize lessons learned from state-of-the-art research works as well as discuss challenges and potential future directions for MEC research.

INDEX TERMS 5G and beyond network, heterogeneous networks, Internet of Things, machine learning, edge computing, non-orthogonal multiple access, testbeds, unmanned aerial vehicle, wireless power transfer and energy harvesting.

ACRONYMS

3GPP	3rd Generation Partnership Project
4C	Communication, Computation, Control, and Caching
5G	Fifth Generation of Mobile Networks
AI	Artificial Intelligence
API	Application Programming Interface
AR	Augmented Reality
BBU	Baseband Unit
BS	Base Station

The associate editor coordinating the review of this manuscript and approving it for publication was Zihuai Lin.

D2D	Device-to-Device
DC	Data Center
DL	Deep Learning
DQN	Deep Q Network
EH	Energy Harvesting
eNB	Evolved Node B
ETSI	European Telecommunications Standards Institute
FDMA	Frequency Devision Multiple Access
FiWi	Fiber-Wireelss
HetNets	Heterogeneous Networks
Het-MEC	Heterogeneous MEC
HD	High Definition
IoT	Internet of Things
LTE	Long-Term Evolution
MDP	Markov Decision Process
MIMO	Multiple-Input and Multiple-Output
MCC	Mobile Cloud Computing
MCS	Mobile Crowdsensing
MEC	Mobile Edge Computing
mmWave	millimeter Wave
ML	Machine Learning
MNO	Mobile Network Operator
NOMA	Non-Orthogonal Multiple Access
NFV	Network Function Virtualization
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RL	Reinforcement Learning
RRH	Remote Radio Head
RSU	Roadside Unit
SBC	Single-Board Computer
SCA	Successive Convex Approximation
SWIPT	Simultaneous Wireless Communication and Power Transfer
OFDM	Orthogonal Frequency-Division Multiplexing
OMA	Orthogonal Multiple Access
SDN	Software-Defined Networking
UAV	Unmanned Air Vehicle
V2X	Vehicle-to-Everything
VM	Virtual Machine
VR	Virtual Reality
WiFi	Wireless Fidelity
WPT	Wireless Power Transfer

I. INTRODUCTION

During the last four decades, the evolution of wireless communication networks has changed every aspect of our lives, society, culture, politics, and economics. Since the commercialization of the first generation (1G) of cellular networks in early 1980's, generations have been launched with enormous differences in terms of the network architectures, key technologies, coverage, mobility, security and privacy, data, spectral efficiency, cost optimality, and so on. The brief summary of wireless communication evolution is shown in Fig. 1. Now, both academic and industry communities are making tremendous efforts to finalize the 5G standardization and commercialization in 2019. 5G communications can be categorized into three categories: enhanced mobile broadband (eMBB), ultra-reliable low-latency communication (URLLC), and massive machine type communications (mMTC). Compared with previous generations, 5G will support not only communication, but also computation, control, and content delivery (4C) functions [1]. Moreover, many new applications and use cases are expected with the advent of 5G, for example, virtual/augmented reality (VR/AR), autonomous vehicle, Tactile Internet, and Internet of Things (IoT) scenarios. These applications are poised to induce a significant surge in demand for not only communication resources but also computation resources. To meet such ever-growing demands, various technological concepts have been developed for 5G in terms of radio access, network resource management, applications, network architectures and scenarios, power supply, and performance improvement [2]. For example, non-orthogonal multiple access (NOMA), dense heterogeneous networks (HetNets), cloud radio access network (C-RAN), unmanned aerial vehicle (UAV), IoT, wireless power transfer (WPT) and energy harvesting (EH), and machine learning (ML), have been considered as key enabling technologies.

The Cisco white paper [3] showed that global data traffic will grow at a compound annual growth rate (CAGR) of 26 percent between 2017 and 2022 (i.e., increase more than threefold) and reach 122 exabytes (EB) per month by 2022. Mobile and wireless networks carried 11.51 EB per month in 2017, 28.56 EB per month in 2019, and 77.49 EB per month at the end of 2022. Moreover, traffic generated by new applications and services will increase at a much higher CAGR, for example, 12-fold for AR and VR, ninefold for Internet gaming, and sevenfold for Internet video surveillance. It is also anticipated that the number of connected things (e.g., sensors and wearable devices) will reach 28.5 billion by 2022, up from 21.5 billion in 2019. However, most connected devices have limited communication and storage resources and finite processing capabilities, which show the mismatch between the stringent requirements for emerging applications and the actual device capabilities. Despite recent advancements in the hardware capability, mobile computing still cannot cope with the demand of many applications that need to generate, process, and store a massive amount of data and require large computing resources. One potential solution to these challenges is to transfer computations to centralized clouds, which can be, however, burdened by many issues, such as network congestion and privacy policies. This has driven the development of mobile edge computing (MEC).

Prior to MEC, there have been some similar computing concepts, for example, mobile cloud computing (MCC), cloudlet, and fog computing. MCC combines cloud computing, mobile computing, and wireless communication networks, thus enabling developers and service providers to support more complex applications by moving the computing

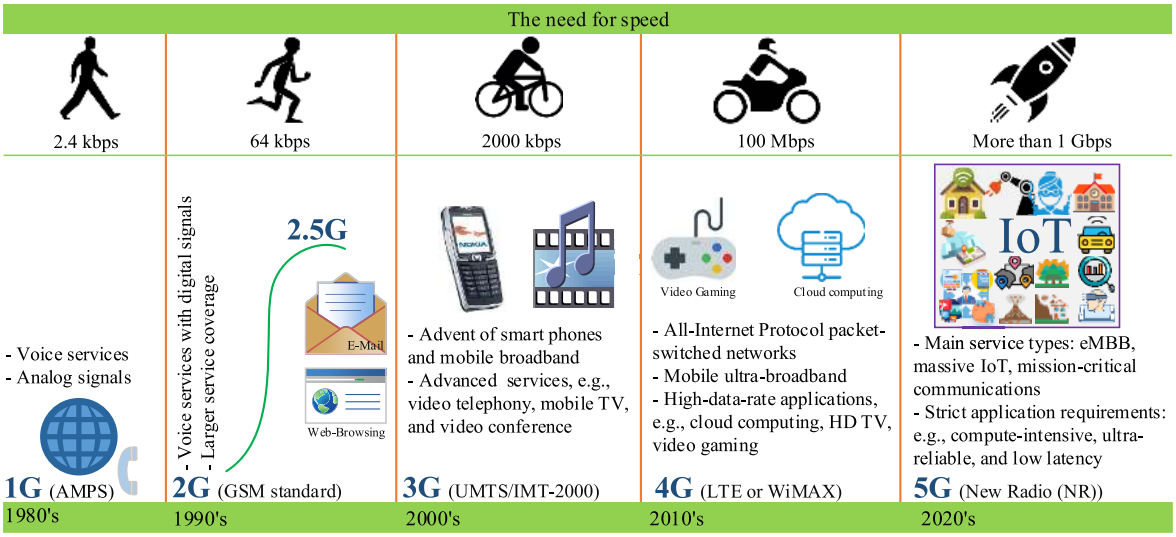

FIGURE 1. Evolution of wireless communication.

capabilities and data storage away from mobile devices and into the cloud [4]. However, MCC suffers from considerable disadvantages, e.g., low scalability, high latency, privacy and security issues, and extreme burden on limited bandwidth. As the very first edge computing concept, Cloudlet, proposed by Satyanarayanan *et al.* in 2009 [5], refers to a trusted and resource-rich computer or a cluster of computers that are located in a strategic location at the network edge and well connected to the Internet. The main purpose of cloudlet is to extend cloud computing to the network edge and support resource-constrained mobile users in running resource-intensive and interactive applications. The WiFi connection between users and cloudlets can be a serious drawback. In particular, users are unable to access cloudlets in the long distance and use both WiFi and cellular connection simultaneously [6], i.e., users have to switch between the mobile network and WiFi when they use cloudlet services. Fog computing, a term put forward by Cisco in 2012, refers to the extension of cloud computing from the core to the network edge, thus it reduces the amount of data needed to transfer to the central cloud [7]. Fog computing plays an important role in many use cases and applications [8], e.g., smart cities, connected vehicles, smart grid, wireless sensor and actuator networks, smart buildings, and decentralized smart building control. However, a fog node cannot act as a self-managed cloud data center (DC) and needs the support of the cloud. The cloudlet and fog computing are similar in that cloudlets and fog nodes are not integrated into the mobile network architecture, thus fog nodes and cloudlets are commonly deployed and owned by private enterprises and it is not easy to provide mobile users with the quality of service (QoS) and quality of experience (QoE) guarantees [9], [10].

In late 2014, the European Telecommunications Standards Institute (ETSI) Mobile Edge Computing Industry Specification Group (MEC ISG) initiated the MEC concept. As a complement of the C-RAN architecture, *MEC aims to unite the telecommunication and IT cloud services to provide the cloud-computing capabilities within radio access networks in the close vicinity of mobile users* [27]. Therefore, MEC enables a wide variety of applications, e.g., driverless vehicles, VR/AR, robotics, and immerse media. In order to reap additional benefits of MEC with heterogeneous access technologies, e.g., 4G, 5G, WiFi, and fixed connection, ETSI ISG officially changed the name of mobile edge computing to mean *multi-access edge computing* in 2017 [28]. After this scope expansion, MEC servers can be deployed by the network operators at various locations within RAN and/or collocated with different elements of the network edge, such as BSs (aka eNB in 4G and gNB in 5G), optical network units, radio network controller sites, and WiFi access points. This transformation pushes intelligence towards the edge so that not only communication functionalities but also computation, caching, and control services can be better facilitated. From this point, the correct name for MEC is multi-access edge computing and this paper uses that name.

Over the last few years, there have been a large number of studies focusing on either technical aspects of MEC architectures or reviews of attributes and application use cases of MEC. Many also consider the importance of MEC in 5G enabling technologies and applications and cover certain research aspects discussed in our article, for example, [1], [9], [11]–[19], [21]–[23]. The previous surveys are summarized as follows. The surveys in [17]–[19] presented a general overview of MEC on definitions, architectures, advantages, deployment scenarios and testbeds, and security and privacy issues. The survey in [9], [11] reviewed several edge computing concepts and focused on computation offloading. The authors in [1] reviewed joint communication and computation resource management in MEC systems. In [12], the authors described four fundamental enabling technologies for MEC including virtual machines and containers, network

TABLE 1. Summary of existing surveys on multi-access edge computing.

Theme	Reference	Major Contribution
Architecture and computation offloading	[9], [11]	- Review of potential MEC architectures and computation offloading.
	[12]	- Introduction of MEC and its key enablers: NFV, SDN, and VM. - Analysis of MEC reference architecture and orchestration deployment scenarios.
Resource allocation	[1]	- Survey of the basic MEC models from the communication perspective. - Review of joint communication and computation resource allocation in MEC systems.
	[13], [14]	- Review of convergence and integration of communication, computation, and caching.
Mathematical frameworks	[15]	- Survey of computation offloading decisions using multi-objective optimization.
	[16]	- Fundamentals of game theory models and MEC. - Review of game theoretical contributions to wireless networks and MEC systems.
General concepts and research directions	[17]	- Fundamentals of MEC, use cases, infrastructure, and security & privacy issues.
	[18]–[20]	- MEC concepts, applications, architectures, and open research challenges.
	[21]–[23]	- Review of how to exploit MEC and other edge computing paradigms for IoT applications.
	[24]–[26]	- Review and analyses of security and resilience of edge computing technologies.
MEC with 5G Technologies	Our Survey	- Survey on integration of MEC with 5G technologies: NOMA, WPT and EH, UAV, IoT, and H-CRAN. - Applications of ML to MEC: 4C optimization, security and privacy, big data analytics, and mobile crowdsensing.

functions virtualization (NFV), software-defined networking (SDN), and network slicing. Moreover, the authors provided analyses of the MEC service orchestration, MEC service mobility, and joint optimization of virtual network functions and MEC services. Several works in [21]–[23] revealed the role of MEC for IoT applications and realization. Recent studies in [13], [14] focused on reviewing the integration of communication, caching, and computation. Mathematical frameworks for optimization of MEC systems were reported in [15], [16]. In particular, the authors in [15] conducted a survey on the computation offloading decisions when multiple challenges, e.g., heterogeneous resources, large amounts of computation and communication, intermittent connectivity and network capacity, are considered (i.e., multi-objective optimization). The authors in [16] reviewed research works that applied theoretical games in addressing problems and challenges of MEC systems. The tutorial in [20] presented three main edge computing concepts: MEC, cloudlet, and fog computing, from the viewpoints of standardization, principles, architectures, and application. In Table 1, we provide a summary of the recently published surveys and reviews on MEC.

Previous surveys addressed important problems in MEC systems, while they have several limitations. These surveys are limited to specific aspects and potential use cases of MEC, for instance, MEC overview [17], [19], architecture and computation offloading [9], resource allocation [1], and mathematical frameworks [15], [16]. Indeed, these articles provide only high-level discussions of the problems and challenges of MEC in 5G. To the best of our knowledge, there is no existing survey to provide a discussion of MEC in the context of other 5G technologies. Furthermore, it is necessary to have an updated survey since MEC has gained popularity in years with a fast-growing research trend and ETSI has released a set of phase 2 specifications, but almost all the articles mentioned in the related work were prepared and/or submitted quite long ago. Therefore, this paper sets to provide a comprehensive survey of the state-of-the-arts which are focused on the integration of MEC and the forthcoming technologies that will be deployed in 5G and beyond network.

In a nutshell, contributions offered by our survey can be summarized as follows:

- We conduct an overview of MEC including fundamentals of MEC (e.g., characteristics, challenges, and market drivers), and MEC integration in the 5G network with potential use cases and applications.
- We discuss the role of MEC in the 5G network architecture and undertake a holistic review of related literature published in the last few years for the integration of MEC with the forthcoming 5G and beyond technologies and scenarios including NOMA, WPT and EH, UAV, IoT, and heterogeneous C-RAN, and ML.
- We provide a concise summary of lessons learned from the state-of-the-art research works and describe potential future directions.

The remaining of this paper is organized as follows. Sections II provides the fundamentals of MEC, including its benefits, integrated architecture in the 5G scenario, and key use cases. The major part of this work is a review of MEC in the context of NOMA (Section III), WPT and EH (Section IV), UAV communications (Section V), IoT (Section VI), heterogeneous C-RAN (Section VII), and machine learning (Section VIII). For each section, we first outline background, then provide motivations for the integration, and finally outline learned lessons and potential directions. The paper is concluded in Section X. For the sake of clarity, Fig. 2 shows the organization of this paper.

II. OVERVIEW OF MEC RESEARCHES

We present fundamentals of MEC by listing the main features and discussing design challenges of MEC and the benefits offered by MEC. We also show the interactions between MEC in the forthcoming 5G technologies and further illustrate MEC use cases with representative examples.

A. FUNDAMENTALS OF MEC

The key idea of MEC is "providing an IT service environment and cloud-computing capabilities at the edge of the mobile network, within the RAN and in close proximity to mobile

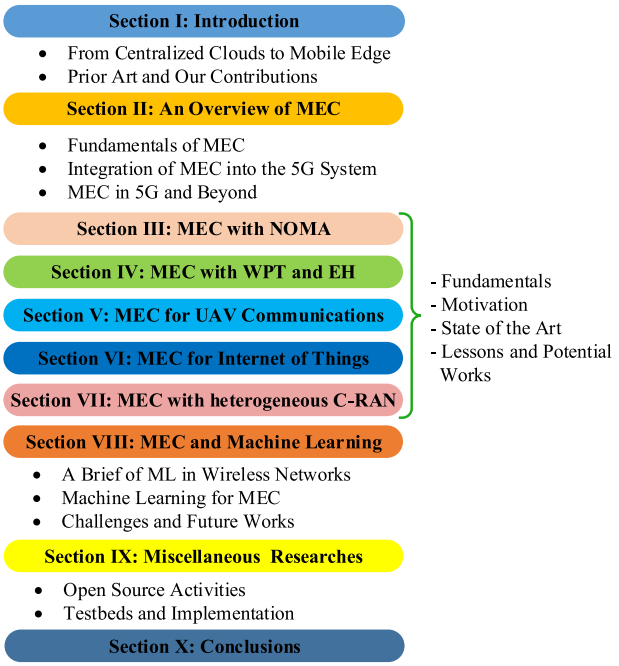

FIGURE 2. Diagrammatic view of the organization.

subscribers'' [29]. The demand for MEC has been driven by many factors, such as the increasing pervasiveness of smart and IoT devices, rapid increase in the data volume and velocity, the increasing need for the rapid development of new high-bandwidth and low-latency applications, introduction of new wireless technologies, and increasing requirement of QoE and QoS. Among those factors, low-latency computing is considered as the primary driven factor for the development of MEC. The demand for low-latency computing is increasing rapidly as low latency is a fundamental metric for network performance and is required by many emerging applications (e.g., VR, interactive gaming, and mission-critical controls). The development of MEC is further fortified by great opportunities for business transformation. On the one hand, mobile network operators (MNOs) need to shorten the time-to-market of new applications and services to maximize the overall revenue. On the other hand, the success and widespread deployment of MEC are guaranteed only when there is the participation of multiple stakeholders (e.g., mobile operators, service providers, vendors, and users) as well as their collaboration. As suggested in [30], the key growth drivers in the MEC market can be classified into four major categories: technical integration, potential use cases, business transformation, and industry collaboration (see Fig. 3). In the foreseeable future, MEC will open up new markets for different industries and sectors by enabling a wide variety of use cases, e.g., IoT, Industry 4.0, Vehicle-to-everything (V2X) communication, smart city, and Tactile Internet. A complete picture of MEC, including challenges, characteristics, use cases and applications, and market drivers, is pictorially illustrated in Fig. 3.

According to the ETSI white paper [27], MEC can be characterized by some features, namely on-premises, proximity, lower latency, location awareness, and network context information. These features can be shortly explained as follows:

- *On-premises*: MEC can operate in standalone environments (i.e., MEC can run isolated from the rest of the network) and has access to local resources.
- *Proximity*: MEC servers are usually positioned in the close vicinity of mobile users, thus MEC can capture information from mobile users for further purposes such as data analytics and big data processing.
- *Lower latency*: although an MEC server has a finite computation power, it is usually sufficient to process emerging compute-intensive applications in real time. MEC has the potential of shortening the communication and propagation latency, which makes MEC a promising enabler for latency-critical 5G applications. MEC also opens up the opportunities to alleviate the burdens on the fronthaul and backhaul links and to accelerate the content and service responsiveness by appropriately caching popular and locally-relevant contents at the network edge.
- *Location awareness*: Due to the close proximity, MEC can utilize signaling information received from end users to estimate their precise locations. This becomes particularly important for MEC location-based services.
- *Network contextual information*: characterized by proximity, MEC can utilize the knowledge of real-time radio network conditions and local contextual information to optimize the network and QoS. For example, real-time and contextual information can be used to improve user experience via personalized services [30].

In spite of several opportunities and potentials, many challenges need to be studied in order to create an edge ecosystem where all network players (i.e., IoT users, service/infrastructure providers, and mobile operators) can benefit from edge services. The discussion can be summarized as follows.

1) *Distributed resource management*: Resource allocation is a key challenge for the success of MEC due to finite resources, growing number of applications, and explosive increase in the mobile traffic [31]. The optimization of resource allocation may be multi-objective that varies in different situations due to diverse nature of applications, heterogeneous MEC servers, various user demands/characteristics, and channel connection qualities. With massive users, the wireless channel would be bottlenecked and the competition among users for scarce computing resources becomes highly intense [32]. Although the centralized approach can achieve competitive performance, it has the weakness of high computational complexity and huge reporting overhead. Therefore, the centralized approach is not suitable for distributed MEC systems [33], [34]. Additionally, there may not exist a dedicated backhaul for information exchange and computation offloading and even if there

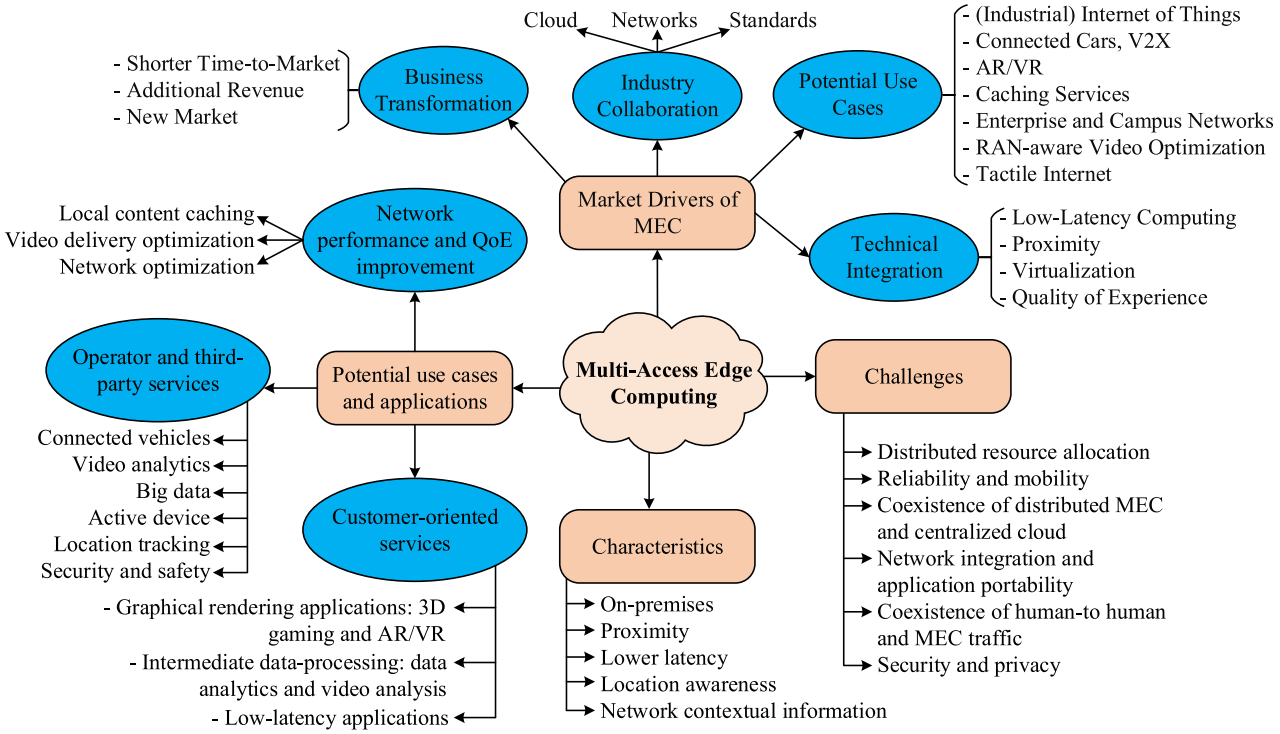

FIGURE 3. An overview of MEC: challenges, characteristics, potential use cases and applications, and market drivers.

is, the wireless backhaul could be congested due to the high burden of huge data sharing [35]. All of these points call for efficient and distributed MEC resource allocation schemes.

2) *Reliability and mobility*: Densification is a key block for the 5G network and is expected to lead to enormous benefits. However, managing mobility and ensuring reliability are quite challenging in such environments. First, under the coverage of multiple small-scale servers, user mobility can cause frequent handovers, which introduce the service disruption problem and affect the overall network performance [36]. Second, users (e.g. vehicles) may move to new locations during the computation offloading period. In such a case, users may not be able to receive the computational result since they already move out of the service coverage of their serving servers. Therefore, efficient computation offloading models are necessary for the application accomplishment. Third, variations in the number of offloading users result in random uplink interference and time-varying computing resources [37]. Finally, ultra-reliability is an important concept in 5G since it initiated the implementation of industrial automation and smart transportation. For instance, AR-based applications usually require the real-time response and ultra-reliable connection between the server and users. While ultra-high reliability on the order of 99.99999% and extremely low latency of 0.1-1ms round-trip time are the communication requirements in industrial control networks or autonomous mobility systems [38]. These requirements would not be well fulfilled under dynamic channel qualities and intermittent connections. There has been a great deal of efforts in utilizing MEC for providing reliable and low-latency services. We invite the interested readers to read the surveys in [39], [40] for more detail.

3) *Network integration and application portability*: Depending on the underlying technologies, technical and business requirements, MEC servers can be deployed at different places within the RAN. Thus, another critical challenge is the seamless integration of MEC into the underlying network architecture and existing interfaces [27]. The existence of MEC and enabled applications should not affect the standard specifications of the core network and end devices. According to [28], the key component of the MEC integration is the ability of MEC to interact with 5G networks in routing the traffic and receiving relevant control information. Furthermore, the application migration necessitates a so-called application portability requirement. This removes the need for app developers to design multiple versions for different MEC platforms.

4) *Coexistence of distributed MEC and centralized cloud*: Cloud DCs, with abundant computing resources, can process big-data applications in near zero time and support a large number of users. However, distributed MEC is highly desired since the computation at the network edge can not only meet the user requirement but also reduce the end-to-end delay caused by the traffic congestion and transmission delay. By analogy to the HetNet architecture, it is highly beneficial to implement MEC

in a hierarchical manner, i.e., user, edge-computing, and cloud-computing layers. In this way, the MEC vendor also injects computing resources to the small-eNBs so that the advantages of HetNets can be exploited for diversifying radio transmissions and spreading computing demands [41]. We note that distributed MEC may not have enough computing resources to process all computation requests and complete reliance on the cloud poses challenges of providing latency-critical services. Therefore, it is intuitive to distribute big-data/latency-critical computations to distributed MEC servers while transferring compute-intensive and delay-tolerant tasks to the cloud DC [42]. The coexistence of distributed MEC and centralized cloud is an important issue and more research is needed for their interactions.

5) *Coexistence of human-to-human and MEC traffic*: Incorporating both conventional Human-to-Human (H2H) traffic (e.g., voice, data, and video) and MEC traffic in 5G is a challenging task due to massive IoT connections coupled with the diverse QoS requirements and unique characteristics of MEC traffic [43]. For instance, the IoT system comprises of human-type devices (HTDs) and machine-type devices (MTDs) that may run different kinds of applications, e.g., MTD with sensors and smart homes, and HTD with video games. While MTDs have a mixed set of QoS requirements, such as latency, reliability, and energy efficiency, HTDs typically require a high-speed rate with the limited energy budget [44]. Similarly, the MEC system should be designed in a way that the QoS requirements of H2H traffic are satisfied while unique characteristics of M2M traffic (e.g., real-time response and context awareness) are maintained.
6) *Security and privacy*: Although MEC has the capability to improve security and privacy compared with MCC, MEC has its own security and privacy challenges. First, MEC can be collocated with different heterogeneous network elements, thus making the conventional privacy and security mechanisms, which have been already operated in MCC, inapplicable to MEC systems. Second, the task offloading over wireless channels may not be secure since computation tasks can be overheard by malicious eavesdroppers. The transfer of compute-intensive applications' data can be secured by encryption at the user side and decryption at the destination server side. This, however, can increase the propagation delay as well as execution delay, thus reducing the application performance [45]. Physical layer security, blockchain, and federated learning have emerged as effective solutions to secure and protect MEC systems [46], [47]. Finally, sharing the same storage and computation resources among multiple mobile users raises issues of private data leakage and loss.

B. INTEGRATION OF MEC INTO THE 5G SYSTEM

After initializing the MEC concept, the ETSI ISG and many members in the value chain have spent a great deal of efforts for the development of MEC specifications based on industry consensus. At the time of writing this paper, there are 68 members and 35 participants in the ETSI consortium,[1] which are not only mobile operators but also manufacturers, service providers, and universities, e.g., Vodafone, IBM, Intel, NTT Corporation, University CarlosIII de Madrid, etc. Their involvement plays a major role in ensuring an open and interoperable MEC environment, and MEC is beneficial to various stakeholders including MNOs, application developers, over-the-top players, independent software vendors, telecom equipment vendors, IT platform vendors, system integrators, and technology providers. The ETSI ISG has published a set of standards and specifications focusing on, for example, framework and reference architecture [48], MEC in the NFV environment [49], and collocating C-RAN and MEC [50]. The 3GPP started including MEC in the 5G network standardization in the technical specification 3GPP TS 23.501 [51]. Recently, in [28] and based on functional enablers defined in [51, clause 5.13], the 3GPP clarified how to deploy MEC in and seamlessly integrate MEC into 5G, which can be illustrated in Fig. 4. The architecture comprises two parts: the 5G service-based architecture (SBA) on the left and an MEC reference architecture on the right.

The network functions defined in the 5G architecture and their roles can be briefly summarized as follows.

- *Access and Mobility Management Function* (AMF): establishes mobility and access procedures, e.g., connection management, reachability management, mobility event notification, termination of the RAN control plane, and access authentication/authorization.
- *Session Management Function* (SMF): performs functionalities related to session management, e.g., session establishment, termination of interfaces towards policy control functions, and downlink data notification.
- *Network Slice Selection Function* (NSSF): executes the allocation of slicing resources and AMF set to serve users.
- *Network Repository Function* (NRF): supports the discovery of network functions and their supported services.
- *Unified Data Management* (UDM): handles user subscription and identification services.
- *Policy Control Function* (PCF): unifies the network policies and provides policy rules to control plane functions.
- *Network Exposure Function* (NEF): acts as a service-aware border gateway for providing secure communication with the services supported by the network functions.
- *Authentication Server Function* (AUSF): performs authentication procedures.
- *User Plane Function* (UPF): provides functionalities to facilitate user plane operations, e.g., packet routing and forwarding, data buffering, and allocation of IP address.

[1]The complete list of MEC members and participants is available at https://portal.etsi.org/TB-SiteMap/MEC/List-of-Members.

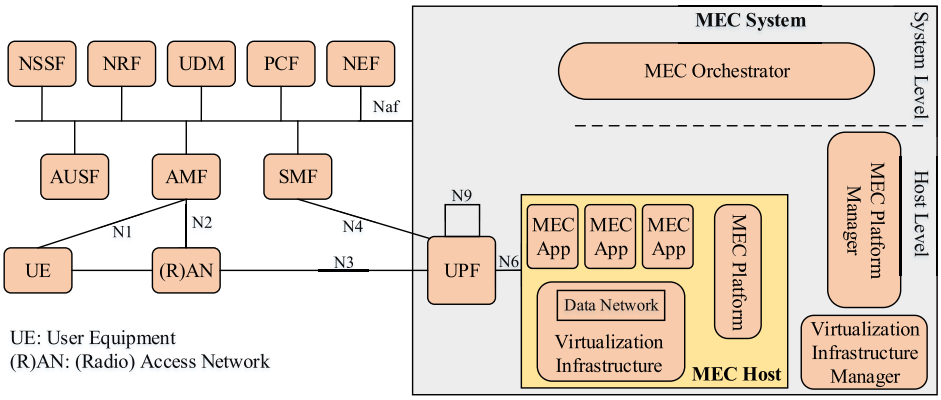

FIGURE 4. MEC integrated architecture in 5G [28].

More details of the SBA and the 5G network functions can be found in 3GPP TS 23.501 [51].

The MEC reference architecture is composed of the MEC system level and host level [48]. The MEC orchestrator (MECO) is the core component of the MEC system level, which maintains information on deployed MEC hosts (i.e., servers), available resources, MEC services, and topology of the entire MEC system. The MECO is also responsible for selecting of MEC hosts for application instantiation, on-boarding of application packages, triggering application relocation, and triggering application instantiation and termination. The host level management consists of the MEC platform manager and the virtualization infrastructure manager (VIM). The MEC platform manager carries out the duties on managing the life cycle of applications, providing element management functions, and controlling the application rules and requirements. The MEC platform manager also processes fault reports and performance measurements received from the VIM. Meanwhile, the VIM is in charge of allocating virtualized resources, preparing the virtualization infrastructure to run software images, provisioning MEC applications, and monitoring application faults and performance. Finally, the MEC host comprises an MEC platform and a virtualization infrastructure. The former includes the set of functionalities needed to run MEC applications on a particular virtualization infrastructure and the latter includes the data plane functionalities of executing the traffic rules received by the MEC platform and steering the traffic among applications and networks.

New functional enablers were defined in [51] to integrate MEC into the 5G SBA, which can be explained as follows.

- *User Plane Reselection and Selection*: The 5G core network supports the UPF (re)selection for selective traffic routing to the data network. Parameters used for the UPF selection mechanism is dependent on the UPF deployment scenario and MEC service operator configuration.
- *Local Routing and Traffic Steering*: The UPF enables various traffic routing schemes for MEC applications in the 5G network. Moreover, application functions (AFs) may affect the UPF (re)selection and make specific traffic routing rules for a particular user.
- *Local Area Data Network* (LADN): The support for LADN is enabled by the flexibility in the UPF location. Then, MEC hosts can be deployed on the N6 interface that is between the UPF and a data network. The user using MEC services may discover LADN availability during the registration procedure based on LADN information received from the AMF.
- *Session and Service Continuity* (SSC): The support for SSC is essential to enable user and application mobility. The 5G architecture allows MEC applications to select one among three SSC modes [51]. Particularly, SSC mode 1 provides the stable network connectivity to the user, SSC mode 2 may release the current connectivity to the user before making a new one, and SSC mode 3 ensures service continuity for the user by changing the new user plane before disconnecting the existing one.
- *Network Capability Exposure*: The 5G architecture allows both direct access to network functions for the authorized MEC and indirect access via the NEF. Examples of exposed capabilities are exposure of user events, exposure of user behavior provisioning to external functions, and exposure of analytics to external parties.
- *QoS and Charging*: The PCF in the 5G SBA defines QoS and charging rules for the user traffic routed to the LADN.

C. MEC IN 5G AND BEYOND

Many services and applications will be supported in 5G and beyond, which can derive substantial benefits from MEC by being executed at the distributed edge servers. No matter what the service is, MEC use cases can be classified under three main categories, namely *consumer-oriented services*, *operator and third-party services*, and *network performance and QoE improvements* (see Fig. 3 and Fig. 5) [29]. The fact is that the MEC ecosystem should support all these categories to create a myriad of new services and applications at the edge of mobile networks. Generally, the classification is dependent on who could reap the advantages and benefits.

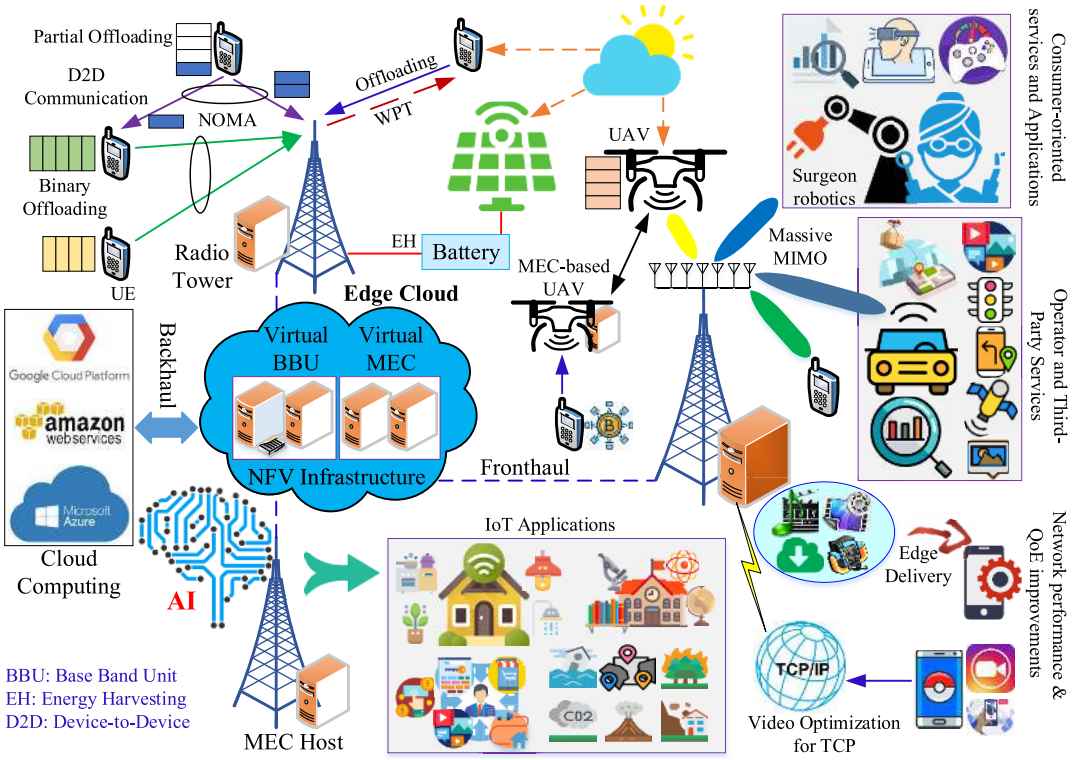

FIGURE 5. Integration of MEC with the forthcoming 5G technologies.

First, the use case "consumer-oriented services" aims to bring direct benefits to users through the capability of running computation-heavy and latency-sensitive applications at the network edge. By means of computation offloading, users can exploit substantial computing resources on the edge server [52]. Applications and services under the first category can include graphical rendering applications (e.g., 3D gaming, AR/VR, assisted reality, and cognitive assistance), intermediate data-processing (e.g., data analytics and video analysis), and low-latency applications (e.g., remote surgery on tactile Internet, AR/VR, video games, and interactive applications), and location-based service recommendation. Under the second category, operators and third parties take advantages of MEC computing and storage facilities to place their own applications and services on the network edge. This is enabled by the "open and interoperable environment" nature of MEC, and is to encourage innovation and development in MEC from multiple parties and overcome obstacles (e.g., deployment difficulties and operational costs) in providing MEC services at the hard-to-reach areas [28]. Applications and services offered by operators and third-party vendors can include V2X applications (e.g., safety, convenience, and driving assistance), big data, active device location tracking, security, safety, data analytics, and indoor precise positioning. Finally, the services under "network performance and QoE improvements" group intend to optimize operations of the network, thus improving the network performance and QoE. Examples of the third category are local content caching at the edge, mobile backhaul optimization, traffic deduplication, video delivery optimization for transmission control protocol (TCP), multi radio access technology (RAT) computation offloading, and network congestion in dense-network environments.

To support the aforementioned applications and services, new architectures and technologies will be introduced in the 5G network. As shown in Fig. 5, the integration of MEC with the forthcoming 5G technologies is necessary to achieve added values in MEC systems. A brief description of MEC in the 5G scenario is given as follows.

1) NOMA, millimeter Wave (mmWave), and massive multiple-input multiple-output (MIMO): As a multiple access technology to meet the demand for massive connectivity, the integration of NOMA into MEC systems is an important research issue, which needs more attention in the years to come. Moreover, the coexistence of MEC with mmWave massive MIMO is necessary to enable massive wireless connectivity with high data rates, low-latency, and large computing capabilities. These schemes are provided in Section III.
2) EH and WPT: Thanks to EH and WPT, the design of self-sufficient and self-sustaining wireless communications (aka green communication) becomes a reality. The combination of EH and/or WPT with MEC in a single system offers great potential to solve fundamental limitations of traditional systems, e.g., limited battery lifetime, unstable grid power supply, and low

computing capability. To understand these issues more clearly, research works on EH and WPT MEC systems are surveyed in Section IV.

3) UAV communications: UAVs can be exploited to enable many potential applications due to their features of flexibility, mobility, maneuverability, and low cost. On the one hand, UAVs can be aerial edge servers to perform heavy computations offloaded from ground users. On the other hand, UAVs can act as aerial users and associate with ground BSs to offload their tasks. The integration of UAV into MEC systems is a promising research topic, which will be summarized in Section V.
4) IoT: IoT devices are quite resource-constrained to run compute-intensive tasks due to their limited computing capability and battery capacity. MEC is a powerful solution to solve these limitations. Inversely, IoT expands MEC services into more scenarios and objects like sensors, actuators, and mechanized agriculture. We will provide a survey on recent MEC-enabled IoT applications in Section VI.
5) H-CRAN: With the realization of NFV, the collocation of MEC and heterogeneous C-RAN (H-CRAN) is expected to bring potential benefits. In such a scenario, the edge host (i.e., MEC server) in MEC and the BBU pool in H-CRAN can be collocated with each other to share the same virtualization infrastructure. We will study the integration of H-CRAN and MEC further in Section VII.
6) Machine Learning: The massive amount of mobile data, together with recent breakthroughs in ML and the non-convexity nature of resource allocation in a complex network, inspires many creative solutions for wireless communications and networking problems. ML plays a central role in the design of MEC mechanisms as we will elaborate in Section VIII.
7) VM, SDN, NFV, and Network Slicing: MEC systems primarily rely on four enablers: VM, SDN, NFV, and network slicing. VM virtualization enables transient customization of MEC infrastructure, while SDN, NFV, and network slicing provide greater flexibility and agility of multi-tenant MEC ecosystems. For more information on these issues, we refer the interested readers to [12], [48], [53] and references therein.

Besides aforementioned ones, integration of MEC with some other technologies such as blockchain and cognitive radio is expected to offer various benefits. A survey of blockchain in the context of cloud of things can be found in [47] and applications of blockchain for 5G and beyond networks were reviewed in [54]. Cognitive radio is a vital technology for efficient spectrum scarcity, which allows to meet the high spectrum demand of many new applications and services, and proliferation of massive IoT connections. studies on cognitive radio MEC may have additional challenges, e.g., how to sense available spectrum bands, how to protect primary users, and how to allocate resources to improve the network performance. Despite these issues, integrating cognitive radio with MEC is expected to offer a number of advantages. For example, the work in [55] showed that cognitive radio edge computing can well support low-latency and compute-intensive industrial applications. In the following sections, we will review a number of studies related to these technologies in the context of MEC systems.

In summary, we focus on the following aspects in MEC systems: radio access (NOMA, mmWave, and massive MIMO), network architectures and scenarios (H-CRAN and UAV), applications (IoT, V2X, and UAV), power supply (EH and WPT), and performance improvement (ML). In the following sections, these researches are discussed in more details.

TABLE 2. Comparison between OMA and NOMA.

	Advantages	Disadvantages
OMA	- Simpler receiver detection	- Lower spectral efficiency - Limited number of users - Unfairness for users
NOMA	- Higher spectral efficiency - Higher connection density - Enhanced user fairness - Lower latency - Supporting diverse QoS	- Increased complexity of receivers. - Higher sensitivity to channel uncertainty.

III. MEC WITH NON-ORTHOGONAL MULTIPLE ACCESS

A. FUNDAMENTALS OF NOMA

Non-orthogonal multiple access (NOMA) has been considered as an essential principle for the design of radio access techniques in the emerging 5G network [56]. The key idea of NOMA is the use of the superposition coding technique at the BS side and interference cancellation techniques (e.g. multiple user detection and successive interference cancellation) at the user side. Compared to the conventional orthogonal multiple access (OMA), NOMA can enable multiple users to share the same time-frequency resource to achieve higher spectral efficiency. There are two main NOMA categories: power-domain NOMA and code-domain NOMA. Power-domain NOMA exploits the channel gain differences between users and multiplexes users in the power-domain while code-domain NOMA uses user-specific sequences for sharing the entire available radio resource [57]. Typical examples of code-domain based access strategies are low-density spreading code division multiple access (CDMA), low-density spreading-based orthogonal frequency-division multiple access (OFDMA), sparse code multiple access (SCMA), and multi-user shared access (MUSA). NOMA has the potential to accommodate more users than the number of available subcarriers, which leads to various potentials, including massive connectivity, lower latency, higher spectral efficiency, and relaxed channel feedback [58]. The comparison between OMA and NOMA is summarized in Table 2 [57].

Although NOMA is able to support a large number of users simultaneously and surpasses OMA in several aspects, various challenging problems associated with NOMA must be addressed before this technology can be employed in real networks. Islam *et al.* in [59] and Dai *et al.* in [60] provided some research directives for NOMA in their survey: dynamic user pairing, the impact of transmission distortion, channel and interference estimation, etc. NOMA can be flexibly combined with many existing wireless technologies and emerging ones including MIMO, massive MIMO, mmWave communications, cognitive and cooperative communications, visible light communications, physical layer security, energy harvesting, wireless caching, and so on [61]. To gain a deeper understanding of the benefits and opportunities that NOMA offers as well as its challenges and application scenarios, the interested readers are recommended to refer to NOMA research works, such as, [57], [59], [60], [62]–[64].

B. MOTIVATION TO COMBINE NOMA AND MEC

Both NOMA and MEC are considered as the key enabling technologies in 5G due to their enormous potentials and wide-range applications. There are many benefits of MEC and NOMA, including supporting massive users, reducing the transmission latency and the energy consumption of end users and providing high performance for more complex network scenarios, i.e. mmWave massive MIMO.

- The combination of MEC and NOMA can significantly improve the user satisfaction and network performance through the provision of golden opportunities. While NOMA offers several advantages at improving the spectral efficiency and cell-edge throughput, relaxing the channel feedback requirement, and reducing the transmission latency, MEC brings considerable benefits to not only users, but also operators and third-parties, and enables to improve overall network performance as well. It is expected that 5G will support a massive increase in device connections, high-speed transmissions of 1–10 Gbps, and greatly reduce latency and high reliability.
- The combination of MEC and NOMA can reinforce the services and applications that are supported by the 5G network. On the one hand, NOMA is expected to vastly increase the number of users in various scenarios where rank deficiency can occur [60]. On the other hand, edge computing in MEC indicates that computing resources are provided for end users in close proximity and at the edge of RANs. Therefore, MEC is capable of widely distributing computing resources from centralized cloud to the network edge and immediately serving a large number of users, hence MEC has the potential to support massive connectivity and distributed computation.
- The combination of MEC and NOMA can provide low-latency transmission. Because the 5G network will not completely rely on a single technology, we must optimize the network from multiple perspectives, e.g., air interface, network architecture, and enabling technologies. To cope with demands for lower latency, MEC and NOMA are two promising solutions. MEC moves the cloud services and functions to the network edge, where data is mostly generated and handled. Hence, MEC empowers the services running at the edge to better meet the lower latency requirements of end users compared to the cloud computing. In a similar sense, flexible scheduling and grant-free access in NOMA enables lower transmission latency for users in the 5G network.
- NOMA and MEC can be flexibly combined with many existing wireless technologies, e.g., MIMO, massive MIMO, mmWave communications, etc., to further increase connectivity, spectral efficiency, energy efficiency and computing capability. For example, massive MIMO can drastically increase the spectral efficiency of wireless networks via excessive spatial multiplexing, thus Massive MIMO-NOMA can support massive connectivity and high spectral efficiency. To support gigabits-per-second data rates, mmWave bands can be used for wireless communications. The large path-losses caused by mmWave can be compensated by high gains, which can be obtained by massive MIMO. As a result, NOMA MEC can be deployed jointly with mmWave massive MIMO to enable multiple mobile devices to offload tasks simultaneously with high uploading/downloading data rates.

Promoted by a variety of opportunities and advantages offered by MEC and NOMA, both academic and industrial communities have conducted extensive researches to design the 5G network with MEC and NOMA [65]–[67]. However, the state-of-the-art MEC researches still have not explored the full potential benefits of NOMA in the context of MEC. NOMA and MEC are both conceived as the bids to fill the gap between IoT devices and IoT applications and services. On the one hand, MEC empowers resource-constrained IoT devices with significant additional computational capabilities through computation offloading, thus bringing new applications and services to IoT devices. Similarly, with IoT, the scope of MEC services and applications is applicable to not only mobile phones, but also a wide range of smart objects ranging from sensors and actuators to smart vehicles. On the other hand, NOMA is capable of substantially improving on system capability since it enables multiple users to transmit using a dedicated orthogonal channel resource. Furthermore, motivated by the benefits of NOMA over OMA, it appears utterly reasonable that one can exploit NOMA to further improve the use of MEC in IoT networks, as compared to the performance of conventional OMA-based MEC approaches.

Apparently, NOMA can be exploited to increase the efficiency and performance of multi-user MEC systems. In the following, we present an overview of research works that have explored the combination of NOMA and MEC and then discuss fundamental challenges and open directions.

TABLE 3. Summary of existing works on NOMA MEC.

Topic	Designed frameworks	References	Contributions
Architecture of NOMA MEC	Uplink NOMA MEC and downlink NOMA MEC	[61], [66]	NOMA MEC outperforms OMA MEC on lower latency and lower energy consumption than the traditional OMA scheme.
Energy consumption minimization	*Partial offloading*	[68], [69]	Resource allocation schemes are proposed to minimize the energy consumption for a uplink and downlink NOMA enabled MEC system.
	Binary offloading	[70]–[72]	A hybrid NOMA MEC system is proposed to minimize energy consumption.
Task delay minimization	*Partial offloading*	[72], [74], [75]	Optimal task and power allocation is proposed to minimize the task delay in NOMA MEC system.
	Binary offloading	[76]	Dinkelbach and Newton's methods are compared to minimize task delay for the hybrid NOMA MEC system.

C. STATE OF THE ART

While the use cases of NOMA or MEC have been widely studied in the literature, there have been some studies on MEC-NOMA scenarios. The advantages of NOMA and MEC have motivated several studies supporting the application of NOMA to MEC [66], [68], [70], [71], [76], [77]. When NOMA uplink transmission is applied to the MEC system, multiple users can offload their tasks to the MEC server simultaneously via the same frequency band. By applying the successive interference cancellation (SIC) technology at the MEC server, the MEC server can remove the interference from the user whose data has been decoded before on the same frequency band. When NOMA downlink transmission is applied to the MEC system, one user can utilize NOMA to offload multiple tasks to multiple MEC servers simultaneously via the same frequency band. The performance comparison of NOMA-MEC and OMA-MEC systems was conducted in [66], which reveals that the NOMA-MEC system can achieve superior performance in reducing latency and energy consumption.

Most existing research works focus on resource allocation i.e., computation resource and communication resource. Specifically, in [68], partial offloading assignment (i.e., each user can partition the computation task into two parts for local computing and offloading) and power allocation were investigated to minimize the weighted sum of the energy consumption for a multi-user NOMA-MEC system. In this work, an efficient algorithm for user's task partitioning, local computing CPU frequency and transmit power allocation was proposed to achieve the minimum energy consumption for multi-user NOMA-MEC networks. Unlike OMA-MEC and pure NOMA-MEC systems (i.e., both the users offload all of their tasks at the same time) proposed in [66], [68], a hybrid NOMA strategy (i.e., a user can first offload parts of its task within time slot allocated to other user and then offload the remaining of its task during a time slot solely occupied by itself) was proposed in [70], in which power allocation and time allocation were optimized to minimize the energy consumption for an MEC-enabled NOMA system. Subsequently, the delay minimization was investigated for the hybrid NOMA-MEC system [76]. The work in [78] defined the objective function balancing the tradeoff between energy consumption and completion delay in hybrid NOMA-MEC systems. A joint power allocation and user clustering problem was investigated in [78], from which power allocation is provided in closed form and user clustering is solved by using a matching theory approach.

Different from partial offloading tasks, the authors in [71] considered that the offloading tasks are independent and non-separate. Then the communication resource (i.e., frequency bands and transmit powers) and the computing resource (i.e., computing resource blocks) were jointly optimized to minimize the energy consumption for the NOMA-MEC system [71], in which an efficient heuristic algorithm of user clustering and frequency and resource block allocation was proposed to address the energy consumption minimization problem per NOMA cluster. In [77], the computing offloading scheme was investigated in the NOMA MEC system where a distributed algorithm based on game theory was proposed to improve the system performance. Moreover, the delay minimization problem was investigated in [74], [75]. In [75] an efficient algorithm of the offloading workload, offloading and downloading duration optimization was proposed to minimize the overall delay of the computation tasks. Another study on NOMA-MEC to minimize the average overall delay can be found in [79], where a joint offloading decision, subchannel assignment, power control, and computing resource allocation problem is investigated and users with differentiated uploading delays are taken into consideration. The energy efficient power allocation, time allocation and task assignment were proposed to minimize the energy consumption for MEC networks [69], [72]. Besides the computational resource, SIC decoding order was optimized to reduce the task delay for NOMA enabled narrowband Internet of Things (NB-IoT) systems [73]. The summary of the existing works on NOMA MEC is provided in Table 3. The work in [80] considered minimizing the total completion time of secondary users in a cognitive NOMA-MEC system. The latency minimization problem is optimized under constraints that the interference at primary users is below an interference threshold and the total computing resources assigned to users cannot exceed the maximum computing capability of the MEC server. Another interesting work on cognitive MEC was studied in [81], where downlink NOMA is applied for the transmissions between secondary users to the MEC servers. A joint offloading decision, local computing capability control, and NOMA power allocation was considered to minimize the system delay. Similar to [80],

the decomposition technique is applied to solve the problem in [81] in an iterative manner.

D. LEARNED LESSONS AND POTENTIAL WORKS

Because of limited researches advocated to coexisting MEC-NOMA scenarios there are many key open problems that must be investigated. The potential works of NOMA and MEC can be viewed from the following four aspects: 1) joint resource optimization; 2) secure communications; 3) cooperative NOMA MEC; 4) coexistence of NOMA MEC and mmWave massive MIMO; 5) low-complexity and online NOMA MEC schemes.

1) JOINT RESOURCE OPTIMIZATION

Resource allocation plays an important role to improve the performance of the wireless network. Thus in MEC-NOMA networks, the communication and computing resource can be jointly optimized to enhance the system performance, i.e., sum rate and energy efficiency. In other words, the scheduler may need to decide the computation load that the user can offload to the MEC server, and the remaining can be computed locally to minimize the latency. Moreover, computation capacity (i.e. processing speed of MEC servers or mobile devices) and communication resource (i.e. transmit power) are also important factors to reduce the computation latency. Joint optimization of these factors presents an open and challenging research problem. When NOMA uplink transmission is applied to the MEC system, multiple users can offload their tasks to the MEC server at the same time. Therefore, the total latency experienced by the multiple users can be investigated. By controlling the offloaded computation load and transmit power of each user, the optimal and suboptimal strategies can be developed to minimize the total latency of the system by considering the total energy consumption. The proposed solution can be extended to the NOMA downlink MEC system. Moreover, user grouping or user association can be another trend in resource optimization of MEC-NOMA systems, where game-theoretic approaches and metaheuristic optimizers [41], [82], [83] can be exploited to group users into different groups which use different subchannels to offload their tasks. Besides, the performance of the SIC technology is sensitive to the availability of channel state information (CSI). Thus another possible direction to address this issue is to rely on partial CSI. The application of partial CSI in downlink NOMA system was investigated in [84], [85], which can be investigated in resource allocation for MEC-NOMA systems.

2) SECURE COMMUNICATION

Security and privacy-preserving communication attract lots of research attention, especially when NOMA is applied to the MEC system. For example, two users are offloading tasks to an MEC server at the same time by using the NOMA principle. When SIC is performed, one user can decode the other user's message. During this period, an eavesdropper or an attacker may attempt to decode the mobile user's message. To address the scenario with external eavesdroppers, the physical layer security can be utilized to cope with this challenge for the NOMA-MEC system [86], [87]. The combination of PLS and NOMA-MEC is a promising research topic.

3) COOPERATIVE NOMA-MEC SYSTEM

To improve the connectivity of the NOMA-MEC network, the cooperative MEC can be adopted to enable computation offloading to the main MEC server. In this scenario, the mobile device transmits the superimposed signals to the primary MEC server and the helper MEC server, which acts as a relay helping MEC server [88], [89]. Considering the local computing capacity of the mobile users and energy consumption constraint, the task assignment and transmit power allocation can be optimized to improve the performance of NOMA MEC system.

4) COEXISTENCE OF NOMA MEC AND mmWave MASSIVE MIMO

Massive MIMO-NOMA is another scenario to support massive connectivity and high spectral efficiency [61]. To further improve the transmission data rate, the mmWave bands (30 GHz to 300 GHz) have been proposed to provide gigabit-per-second data rates. Therefore, the integration of NOMA MEC into mmWave MIMO based wireless networks can improve the computing capability, spectrum efficiency and reduce the task delay, where multiple mobile devices can transmit tasks simultaneously via the mmWave bands. Inspired by the challenges of the traditional MIMO transmission scheme, an efficient approach of joint beamforming design and communication and computing resource allocation will be a major challenge to tackle. Moreover, the user grouping needs to be well investigated to further enhance the system performance. Very recently, there have been some research studies pertaining to MEC with massive MIMO and mmWave like [90], [91]. For example, the work in [90] considered a cell-free massive MIMO system with a cloud DC and a number of access points (APs), and further derived the successful edge computing probability (SECP). This work showed an interesting observation that for a given SECP, the system becomes more energy-efficient with higher AP density and less antennas at each AP, rather than with smaller AP density and larger number of antennas.

5) LOW-COMPLEXITY AND ADAPTIVE NOMA-MEC

It is widely known that the NOMA computational complexity (e.g., user clustering, signal decoding, and CSI acquisition) is the main barrier against the NOMA practicality [92]. Most of the existing studies consider using convex optimization and game theory approaches for solving resource management problems in NOMA MEC systems; however, they typically have high complexity. Therefore, low-complexity NOMA-MEC techniques are of importance for the NOMA practicality. Moreover, to cope with the dynamics of MEC wireless environments and computation tasks, designing

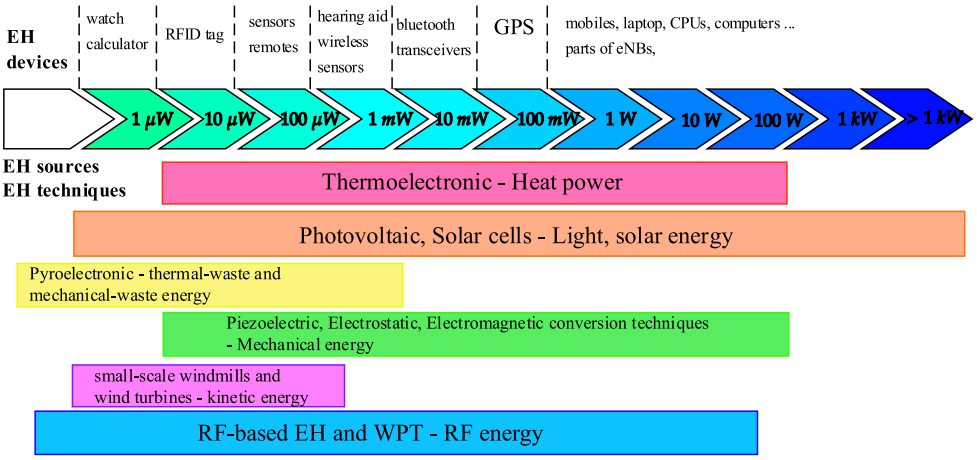

FIGURE 6. EH technologies with generated intermittent power and power consumption for various devices, adapted from [94]–[96].

algorithms that are well adaptive to the system dynamics and for online implementation is necessary. Besides, investigation of MEC systems employing other multiple access schemes, e.g., rate splitting multiple access (RSMA), is a promising direction. It is necessary since RSMA has shown considerable benefits over NOMA and OMA [93].

IV. MEC WITH ENERGY HARVESTING AND WIRELESS POWER TRANSFER

A. FUNDAMENTAL OF EH AND WPT

The current industrial landscape is becoming increasingly aware of the need to optimize energy use and management for all domains, including telecommunications. Among others, EH, also known as energy scavenging or power harvesting, is a promising technique for 5G systems since EH is an alternative solution to traditional energy supply sources [97]. The basic concept of EH is to capture various available energy from different sources to power the energy-constrained devices for prolonging their lifetime [98]. Together with the traditional energy grid, EH can help to fulfill the energy requirements of the different tiers of 5G networks including the sensors in IoTs, the mobile devices, the eNBs in HetNets, assisting relays in D2D systems, and the computing servers [94]. Additionally, the recent development in advanced materials and hardware designs helps realize the EH circuits for small portable consumer electronic devices which accelerate the adoption of EH for the IoTs [95].

EH is simple in concept, but more complex in implementation which strongly depends on the type of EH power sources. The harvestable energy can be scavenged from natural or human-made sources which are controllable or uncontrollable [99]. As illustrated in Fig. 6, various EH techniques (e.g., thermoelectrics, photovoltaic conversion, pyroelectrics, piezoelectrics, electrostatics, and radio frequency (RF)-based EH and WPT) can be employed to leverage the corresponding sources of energy [94], [96]. Besides, different devices may have different energy harvesting capabilities, for example, a wearable device and a smart boot may harvest a power value of 1 mW and 100 mW, respectively [94]. Compared with the traditional natural energy sources, RF signals are less affected by weather or other external environmental conditions. As a result, these signals can be efficiently controlled and designed, so RF-based EH has great potential to provide stable energy to low-power energy-constrained networks including wireless sensor networks (WSNs), IoTs, and extremely remote area communication (eRAC) use cases in 5G networks [100]. Specifically, RF-EH can be employed in indoor, hostile, and harsh environments, e.g., sensors inside a building or human body, toxic environment, and so on [100]. RF-EH can scavenge wireless energy from 1) ambient sources (e.g., WiFi, AM, and FM) which can be predictable or unpredictable or 2) dedicated sources which are deployed to provide an energy supply. RF-EH exploiting the ambient sources normally requires an intelligent process to monitor the communication frequency bands and time periods for harvesting opportunities. RF-EH with proper management of dedicated energy sources between the emitters and the harvesters can be considered as WPT.

WPT was first proposed by Nikola Tesla in 1899 [99] and continuously studied by both industry and academic communities. Existing WPT technologies can be categorized into three classes: inductive coupling, magnetic resonant coupling, and RF-based WPT. The first two technologies rely on near-field electromagnetic (EM) waves, which cannot support mobility for the energy-limited wireless communication devices due to the limited wireless charging distances (a few meters) and the required alignment of the EM field with the EH circuits [101]. In contrast, RF-based WPT exploits the far-field properties of EM waves over long distances (hundreds of meters). In the RF-based WPT system, embedding the modulated information (e.g., phase-embedded information) into the RF-based WPT signals forms the concept of simultaneous wireless communication and power transfer (SWIPT) which was proposed and studied in [102] from an information theoretical perspective.

Recently, [94], [97] have demonstrated that integrating EH/WPT into typical 5G systems including IoTs, device-to-device (D2D) networks, HetNets, and cognitive radio networks (CRNs), can bring benefits in improving energy and spectral efficiency. However, integration of EH/WPT in 5G architecture also raises some technical challenges as follows.

- How to cover the unstable and intermittent characteristics of the ambient resources, i.e, power, spectrum, periods, is a challenging problem which should be considered in designing an EH systems.
- How to allocate the network resources to well balance between harvested energy and consumed power is another issue. Towards this end, one must well understand the generating environment and the characteristics of energy source, the power consumption properties of different elements in the system, the coverage area, the communication distance, the data rate, and underlying application specifics.
- In WPT systems, since the energy harvesting process may affect the modulated information, joint resource allocation for the EH and data transmission should be investigated to improve the network performance.

An in-depth study of these challenges is required to design an efficient wireless network, which must consider different factors, like features of power generators, transducers, power storage, power management methods and application requirements. For deeper understanding of the benefits and opportunities offered by EH and WPT techniques as well as their challenges and potential application scenarios, the interested readers are recommended to refer to EH and WPT surveys, such as [98], [103], [104].

B. MOTIVATION

Both EH/WPT and MEC have been considered as promising technologies for the 5G networks, which can improve the energy efficiency of mobile/edge devices and prolong their battery lifetime of communication nodes at remote areas. While MEC enables to detach the end devices from heavy computation workloads for saving their energy consumption, EH/WPT techniques allow them to exploit the energy in their surrounding environment for re-charging their batteries. Hence, integrating these two technologies in the future wireless communication systems can significantly improve network performance by leveraging the strengths of both underlying technologies. EH, WPT and MEC technologies will lead to the following benefits:

- EH/WPT techniques can power the edge devices in the MEC systems to enlarge the set of options for computation offloading which will result in improving the network performance [105]. Specially in the IoT context, important use cases of MEC-enabled 5G networks, scavenging the ambient environment and utilizing WPT, provide promising solutions to perpetually support the massive number of electronic sensors [95]. Moreover, EH/WPT modules by leveraging green energy (e.g. solar and/or wind) can be employed to power MEC servers. Especially, EH/WPT provides a great solution for the eRAC use cases of MEC-based 5G where the MEC servers and mobile edge devices can be located outside the coverage of the electric grid for reasons such as deployment constraints, reliability requirements, carbon footprint, weather, disasters, and maintenance expenses.
- The distributed computing power of MEC systems can be leveraged to learn time-varying properties of the energy source for optimizing the network performance [106].
- A MEC server can be deployed to support a cluster of mobile/sensor nodes in EH-enabled wireless networks. At the node level, MEC can help each EH device reduce processing time and reserve more time for EH by offloading its heavy workloads to fog servers [107]. At the network-level, MEC can allow deploying a centralized EH strategy to tune the functionality of all devices for better EH and performance [108].

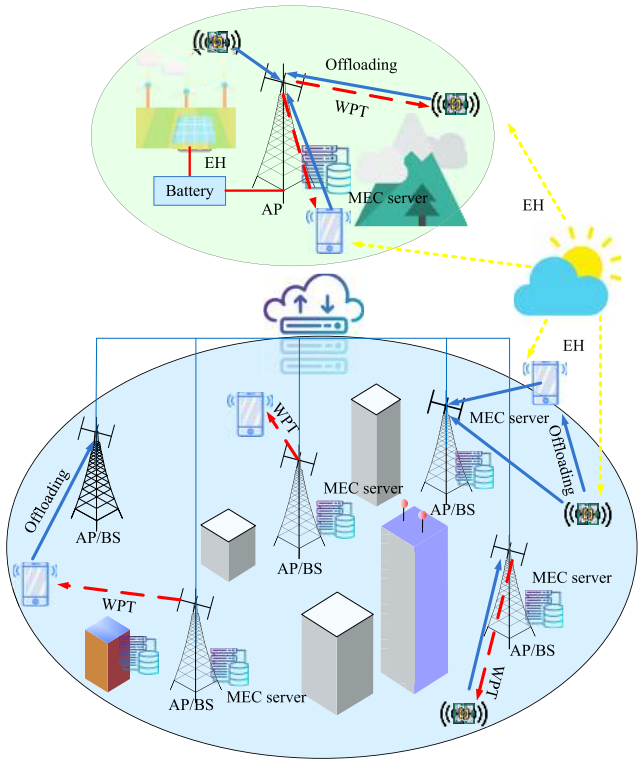

FIGURE 7. EH/WPT enabled MEC access networks.

A simple EH-enabled and wireless powered MEC system is illustrated in Fig. 7, where the EH and WPT are employed at MEC servers and mobile devices. For example, a batteryless user can utilize WPT to harvest energy from the BS, and then uses the harvested energy to offload computation tasks to the MEC server for remote computing. While EH and WPT bring many benefits as discussed above, the MEC system also faces many challenges including communication resource allocation, computing resource allocation, latency minimization, and security problem. In the following, we describe

TABLE 4. Summary of EH/WPT-MEC works.

	EH-enabled devices	EH-enabled MEC server	WPT	SWIPT
Resource allocation	[109], [110]	[111]–[113]	[114]–[117]	[118], [119]
Offloading designs	[109], [110], [120]		[114]–[117]	[118], [119]
Load balancing		[111]–[113]		

some major challenges which must be addressed for efficient integration of EH/WPT into the MEC system.

- In general, mobile devices have limited battery size and computation capability. Integrating EH/WPT into MEC-based wireless networks facilitates mobile devices with an external power source for processing heavy workload but this requires additional processing workload on controlling the EH function. Thus, such integrated system must cope with a more complicated resource allocation problem. Major research issues along this line include resource allocation and offloading design to well balance between harvested energy and consumed energy consumption. Specifically, how to perform the energy-efficient computation offloading in EH/WPT-MEC system considering practical constraints in the harvesting process remains a challenging issue.
- In the scenarios where MEC servers are primarily powered by renewable energy, the availability of energy source in the space and time domains would follow a unstable and non-uniform distribution. Moreover, these harvested energy level may vary over space which leads to load imbalance among servers. Hence, load balancing among all edge servers is also an interesting research problem which should be addressed in engineering EH/WPT-based MEC systems.

C. STATE OF THE ART

This section provides a survey on some recent works for efficient integration of EH/WPT into MEC systems which are summarized in Table 4. Existing works on combining EH/WPT and MEC mainly consider three schemes. In the first two schemes, the EH and WPT techniques are implemented at mobile devices in MEC-enabled wireless communication networks. These schemes can be applicable to WSNs, IoTs, eRAC, and D2D systems in the 5G network which support a massive number of small battery-operated devices connecting wirelessly to MEC servers for offloading and data processing. Because these devices typically have very limited batteries to supply power for their communication, EH and WPT technologies can be employed to provide valuable additional powers for their long-term operations such as sensing, reporting data, or offloading the heavy computation load. To do so, the edge devices need to estimate the power and time consumed by their operation. The resource allocation and offloading decision designs for these devices become more complicated due to the additional energy harvesting stages in EH/WPT enabled MEC systems which are promising research issues. The third scheme focuses on the scenarios in which connecting the MEC servers of MEC-enabled systems to the electric grid is costly and even impossible in certain situations such as natural disasters, remote locations. Then, on-site renewable energy is mandated as a major or even sole power supply source for these MEC servers [112]. In these cases, efficient load balancing design among all MEC servers under the unpredictable and unstable harvested energy has attracted a lot of research attention.

1) OFFLOADING AND RESOURCE ALLOCATION FOR MEC-ENABLED SYSTEMS USING EH TECHNIQUE

Recently, [109] considered the multi-user multi-task computation offloading problem which aims to maximize the overall revenue of the wireless EH-enabled devices. The Lyaponuv optimization approach was adopted in this work to devise the energy harvesting policy and the task offloading schedule. The tradeoff between energy consumption and execution delay for the MEC system with EH capability was studied in [110] in which the authors proposed an online dynamic task scheduling to minimize the average weighted energy consumption and execution delay subject to constraints on the stability of buffer queues and battery level. Employing the game theoretic approach, authors in [120] studied the impact of the EH technique at mobile devices in the computation offloading design. The work aimed to minimize the social group execution cost. Different queue models are applied to model the energy cost and delay performance, based on which a dynamic computation offloading scheme was designed. Using the deep learning (DL) approach, [105] proposed a reinforcement learning offloading scheme, where each EH-based IoT device selects its MEC server and the offloading rate without knowledge of the MEC model based on the current battery level, the previous radio transmission rate to each server, and the predicted harvestable energy.

2) OFFLOADING AND RESOURCE ALLOCATION FOR MEC-ENABLED SYSTEMS USING WPT TECHNIQUE

Considering the WPT-enabled MEC systems, [114] aimed to maximize the (weighted) sum computation rate of all wireless devices in the network by jointly optimizing the individual offloading decision and the time allocation for transmission. Similarly, [115] considered the time division strategy for the two-way data exchange between the fog node and the mobile user in WPT-based MEC systems. The closed-form average age of information for both directions as well as the achievable data rate of the mobile user was described in this paper, based on which the trade-off between the downlink and uplink performance was investigated. The cooperation among edge users was studied in [116], [117]. Specifically, the work [116] aimed to maximize energy efficiency (EE) to ensure the fairness of users by encouraging the near user (NU) forwarding the far user's (FU) tasks to the edge cloud. While [117] enabled the surrounding idle devices as the helpers to use their opportunistically scavenged wireless

energy to help remotely execute active users' computation tasks. The work tried to maximize the computation rate by jointly optimizing the transmit energy beamforming at the ET, as well as the communication and computation resource allocations at both the user and its helpers. Reference [121] considered a WPT-based UAV-assisted MEC system in which a UAV acts as an MEC-enabled BS offering WPT and offloading services to a number of EH-enabled ground mobile devices. The work aimed to maximize the system computation rate under both partial and binary computation offloading modes, subject to the energy-harvesting causal constraint and the UAV's speed constraint. On another approach, [118] investigated the power splitting problem for information transmission and power transfer in the SWIPT-based MEC system. Specifically, the authors proposed a new algorithm to minimize the required energy under the constraints on required information transmissions and processing rates. The work in [122] studied imperfect spectrum sensing in cognitive radio MEC with WPT. Specifically, a joint CPU frequency control, time power allocation problem was formulated and solved via a number of techniques, including dual decomposition, 1-dimensional search, bisection and subgradient method.

3) LOAD BALANCING DESIGN FOR MULTIPLE EH-BASED MEC SERVERS

In the EH-based MEC systems where the computation servers are mainly powered by the uncontrollable and unpredictable energy sources (e.g, solar, wind), individual MEC servers may be overloaded at any moment due to the limited harvested power and computing capacity [111]. Hence, energy prediction and load balancing among all EH-based servers are important research issues which must be tackled to achieve effective MEC operations. In particular, [112] considered a joint geographical load balancing and admission control for EH-based MEC networks which aims to minimize the long-term system cost due to violating the computation delay constraint and dropping data traffic. To deal with this geographically load balancing (GLB) optimization problem, Xu *et al.* developed an algorithm, called GLOBE, by leveraging the Lyapunov stochastic optimization technique. In particular, the algorithm enables MEC-enabled BSs to make GLB decisions without requiring future system information. Integrating the EH into MEC-enabled HetNets, authors in [113] investigated the joint load management and resource allocation problem that maximizes the number of offloading users utilizing the limited energy and computation resources, via managing the load and distributing the resources to the users. To solve the underlying complicated problem, a distributed three-stage iterative algorithm was proposed to obtain the joint load balancing and resource allocation solution.

D. LEARNED LESSONS AND POTENTIAL WORKS

Due to the great benefits offered by MEC and EH/WPT as well as their complementary properties, it is convinced that the combination of MEC and EH/WPT is beneficial in the future. Although various problems and issues in EH/WPT-MEC systems have been intensively studied, there are still several challenges. In the following, we discuss some challenges and outline the open research directions.

1) ENERGY PREDICTION

Most of the renewable energy sources are unpredictable. For example, clouds can appear or disappear which can affect the solar harvesting process. Other kinds of harvestable energy sources, e.g., wind, heat, and vibration, vary over time. In the WPT systems, channel characteristics practically vary depending on the environment in which the level of interference and the number of paths cannot be known in advance. Thus, understanding the surrounding ambient environment is critical for efficient implementation of the EH and WPT techniques. Recently, advanced machine-learning and deep-learning methods have been utilized to predict the arrival energy based on the historical and geographic data. Notwithstanding considerable benefits, ML/DL mechanisms and big data analytics raise some several challenging issues for implementation, such as, collecting data, large computation resources required to process the high-dimensional big data, which can be overcome by employing the MEC concept. Exploiting learning at MEC servers to extract useful information collected by all EH-enabled devices can reduce the time caused by sending the data to a remote cloud server; hence, the predicted information can be achieved on-time for high efficient EH, which can extend the capability of EH-enabled devices.

2) EH/WPT-BASED MEC FOR IoT/DENSE NETWORKS

An IoT network aims at supporting massive connections from machine-type devices which are small, fabricated and deployed at very low cost, and are expected to operate in a self-sufficient manner for a long time. The large number of connecting devices and their low power operation require an advanced wireless access networks, such as, dense access points or multi-hop data transmissions. MEC systems can play a relevant role in this scenario to manage functionality of individual nodes in terms of synchronization, reliability, efficiency of utilizing channel resource and energy, to exploit the available harvestable energy source, to cooperate with others for WPT, data transmission and offloading. The other challenge in successful large-scale deployment of devices in an IoT infrastructure is to minimize their impact on human-body and the environment [123]. The presence of multiple devices implementing various EH technologies corresponding to different kinds of energy sources, WPT and SWIPT over different frequency bands in the dense-users networks also require efficient and scalable offloading and resource allocation designs.

V. MEC FOR UAV COMMUNICATIONS

A. FUNDAMENTALS OF UAV

Historically, UAVs have been considered as enablers of various applications including military, surveillance and

monitoring, telecommunications, delivery of medical supplies, and rescue operations, owing to their autonomy, flexibility and broad range of coverage [124]. However, in those applications, UAVs mainly focused on navigation, control, and autonomy. As a result, the communication challenges of UAVs have typically been either neglected or considered as part of the control and autonomy components [125]. UAVs are commonly known as drones or remotely piloted aircrafts, and have several key potential applications in wireless communication systems due to its high mobility, flexibility, adaptive altitude and low cost [126]. Specifically, small UAVs are more easily accessible to the public recently due to its continuous cost reduction and device miniaturization, thus small UAVs can be used in weather monitoring, forest fire detection, traffic control, emergence search and rescue, cargo transport etc. In recent years, UAV-based wireless communication systems attract lots of attention thanks to their cost-effective wireless connectivity in scenarios without infrastructure coverage, which is caused by severe shadowing by urban or mountainous terrain, or damage to the communication infrastructure caused by natural disasters [127]. Among the UAV applications in wireless communication systems, UAV mainly serves as two important roles: 1) aerial BS and 2) flying mobile terminals. In the first scenario, when UAV serves as an aerial BS, it can provide communications in emergency and public safety situations to enhance coverage, capacity, reliability and energy efficiency of the wireless networks. In the second scenario, UAV can serve as a flying mobile terminal within the cellular networks to deliver real time video stream.

For UAV classifications, several factors such as outlook and application goals, need to be taken into account. The different types of UAVs depend on their functions, and capabilities. From their outlook characteristics, UAVs can be broadly classified into two categories: fixed-wing UAVs and rotary-wing UAVs. From the UAV application and goals, one alternative classification of UAVs can be done to meet various QoS requirements, the nature of the operation environment and federal regulations. To properly classify the applications and use of UAVs, UAVs' flying altitude and capabilities can be taken into account. Among these factors, flying altitude can be utilized for UAVs classification: high altitude platforms (HAPs) and low altitude platforms (LAPs) [126]. HAPs, e.g., balloons, usually operate in the stratosphere that is 17 km above the Earth's surface. On the contrary, LAPs, flying at altitudes not exceeding several kilometers, have several important advantages: fast movement and more flexibility compared to LAPs. The benefits of UAVs application in wireless communications can be summarized as follow:

- *Cost-effective, fast, flexible and efficient deployment:* UAVs can provide cost effective wireless communications and can be more flexibly deployed for unexpected or limited-duration missions. One of the main applications is that UAVs can serve as aerial BS. It is well known that building a conventional terrestrial BS, including radio towers and infrastructure deployment, is very expensive. In this case, UAV aided BS can provide on-the-fly communications at low cost since UAVs do not require highly constrained and expensive infrastructures.
- *Line-of-sight (LoS) link:* Compared with conventional terrestrial BSs, a UAV-aided flying BS is able to offer on-the-fly communications and to establish LoS communication links to ground users. Especially in low-altitude UAVs, the established LoS communication links can improve the network performance significantly. LoS communication can facilitate high frequency (e.g., mmWave). Combined with other 5G and beyond technologies, e.g., mmWave, MIMO, and visible light communications, UAV aided BSs can establish LoS communication links so as to achieve high data rates [64].
- *Coverage and capacity enhancement:* In the downlink communications, UAV aided flying BSs can rapidly reconfigure UAV-to-ground user links to provide a large coverage network due to its maneuverability. Specifically, in the uplink communications, the UAV-aided flying BS can also collect delay-tolerant information from the distributed wireless devices within the coverage. Since UAVs experience good channels, e.g., LoS link, they can provide higher transmit data rates. Moreover, the speed of UAVs can be manually adjusted to support wireless connectivity to the ground terminals. The benefits of large coverage and capacity improvement make UAV-aided wireless communication a promising integral component of the 5G wireless systems and beyond.
- *Complementary network for emergency situations and disaster relief, search and rescue:* Compared to the traditional network scenarios (e.g., 4G long term evolution (LTE) and WiFi), UAV aided wireless communication networks can provide a complementary network to the existing networks in emergency situations. For example, UAVs can act as hotpots for an ultra dense network, where the ground BS is overloaded. When the ground BS is damaged or even completely destroyed by natural disasters (e.g., earth quake, floods, severe hurricanes and snow storms), UAV aided wireless networks enable to provide effective communications and help rescue lives.

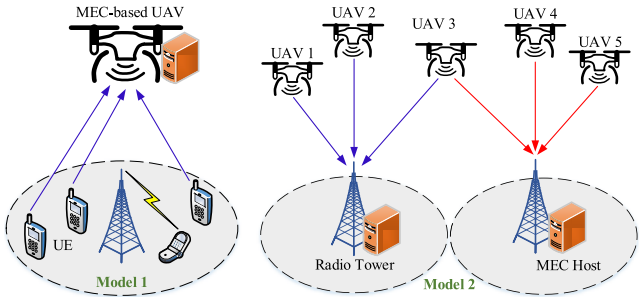

FIGURE 8. MEC-enabled UAV networks architecture.

B. MOTIVATION TO COMBINE MEC AND UAV

Due to the features of UAV, such as mobility, maneuverability, and flexible development, UAVs can be integrated into wireless communication systems to provide seamless, reliable, low delay and cost-effective communication [128]. To further improve the computation capacity, the combination of UAV and MEC has been proposed in existing works. There are two typical scenarios as shown in Fig. 8, including UAV-assisted communications and cellular-connected UAVs. In Mode 1 of Fig. 8, UAVs serve as aerial BSs [129]. In this scenario, UAV can be equipped with an MEC server. Thus, MEC-enabled UAV servers provide opportunities for ground mobile users to offload heavy computation tasks. After computation, the mobile users can download the computation results from UAV based MEC servers via reliable, cost-effective wireless communication links. In Mode 2 of Fig. 8, UAVs serve as new aerial mobile users of the *cellular-connected* network, where the MEC server based BS is able to provide the seamless and reliable wireless communications for UAVs to improve the computation performance. MEC has strong computing capability which can be complementary to the UAVs enabled wireless communications systems. The combination of UAV and MEC technology will lead to the following benefits:

- *UAV based MEC server*: In this scenario, UAVs can be used as mobile cloud computing systems, in which the UAV based MEC server can provide offloading opportunities to ground mobile users. Due to its flexibility and mobility, UAVs can receive the offloaded tasks especially when the territorial MEC servers are not available. For example, when the emergency relief or disaster happened, the mobile device with limited processing capability can benefit from the moving UAV aided MEC server to execute tasks, e.g., analyzing assessment of the status of victims, enemies and hazardous terrain [129]. Thanks to LoS links between UAVs and ground mobile users, the offloading and downloading capacity can be largely enhanced. Moreover, the coverage can be improved by the UAVs based MEC communication system.
- *UAV-UE MEC system*: Different from the traditional scenario where the mobile user is associated with a fixed GBS over the complex fading channel, the UAV-UE MEC system enables the high-mobility UAV-UEs to offload their computation tasks to the number of optimized GBSs simultaneously leveraging more reliable LoS links. There are two advantages of this scenario. On the one hand, the trajectory of the UAV can be jointly designed with the resource allocation (offloading task scheduling) as it has controllable mobility in 3D airspace. On the other hand, UAVs are associated with a group of GBSs simultaneously over LoS links to exploit their distributed computing resources to improve the computation capability.

Despite the promising benefits from the combination of UAV and MEC, there are several technical challenges existing in the MEC-enabled UAV systems. On the one hand, the main challenges in the UAV-BS scenario include the optimal 3D deployment of UAVs, the flight time optimization and the trajectory optimization. On the other hand, the challenges faced in MEC including communication resource allocation, computing resource allocation and security problem, need to be addressed. Therefore, combining UAV with the MEC system may raise the following challenges:

- *Mobility control and trajectory optimization*: Since UAV has limited flight time, the optimal path planning for UAVs MEC systems is an important research issue. For the UAV-based MEC server, the location and flying path must be optimized to provide better offloading opportunities for the mobile devices. Similar with the UAV-UE scenario, the location and flying path must be optimized to better offload computation tasks to a group of GBSs to provide seamless communication with other UAVs. In both scenarios, the mobility control has a significant impact on the quality of the network. It is challenging to optimize the trajectory of UAV as it typically requires to solve non-convex continuous optimization problems. The channel variation and energy consumption and maximum flying speed are required in this design. In addition, coupled with other optimization factors, such as QoS metric, the trajectory optimization is challenging to tackle.
- *Communication and computation resource optimization*: In the UAV based MEC server communication system, UAVs act as flying BSs equipped with MEC servers. The communication resource (i.e., offloading power) and computation resource (i.e., task offloading ratio) need to be jointly optimized considering potentially different objectives, e.g., relay minimization and energy consumption minimization. In the UAV-UE MEC system, UAVs act as high-mobility relay users to offload their computation-intensive tasks to the MEC server deployed at GBSs for remote execution. In this case, the trajectory of UAVs can be jointly optimized with the communication and computation resource allocation, which would be more challenging compared with the fixed user and BS cases.

C. STATE OF THE ART

There are two scenarios for which UAVs can be combined with MEC in communication systems. In the first scenario, UAVs act as flying BSs equipped with MEC servers offering offloading opportunities for the users on the ground [129]. This scenario is quite common in practice. For example, the moving MEC enabled UAV plays an important role in disaster response and emergence scenario, in which the ground BS (GBS) cannot provide any service due to the damages caused by a sudden disaster, e.g. earthquake. Mobile devices with limited processing capabilities can benefit from the UAV based MEC server. In the scenario of UAV-based MEC server [129], the UAV can act as a moving MEC server in the sky to help execute the computation tasks offloaded

TABLE 5. Summary of existing works on UAV MEC.

Topic	References	Scenarios or design objectives
UAV-based MEC server	[130]–[133]	UAV acts as flying BS with MEC server.
BS-based MEC server	[134], [135]	UAV is served by multiple ground BSs with MEC servers
Energy efficient design	[129], [130], [135]	To minimize the energy consumption of UAV MEC systems.
Delay minimization	[131]–[134]	To minimize the flying time of UAV.
UAV flight design	[136]	To optimize UAV trajectory, e.g., minimize the total flying distance of UAV.

by multiple ground users. This work aimed to minimize the total energy consumption considering the QoS requirement. By means of successive convex approximation (SCA) methods, the bit allocation was studied to minimize the mobile energy for OMA uplink and NOMA downlink in the UAV based MEC system. An energy consumption minimization problem was investigated for the UAV-enabled MEC system in [130]. To address the limited computing capacity and finite battery life time of the mobile device, the UAV based MEC server was proposed to provide offloading opportunities to the mobile device. An alternative algorithm was proposed to minimize the UAV's energy consumption by optimizing the offloaded computation bits and the CPU frequency of the users and the trajectory of the UAV with the maximum speed limitation. Simulation results in this work showed that the proposed scheme outperforms the benchmark schemes. In [131], the computing resource allocation and UAV hovering time were optimized to minimize the total energy consumption of the UAV. Moreover, the CPU's computational speed was considered in the optimization of UAV's trajectory and task assignment to minimize the energy consumption in [132]. Delay minimization is also an important issue for UAV-MEC communication. In [133], the delay minimization among all users was studied by jointly optimizing the UAV trajectory, the ratio of offloading tasks and the user scheduling.

In the second scenario, a cellular-connected UAV is served by multiple ground BSs that are equipped with MEC servers [134]. In this scenario, UAV needs to complete certain computation tasks during the flying time over some given locations. Thus the tasks can be offloaded to some selected ground BS. The work [134] aimed to minimize the UAV's mission completion time by jointly optimizing its trajectory and computation task scheduling considering the maximum speed constraint of the UAV and the computation capacity of the GBSs. It turns out that the formulated problem is nonconvex, thus it is difficult to find the global optimal solution in polynomial time. Therefore, the alternating optimization and SCA were exploited to obtain a high-quality suboptimal solution. In [136], the total travel distance of UAV was minimized and two different solutions were proposed, i.e., MEC-ware UAV's path planning (MAUP) based integer linear programming and accelerated MAUP. Physical-layer security was investigated in [135], where the optimal solutions based on the condition of three offloading options and the computational overload event from a physical-layer perspective were provided. The summary of exiting works on UAV MEC is provided in Table 5.

D. LEARNED LESSONS AND POTENTIAL WORKS

Thanks to the great benefits from the combination of MEC and UAVs as well as their limited resource, it can be concluded that MEC-UAV is an inevitable trend in the future wireless communication systems. Although some existing works have been done to engineer MEC-UAV systems, there are still several challenges to address. In the following, we discuss key open problems in MEC-UAV systems:

Performance analysis of UAV-MEC systems: A fundamental performance analysis is required for the UAV-MEC system. In particular, the coverage probability, throughput, delay or reliability can be investigated to evaluate the impact of each design parameter on the overall system performance. Due to the 3D development and short flight duration of UAVs and the delay awareness of MEC, the performance analysis for the UAV-MEC system is challenging.

Energy-aware resource allocation: The flying time and the resource of UAVs are limited because UAVs typically have small sizes, weight and limited power. Thus, the trajectory and resource allocation (i.e., communication and computation) need to be optimally designed to reduce the energy consumption. However, most existing works only considered designing trajectory and optimizing resource allocation separately, which cannot achieve the highest network performance. Hence, jointly optimizing the path planning and resource allocation for MEC-UAV system is an open challenging problem. It becomes more challenging when other factors, such as, QoS requirement, offloading power allocation and task assignment together with the channel variation, delay constraint and maximum flying speed, are considered in such design.

User grouping and UAV association: In the UAV based MEC server communication system, each UAV acts as a flying MEC-enabled BS. The ground users need to offload their tasks to one UAV or multiple UAVs simultaneously. Thus the user group problem must be solved by using suitable approaches, e.g., matching theory, game theory and convex optimization methods. On the contrary, in the UAV MEC systems, UAVs need to offload tasks to GBSs for remote computation. The subchannel allocation and UAVs association can be investigated.

VI. MEC FOR INTERNET OF THINGS

A. FUNDAMENTALS OF IoT

Thanks to significant advancement in computation and storage technologies, and communication networks, billions of devices with their every domain-specific applications are able

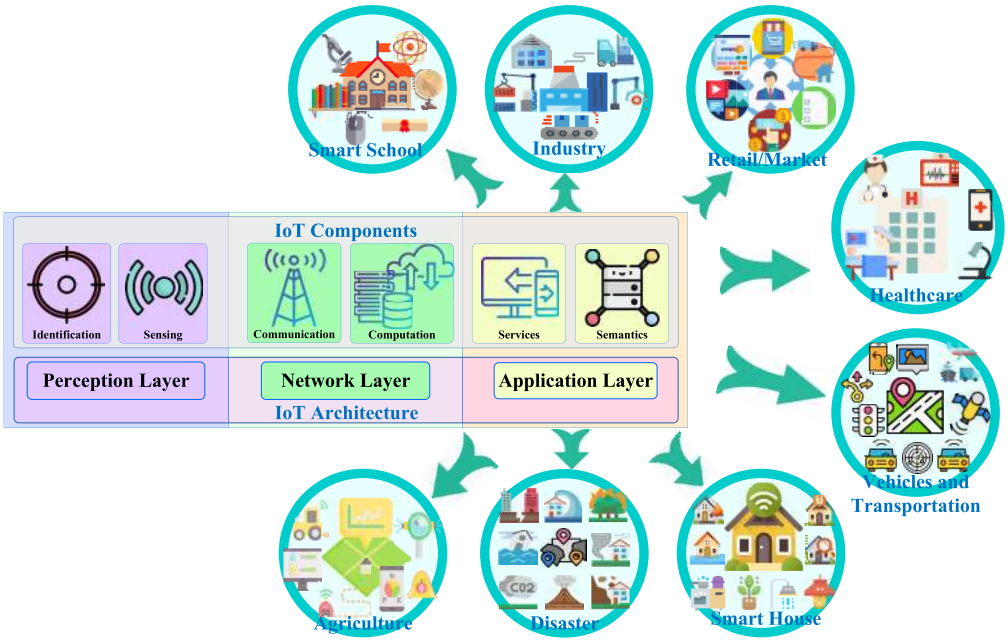

FIGURE 9. The overall picture of IoT applications and architecture.

to connect to the Internet to generate/collect data, to exchange important messages amongst themselves, and to coordinate decisions via complex communication networks [137]. This phenomenon has opened a new era of Internet, the so-called the IoT [137]. The basic concept of IoT is that anything can be interconnected with the global information and communication infrastructure at any time and any place [138]. Things can be physical things existing in the physical world or virtual things existing in the information world. IoT has been playing a significant role in solving various challenges of modern society effectively and improving the quality of human life, such as, safer, healthier, more productive, and more comfortable [139]. The fundamental characteristics of IoT can be condensed as follows: *1) inter-connectivity, 2) things-related services, 3) heterogeneity, 4) dynamic changes*, see [138] for details. IoT is also one of the main motivations for developing the promising 5G technologies to allow the massive connections from a large numbers of ''things'' to the Internet via wireless networks. Inversely, 5G is considered a basic platform to facilitate emerging IoT applications [140]. As expected, manifold data traffic (typically of Gbps order), low latency transmission can be provided by 5G communication networks which can support a tremendous increase in dense connected ''things'' in wireless networks, including high-mobility IoT/UEs, embedded sensors in the human body (or clothing), wearable devices, equipment for monitoring biometrics, or even autonomous cars (also called V2X communications). Furthermore, by exploiting spectrum resources in high-frequency bands and providing the coexistence of multiple numerologies, 5G networks can realize Tactile Internet requiring ultra-low latency with extremely high availability, reliability, and security [141]. For more information on the techniques and future trends of IoT, we invite the readers to further refer to the following references [142]–[144].

A basic architecture of IoT as well as its specific every-domain applications can be summarized in Fig. 9. In particular, the IoT basic architecture consists of three layers: Perception, Network, and Application [142], [143]. In the first layer, the physical sensors collect useful information/data from things or the environment which are then transformed into digital form and it marks all objects with a unique address identification. The principal responsibility of the second layer is to help and secure data transmission between the perception and the application layers [144]. The third layer is to provide the personalized based services according to users' relevant needs and to link the major gap between users and applications. It combines the industry to attain the high-level intelligent solutions for IoT specific every-domain applications such as the disaster monitoring, healthcare, smart house, transposition, production controlling, health care, retail, education. In other aspects, the third layer can be further divided into three sub-layers: *1) The service management layer, 2) The application layer, 3) The Business layer*. Due to the high-level requirement of some applications and services, one more layer has been potentially added between the application and network layers which consists of MEC and fog computing servers to perform some specific distributed computation duty or pre-data processing.

B. MOTIVATION TO USE MEC FOR IoT AND CHALLENGES

ETSI, in its report [145], has distinguished IoT as one of the most important MEC application instances. There are many benefits of employing MEC into IoT systems, including but

TABLE 6. Summary of MEC-enabled IoT papers on different application scenarios and technical aspect.

	Smart city	Healthcare	V2X	Industrial Internet	Wearable IoT/ AR and VR	Mechanized Agriculture
Offloading/ Resource allocation	[148], [149]	[150]	[151]	[152]–[157]	[158]–[162]	[163], [164]
Energy Management	[148], [165]	[150]	[151]	[152], [153], [166]	[158], [161]	
Safety	[167]–[170]		[171]		[172]	
Security	[167], [168], [173]	[174], [175]	[171]			
Privacy	[173]		[176]			
Convenience	[169], [170], [177]	[175]	[171]	[153], [156]		
Monitoring/controlling	[177]		[151]	[154], [155], [166]		[163], [164], [172]
Reliability		[150], [175], [178]	[171]	[146], [154], [157], [179]–[181]	[158], [162]	[164]
Latency		[150], [178]	[151]	[146], [153], [154], [180]–[182]	[159]–[162]	
Scalibility				[154]		

not limited to, lowering the amount of traffic passing through the infrastructure and reducing the latency for applications and services [12]. Among these, the most significant is the low latency introduced by MEC which is suitable for 5G Tactile Internet applications requiring round-trip latency in the millisecond range [146]. MEC technologies are envisioned to work as gateways placed at the middle layer of IoT architecture which can aggregate and process the small data packets generated by IoT services and provide some additional special edge functions before they reach the core network; hence, the end-to-end delay can be reduced. Additionally, these techniques are also able to lower the energy consumption of small-size IoT devices and prolong their battery-life by supporting significant additional computational capabilities through intelligent computation offloading strategies. Furthermore, MEC platforms will be offered and deployed by the network operator at any tiers of 5G networks, e.g., eNBs, multi-RAT aggregation points, neighbor mobile devices, which can be made open to authorised developers and content providers to deploy versatile and uninterrupted services on IoT applications [9]. In addition, based on the context and platforms of MEC, artificial intelligence (AI) on the edge can gain the huge benefit to realize distributed IoT applications and intelligent system management, which is now considered as a part of beyond 5G standardization [147]. Inversely, IoT also energizes MEC with mutual advantages. In particular, IoT expands MEC services to all types of smart objects ranging from sensors and actuators to smart vehicles. Integrating MEC capabilities to the IoT systems come with an assurance of better performance in terms of quality of service and ease of implementation.

C. STATE OF THE ART-MEC-ENABLED IoT APPLICATION SCENARIOS

This section focuses on providing a survey on recent MEC-enabled IoT works in application scenarios related to 5G uses cases. The technical aspects and application scenarios of these works are summarized in Table 6.

1) SMART HOME AND SMART CITY

One of the most important use cases of IoT is smart city and its important subset smart home/building [183]. Recently, the MEC contexts and novel 5G technologies have been enabled to emerge the judicious edge big data analysis and wireless access for IoT systems to further improve the urban quality of life for citizen with many aspects including security, privacy, energy management, safety, convenient life, ect.. For energy management, an fog-based IoT automation mechanism was validated in [148] to optimize the resource management for smart building systems. By leveraging the fog-enabled cloud computing environments, the novel implemented smart home systems can reduce 12% utilized network bandwidth, 10% response time, 14% latency and 12.35% in energy consumption. For monitoring and controlling the smart home/buildings, innovative analytics on IoT captured data from smart homes was presented in [149] employing the fog computing nodes. This fog-based IoT system can address the challenges of complexities and resource demands for online and offline data processing, storage, and classification analysis in home/building environment. The MEC-enabled IoT frameworks in [167], [168] focus on behaviour features by monitoring the student's location and activities in school environment for safety aspect. In particular, [167] designed a platform to identify any student activities that occur at the classroom level in which the raw indoors environment data is processed at an edge computing server (Raspberry Pi) for detecting the presence of individuals in a classroom while [168] exploited the DL algorithms in an MEC-enabled IoT smart classroom for person recognition.

For the smart city use cases, the security and privacy aspects were considered in [173] where a blockchain-based smart contract services for the sustainable IoT-enabled economy is proposed for smart cities by employing AI solutions in processing and extracting significant event information at the fog nodes, and then utilizing blockchain algorithms to save and deliver results. Recent work in [165] studied the energy management aspect in smart city where the deep reinforcement learning methods were employed into MEC-enabled IoT system to manage the energy grid efficiently. References [169] and [177] both considered the safety and convenience aspects where Pratam *et al.* [169] implemented a Raspberry Pi-based MEC system on school shuttle buses for tracking the locations of students and vehicles while [177] developed a smart routing for crowd management based on

deep reinforcement learning algorithms to satisfy the latency constraints of service requests from the people. A platform to detect potholes and road monitoring was studied in [170] to cope with flooding on the roads in rainy seasons for traffic safety.

2) HEALTHCARE

Healthcare solutions with more intelligent and prediction capabilities have been developed and implemented based on the rapid developments of IoT and cyber physical systems [184]. MEC-enabled IoT has shown a huge potential in improving the performance of healthcare systems which includes but not limited to the mobile monitoring healthcare scheme. In this system, the MEC-enabled gateways can offer several higher-level services such as local storage, real-time local data processing, embedded data mining, etc. beside controlling the data transmission [185]. These enable to empower the system to deal with many challenges of managing the remote devices, i.e., security, reliability, latency, energy efficiency issues. Freshly, Li *et al.* in [174] considered the security issue in mobile healthcare systems by proposing a secure and efficient data management system named EdgeCare in which healthcare data and facilitating data trading are processed at edge servers with security considerations. Focusing on improvement of latency and reliability performance, [178] proposed BodyEdge, a novel body healthcare architecture consisting of a tiny mobile client module and an edge gateway for collecting and locally processing data coming from different scenarios. Sharing the same view, [150] implemented an accurate and lightweight classification mechanism employing the edge computing to detect the seizure at network edge based on the information extracted from the vital signs with precise classification accuracy and low computational requirement. The implementation results show that the proposed system outperforms conventional non-MEC remote monitoring systems by: 1) achieving 98.3% classification accuracy for seizures detection, 2) extending battery lifetime by 60%, and 3) decreasing average transmission delay by 90%. For emergency department systems, Oueida *et al.* [175] proposed a resource preservation net framework integrated with cloud and edge computing where the key performance indicators such as patient length of stay, resource utilization rate and average patient waiting time are modeled and optimized considering high reliability, efficiency and security.

3) VEHICLE-TO-EVERYTHING (V2X) IoT

In [186], 3GPP has identified 5G as the key technology supporting the V2X concepts in several use cases: Information (state map, environment, traffics) sharing, vehicle platooning, remote driving, grouping-based cooperative driving, communication between vehicles, cooperative collision avoidance, dynamic ride sharing. The QoS requirements in data rate and communication range may vary in different V2X applications [187]. However, the crucial factors such as ultra low latency, high reliability, and security have to be improved due to the safety in most use cases, which can be fulfilled by employing MEC technologies [188]. Recently, the security aspects in V2X were considered in [171] which enabled a cooperative intelligent transportation system by deploying MEC-equipped cell towers hosting local communication to increase the safety on roads and the traffic efficiency with smoother flow. Reference [151] focused on the latency in MEC based dense mmWave V2X networks by optimizing the offloaded computing tasks and transmit power of vehicles and road side units to minimize the energy consumption under delay constraint resulting from vehicle mobility. The work in [176] enabled the object recognition enhancement with DL algorithms at the edge side with MEC deployment in V2X networks to improve the information sharing and communication performance. Specifically, an Intel Movidius Neural Compute Stick along with Raspberry Pi 3 Model B is used as an edge computing server to analyze the objects contained in real-time images and videos.

4) INDUSTRIAL INTERNET

MEC yields a significant paradigm shift in industrial Internet of Things (IIoT), well-known as Industry 4.0 - a use-case of 5G technologies, by bringing computing resources close to the lightweight IIoT devices in IIoT domain [152], [189]. In IIoT, there are many application scenarios such as, factory automation, process automation, human-machine interfaces, production IT, logistics and warehousing, monitoring and maintenance. Intelligently managing the edge resources, MEC enables to power the IIoT system to address some significant technical issues, e.g. latency, resilience, connectivity, and security.

To make MEC an enabler for latency-critical IIoT applications, time-sensitive networking (TSN)[2] is a vital solution. Reference [190] proposed TSN-based configuration architectures of MEC that can support real-time IIoT applications. Considering system resources, [152] reported that enabling MEC in IIoT systems can improve the system efficiency by jointly designing resource allocation and offloading based on an auction-based method where both claimed bids and asked prices were given by the MEC servers. Additionally, Li *et al.* in [153] employed MEC servers in SDN for IIoT systems to dynamically optimize the routing path considering the aggregation of time deadline, traffic load balances, and energy consumption to provide a better solution for IIoT data transmission in terms of average time delay, throughput, energy efficiency, and download time. Reference [154] proposed a service popularity-based smart resource partitioning scheme for fog computing-enabled IIoT. By demonstrating the notable performance improvements on delay time, successful response rate and fault tolerance, the authors confirm the significant benefit of enabling fog computing to cope with the large-scale IIoT services. While [182] implemented DL at the edge servers to enhance the range and

[2]TSN includes a set of protocols to provide timing guarantees for latency-critical applications. The IEEE 802.1 TSN's home page is available at https://1.ieee802.org/tsn/ and its overview paper can be found in [40].

computational speed of IIoT devices remarkably in the MEC-based IIoT framework for increasing the energy efficiency and battery lifetime at acceptable reliability (around 95 %). Reference [179] focused on obtaining higher reliability of network interactions by proposing a deadlock avoidance resource provisioning algorithm for Industrial IoT devices using MEC platforms.

Aiming at improving the quality of industrial production, [155] implemented parallel MEC to improve the efficiency of equipment identification. In particular, adopting the long short-term memory to analyze big data features and build a non-intrusive load monitoring system with MEC can enlarge the average recognition rate to over 80%. MEC can also be applied for smart IoT-based manufacturing to improve performance of edge-equipment network, information fusion, and cooperative mechanism, based on which the excellent real-time, satisfaction degree and energy consumption performance of the manufacturing system can be significantly improved [166]. On another view, to achieve higher goodput, [157] enabled the MEC platform to improve the caching management for IIoT system. For the security purpose, [156] employed a smart blockchain-based platform with many MEC servers in IIoT systems to effectively solve the network congestion caused by transferring raw data (e.g., pictures or video clips) between a publisher and workers.

5) WEARABLE IoT, AR AND VR

The newly emerging applications corresponding to mobile AR, VR, and wearable devices, e.g., smart glasses and watches, are anticipated to be among the most demanding applications over wireless networks so far, but there is still lack of sufficient capacities to execute sophisticated data processing algorithms. To overcome such challenges, the emergence of MEC and 5G techniques would pose the longer battery lifetime, powerful set of computing and storage resources, and low end-to-end latency [160], [191]. Sharing this view, [158] presented Outlet system to explore the available computing resources from users' ambiance, e.g., from nearby smart phones, tablets, computers, Wi-Fi APs, to form an MEC platform for executing the offloading tasks from wearable devices. Promising performance achieved by Outlet, e.g., mostly within 97.6% to 99.5% closeness of the optimal performance, has demonstrated the advantage of edge computing into wearable IoT systems. Applying MEC on VR devices, [159] presented an effective solution to deliver VR videos over wireless networks minimizing the communication-resource consumption under the delay constraint. This work also demonstrated the interesting tradeoffs among communications, computing, and caching. In [160], a novel delivery framework enabling field of views caching and post-processing procedures at the mobile VR device was proposed to save communication bandwidth while meeting low latency requirement. Impressively, an implementation of MEC concepts over Android OS and Unity VR application engine in [161] enabled to reduce more than 90% computation burden and more than 95% of the VR frame data. On a different view, Liu *et al.* in [162] illustrated the advantage of implementing MEC in panoramic VR system to maintain the high quality of the video streaming by intelligent balancing the link adaptation, transcoding-based chunk quality adaptation, and viewport rendering offloading.

6) MECHANIZED AGRICULTURE WITH IoT

IoT emerging the use of low-cost hardware (sensors/microcontrollers) and 5G communication technologies for eRAC has opened a new era for cultivating soil, namely *"smart agricultural"* [192]. Many advanced abilities, e.g., predictive analytic, weather forecasting for crops or smart logistics and warehousing, can be offered by enabling MEC technologies in this scenario [193]. Recently, there are some works on emergence of MEC and IoT in agriculture. In particular, [163] proposed an intelligent agricultural water monitoring system with advanced MEC technology to effectively manage the data collected by the sensors. As a part of EU DrainUse project, [164] presented a local/edge/cloud three-tier platform for monitoring and managing soil-less agriculture in full re-circulation greenhouses using moderately saline water. In this platform, the edge plane is deployed to increase system reliability against network access failures while the data analytic modules are located in the cloud. To protect the plant on vineyard fields, [172] implemented a disease alerting platform using a low-cost sensors in the municipality of Vilafamés (Castelló, Spain). In this platform, the edge computing is deployed to improve the capability of monitoring meteorological phenomena collect (e.g. temperature, humidity) based on that an alert disease model was developed for improving the product quality.

7) TACTILE INTERNET

Tactile Internet is defined by the International Telecommunication Union (ITU) as the next evolution of IoT that combines ultra low latency with extremely high availability, reliability and security [194]. Encompassing human-to-machine and machine-to-machine interaction, Tactile Internet will combine multiple technologies including 5G and MEC, i.e., 5G may be employed for the data transmission with low delay and high reliability while MEC efficiently exploit computing resources close to the end users for better QoE. The applications related to Tactile Internet can be automation, robotics, tele-presence, tele-operation, AR, VR [141], [194]. The works employing MEC in these scenarios considering low latency and high reliability can be found in Table 6 and introduced in the previous parts. The following summarizes the recent works focusing on the technical aspects involving to the MEC implementation in Tactile Internet. Reference [146] considered an energy-efficient design of fog computing networks that supports low service response time of end-users in Tactile Internet applications and efficiently utilizes the power of fog nodes. The trade-off between the latency and required power was presented and then extended to fog computing networks leveraging cooperation between fog nodes. Reference [180] exploited the MEC systems

including cloud, decentralized cloudlets, and neighboring robots equipped with computing resource collaborative nodes for computation offloading in support of a host robot's task execution. Then, a proper task allocation strategy by combining suitable host selection and computation task offloading was proposed to meet the required task execution time. The work also showed that the MEC-based collaborative task execution scheme outperforms the non-collaborative scheme in terms of task response time and energy consumption efficiency. Recently, Xu *et al.* in [181] designed a hybrid edge caching scheme for Tactile Internet which can reduce latency and achieve better performance in overall energy efficiency than existing ones.

D. LEARNED LESSONS AND POTENTIAL WORKS

Several research works and implementations in the literature have demonstrated that MEC is an ideal solution for IoT systems. In many applications and use cases, exploiting MEC resources for managing the data collection or pre-processing the massive data at the edge networks is able to lead to significant advantages. These advantages include but not limited to reducing the radio resource consumption (i.e., 12% in [148]), shortening the reaction time (i.e., 10% in [148]), lessening the system latency (i.e., 14% in [148], 90% in [150]), and diminishing the overall energy consumption (i.e., 12.35% in [148]). In addition, MEC also helps offload the computational burden at IoT devices, which results in prolonging their battery life (i.e., 60% in [150]), increasing the accuracy rate of task processing (i.e., improving the seizures detection rate over 98% in [150]), mitigating the amount of transmission data (i.e., 95% in [161]), and lowering the computation load (i.e., 90% in [161]). However, to maximize benefits of MEC in IoT applications, one requires the more efficient management of the MEC resources and access networks, and capacities as well as abilities of the IoT components or elements. These demands open many potential research directions to effectively governance MEC in IoT systems. The future works considering technical aspects of IoT and MEC, i.e., scalability, communication, computation offloading and resource allocation, mobility management, security, privacy, and trust management, have been well indicated and manifested in some recent MEC-IoT surveys, such as, [195], [196] to which the interested readers are recommended to refer. In the following, we discuss key open problems in MEC IoT systems which are different to the mentioned challenging technical aspects.

1) EFFECTIVE COOPERATION IN DENSE MEC-BASED IoT NETWORKS

Currently, each MEC server is deployed by the infrastructure providers to supply the computing and radio access services to a specific set of distributed edge IoT nodes at the IoT network edge. In addition, a provisioning set of computation or networking functions including data analyzing, compressing, caching, routing, etc., are installed at a distributed MEC server to serve its set of devices from the aspect of their applications. In dense IoT-based smart cities, massive heterogeneous IoT devices running diversely advanced services corresponding to various domains of city life [197]. This leads to a huge number of devices with diverse service requirements from different infrastructure providers locating in a same geographical area. Although a new service (i.e., out-set-of-function service) can be supported by MEC by offloading raw data to cloud for processing, this may lead to huge cost of energy and time. In addition, non-cooperative edge servers deployed by different infrastructure providers may result in severe under-utilization of resources. Hence, enabling cooperative edge computing environment can open the resource of many types of edge computing servers for serving the diverse requirements in the dense IoT networks. However, to realize the cooperation among the edge nodes to maximize their benefits, several particular challenges should be solved: The trade-off between the cloud and the edge; The optimization of the service placement on distributed and limited edge resources; The contradiction between the computation-intensive edge services and the limited edge resources [198].

2) EMPLOYING AI TECHNIQUES IN MEC-BASED IoT SYSTEMS

Recently, AI techniques with ML/DL have been considered as important tools for processing big data in the IoT-based environment. The integration of ML/DL and AI algorithms at the network edge can provide efficient data analysis, make accurate decisions, predict tasks at the network edge, optimize the mobile edge caching, computation offloading, and preserve network security and data privacy. In addition, adopting AI techniques for MEC-enabled IoT system can extract the behaviors of physical/networking resources and users in different time and scenarios, dynamically monitor and adjust the configuration of network resources, and realize real-time data collection of loT, efficient processing of computation, based on which the intelligent services for heterogeneous IoT devices can be optimized [199]. However, to apply the AI technology regularly requiring big data processing at the edge nodes which are commonly equipped limited computation, storage resource, one needs novel ML/DL-based algorithm with distributed computing and data access which is an challenging issue for the future works. For more detail on ML/DL for MEC applications, we invite the readers to refer to Section VIII.

VII. MEC WITH HETEROGENEOUS CLOUD RADIO ACCESS NETWORK

A. FUNDAMENTALS OF HETEROGENEOUS C-RANs

To meet the unprecedented increase in the network traffic volume and the massive number of connected devices, network densification has become the cornerstone of the 5G networks, where more base stations and access points are added and spatial spectrum reuse is exploited. HetNet is defined as an integration of higher-tier macrocells and lower-tier small cells, for example, picocells, femtocells, and relay nodes [200].

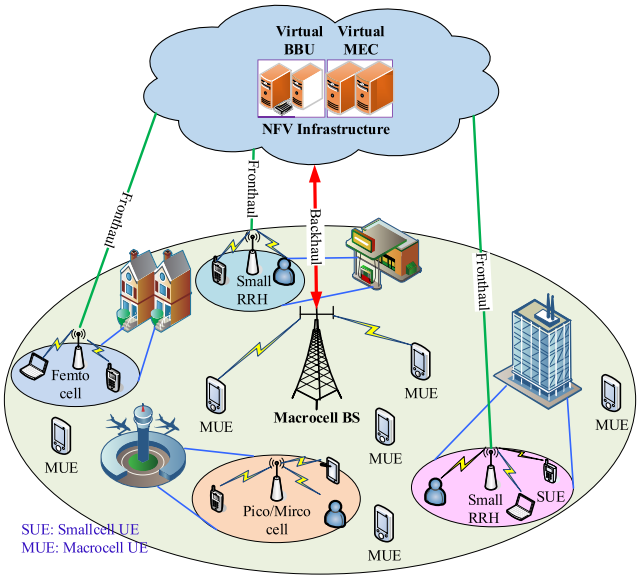

FIGURE 10. H-CRAN MEC architecture.

HetNets have been developed because of its following benefits: 1) better coverage and capacity, 2) improved macrocell reliability, cost benefits, and 3) reduced cost and subscriber turnover [41], [201]. However, the deployment of dense HetNets has several challenges: 1) severe interference, 2) unsatisfactory energy efficiency, and 3) inflexibility and unscalability. To overcome these challenges, another new promising network infrastructure, C-RAN, is proposed to provide a high transmission data rate and high energy efficiency performance, which attracts a lot of attention from academic and industrial communities [202]. In [203], the challenges and requirements of C-RAN were studied to enable network densification and centralized operation of the radio access network over heterogeneous backhaul networks. In C-RANs, shown in Fig. 10, a large number of low-cost low-power RRHs connecting to the BBU pool through the fronthaul links, are randomly deployed to enhance the wireless capacity in hotspots. RRHs operate as soft relay by compressing and forwarding the received signals from users to the BBU pool via wire/wireless fronthaul links. As a result, the combination of HetNets and C-RANs, known as heterogeneous C-RANs (H-CRANs), is proposed as a potential solution to provide high spectral and energy efficiency [204]. In order to support more 5G applications and reduce the investment cost of MEC deployment, MEC was proposed to be combined with CRAN in [50], where MEC services enable to exploit C-RAN by using the planned BBU pool. Even though CRANs and MEC can be perfectly paired to provide low latency for the IoT applications in HeNets, the co-location of MEC and C-RAN results in some challenges (e.g., network management), especially in HetNets.

B. MOTIVATIONS AND CHALLENGES

H-CRANs can provide large coverage and high energy efficiency, while MEC can provide the considerable computing capability for the low-latency applications. Collocating these two key technologies can help support more applications in 5G. Considering the computational and storage resources in the BBU pool and the distribution of the RRHs, H-CRAN can be combined with MEC to facilitate the implementation of the MEC system. Therefore, the combination of MEC with H-CRANs can bring the following benefits:

- The investment of MEC deployment can be significant reduced by collocating MEC and H-CRAN. As we all know, it is a significant investment to deploy a sufficiently extensive MEC network. One way to mitigate the investment cost is to bootstrap MEC deployment to the C-RAN deployment. In this case, the cost of providing additional task calculation across the existing BBU pool or RRHs will be reduced.
- The combination of MEC and H-CRAN can provide operational flexibility and network re-configurability, which can be offered by virtualization of H-CRAN. The H-CRAN can facilitate a faster radio deployment by reducing the time needed in the conventional deployments, e.g., standard General-Purpose Processors. Since CRAN virtualizes much of the RAN functions, thus MEC can also benefit large coverage, the energy savings, network simplicity and high security from H-CRAN.
- H-CRAN MEC can be flexibly deployed across different locations. For example, C-RAN can process the task signals any locations, e.g., cell-tower co-located hut. Since H-CRAN deployment requires a substantial amount of processing power, it can automatically becomes an MEC server to calculate the tasks from the mobile users.

In addition to the above benefits, there exist several challenges in H-CRAN MEC systems that can be induced by co-location of MEC and H-CRAN, e.g., deployment scenarios design. In the following, the major challenges of H-CRAN MEC systems are discussed [50], [205].

- In the H-CRAN MEC system, the balance of the deployment and the network performance should be well investigated. Since H-CRAN supports a dynamic capacity of the H-CRAN, how far the C-RAN/MEC site is located to cell-sits will affect the performance of MEC systems, e.g., how well it can support the applications. For example, locating CRAN/MEC site in a central office can reduce the cost significantly but it causes high latency [50]. In this case, use-cases should be carefully studied to run which applications at which sits.
- Most resource management methods for MEC consider the computation resource at MEC servers [206], [207] and thus can be applied in H-CRAN MEC directly. However, it is still challenging to jointly optimize computing resource and scheduling network resource in H-CRAN [67]. Especially in HetNets, the cross-layer and inter-cell interference needs to be considered. Moreover, based on NFV of C-RAN, the dynamic resource management scheme may need to be redesigned to

elastically schedule virtual computation resources under different network sizes and task arrival rates.

- Security is another issue to be addressed in H-CRAN MEC systems. Since MEC service supports various kinds of applications, such as third party applications, which are not controlled by mobile network operators directly. There may be risks that these applications will exhaust resources or offer hackers to affect the functions of the network. Therefore, the service of performing integrity assurance checks on applications should be considered at installation or upgradation.
- Due to the existence of inter-carrier interference, the resource allocation problem in H-CRAN MEC networks is much more challenging than that in traditional MEC systems [67]. To mitigate this effect, the spectrum resource within each cell can be divided into orthogonal subchannels, which should be efficiently allocated to mobile users (i.e., which subchannel a user should use to offload its computation task to the MEC server). In H-CRAN MEC networks, various types of resources need to be considered to reduce the inter-cell interference, including not only conventional wireless resources (e.g., subchannel, transmit power, time, and space) but also contra costs (e.g., backhaul spectrum, harvested energy, computing capabilities, and caching storage). The major challenges of dense H-CRAN MEC systems are user association, computation offloading, interference management, and resource allocation. More importantly, these problems are tightly coupled and must be solved jointly.
- On the one hand, it is foreseeable that a massive number of MEC servers will be widely deployed in the near future, which can be distinctly different in sizes (computing units) and configurations (computational speeds). On the other hand, the association between users and MEC servers (BBUs) greatly depends on the deployment locations of the MEC servers (BBUs). User mobility can be ignored whenever the UE moves inside the geographical area covered by the centralized BBUs. The type of BBU centralization determines the system efficiency and the user experience.

C. STATE OF THE ART

The majority of the existing studies have focused on Heterogeneous MEC (Het-MEC) and C-RAN MEC. For Het-MEC network, there are several papers working on interference management in dense Het-MEC systems [208]–[213]. In [208], the authors investigated a joint problem of radio and computational resources to minimize the total energy consumption of all mobile users under transmit power budget, latency, and maximum computing capability constraints. Similarly, Al-Shuwaili *et al.* in [209] considered several issues in single-server multi-cell Het-MEC systems: (1) the management of uplink and downlink interference, (2) the allocation of backhaul capacity for task offloading, and (3) the allocation of computing capabilities at the cloud for offloading users. Moreover, the joint optimization of offloading decisions and resource allocation has been extensively investigated to improve the network performance [210], [211]. In order to realize the potential benefits of dense Het-MEC networks, a new technical challenge is mobility management. According to [214], [215], there are several key issues for mobility management in Het-MEC systems. First, users may experience frequent handover when they move across different small-size and small-coverage smallcells/ MEC servers, thus increasing the overhead and interrupting the MEC services [216]. Second, continuously performing handover measurements and processing, which is needed to discover new target MEC servers in dense Het-MEC systems, is power- and radio resource-consuming, especially for battery-limited users. Third, in traditional dense HetNets, handover decision is mainly based on the quality of radio signals between users and potential eNBs. In addition, due to the lack of future information, e.g., channel conditions, available computing resources, task arrivals, the offloading and handover decisions should be known without prior information and be optimized in a long-term manner [213]. Due to its critical importance, an extensive body of work has appeared in the literature to address the challenges of mobility management in conventional dense HetNets [214]–[221]. For example, two localized mobility management schemes for dense HetNets were proposed in [214], a cache-enabled mobility management framework in mmWave-microwave HetNets was studied in [215], various energy-efficient cell discovery techniques were discussed in [217], a comprehensive review of mobility management was provided in [218], and the adoption of distributed mobility management was presented in [219]. Although interesting, the body of work in [214]–[221] solely focused on mobility management in HetNets. Taking challenges of mobility management in dense Het-MEC systems into consideration, the study in [213] optimized the association (which MEC server is selected for remote execution) and handover (i.e., when task migration is needed) decisions to minimize the average delay with the long-term energy budget constraint. Simulation in [213] indicated that without complete future information, the proposed algorithm for energy-efficient mobility management can still achieve close-to-optimal performance while guaranteeing the long-term energy budget constraint.

There are several research works on the combination of C-RAN MEC systems [206], [207], [222]. In [222], the authors focused on C-RAN MEC systems to minimize energy by the proposed two algorithms, i.e., decentralized local decision algorithm and centralized decision and resource allocation algorithm. To deal with the resource-limited mobile user with computation intensive tasks, C-RAN with MCC was combined to provide high energy efficiency performance [206], in which a joint computational resource and transmit power allocation allocation scheme was proposed to minimize the energy consumption under the constraints of task latency, and fronthaul capacity.

To further enhance the capabilities of mobile devices, C-RAN with MEC was proposed to be combined with each other to efficiently address the increasing mobile traffic issue [207]. Different from previous work, in [207], a resource framework was proposed for power-performance tradeoff of mobile service provider. In this work, Lyapunov technique was exploited to dynamically make online decisions in consecutive time slots for task request. The proposed algorithm can achieve close to optimal performance. In [223], the profit function based on revenue and cost analysis was maximized by jointly optimization of offloading strategy, communication and computation resource. MEC was applied to ultra dense networks (UDNs) [224], where the authors investigated the task offloading policy in MEC-enabled UDN and introduced the software defined networking technology to manage the computation resource in edge cloud with centralized controller. Furthermore, there are other resource allocation schemes were proposed for other C-RAN MEC scenarios, i.e., Vehicular Fog-RANs [225], Near-Far Computing Enhanced C-RAN and [226].

D. LEARNED LESSONS AND POTENTIAL WORKS

Due to the great benefits offered by MEC and H-CRAN, it is envisioned that the combination of MEC and H-CRAN is unavoidable in the future. Although various problems and issues in H-CRAN MEC systems have been intensively studied, there are still several challenges. In the following, we discuss some challenges in dense H-CRAN MEC systems and outline the open research directions.

1) COMPUTATIONAL COMPLEXITY AND SIGNALING OVERHEAD

It is obvious that the centralized optimization is usually easy to implement compared to distributed approaches and can provide the optimal/near-optimal solution with the desired performance guarantee. However, in H-CRAN MEC systems such centralized approaches are not scalable due to the explosive increase in the numbers of mobile users, eNBs, and MEC servers. As a result, there is a need for lightweight and effective algorithms. In these schemes, distributed approaches can offer many benefits as they do not need any central entity and the algorithms are based on only local information or small amounts of signaling overhead. However, it is hard to guarantee the solution optimality with distributed approaches due to the lack of complete information. Therefore, one needs to tradeoff between the computational complexity and solution optimality. An effective way is to decompose the entire network into several regions and assign the responsibility for executing the algorithm to distributed MEC servers, that is the underlying problem is decomposed into subproblems, which are executed distributively at different MEC servers. This would significantly reduce the amount of information which need to be exchanged between the central entity and all users; hence, the network overhead can be also degraded.

2) MOBILITY MANAGEMENT

Ensuring the benefits of mobile users through computation offloading while taking into account user mobility is a challenging issue. Most existing studies in (Het-/CRAN-) MEC systems ignore the effect of user mobility due to its difficulty and intractability. In the proposed H-CRAN MEC systems, users may change their positions while using MEC services, e.g., they can move out of the coverage area of their source MEC servers and are in the serving coverage of other ones. This will result in user association in H-CRAN MEC since the scheduler may need to re-associate the user to a different RRH and then the offloaded task can be calculated by BBU pool with MEC server. In this case, the scheduler (BBU pool) needs to be aware of user mobility in order to maintain service continuity. Thus the dynamic user association and resource allocation can be well studied in the future work. For example, some ML algorithms can be exploited to address the user mobility issue in the resource allocation for H-CRAN MEC. Another potential solution to deal with user mobility is enabling MEC servers to continuously update the user context and then designing context-aware algorithms. Instead of using one-shot optimization, long-term optimization can be used to tackle the challenges of user mobility. To illustrate this point, we consider the following example with a mobile user, which is located far from the MEC server. The short-term optimization for computation offloading decision is not offloading, that is local execution. However, fixing this short-term decision is not always optimal since the user can move to a new position with better channel quality. Moreover, the short-term offloading decision affects not only the instant performance but also the long-term energy budget. In summary, there is a big room for researches into mobility management in dense Het-MEC systems.

3) INTERFERENCE MANAGEMENT AND JOINT RESOURCE ALLOCATION

Inherited from dense HetNets, the spectrum reuse among cells incurs severe mutual interference, which may significantly reduce the expected system spectrum and energy efficiency. Therefore, the challenges for interference management in H-CRAN MEC systems remain to be solved for many reasons. Heterogeneity of mobile users and BBU pool with the MEC server makes the interference problem more challenging due to various transmit power budgets of users in the uplink. Moreover, the network scheduling resource, communication resource and computing resource at BBU pool are coupled with each other, which makes the resource allocation more challenging. The various computation task characteristics require different priorities for users in accessing radio and MEC resources. Finally, interference management is highly coupled with other domains, such as resource allocation and network planning. Hence, more sophisticated interference management schemes incorporating features of H-CRAN MEC systems would be highly required for improving the users' QoS with MEC services.

4) WIRELESS BACKHAUL LIMITATION

In H-CRAN MEC scenarios, the capacity of backhaul and fronthaul is of an important issue. For example, in case that backhaul is limited, the transmission time via backhaul links should be taken into consideration, thus affecting the offloading decisions of users (and other optimization variables as well). Most research works assume that small cells are connected with the central location (where vBBU and MEC servers are located) through high-speed wired links, e.g., fiber links [227]. As a result, the scenario with wired backhaul/fronthaul may be simple and limited to implement for H-CRAN MEC networks, and then discuss their proposed approach in such network settings. The wireless bachhaul and fronhaul can be further investigated to enhance the networks performance. For example, the authors in [228] focused on MEC with wireless backhaul; however, the network setting in this literature is simple, comprising a small-eNB and an MEC server collocated at the macro-eNB. This work is served as a fundamental study for more complex frameworks, e.g., the extension to dense Het-MEC systems and the consideration of mixed wireless and wired backhaul links.

5) PHYSICAL SECURITY

In H-CRAN MEC networks, security will be a significant issue since MEC applications will run on the same physical platforms as some network functions. Therefore, to reduce the risk of that the external eavesdroppers/hackers who may affect the network functions, the physical layer security can be studied for H-CRAN MEC systems, which will be a promising research topic.

VIII. MEC AND MACHINE LEARNING

This section reviews the fundamentals and applications of ML in addressing various MEC problems: edge caching, computation offloading, joint optimization, security and privacy, big data analytics, and mobile crowdsensing. We also identify challenges and potential directions to energize further studies on applications of ML in MEC.

A. A BRIEF REVIEW OF MACHINE LEARNING IN WIRELESS NETWORKS

ML has been applied in a myriad of applications, for example, virtual personal assistants, video surveillance, social media services, email Spam and malware filtering, search engine result refining, and product recommendation. There are several reasons why ML algorithms are increasingly being used: 1) ML enables systems that can automatically adapt and customize themselves to individual users, 2) ML can discover new knowledge from large databases, 3) ML can mimic human and replace certain monotonous tasks, which requires some intelligence, 4) ML can develop systems that are difficult and expensive to construct manually because they require specific detailed skills or knowledge tuned to a specific task, and finally 5) there is a vast increase in computational power, growing progress in available

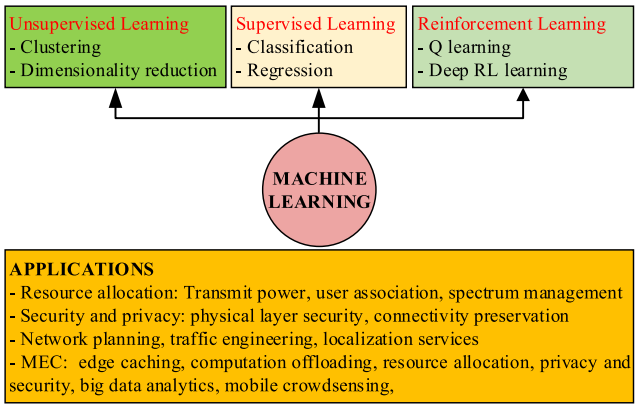

FIGURE 11. Classification and applications of ML in mobile and wireless networking.

algorithms and theory developed by researchers, and increasing support from industries. Generally, ML is divided into three core types: supervised learning, unsupervised learning, and reinforcement learning (RL), while DL has been introduced as a breakthrough technique and a huge step forward in ML, which can achieve higher-level representations based on simpler ones. The classification and applications of ML in mobile and wireless networking, also in MEC and other edge computing paradigms, are illustrated in Fig 11. Recently, the ITU Telecommunication Standardization Sector proposed a unified architecture for ML in future networks, where MEC is expected to play crucial roles as source, collector, pre-processor, model, policy, distributor, and sink [229]. For example, MEC can collect data from end users, then perform data preprocessing, and execute an ML model to extract necessary information before sending the output to the central cloud for further training. Moreover, some surveys and tutorials on ML, DL, (deep) RL, as well as their applications in communications and networking [230]–[232] have come out, and readers can refer to these literature for more details.

Due to the rapid evolution of wireless communications and networks, it is believed that artificial intelligence in general and ML in particular will play vital roles in beyond 5G and 6G [233]. In general, ML can provide the following advantages:

- First, the most natural advantage of ML is the ability to learn from big data to improve the network operation and performance, which can be done without any hand-crafting feature. The importance of learning arises naturally in wireless networks since 1) mobile data is massive, 2) mobile data increases at exponential rates, 3) mobile data is non-stationary (i.e., the time duration for data validity can be relatively short), 4) mobile data quality is not guaranteed (i.e., data collected can be low-quality and noisy), and 5) mobile data is heterogeneous (i.e., data can be generated from many sources, such as mobile users and IoT devices, and in different types) [234].
- Second, the design and optimization of wireless networks are sufficiently challenging without known

channel and mobility models. Conventional optimization techniques are usually performed in an offline, heuristic, or iterative manner, which cannot guarantee the performance optimality or is not suitable for dynamic and time-varying systems. ML is a promising tool such that the network operation can be optimized over time, thus continuously improving the network performance. For example, ML showed a noticeable improvement in uplink data rate by managing uplink interference in cellular networks [235].

- Third, joint 4C optimization in 5G and beyond is immensely complicated due to large state and action spaces, heterogeneous network devices, and various QoS requirements. In such a case, ML is capable of providing online and/or fully-distributed algorithms. Moreover, model-free wireless networks introduce various issues of channel modeling, problem formulation, and closed-form solution, which, however, can be efficiently solved by ML.
- Next, ML should be deployed at the IoT device level and on large-scale distributed networks without violating user data privacy. In 2017, Google introduced an additional ML approach, called ''federated learning'' that enables individual devices collaboratively learn a shared prediction model while keeping their own data locally, thus improving the training efficiency and data privacy. As the network will be highly dense and heterogeneous, federated learning is expected to be a major tool of beyond 5G. Motivated by the application of federated learning in Google board in Android [236], there have been a wide range of applications and problems in wireless networks that can adopt federated learning.
- Last, since edge computing will play an important role in providing low-latency actions and the majority of intelligent applications will be deployed at the network edge, the emergence of edge learning is unavoidable. On the one hand, exploiting edge learning to extract useful information from a massive amount of mobile data can extend the capability of small IoT devices and enable the deployment of compute-intensive and low-latency applications at the edge [237]. On the other hand, edge learning can circumvent drawbacks of cloud AI and on-device AI through the tradeoff between the learning model complexity and the training time [238].

B. MACHINE LEARNING FOR MULTI-ACCESS EDGE COMPUTING

Optimizing MEC faces several challenges of caching placement, allocation of radio and computing resources, assignment of computation tasks, and joint 4C optimization. The existing literature has studied a number of problems in MEC systems, including *computation offloading* [105], [106], [233], [239]–[243], *caching* [244]–[247], *joint 4C optimization* [247]–[252], *security and privacy* [253]–[259], *big data analytics* [234], [257], [260], and *mobile crowd sensing* [261]. In what follows, we summarize the sate-of-the-art related to applications of ML approaches in these aspects.

1) EDGE CACHING

Studies on mobile edge caching have focused on three main issues that are where to cache, what to cache, and how to cache [13], [262]. In terms of caching places, the state-of-the-art showed that the requested content can be cached at macro-eNBs, small-eNBs, and/or end users, where the storage resource of nearby mobile devices is exploited for content caching and D2D communication is used for content retrieving [263]. To decide what to cache, one popular metric is the content popularity, which is defined as the ratio of the number of requests for a particular content to the total number of requests from all users within a specific region during a period of time. The survey paper [13] showed that there are five main algorithms: content replacement policies such as the least frequently used (LFU) and least recently used (LRU), user preference based policies, learning based policies, non-cooperative caching, and cooperative caching.

As the content popularity is time-varying and cannot be known in advance, many studies have focused on ML based caching strategies. Most of the existing works focus on applications of deep RL (DRL) for proactive caching since DRL is able to learn caching policies automatically without any predefined network model and explicit assumption. The authors in [244] explored the key challenges of edge caching and reviewed the state-of-the-art related to learning-based caching policies and algorithms. They showed that mobile edge caching schemes can be classified into two main approaches: *popularity-prediction-based approach*, where the popularity estimation and caching policy are learned separately, and *RL based approach*, where these two terms are learned simultaneously. Other studies on (deep) RL based caching algorithms can be found in [246], [247]. It is a widely held axiom that besides the historical data, the correlation between social and geographic data of mobile users can be utilized to provide more accurate content popularity prediction. Thus, the authors in [264] proposed using big data analytics techniques to advance edge caching designs and proved the effectiveness of these techniques via two case studies of eNB caching and device caching. However, big data analytics, particularly ML/DL mechanisms, has several challenging issues for implementation [265]: huge computation resources required to process the high-dimensional big data, lack of an appropriate prediction model for various types of DL models, optimization of DL parameters, e.g., the depth of deep neural networks and learning rate.

2) COMPUTATION OFFLOADING

Due to the importance of computation offloading from the user perspective, recent years have seen many research works pertaining to computation offloading. In [240], the authors formulated the computation offloading decision problem of a user in ad-hoc mobile clouds as an MDP. More specifically, both channel gains between the user and cloudlets

and the user's and cloudlets' queue states are considered in the system state, the action is the task distribution decision (i.e., how many tasks to process locally and how many tasks to offload to each cloudlet), and the reward function is defined to maximize the user utility and minimize the cost of required payment, energy consumption, delay and, task loss probability. Simulation results showed that the DQN based offloading decision algorithm performed well under various task arrival rates. In [105], the combination of a "hotbooting" Q-learning,[3] computation task queue, user association, and channel gain quality, and the immediate reward is the weighted sum of satisfaction of the task execution delay and computation task drops, the task queuing delay, the penalty of failing to execute a computation task, and the payment of accessing the MEC service. The work in [243] utilized RL to jointly consider traffic and computation offloading for industrial applications in fog computing.

3) JOINT OPTIMIZATION

Due to the facts that 1) the joint 4C optimization is needed for improving the network performance and 2) conventional approaches cannot efficiently solve the optimization problems with large action and state spaces, recent studies on MEC have addressed various problems pertaining to joint 4C optimization. For example, the authors in [247] investigated two deep Q-learning models for mobile edge caching and computing in vehicular networks. To reduce the computational complexity of the original problem and circumvent the high mobility constraint of vehicles, the authors further proposed deploying two DQN models at two distinct timescales. In particular, each epoch is divided into several time slots and then the large timescale deep Q-learning model is executed at every epoch while the small timescale model is performed at every time slot. We note that the concept of multi-timescale control has been applied for some existing research works, e.g., cross-layer optimization [266], [267]. The authors in [248] investigated two learning models, classical Q-learning and DQN method, for joint optimization of offloading decision and computation resource allocation in single-server MEC systems. Since DRL with discretized states suffers from the curse of dimensionality and slow convergence when a high quantization accuracy is required, a continuous control with DRL based framework of computation offloading and resource allocation in wireless powered MEC systems was studied in [249]. As shown in [249, Fig. 3], the proposed algorithm is composed of two alternating phases: i) offloading action generation to quantize the relaxed offloading decision as a set of binary actions, and ii) offloading policy update to select the best offloading action among quantized ones. Similarly, the authors in [268] extended the framework proposed in [249] for multi-carrier NOMA based MEC systems.

More recently, there have been some works that study the joint optimization of computation, caching, and communication. The work in [250] studied the joint optimization of resource allocation in hierarchical networks of fog-enabled IoT with edge caching and computing capability. In [251], the authors proposed an integrated framework of networking, caching, and computing for connected vehicle networks and showed that the proposed DRL based algorithm is superior to the existing static scheme and those without virtualization, MEC offloading, or edge caching. Besides the integration of edge computing, in-network caching, and D2D communication, the literature [252] also took into consideration the social relationships among mobile users so as to improve the reliability and efficiency of resource sharing and delivery in mobile social networks.

4) SECURITY AND PRIVACY

The following reasons explain why security and privacy are the greatest challenges [24]. First, since there are many enabling technologies of MEC, it is necessary to not only protect individual enabling technology, but also orchestrate the diverse security algorithms. Second, the distributed nature of MEC causes many new network situations (e.g., heterogeneous computing capabilities and collaboration between edge devices), which call for new security mechanisms. Third, it is possible that a large-scale edge computing system can be severely affected by the security threats of just a network component. Finally, there are many scenarios and aspects that can be influenced by privacy and security threats, e.g., private data generated by in-car sensors and critical emergency systems. In edge computing paradigms, there are numerous security and privacy threats, for example, wireless jamming, denial of service, man-in-the-middle, spoofing attacks, privacy leakage, virtual machine manipulation, and injection of information [24], [255].

Recently, ML-based security and privacy in MEC have been studied from various perspectives. The use of DL for cyber-attack detection in edge networks was considered in [254], where the experiments demonstrate that the DL based model is better than that with a shallow model in terms of learning accuracy, detection rate, and false alarm rate. The authors in [255] proposed different RL based edge caching security mechanisms of anti-jamming mobile offloading, physical authentication, and friendly jamming. Taking the randomness and variation of wireless channels

[3]The hotbooting Q-learning technique exploits experiences in similar scenarios to initialize the Q-function value so as to save the exploration time at the beginning of learning. and DL was adopted to find the computation offloading decision and offloading rate in IoT with energy harvesting. Another study on computation offloading in IoT with energy harvesting can be found in [259]. The work in [241] formulated the offloading decision problem as a multi-label classification problem and then utilized the deep supervised learning to minimize the computation and offloading overhead. Simulation results demonstrated that the proposed scheme can reduce the system cost in average by 49.24%, 23.87%, 15.69%, and 11.18% compared to the no offloading, random offloading, total offloading, and multi-label linear classifier-based offloading schemes, respectively, and can achieve a higher offloading accuracy. The literature [106] jointly studied the offloading and autoscaling policy (i.e., the number of MEC servers is activated) in energy harvesting MEC systems, which was learned by a post-decision RL algorithm. In [242], the DQN was deployed to learn the offloading decision and energy allocation of a representative mobile user in ultra-dense sliced RAN. The state is characterized by the energy level.

between mobile users and fog nodes, the literature [256] studied Q-learning based physical layer security in fog computing to improve the impersonation detection attack and the accuracy of receivers by learning from the dynamic environment. The work in [258] investigated a new ML based privacy-preserving multifunctional data aggregation framework in order to overcome drawbacks of existing methods, which are high computation overhead, communication efficiency, and single aggregation function calculation. In [259], privacy-aware computation offloading in MEC-enabled IoT was studied, where the post-decision learning is used in conjunction with the standard DQN to accelerate the learning speed.

5) BIG DATA ANALYTICS

As aforementioned, there are three main challenges of mobile big data (MBD) analytics: *large-scale and high-speed mobile networks* which reflect MBD volume and velocity, *portability* which causes MBD volatility, and *crowdsensing* which introduces MBD veracity and variety. Big data analytics enable the design of many smart applications, such as smart city, smart building, and smart manufacturing [269]. Intelligence at the edge is expected to play a major role in data analytics applications. In [234], DL is considered as an attractive solution for MBD analytics by leveraging several advantages: 1) DL scores highly accurate results, 2) DL can automatically generate intrinsic features from MBD, 3) DL does not require labeled samples as the input training data, and 4) multimodal DL allows the learning from heterogeneous data sources. MEC is highly suitable for big data processing. However, there are several challenges [270]: 1) how to distribute big data to distributed resource-finite servers, 2) collaborative MEC for resource sharing and optimization is needed, 3) 4C resources are tightly coupled, and 4) privacy is a critical issue due to the lack of a central management entity.

Some recent studies have utilized ML to address various problems pertaining to MEC big data. The work in [257] divided big data processing into three steps: data collection, aggregation, mining and analysis. Moreover, the authors proposed two privacy-preserving methods, namely output perturbation (OPP) and objective perturbation methods (OJP). In particular, training data privacy can be achieved by adding randomization noise to aggregated query results in the OPP method and to the objective function in the OJP method. Experimental results showed the high accuracy and data utility of OPP and OJP algorithms. In [260], the authors tried to provide users with better QoE in pervasive edge computing environments. The authors first deployed a Tensor-Fast convolutional neural network (TF-CNN) algorithm to guarantee accuracy and increase training speed with big data and next managed high-dimensional big data by using different accurate data transmission rates. It was shown that the proposed TF-CNN algorithm can achieve a higher QoE performance than the state-of-the-art training model.

6) MOBILE CROWDSENSING

While mobile crowdsensing (MCS) has been widely studied in the literature, there are only a handful of studies on edge computing empowered MCS. There are several benefits of MEC in the context of MCS as follows [271]. First, MEC enables the parallelization and partitioning of the centralized and large-scale problem, where MEC servers are responsible for controlling the sensing process on mobile devices located within their deployment area and manage MCS tasks within the same area. Second, the immense computational complexity of the central cloud that is caused by a large number of mobile users participating in MCS tasks with frequent context changes can be greatly reduced because of the distributed deployment of MEC. Third, MEC can reduce the latency of data and information propagation, that is suitable for real-time MCS services. Next, intensive computations can be offloaded from both mobile users and cloud servers to the edge and then being processed therein. Finally, MEC can reduce privacy threats since privacy-sensitive data can be distributed and handled across MEC servers. Recently, the work in [261] proposed a framework that integrates DL and MEC for robust MCS services. In particular, the proposed framework can be implemented by firstly designing an auction mechanism for participant recruitment, then using DL for data validation, and finally implementing data processing at the network edge. In [261], the authors also discussed several open research problems, including how to leverage DL to detect privacy and security threats, how to reduce computational overhead in vastly and rapidly changing environments, and how to implement DL in mobile users for energy and cost efficiency. A hierarchical computing architecture for task allocation was proposed in [272], where the cloud layer does learning of participants' reputation and the edge layer communicates with participants for data collection and optimization.

C. CHALLENGES AND FUTURE WORKS

Clearly, ML techniques will be an important tool for various problems in wireless networks and at the network edge so as to optimize edge caching, computation, enhance big data analytics, and improve security and data privacy. A summary of key problems solved by ML techniques in MEC is presented in Table. 7 along with major challenges. Despite many studies on ML MEC, there are still several key open problems that could be investigated in the future.

1) MACHINE LEARNING BASED FRAMEWORKS OF ULTRA-DENSE MEC SYSTEMS

It is widely expected that both wireless and wired backhaul solutions will coexist in future wireless networks. The simulation results in [228] showed that the bandwidth allocation between wireless access and wireless backhaul plays a major role in the achievable performance. In this case, ML approaches can be deployed at the macro-eNB to predict the appropriate bandwidth partitioning factor based on user

TABLE 7. Summary of key MEC problems that can be solved by machine learning techniques.

Applications	Existing Works	Proposed Framework	Challenges
Edge caching	DRL-based caching strategies	[244]–[247]	- Combination of transfer learning and DRL to exploit knowledge from other domains, e.g., content distribution in mobile social networks can be used to learn caching strategies in D2D. - Tradeoff between exploration and exploitation due to edge dynamics. - Competition and collaboration between caching nodes (e.g. eNB or device caching).
	DL-based caching	[265]	- Determining the suitable model among various types of deep learning models. - Configuring hyperparameter settings.
	Big data analytics based caching	[264]	- Utilization of different features of the previously requested data. - Time-varying and spatio-temporal user behaviors.
Computation offloading	DRL-based computation offloading	[105], [232], [238]–[243]	- Dependence on statistical information of channel quality and task arrival rates. - Time-varying user behaviors and unknown MEC network model.
Joint resource optimization	ML for caching, computation, communication, and control	[247]–[252]	- The complexity of joint optimization problems, and immense action and state spaces due to the combination of couple of different resource types. - Real-time learning training model for time-varying and dynamic MEC systems. - High overhead of signaling transmission and information exchange for Generation of the network state and action spaces, especially in ultra-dense networks.
Privacy and Security	DRL-based privacy and security	[254]	- Lack of massive and high-quality training, validation, and test datasets, which is caused by heterogeneity of wireless networks, mobile devices, and edge nodes. - Limited storage and computation for training DL models.
	DL-based privacy and security	[254], [256], [259]	- Inaccurate and delayed state information, e.g., CSI and energy state information. - Reward function evaluation that is usually estimated according to the security/privacy gain and the protection cost (e.g., computation and communication delay, and energy cost). - Bad security policies at the beginning of learning (the basis of the trial and error methods), which can be effectively addressed by transfer learning techniques. - Tight coupling between privacy and performance gain, thus requiring the optimization of privacy-aware computation offloading and resource allocation schemes.
Big data analytics	ML-based big data processing	[233], [257], [259]	- Storage and computation burdens due to the curse of big data dimensionality. - Tradeoff between the resource-limited MEC servers and the large-scale DL models.
Mobile crowdsensing	DL based MCS	[261]	- Lack of privacy and security protection schemes for crowdsensing data. - High computation overhead for collecting training data, i.e., the DL model requires a large amount of data to retrain the learning model due to edge dynamics. - Lack of efficient DL approaches to be deployed at the lower-tier devices and to detect contaminated and/or fake data.

CSI and task characteristics. Moreover, a critical issue in ultra-dense MEC system is user association and its joint optimization with other aspects such as computation offloading and resource allocation. However, the joint problem of user association, offloading decision, and resource allocation are typically NP-hard non-convex, which are further exacerbated in time-varying and dynamic environments. In such networks, DRL can be used to provide fast and near-optimal solutions.

2) DISTRIBUTED AND COLLABORATIVE ML IMPLEMENTATION IN HIERARCHICAL AND HETEROGENEOUS MEC

The central implementation of ML algorithms faces serious challenges, such as learning complexity, storage and computation resources, and non-suitability for pervasive computing applications and large-scale systems. A potential solution is distributed ML, where the computation of a learning algorithm is divided into smaller parts and then these computations are allocated to distributed MEC servers. However, a number of questions need to be exhaustively answered when distributed ML is used: which computation parts can be divided, how to divide the computation to subtasks, how to synchronize the output among different MEC servers, and how to integrate the outputs from subparts into the output of the master model? Distributed ML becomes particularly important when a learning agent (e.g., MEC server) cannot observe the global state and action, and is merely aware of its local state, reward, and action.

Actually, there is a tradeoff between the computation capability and learning efficiency when ML mechanisms are centralizedly implemented at resource-limited MEC servers. Thus, it is hard to efficiently implement a ML algorithm at MEC server with a very large number of users and an enormous amount of training data. Due to the fact that an artificial neural network (ANN) is composed of many layers (e.g., input, hidden, and output layers) [233], the ANN model and the hierarchical MEC architecture are supposed to fit together, where an immediate layer of the entire ANN model can be offloaded to and performed by MEC layers (e.g., MEC at macro-eNBs and at small-eNBs) and the output of the edge learning is then transferred to higher-tier clouds for further processing. The collaborative learning offers considerable benefits from the reduction of training data size, the exploitation of ubiquitous computing, and the preservation of user data privacy. Moreover, DL approaches can be deployed at the MEC servers to detect contaminated and/or fake data, thus improving the data quality. For instance Li *et al.* in [273] considered a two-layer DL model for video recognition with IoT devices. Due to resource-limited MEC compared, the authors proposed determining the maximum number of computation tasks that can be handled at the edge layer.

3) FEDERATED LEARNING AND APPLICATIONS FOR MEC

Conventional ML approaches are not a suitable way to preserve data privacy. Federated learning leaves the training data distributed across individual users, thus enabling them to collaboratively learn a shared model while keeping their own data locally. Moreover, federated learning is able to address major drawbacks of distributed learning [274], which are 1) lack of time and training data, 2) low performance due to heterogeneous user capabilities and network states, 3) unbalanced number of training data samples, and 4) non-independent and identically distributed data among users. Federated learning is expected to be a sharp tool for various problems in MEC. Take the computation offloading problem as an example, where massive users are trying to offload their computations to an MEC server for remote execution. Conventionally, to determine the offloading decision, users need to report their information such as channel gain, current battery level, and computation characteristics, to the MEC server [242], [249]; however, such information can be revealed by eavesdroppers and can be used illegally to predict the user location. Applying federated learning, each user needs to download the master model from the MEC server and then learns the offloading decision based on its local information only, and the MEC server is merely responsible for updating the master model according to updates from individual users. In such way, federated learning can preserve data privacy and provide distributed offloading decisions, thus being suitable for large-scale MEC systems. Recently, the authors in [275] applied federated learning to estimate the tail distribution of the queues in URLLC vehicle communications and the works in [276] proposed a new adaptive federated learning protocol in heterogeneous MEC systems.

IX. MISCELLANEOUS RESEARCHES

In this section, we first focus on recent open source activities. Then, we look at studies denoted to the testbed and implementation of MEC systems.

A. OPEN SOURCE ACTIVITIES

The ETSI ISG has created a new group, namely Deployment and Ecosystem Development working group (WG DECODE) to accelerate the adoption and implementation of MEC services in the industry.[4] The group is expected to play a leading role in pursuing research activities defined in Phase 2 specifications.

To achieve its objectives, the WG DECODE first exposes MEC descriptions based APIs to increase the adoption of MEC specifications and develop a strong MEC ecosystem. The set of open APIs (e.g., bandwidth management service API and radio network information API) are publicly available at https://forge.etsi.org/rep/mec. Moreover, the WG DECODE promotes the initiation of open source initiatives and facilitates the implementation of open source solutions for MEC applications. For instance, the Open Edge Computing Initiative[5] was introduced in Jun. 2015 by Carnegie Mellon University and industry partners (e.g., Intel, Vodafone, and T-Mobile). Recently, the Open Edge and HPC Initiative[6] was launched in Nov. 2018 by Atos, E4, Forschungszentrum Jʹulich, Fraunhofer FOKUS, Huawei, Mellanox, and SUSE. But the availability of many platforms can cause edge market fragmentation, thus it leads to the interoperability problems and limits the industry collaboration. To circumvent these issues, the Linux Foundation started LF Edge in Jan 2019 to establish an open and interoperable framework, which currently includes five projects: Akraino Edge Stack, EdgeX Foundry, Open Glossary of Edge Computing, Home Edge, and Edge Virtualization Engine.[7] Due to the importance of edge computing, we believe that there will be many more groups and frameworks. More importantly, harmonizing open source platforms for MEC necessitates closer cooperation between ETSI and other edge organizations/standards like Open Edge Computing, LF Edge, OpenFog, and OpenStack in the future.

B. TESTBED AND IMPLEMENTATION

1) SINGLE-BOARD COMPUTER BASED EDGE CLOUD

There are many ways to create an edge server; however, the implementation of single-board computers as edge clouds has been considered as an efficient and cost-effective solution. The increase in popularity of single-board computers (SBCs) (e.g., Raspberry Pi (RPi), Asus Tinker Board S, and Arduino Mega 2560) is due to their low cost, low energy, enough resource for various applications in not only education, but also in industry, hobbyists, prototype builders, and gamers [277], [278]. The availability of SBCs has introduced a new concept, *disposable computing*, such that SBCs are deployed as edge servers at any location where the edge service is not available or the current edge server is discarded and needs to be replaced by a new one. Another advantage is its potential use in emergency applications and security crises. For example, SBCs, built as edge servers, can be used for rescue missions in the area, where the underlying infrastructure has been destroyed by natural disasters, e.g., earthquakes and windstorms.

Elkhatib *et al.* [279] considered the concept of ''micro-cloud'' and examined the suitability and performance trade-offs of RPi-based micro-clouds using four metrics: serving latency, hosting capability, the cost of memory writing/reading, and booting time. Experimental evaluations in [279] demonstrated that RPi clouds can serve a large number of users with low latency and booting time, and can further reduce the cost compared with that of Amazon EC2. In [280], the authors proposed an IoT-edge cloud framework for a smart healthcare information system using SBCs. The authors in [281] implemented an MEC framework with the

[4] Announcement was issued at www.etsi.org/newsroom/press-releases

[5] https://www.openedgecomputing.org/

[6] http://www.open-edge-hpc-initiative.org/

[7] https://www.lfedge.org/

OpenAirInterface[8] and evaluated their prototype framework with a streaming face detection application. Other studies have been conducted to realize SBCs for various applications: fast and accurate object analysis for AR applications [282], real-time image-based object tracking from live videos [283], social sensing applications [284], and latency-aware video analytics [285].

2) LIGHTWEIGHT PLATFORMS FOR EDGE COMPUTING

As MEC and D2D communication are both applications of the offloading concept [9], [286], [287], the authors in [288], [289] proposed different MEC architecture to further improve the network performance compared with the standard MEC. A D2D-based MEC architecture was proposed in [288], where each relay gateway can act as a local cloud. Further, D2D communication is used to establish direct connections between a relay gateway and users so as to provide edge services and between two neighbor relay gateways to balance the traffic and computation demands among them. The work in [289] introduced a concept of ''MEC D2D''. Concretely, *D2D MEC* enables the direct link between users and the MEC server, *neighboring D2D* helps users to connect with the other server if they are not satisfied with the local MEC sever, *cooperative relay* can extend the MEC service, *conventional MEC* provides service to users via the collocated eNB, and *remote cloud* let all users with Internet access use cloud services.

Wang *et al.* [290] proposed a lightweight edge computing platform that is based on SBCs, lightweight virtual switching, and lightweight container virtualization. Taking into account both the QoS requirements of edge services and the deployment cost and status of the underlying hardware, a lightweight platform for service deployment at the network edge was considered in [291]. To evaluate performance of the proposed platform, the authors developed RPis as edge servers and identified a set of the system parameters, such as, the number of services to be deployed and the number of supported users per service. The work in [292] proposed an open carrier interface to offer a fair pay-on-use business model and to provide edge services in a distributed and autonomous manner.

3) MIDDLEWARE FOR EDGE COMPUTING

The very first context-adaptive middleware, named CloudAware, for computation offloading was proposed in [293]. CloudAware is able to predict arbitrary context attributes, thus supporting a wide range of applications with dynamics of the underlying network. The evaluation showed that compared with local computing only, CloudAware can reduce the execution time by 276% while maintaining the same level of offloading success rate. More recently, there have been a number of other studies on messaging middleware for edge computing applications. The middleware investigated in [294] optimized diverse user QoS requirements and orchestrated connections between users and brokers, [295] leveraged SDN to monitor network conditions for resilient data exchange of mission-critical applications, and the messaging middleware proposed in [296] enabled the development and deployment of emerging applications in distributed and heterogeneous edge computing systems.

In [297], the author proposed a middlebox approach to implement the MEC paradigm in 4G LTE networks. Some critical issues are needed to implement the proposed approach without the need to modify the underlying infrastructure: 1) how to intercept and forward the data packets, 2) how to serve the data packets by the MEC servers, 3) how to redirect data traffic to the MEC servers and to the centralized clouds, and 4) how to identify the tunnel for specific users? To solve these issues, the authors in [297] proposed implementing the MEC middlebox between the LTE eNB and the core network, and utilizing some novel design principles, for example, tunnel stateful tracking and traffic redirection.

X. CONCLUSION AND DISCUSSION

This paper covers both fundamentals of MEC and a review of up-to-date research on ''integration of MEC with the forthcoming 5G technologies''. In each section, we have presented a brief background, motivations, and overview in combining the corresponding individual technology in MEC systems. Moreover, we have outlined and discussed the lessons learned, open challenges, and future directions. A number of lessons have been learned from this survey paper:

- There have been enormous efforts from academia and industry to realize MEC as the key enabler for applications and services (e.g., V2X, Tactile Internet, AR/VR, and big data) in the 5G and beyond network. MEC provides a great number of opportunities and potentials; however, some challenges exist and need to be further studied and tackled, e.g., distributed resource management, reliability and mobility, network integration and application portability, the coexistence of heterogeneous (i.e., H2H and MEC) traffic, data privacy, and security.
- There are three main types of MEC use cases: consumer-oriented services, operator and third-party services, and network performance and QoE improvements. To support these categorizations, the integration of MEC with the key enabling technologies in the 5G and beyond network is essential. Moreover, to enable a seamless integration of MEC into the 5G network architecture, the 3GPP has introduced several new functional enablers, namely user plane (re)selection, data network interface, local routing and traffic steering, session and service continuity, network capability expose, and QoS and charging.
- By integrating with other 5G technologies, MEC systems can support massive IoT (NOMA), maintain the system self-sustainability and self-sufficiency (ET and WPT), improve the network performance, adaptability, and scalability (ML), improve the connectivity and coverage of terrestrial cellular networks (UAV), and help

[8] http://www.openairinterface.org/

service/infrastructure providers make the economics of MEC services (collocation with C-RAN).
- To accelerate the adoption of MEC services, the ETSI ISG has defined and exposed a set of open APIs, and further participated in open source activities. Moreover, there have been many efforts and solutions for MEC testbeds and implementation.

For the sake of achieving the seamless integration of MEC in the 5G and beyond network, a number of potential works have been given before. Here, we outline some open problems and challenges which need to be further studied and tackled.

- *Higher-Level Integration*: Although existing research integrates MEC with several enabling technologies, in fact, they are completely independent of each other. Therefore, it is possible to combine more than one of these technologies into a single MEC system. For example, IoT devices first harvest energy from a power source and then follow the NOMA principle to offload their computation tasks to a flying BS equipped with computing capability, where a DRL model is trained to determine the UAV's trajectory and adapt to the underlying dynamic network.
- *Coexistence of Multiple MEC Designs*: This issue becomes crucial when a number of proposals for the same problem of MEC systems are simultaneously proposed, e.g., offloading decision and resource allocation. There has been no answer for how different proposals can be integrated into a unique framework. One possible solution to overcome this issue is that different proposals are classified to find their common viewpoints and then a standard solution should be investigated to support MEC systems with these viewpoints.
- *More Opportunities and Challenges from 6G*: While the 5G standards are not well established yet, there have been some speculative studies for 6G wireless systems to circumvent limitations of the 5G network. For example, a wireless system must support ultra reliability, low latency, high data rate simultaneously, which cannot be fulfilled in the 5G system [298]. It is expected that 6G will include new use cases like haptic communications for eXtended Reality (XR) services, massive IoT for smart city applications, automation and manufacturing. To support these new services, various promising technologies have been speculated and discussed recently, including pervasive and collective AI, radar-enabled communications, metamaterials and intelligent structures, cell-free networks, visible light communication, quantum computing and communications, and tiny cells with THz spectrum [298]. It is inevitable that besides many more use cases and scenarios, new 6G technologies and application requirements also introduce hurdles in MEC and tremendous efforts need to be paid in the future.
- *More challenges and opportunities with distributed learning and FL*: To cope with stringent security requirements, data privacy concerns, massive connectivity, and network heterogeneity, enabling learning techniques (e.g., distributed and FL) in mobile edge networks is of crucial importance. Despite their considerable advantages, there are still many challenges and issues. In recent review articles [299], several challenges and issues of deploying FL in mobile edge networks are discussed, which include, participant selection, trade-off between privacy protection level and system performance, beyond supervised learning, interference management, communication security, incentive mechanism designs, and asynchronous FL approaches. Moreover, promising research directions, e.g., convergence guarantees for the non-convex loss function, heterogeneity diagnostics, and mobile crowdsensing for FL are outlined. In summary, providing solutions to these problems and enabling more applications of FL in MEC systems require interdisciplinary efforts from a variety of research communities.

We strongly believe that this survey can help the readers to deeply understand MEC and its interactions with the enabling technologies in 5G and beyond. We also hope that this survey will stimulate further 5G and MEC research activities.

REFERENCES

[1] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, 4th Quart., 2017.
[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
[3] Cisco. (Feb. 2019) *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022, White Paper*. [Online]. Available: www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-netw% orking-index-vni/white-paper-c11-741490.html
[4] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, "A survey of mobile cloud computing: Architecture, applications, and approaches," *Wireless Commun. Mobile Comput.*, vol. 13, no. 18, pp. 1587–1611, Dec. 2013.
[5] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, "The case for VM-based cloudlets in mobile computing," *IEEE Pervas. Comput.*, vol. 8, no. 4, pp. 14–23, Oct. 2009.
[6] S. Barbarossa, S. Sardellitti, and P. Di Lorenzo, "Communicating while computing: Distributed mobile cloud computing over 5G heterogeneous networks," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 45–55, Nov. 2014.
[7] M. Chiang and T. Zhang, "Fog and IoT: An overview of research opportunities," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 854–864, Dec. 2016.
[8] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proc. 1st Workshop Mobile cloud Comput. MCC*, 2012, pp. 13–16.
[9] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1628–1656, 3rd Quart., 2017.
[10] M. Mukherjee, L. Shu, and D. Wang, "Survey of fog computing: Fundamental, network applications, and research challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1826–1857, 3rd Quart., 2018.
[11] C. Jiang, X. Cheng, H. Gao, X. Zhou, and J. Wan, "Toward computation offloading in edge computing: A survey," *IEEE Access*, vol. 7, pp. 131543–131558, 2019.
[12] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1657–1681, 3rd Quart., 2017.
[13] 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.

[14] C. Wang, Y. He, F. R. Yu, Q. Chen, and L. Tang, "Integration of networking, caching, and computing in wireless systems: A survey, some research issues, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 7–38, 1st Quart., 2018.
[15] H. Wu, "Multi-objective decision-making for mobile cloud offloading: A survey," *IEEE Access*, vol. 6, pp. 3962–3976, 2018.
[16] J. Moura and D. Hutchison, "Game theory for multi-access edge computing: Survey, use cases, and future trends," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 260–288, 1st Quart., 2019.
[17] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 450–465, Feb. 2018.
[18] W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, and A. Ahmed, "Edge computing: A survey," *Future Gener. Comput. Syst.*, vol. 97, pp. 219–235, Aug. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X18319903
[19] H. Liu, F. Eldarrat, H. Alqahtani, A. Reznik, X. de Foy, and Y. Zhang, "Mobile edge cloud system: Architectures, challenges, and approaches," *IEEE Syst. J.*, vol. 12, no. 3, pp. 2495–2508, Sep. 2018.
[20] Y. Ai, M. Peng, and K. Zhang, "Edge computing technologies for Internet of Things: A primer," *Digit. Commun. Netw.*, vol. 4, no. 2, pp. 77–86, 2018.
[21] P. Porambage, J. Okwuibe, M. Liyanage, M. Ylianttila, and T. Taleb, "Survey on multi-access edge computing for Internet of Things realization," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2961–2991, 4th Quart., 2018.
[22] G. Premsankar, M. Di Francesco, and T. Taleb, "Edge computing for the Internet of Things: A case study," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1275–1284, Apr. 2018.
[23] W. Yu, F. Liang, X. He, W. Grant Hatcher, C. Lu, J. Lin, and X. Yang, "A survey on the edge computing for the Internet of Things," *IEEE Access*, vol. 6, pp. 6900–6919, 2018.
[24] R. Roman, J. Lopez, and M. Mambo, "Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges," *Future Gener. Comput. Syst.*, vol. 78, pp. 680–698, Jan. 2018.
[25] N. Makitalo, A. Ometov, J. Kannisto, S. Andreev, Y. Koucheryavy, and T. Mikkonen, "Safe, secure executions at the network edge: Coordinating cloud, edge, and fog computing," *IEEE Softw.*, vol. 35, no. 1, pp. 30–37, Jan. 2018.
[26] S. N. Shirazi, A. Gouglidis, A. Farshad, and D. Hutchison, "The extended cloud: Review and analysis of mobile edge computing and fog from a security and resilience perspective," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 11, pp. 2586–2595, Nov. 2017.
[27] M. Patel, B. Naughton, C. Chan, N. Sprecher, S. Abeta, and A. Neal, "Mobile-edge computing introductory technical white paper," ETSI White Paper, Sep. 2014. [Online]. Available: https://portal.etsi.org/Portals/0/TBpages/MEC/Docs/Mobile-edge_Computing_-_Introductory_Technical_White_Paper_V1%2018-09-14.pdf
[28] S. Kekki, W. Featherstone, Y. Fang, P. Kuure, A. Li, A. Ranjan, D. Purkayastha, F. Jiangping, D. Frydman, G. Verin, K.-W. Wen, K. Kim, R. Arora, A. Odgers, L. M. Contreras, and S. Scarpina, "MEC in 5G networks," *ETSI White Paper*, no. 28, pp. 1–28, Jun. 2018.
[29] *Multi-Access Edge Computing (MEC); Phase 2: Use Cases and Requirements*, Standard ETSI GS MEC 002 V2.1.1, Oct. 2018.
[30] 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, Sep. 2015.
[31] Q.-V. Pham and W.-J. Hwang, "Resource allocation for heterogeneous traffic in complex communication networks," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 63, no. 10, pp. 959–963, Oct. 2016.
[32] X. Lyu, H. Tian, C. Sengul, and P. Zhang, "Multiuser joint task offloading and resource optimization in proximate clouds," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3435–3447, Apr. 2017.
[33] Q.-V. Pham, T. Leanh, N. H. Tran, B. J. Park, and C. S. Hong, "Decentralized computation offloading and resource allocation for mobile-edge computing: A matching game approach," *IEEE Access*, vol. 6, pp. 75868–75885, 2018.
[34] Q.-V. Pham and W.-J. Hwang, "Fairness-aware spectral and energy efficiency in spectrum-sharing wireless networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10207–10219, Nov. 2017.
[35] X. Ge, H. Cheng, M. Guizani, and T. Han, "5G wireless backhaul networks: Challenges and research advances," *IEEE Netw.*, vol. 28, no. 6, pp. 6–11, Nov. 2014.
[36] N. Woon Sung, N.-T. Pham, T. Huynh, and W.-J. Hwang, "Predictive association control for frequent handover avoidance in femtocell networks," *IEEE Commun. Lett.*, vol. 17, no. 5, pp. 924–927, May 2013.
[37] Y. Dong, Z. Chen, P. Fan, and K. B. Letaief, "Mobility-aware uplink interference model for 5G heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2231–2244, Mar. 2016.
[38] G. Brown, "Ultra-reliable low-latency 5G for industrial automation," Qualcomm, San Diego, CA, USA, Tech. Rep. [Online]. Available: https://www.qualcomm.com/media/documents/files/read-the-white-paper-by-heavy-reading.pdf
[39] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai, "A survey on low latency towards 5G: RAN, core network and caching solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3098–3130, 4th Quart., 2018.
[40] A. Nasrallah, A. S. Thyagaturu, Z. Alharbi, C. Wang, X. Shao, M. Reisslein, and H. ElBakoury, "Ultra-low latency (ULL) networks: The IEEE TSN and IETF DetNet standards and related 5G ULL research," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 88–145, 1st Quart., 2019.
[41] L. N. T. Huynh, Q.-V. Pham, X.-Q. Pham, T. D. T. Nguyen, M. D. Hossain, and E.-N. Huh, "Efficient computation offloading in multi-tier multi-access edge computing systems: A particle swarm optimization approach," *Appl. Sci.*, vol. 10, no. 1, pp. 1–17, 2020.
[42] L. Zhang, K. Wang, D. Xuan, and K. Yang, "Optimal task allocation in near-far computing enhanced C-RAN for wireless big data processing," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 50–55, Feb. 2018.
[43] M. Levesque, F. Aurzada, M. Maier, and G. Joos, "Coexistence analysis of H2H and M2M traffic in FiWi smart grid communications infrastructures based on multi-tier business models," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3931–3942, Nov. 2014.
[44] N. Abuzainab, W. Saad, C. S. Hong, and H. V. Poor, "Cognitive hierarchy theory for distributed resource allocation in the Internet of Things," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 7687–7702, Dec. 2017.
[45] E. Ahmed and M. H. Rehmani, "Mobile edge computing: Opportunities, solutions, and challenges," *Future Gener. Comput. Syst.*, vol. 70, pp. 59–63, May 2017.
[46] R. Khan, P. Kumar, D. N. K. Jayakody, and M. Liyanage, "A survey on security and privacy of 5G technologies: Potential solutions, recent advancements, and future directions," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 196–248, 1st Quart., 2020.
[47] D. C Nguyen, P. N Pathirana, M. Ding, and A. Seneviratne, "Integration of blockchain and cloud of things: Architecture, applications and challenges," 2019, *arXiv:1908.09058*. [Online]. Available: http://arxiv.org/abs/1908.09058
[48] *Multi-Access Edge Computing (MEC); Framework and Reference Architecture*, Standard ETSI GS MEC 003 v2.1.1, Jan. 2019.
[49] *Mobile-Edge Computing (MEC); Deployment of Mobile Edge Computing in an NFV Environment*, Standard ETSI GR MEC 017 V1.1.1, Feb. 2018.
[50] A. Reznik, L. M. C. Murillo, Y. Fang, W. Featherstone, M. Filippou, F. Fontes, F. Giust, Q. Huang, A. Li, C. Turyagyenda, C. Wehner, and Z. Zheng, "Cloud RAN and MEC: A perfect pairing," *ETSI White Paper*, no. 22, pp. 1–24, Feb. 2018.
[51] *3GPP Technical Specification Group Services and System Aspects; System Architecture for the 5G System*, Standard 3GPP TS 23.501 v16.1.0, Jun. 2019.
[52] J.-W. Ryu, Q.-V. Pham, H. N. T. Luan, W.-J. Hwang, J.-D. Kim, and J.-T. Lee, "Multi-access edge computing empowered heterogeneous networks: A novel architecture and potential works," *Symmetry*, vol. 11, no. 7, p. 842, Jul. 2019.
[53] A. C. Baktir, A. Ozgovde, and C. Ersoy, "How can edge computing benefit from software-defined networking: A survey, use cases, and future directions," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2359–2391, 4th Quart., 2017.
[54] D. C Nguyen, P. N Pathirana, M. Ding, and A. Seneviratne, "Blockchain for 5G and beyond networks: A state of the art survey," *J. Netw. Comput. Appl.*, vol. 166, Sep. 2020, Art. no. 102693.
[55] P. Si, H. Liang, W. Wu, and Y. Zhang, "Joint resource management in cognitive radio and edge computing based industrial wireless networks," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–6.
[56] Q.-V. Pham and W.-J. Hwang, ""α-fair resource allocation in non-orthogonal multiple access systems," *IET Commun.*, vol. 12, no. 2, pp. 179–183, Jan. 2018.
[57] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple Access Techn. for 5G Wireless Networks and Beyond*. Cham, Switzerland: Springer, 2019.

[58] W. Shin, M. Vaezi, B. Lee, D. J. Love, J. Lee, and H. V. Poor, "Non-orthogonal multiple access in multi-cell networks: Theory, performance, and practical challenges," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 176–183, Oct. 2017.

[59] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[60] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 3rd Quart., 2018.

[61] M. Vaezi, G. A. A. Baduge, Y. Liu, A. Arafa, F. Fang, and Z. Ding, "Interplay between NOMA and other emerging technologies: A survey," *IEEE Trans. Cognit. Commun. Netw.*, vol. 5, no. 4, pp. 900–919, Dec. 2019.

[62] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[63] F. Fang, J. Cheng, and Z. Ding, "Joint energy efficient subchannel and power optimization for a downlink NOMA heterogeneous network," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1351–1364, Feb. 2019.

[64] Q.-V. Pham, T. Huynh-The, M. Alazab, J. Zhao, and W.-J. Hwang, "Sum-rate maximization for UAV-assisted visible light communications using NOMA: Swarm intelligence meets machine learning," *IEEE Internet Things J.*, early access, Apr. 21, 2020, doi: 10.1109/JIOT.2020.2988930.

[65] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[66] Z. Ding, P. Fan, and H. V. Poor, "Impact of non-orthogonal multiple access on the offloading of mobile edge computing," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 375–390, Jan. 2019.

[67] T. X. Tran, A. Hajisami, P. Pandey, and D. Pompili, "Collaborative mobile edge computing in 5G networks: New paradigms, scenarios, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 4, pp. 54–61, Apr. 2017.

[68] F. Wang, J. Xu, and Z. Ding, "Optimized multiuser computation offloading with multi-antenna NOMA," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2017, pp. 1–7.

[69] Y. Pan, M. Chen, Z. Yang, N. Huang, and M. Shikh-Bahaei, "Energy-efficient NOMA-based mobile edge computing offloading," *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 310–313, Feb. 2019.

[70] Z. Ding, J. Xu, O. A. Dobre, and H. V. Poor, "Joint power and time allocation for NOMA–MEC offloading," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6207–6211, Jun. 2019.

[71] A. Kiani and N. Ansari, "Edge computing aware NOMA for 5G networks," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1299–1306, Apr. 2018.

[72] X. Cao, F. Wang, J. Xu, R. Zhang, and S. Cui, "Joint computation and communication cooperation for energy-efficient mobile edge computing," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4188–4200, Jun. 2019.

[73] L. P. Qian, A. Feng, Y. Huang, Y. Wu, B. Ji, and Z. Shi, "Optimal SIC ordering and computation resource allocation in MEC-aware NOMA NB-IoT networks," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2806–2816, Apr. 2019.

[74] F. Fang, Y. Xu, Z. Ding, C. Shen, M. Peng and G. K. Karagiannidis, "Optimal task partition and power allocation for mobile edge computing with NOMA," in *Proc. IEEE Globecom*, Dec. 2019, pp. 1–6.

[75] Y. Wu, L. P. Qian, K. Ni, C. Zhang, and X. Shen, "Delay-minimization nonorthogonal multiple access enabled multi-user mobile edge computation offloading," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 392–407, Jun. 2019.

[76] Z. Ding, D. W. K. Ng, R. Schober, and H. V. Poor, "Delay minimization for NOMA-MEC offloading," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1875–1879, Dec. 2018.

[77] Q.-V. Pham, H. T. Nguyen, Z. Han, and W.-J. Hwang, "Coalitional games for computation offloading in NOMA-enabled multi-access edge computing," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1982–1993, Feb. 2020.

[78] J. Zhu, J. Wang, Y. Huang, F. Fang, K. Navaie, and Z. Ding, "Resource allocation for hybrid NOMA MEC offloading," *IEEE Trans. Wireless Commun.*, early access, Apr. 27, 2020, doi: 10.1109/TWC.2020.2988532.

[79] M. Sheng, Y. Dai, J. Liu, N. Cheng, X. Shen, and Q. Yang, "Delay-aware computation offloading in NOMA MEC under differentiated uploading delay," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2813–2826, Apr. 2020.

[80] Y. Cheng, Z. Liu, Q. Chen, and C. Liang, "Edge computing and power control in NOMA–enabled cognitive radio networks," *Trans. Emerg. Telecommun. Technol.*, vol. 31, no. 3, Mar. 2020, Art. no. e3842.

[81] A. Chen, Z. Yang, B. Lyu, and B. Xu, "System delay minimization for NOMA-based cognitive mobile edge computing," *IEEE Access*, vol. 8, pp. 62228–62237, 2020.

[82] F. Fang, H. Zhang, J. Cheng, and V. C. M. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3722–3732, Sep. 2016.

[83] Q.-V. Pham, S. Mirjalili, N. Kumar, M. Alazab, and W.-J. Hwang, "Whale optimization algorithm with applications to resource allocation in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4285–4297, Apr. 2020.

[84] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, "On the performance of non-orthogonal multiple access systems with partial channel information," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 654–667, Feb. 2016.

[85] F. Fang, H. Zhang, J. Cheng, S. Roy, and V. C. M. Leung, "Joint user scheduling and power allocation optimization for energy-efficient NOMA systems with imperfect CSI," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2874–2885, Dec. 2017.

[86] Y. Zhang, H.-M. Wang, Q. Yang, and Z. Ding, "Secrecy sum rate maximization in non-orthogonal multiple access," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 930–933, May 2016.

[87] Y. Liu, Z. Qin, M. Elkashlan, Y. Gao, and L. Hanzo, "Enhancing the physical layer security of non-orthogonal multiple access in large-scale networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1656–1672, Mar. 2017.

[88] Z. Ding, H. Dai, and H. Vincent Poor, "Relay selection for cooperative NOMA," *IEEE Wireless Commun. Lett.*, vol. 5, no. 4, pp. 416–419, Aug. 2016.

[89] G. Liu, X. Chen, Z. Ding, Z. Ma, and F. R. Yu, "Hybrid half-duplex/full-duplex cooperative non-orthogonal multiple access with transmit power adaptation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 506–519, Jan. 2018.

[90] S. Mukherjee and J. Lee, "Edge computing-enabled cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2884–2899, Apr. 2020.

[91] C. Zhao, Y. Cai, A. Liu, M. Zhao, and L. Hanzo, "Mobile edge computing meets mmWave communications: Joint beamforming and resource allocation for system delay minimization," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2382–2396, Apr. 2020.

[92] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, "A survey of NOMA: Current status and open research challenges," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 179–189, 2020.

[93] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting multiple access for downlink communication systems: Bridging, generalizing, and outperforming SDMA and NOMA," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, p. 133, May 2018.

[94] M. Imran, L. U. Khan, I. Yaqoob, E. Ahmed, M. A. Qureshi, and A. Ahmed, "Energy harvesting in 5G networks: Taxonomy, requirements, challenges, and future directions," 2019, *arXiv:1910.00785*. [Online]. Available: http://arxiv.org/abs/1910.00785

[95] D. Ma, G. Lan, M. Hassan, W. Hu, and S. K. Das, "Sensing, computing, and communications for energy harvesting IoTs: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 1222–1250, 2nd Quart., 2020.

[96] F. Unlu and L. Wawrla. (Jul. 2018). *Energy Harvesting Technologies for IoT Edge Devices*. Accessed: Nov. 17, 2019]. [Online]. Available: https://www.iea-4e.org/document/417/energy-harvesting-technologies-for-%iot-edge-devices

[97] Y. Liu, Y. Zhang, R. Yu, and S. Xie, "Integrated energy and spectrum harvesting for 5G wireless communications," *IEEE Netw.*, vol. 29, no. 3, pp. 75–81, May 2015.

[98] X. Lu, P. Wang, D. Niyato, D. I. Kim, and Z. Han, "Wireless networks with RF energy harvesting: A contemporary survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 757–789, 2nd Quart., 2015.

[99] W. Lumpkins, "Nikola Tesla's dream realized: Wireless power energy harvesting.," *IEEE Consum. Electron. Mag.*, vol. 3, no. 1, pp. 39–42, Jan. 2014.

[100] C. R. Valenta and G. D. Durgin, "Harvesting wireless power: Survey of energy-harvester conversion efficiency in far-field, wireless power transfer systems," *IEEE Microw. Mag.*, vol. 15, no. 4, pp. 108–120, Jun. 2014.

[101] D. W. K. Ng, T. Q. Duong, C. Zhong, and R. Schober, *Wireless Information and Power Transfer: Theory and Practice*. Hoboken, NJ, USA: Wiley, 2019.

[102] X. Zhou, R. Zhang, and C. K. Ho, "Wireless information and power transfer: Architecture design and rate-energy tradeoff," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4754–4767, Nov. 2013.

[103] G. Piro, M. Miozzo, G. Forte, N. Baldo, L. A. Grieco, G. Boggia, and P. Dini, "HetNets powered by renewable energy sources: Sustainable next-generation cellular networks," *IEEE Internet Comput.*, vol. 17, no. 1, pp. 32–39, Jan. 2013.

[104] B. Clerckx, R. Zhang, R. Schober, D. W. K. Ng, D. I. Kim, and H. V. Poor, "Fundamentals of wireless information and power transfer: From RF energy harvester models to signal and system designs," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 4–33, Jan. 2019.

[105] M. Min, L. Xiao, Y. Chen, P. Cheng, D. Wu, and W. Zhuang, "Learning-based computation offloading for IoT devices with energy harvesting," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1930–1941, Feb. 2019.

[106] J. Xu, L. Chen, and S. Ren, "Online learning for offloading and autoscaling in energy harvesting mobile edge computing," *IEEE Trans. Cognit. Commun. Netw.*, vol. 3, no. 3, pp. 361–373, Sep. 2017.

[107] V. Balasubramanian, N. Kouvelas, K. Chandra, R. V. Prasad, A. G. Voyiatzis, and W. Liu, "A unified architecture for integrating energy harvesting IoT devices with the mobile edge cloud," in *Proc. IEEE 4th World Forum Internet Things (WF-IoT)*, Feb. 2018, pp. 13–18.

[108] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: Survey and implications," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 3, pp. 443–461, 3rd Quart., 2011.

[109] W. Chen, D. Wang, and K. Li, "Multi-user multi-task computation offloading in green mobile edge cloud computing," *IEEE Trans. Services Comput.*, vol. 12, no. 5, pp. 726–738, Sep. 2019.

[110] G. Zhang, W. Zhang, Y. Cao, D. Li, and L. Wang, "Energy-delay tradeoff for dynamic offloading in mobile-edge computing system with energy harvesting devices," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4642–4655, Oct. 2018.

[111] S. Ulukus, A. Yener, E. Erkip, O. Simeone, M. Zorzi, P. Grover, and K. Huang, "Energy harvesting wireless communications: A review of recent advances," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 3, pp. 360–381, Mar. 2015.

[112] H. Wu, L. Chen, C. Shen, W. Wen, and J. Xu, "Online geographical load balancing for energy-harvesting mobile edge computing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

[113] F. Guo, L. Ma, H. Zhang, H. Ji, and X. Li, "Joint load management and resource allocation in the energy harvesting powered small cell networks with mobile edge computing," in *Proc. IEEE INFOCOM Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2018, pp. 299–304.

[114] S. Bi and Y. J. Zhang, "Computation rate maximization for wireless powered mobile-edge computing with binary computation offloading," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4177–4190, Jun. 2018.

[115] Y. Dong, Z. Chen, and P. Fan, "Timely two-way data exchanging in unilaterally powered fog computing systems," *IEEE Access*, vol. 7, pp. 21103–21117, 2019.

[116] L. Ji and S. Guo, "Energy-efficient cooperative resource allocation in wireless powered mobile edge computing," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4744–4754, Jun. 2019.

[117] D. Wu, F. Wang, X. Cao, and J. Xu, "Wireless powered user cooperative computation in mobile edge computing systems," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2018, pp. 1–7.

[118] H. Zheng, K. Xiong, P. Fan, Z. Zhong, and K. B. Letaief, "Fog-assisted multiuser SWIPT networks: Local computing or offloading," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5246–5264, Jun. 2019.

[119] N. Janatian, I. Stupia, and L. Vandendorpe, "Optimal resource allocation in ultra-low power fog-computing SWIPT-based networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1–6.

[120] L. Liu, Z. Chang, and X. Guo, "Socially aware dynamic computation offloading scheme for fog computing system with energy harvesting devices," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1869–1879, Jun. 2018.

[121] F. Zhou, Y. Wu, R. Q. Hu, and Y. Qian, "Computation rate maximization in UAV-enabled wireless-powered mobile-edge computing systems," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 9, pp. 1927–1941, Sep. 2018.

[122] B. Liu, W. Li, Y. Ma, J. Wang, and G. Lu, "Wireless powered cognitive-based mobile edge computing with imperfect spectrum sensing," *IEEE Access*, vol. 7, pp. 80431–80442, 2019.

[123] P. Kamalinejad, C. Mahapatra, Z. Sheng, S. Mirabbasi, V. C. M. Leung, and Y. L. Guan, "Wireless energy harvesting for the Internet of Things," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 102–108, Jun. 2015.

[124] R. W. Beard and T. W. McLain, *Small Unmanned Aircraft: Theory and Practice*. Princeton, NJ, USA: Princeton Univ. Press, 2012.

[125] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2334–2360, 3rd Quart., 2019.

[126] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, May 2016.

[127] A. Merwaday and I. Guvenc, "UAV assisted heterogeneous networks for public safety communications," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Mar. 2015, pp. 329–334.

[128] Q. Wu and R. Zhang, "Common throughput maximization in UAV-enabled OFDMA systems with delay consideration," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6614–6627, Dec. 2018.

[129] S. Jeong, O. Simeone, and J. Kang, "Mobile edge computing via a UAV-mounted cloudlet: Optimization of bit allocation and path planning," *IEEE Trans. Veh. Technol.*, vol. 67, no. 3, pp. 2049–2063, Mar. 2018.

[130] F. Zhou, Y. Wu, H. Sun, and Z. Chu, "UAV-enabled mobile edge computing: Offloading optimization and trajectory design," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.

[131] Y. Du, K. Yang, K. Wang, G. Zhang, Y. Zhao, and D. Chen, "Joint resources and workflow scheduling in UAV-enabled wirelessly-powered MEC for IoT systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 10187–10200, Oct. 2019.

[132] L. Fan, W. Yan, X. Chen, Z. Chen, and Q. Shi, "An energy efficient design for UAV communication with mobile edge computing," *China Commun.*, vol. 16, no. 1, pp. 26–36, Jan. 2019.

[133] Q. Hu, Y. Cai, G. Yu, Z. Qin, M. Zhao, and G. Y. Li, "Joint offloading and trajectory design for UAV-enabled mobile edge computing systems," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1879–1892, Apr. 2019.

[134] X. Cao, J. Xu, and R. Zhang, "Mobile edge computing for cellular-connected UAV: Computation offloading and trajectory optimization," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.

[135] T. Bai, J. Wang, Y. Ren, and L. Hanzo, "Energy-efficient computation offloading for secure UAV-Edge-Computing systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6074–6087, Jun. 2019.

[136] O. Bekkouche, T. Taleb, M. Bagaa, and K. Samdanis, "Edge cloud resource-aware flight planning for unmanned aerial vehicles," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–7.

[137] V. R, M. Alazab, S. Srinivasan, Q.-V. Pham, S. K. Padannayil, and K. Simran, "A visualized botnet detection system based deep learning for the Internet of Things networks of smart cities," *IEEE Trans. Ind. Appl.*, early access, Feb. 6, 2020, doi: 10.1109/TIA.2020.2971952.

[138] *Overview Internet Things*, document Recommendation ITU-T Y.2060, International Telecommunication Union, Geneva, Switzerland, 2012.

[139] Y. Gu, Z. Chang, M. Pan, L. Song, and Z. Han, "Joint radio and computational resource allocation in IoT fog computing," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7475–7484, Aug. 2018.

[140] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, "A survey on 5G networks for the Internet of Things: Communication technologies and challenges," *IEEE Access*, vol. 6, pp. 3619–3647, Feb. 2018.

[141] M. Maier, A. Ebrahimzadeh, and M. Chowdhury, "The tactile Internet: Automation or augmentation of the human?" *IEEE Access*, vol. 6, pp. 41607–41618, 2018.

[142] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.

[143] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on Internet of Things: Architecture, enabling technologies, security and privacy, and applications," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1125–1142, Oct. 2017.

[144] B. N. Silva, M. Khan, and K. Han, "Internet of Things: A comprehensive review of enabling technologies, architecture, and challenges," *IETE Tech. Rev.*, vol. 35, no. 2, pp. 205–220, Mar. 2018.

[145] 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, pp. 1–16, Sep. 2015.

[146] Y. Xiao and M. Krunz, "Distributed optimization for energy-efficient fog computing in the tactile Internet," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 11, pp. 2390–2400, Nov. 2018.
[147] T. Leppanen, "Distributed artificial intelligence with multi-agent systems for MEC," in *Proc. 28th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul. 2019, pp. 1–8.
[148] S. S. Gill, P. Garraghan, and R. Buyya, "ROUTER: Fog enabled cloud based intelligent resource management approach for smart home IoT devices," *J. Syst. Softw.*, vol. 154, pp. 125–138, Aug. 2019.
[149] A. Yassine, S. Singh, M. S. Hossain, and G. Muhammad, "IoT big data analytics for smart homes with fog and cloud computing," *Future Gener. Comput. Syst.*, vol. 91, pp. 563–573, Feb. 2019.
[150] A. Awad Abdellatif, A. Emam, C.-F. Chiasserini, A. Mohamed, A. Jaoua, and R. Ward, "Edge-based compression and classification for smart healthcare systems: Concept, implementation and evaluation," *Expert Syst. Appl.*, vol. 117, pp. 1–14, Mar. 2019.
[151] J. Zhao, L. Wang, K.-K. Wong, M. Tao, and T. Mahmoodi, "Energy and latency control for edge computing in dense V2X networks," 2018, *arXiv:1807.02311*. [Online]. Available: http://arxiv.org/abs/1807.02311
[152] W. Sun, J. Liu, Y. Yue, and H. Zhang, "Double auction-based resource allocation for mobile edge computing in industrial Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4692–4701, Oct. 2018.
[153] X. Li, D. Li, J. Wan, C. Liu, and M. Imran, "Adaptive transmission optimization in SDN-based industrial Internet of Things with edge computing," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1351–1360, Jun. 2018.
[154] G. Li, J. Wu, J. Li, K. Wang, and T. Ye, "Service popularity-based smart resources partitioning for fog computing-enabled industrial Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4702–4711, Oct. 2018.
[155] C.-F. Lai, W.-C. Chien, L. T. Yang, and W. Qiang, "LSTM and edge computing for big data feature recognition of industrial electrical equipment," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2469–2477, Apr. 2019.
[156] J. Xu, S. Wang, B. K. Bhargava, and F. Yang, "A blockchain-enabled trustless crowd-intelligence ecosystem on mobile edge computing," *IEEE Trans. Ind. Informat.*, vol. 15, no. 6, pp. 3538–3547, Jun. 2019.
[157] X. Li and J. Wan, "Proactive caching for edge computing-enabled industrial mobile wireless networks," *Future Gener. Comput. Syst.*, vol. 89, pp. 89–97, Dec. 2018.
[158] L. Tao, Z. Li, and L. Wu, "Outlet: Outsourcing wearable computing to the ambient mobile computing edge," *IEEE Access*, vol. 6, pp. 18408–18419, 2018.
[159] X. Yang, Z. Chen, K. Li, Y. Sun, N. Liu, W. Xie, and Y. Zhao, "Communication-constrained mobile edge computing systems for wireless virtual reality: Scheduling and tradeoff," *IEEE Access*, vol. 6, pp. 16665–16677, 2018.
[160] Y. Sun, Z. Chen, M. Tao, and H. Liu, "Communications, caching, and computing for mobile virtual reality: Modeling and tradeoff," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7573–7586, Nov. 2019.
[161] Y. Li and W. Gao, "MUVR: Supporting multi-user mobile virtual reality with resource constrained edge cloud," in *Proc. IEEE/ACM Symp. Edge Comput. (SEC)*, Oct. 2018, pp. 1–16.
[162] Y. Liu, J. Liu, A. Argyriou, and S. Ci, "MEC-assisted panoramic VR video streaming over millimeter wave mobile networks," *IEEE Trans. Multimedia*, vol. 21, no. 5, pp. 1302–1316, May 2019.
[163] D. H. Fan and S. Gao, "The application of mobile edge computing in agricultural water monitoring system," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 191, Nov. 2018, Art. no. 012015.
[164] M. A. Zamora-Izquierdo, J. Santa, J. A. Martínez, V. Martínez, and A. F. Skarmeta, "Smart farming IoT platform based on edge and cloud computing," *Biosyst. Eng.*, vol. 177, pp. 4–17, Jan. 2019.
[165] Y. Liu, C. Yang, L. Jiang, S. Xie, and Y. Zhang, "Intelligent edge computing for IoT-based energy management in smart cities," *IEEE Netw.*, vol. 33, no. 2, pp. 111–117, Mar. 2019.
[166] X. Li, J. Wan, H.-N. Dai, M. Imran, M. Xia, and A. Celesti, "A hybrid computing solution and resource scheduling strategy for edge computing in smart manufacturing," *IEEE Trans. Ind. Informat.*, vol. 15, no. 7, pp. 4225–4234, Jul. 2019.
[167] Z. Samia, R. Khaled, and Z. Warda, "Multi-agent systems and ontology for supporting management system in smart school," in *Proc. 3rd Int. Conf. Pattern Anal. Intell. Syst. (PAIS)*, Oct. 2018, pp. 1–8.
[168] A. Pacheco, P. Cano, E. Flores, E. Trujillo, and P. Marquez, "A smart classroom based on deep learning and osmotic IoT computing," in *Proc. Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI)*, Oct. 2018, pp. 1–5.
[169] A. Y. N. Pratama, A. Zainudin, and M. Yuliana, "Implementation of IoT-based passengers monitoring for smart school application," in *Proc. Int. Electron. Symp. Eng. Technol. Appl. (IES-ETA)*, Sep. 2017, pp. 33–38.
[170] A. Mochamad Rifki Ulil, Fiannurdin, S. Sukaridhoto, A. Tjahjono, and D. K. Basuki, "The vehicle as a mobile sensor network base IoT and big data for pothole detection caused by flood disaster," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 239, Feb. 2019, Art. no. 012034.
[171] N. Bissmeyer, J. V. Dam, C. Zimmermann, and K. Eckert, "Security in hybrid vehicular communication based on ITS-G5, LTE-V, and mobile edge computing," in *Proc. AmE Automot. Meets Electron., 9th GMM-Symp.*, Dortmund, Germany, Mar. 2018, pp. 1–6.
[172] S. Trilles, J. Torres-Sospedra, Ó. Belmonte, F. J. Zarazaga-Soria, A. González-Pérez, and J. Huerta, "Development of an open sensorized platform in a smart agriculture context: A vineyard support system for monitoring mildew disease," *Sustain. Comput., Informat. Syst.*, early access. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S2210537918302270
[173] M. A. Rahman, M. M. Rashid, M. S. Hossain, E. Hassanain, M. F. Alhamid, and M. Guizani, "Blockchain and IoT-based cognitive edge framework for sharing economy services in a smart city," *IEEE Access*, vol. 7, pp. 18611–18621, 2019.
[174] X. Li, X. Huang, C. Li, R. Yu, and L. Shu, "EdgeCare: Leveraging edge computing for collaborative data management in mobile healthcare systems," *IEEE Access*, vol. 7, pp. 22011–22025, 2019.
[175] S. Oueida, Y. Kotb, M. Aloqaily, Y. Jararweh, and T. Baker, "An edge computing based smart healthcare framework for resource management," *Sensors*, vol. 18, no. 12, p. 4307, Dec. 2018.
[176] J. Hochstetler, R. Padidela, Q. Chen, Q. Yang, and S. Fu, "Embedded deep learning for vehicular edge computing," in *Proc. IEEE/ACM Symp. Edge Comput. (SEC)*, Oct. 2018, pp. 341–343.
[177] L. Zhao, J. Wang, J. Liu, and N. Kato, "Routing for crowd management in smart cities: A deep reinforcement learning perspective," *IEEE Commun. Mag.*, vol. 57, no. 4, pp. 88–93, Apr. 2019.
[178] P. Pace, G. Aloi, R. Gravina, G. Caliciuri, G. Fortino, and A. Liotta, "An edge-based architecture to support efficient applications for healthcare industry 4.0," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 481–489, Jan. 2019.
[179] E. E. Ugwuanyi, S. Ghosh, M. Iqbal, and T. Dagiuklas, "Reliable resource provisioning using Bankers' deadlock avoidance algorithm in MEC for industrial IoT," *IEEE Access*, vol. 6, pp. 43327–43335, 2018.
[180] M. Chowdhury and M. Maier, "Collaborative computing for advanced tactile Internet human-to-robot (H2R) communications in integrated FiWi multirobot infrastructures," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 2142–2158, Dec. 2017.
[181] J. Xu, K. Ota, and M. Dong, "Energy efficient hybrid edge caching scheme for tactile Internet in 5G," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 2, pp. 483–493, Jun. 2019.
[182] A. H. Sodhro, S. Pirbhulal, and V. H. C. de Albuquerque, "Artificial intelligence-driven mechanism for edge computing-based industrial applications," *IEEE Trans. Ind. Informat.*, vol. 15, no. 7, pp. 4235–4243, Jul. 2019.
[183] L. U. Khan, I. Yaqoob, N. H. Tran, S. M. A. Kazmi, T. N. Dang, and C. S. Hong, "Edge computing enabled smart cities: A comprehensive survey," *IEEE Internet Things J.*, early access, Apr. 13, 2020, doi: 10.1109/JIOT.2020.2987070.
[184] P. P. Ray, D. Dash, and D. De, "Edge computing for Internet of Things: A survey, e-healthcare case study and future direction," *J. Netw. Comput. Appl.*, vol. 140, pp. 1–22, Aug. 2019.
[185] A. M. Rahmani, T. N. Gia, B. Negash, A. Anzanpour, I. Azimi, M. Jiang, and P. Liljeberg, "Exploiting smart e-health gateways at the edge of healthcare Internet-of-things: A fog computing approach," *Future Gener. Comput. Syst.*, vol. 78, pp. 641–658, Jan. 2018.
[186] 3GPP. (Dec. 2018). *Study on Enhancement of 3GPP Support for 5G V2X Services*. Accessed: May 6, 2019. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDeta%ils.aspx?specificationId=3108
[187] F. Giust, V. Sciancalepore, D. Sabella, M. C. Filippou, S. Mangiante, W. Featherstone, and D. Munaretto, "Multi-access edge computing: The driver behind the wheel of 5G-connected cars," *IEEE Commun. Standards Mag.*, vol. 2, no. 3, pp. 66–73, Sep. 2018.
[188] Y. Kabalci, *5G Mobile Communication Systems: Fundamentals, Challenges, and Key Technologies*. Singapore: Springer, 2019, pp. 329–359.

[189] P. Varga, J. Peto, A. Franko, D. Balla, D. Haja, F. Janky, G. Soos, D. Ficzere, M. Maliosz, and L. Toka, "5G support for industrial IoT applications—Challenges, solutions, and research gaps," *Sensors*, vol. 20, no. 3, p. 828, Feb. 2020.
[190] P. Pop, M. L. Raagaard, M. Gutierrez, and W. Steiner, "Enabling fog computing for industrial automation through time-sensitive networking (TSN)," *IEEE Commun. Standards Mag.*, vol. 2, no. 2, pp. 55–61, Jun. 2018.
[191] M. Erol-Kantarci and S. Sukhmani, "Caching and computing at the edge for mobile augmented reality and virtual reality (AR/VR) in 5G," in *Proc. Int. Conf. Ad Hoc Netw.*, Niagara Falls, ON, Canada, Jan. 2018, pp. 169–177.
[192] A. Jukan, F. Carpio, X. Masip, A. J. Ferrer, N. Kemper, and B. U. Stetina, "Fog-to-cloud computing for farming: Low-cost technologies, data exchange, and animal welfare," *Computer*, vol. 52, no. 10, pp. 41–51, Oct. 2019.
[193] X. Shi, X. An, Q. Zhao, H. Liu, L. Xia, X. Sun, and Y. Guo, "State-of-the-art Internet of Things in protected agriculture," *Sensors*, vol. 19, no. 8, p. 1833, Apr. 2019.
[194] *The Tactile Internet*, document ITU-T Technology Watch Report, Aug. 2014, pp. 1–18.
[195] C. Puliafito, E. Mingozzi, F. Longo, A. Puliafito, and O. Rana, "Fog computing for the Internet of Things: A survey," *ACM Trans. Internet Technol.*, vol. 19, no. 2, p. 18, 2019.
[196] J. Ni, X. Lin, and X. S. Shen, "Toward edge-assisted Internet of Things: From security and efficiency perspectives," *IEEE Netw.*, vol. 33, no. 2, pp. 50–57, Mar. 2019.
[197] L. Zhao, J. Wang, J. Liu, and N. Kato, "Optimal edge resource allocation in IoT-based smart cities," *IEEE Netw.*, vol. 33, no. 2, pp. 30–35, Mar. 2019.
[198] X. Cao, G. Tang, D. Guo, Y. Li, and W. Zhang, "Edge federation: Towards an integrated service provisioning model," 2019, *arXiv:1902.09055*. [Online]. Available: http://arxiv.org/abs/1902.09055
[199] K. Guo, Y. Lu, H. Gao, and R. Cao, "Artificial intelligence-based semantic Internet of Things in a user-centric smart city," *Sensors*, vol. 18, no. 5, p. 1341, Apr. 2018.
[200] Q.-V. Pham and W.-J. Hwang, "Energy-efficient power control for uplink spectrum-sharing heterogeneous networks," *Int. J. Commun. Syst.*, vol. 31, no. 14, p. e3717, Jul. 2018.
[201] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: A survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, Sep. 2008.
[202] C.-L. I, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, "Toward green and soft: A 5G perspective," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 66–73, Feb. 2014.
[203] P. Rost, C. J. Bernardos, A. D. Domenico, M. D. Girolamo, M. Lalam, A. Maeder, D. Sabella, and D. Wübben, "Cloud technologies for flexible 5G radio access networks," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 68–76, May 2014.
[204] Y. Li, T. Jiang, K. Luo, and S. Mao, "Green heterogeneous cloud radio access networks: Potential techniques, performance trade-offs, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 33–39, Nov. 2017.
[205] Y. Teng, M. Liu, F. R. Yu, V. C. M. Leung, M. Song, and Y. Zhang, "Resource allocation for ultra-dense networks: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2134–2168, 3rd Quart., 2019.
[206] K. Wang, K. Yang, and C. S. Magurawalage, "Joint energy minimization and resource allocation in C-RAN with mobile cloud," *IEEE Trans. Cloud Comput.*, vol. 6, no. 3, pp. 760–770, Sep. 2018.
[207] X. Wang, K. Wang, S. Wu, S. Di, H. Jin, K. Yang, and S. Ou, "Dynamic resource scheduling in mobile edge cloud with cloud radio access network," *IEEE Trans. Parallel Distrib. Syst.*, vol. 29, no. 11, pp. 2429–2445, Nov. 2018.
[208] S. Sardellitti, G. Scutari, and S. Barbarossa, "Joint optimization of radio and computational resources for multicell mobile-edge computing," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 1, no. 2, pp. 89–103, Jun. 2015.
[209] A. Al-Shuwaili, O. Simeone, A. Bagheri, and G. Scutari, "Joint uplink/downlink optimization for backhaul-limited mobile cloud computing with user scheduling," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 3, no. 4, pp. 787–802, Dec. 2017.
[210] K. Zhang, Y. Mao, S. Leng, Q. Zhao, L. Li, X. Peng, L. Pan, S. Maharjan, and Y. Zhang, "Energy-efficient offloading for mobile edge computing in 5G heterogeneous networks," *IEEE Access*, vol. 4, pp. 5896–5907, 2016.
[211] J. Zhang, W. Xia, F. Yan, and L. Shen, "Joint computation offloading and resource allocation optimization in heterogeneous networks with mobile edge computing," *IEEE Access*, vol. 6, pp. 19324–19337, 2018.
[212] Y. Zhao, V. C. M. Leung, H. Gao, Z. Chen, and H. Ji, "Uplink resource allocation in mobile edge computing-based heterogeneous networks with multi-band RF energy harvesting," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.
[213] Y. Sun, S. Zhou, and J. Xu, "EMM: Energy-aware mobility management for mobile edge computing in ultra dense networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 11, pp. 2637–2646, Nov. 2017.
[214] H. Wang, S. Chen, M. Ai, and H. Xu, "Localized mobility management for 5G ultra dense network," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8535–8552, Sep. 2017.
[215] O. Semiari, W. Saad, M. Bennis, and B. Maham, "Caching meets millimeter wave communications for enhanced mobility management in 5G networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 779–793, Feb. 2018.
[216] D. Lopez-Perez, I. Guvenc, and X. Chu, "Mobility management challenges in 3GPP heterogeneous networks," *IEEE Commun. Mag.*, vol. 50, no. 12, pp. 70–78, Dec. 2012.
[217] A. Prasad, O. Tirkkonen, P. Lundén, O. Yilmaz, L. Dalsgaard, and C. Wijting, "Energy-efficient inter-frequency small cell discovery techniques for LTE-advanced heterogeneous network deployments," *IEEE Commun. Mag.*, vol. 51, no. 5, pp. 72–81, May 2013.
[218] D. Xenakis, N. Passas, L. Merakos, and C. Verikoukis, "Mobility management for femtocells in LTE-advanced: Key aspects and survey of handover decision algorithms," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 64–91, 1st Quart., 2014.
[219] F. Giust, L. Cominardi, and C. Bernardos, "Distributed mobility management for future 5G networks: Overview and analysis of existing approaches," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 142–149, Jan. 2015.
[220] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, and V. C. M. Leung, "Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 138–145, Aug. 2017.
[221] G. Qiao, S. Leng, K. Zhang, and K. Yang, "Joint deployment and mobility management of energy harvesting small cells in heterogeneous networks," *IEEE Access*, vol. 5, pp. 183–196, 2017.
[222] K. Wang, P.-Q. Huang, K. Yang, C. Pan, and J. Wang, "Unified offloading decision making and resource allocation in ME-RAN," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8159–8172, Aug. 2019.
[223] Z. Jian, W. Muqing, and Z. Min, "Joint computation offloading and resource allocation in C-RAN with MEC based on spectrum efficiency," *IEEE Access*, vol. 7, pp. 79056–79068, 2019.
[224] M. Chen and Y. Hao, "Task offloading for mobile edge computing in software defined ultra-dense network," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 3, pp. 587–597, Mar. 2018.
[225] K. Xiong, S. Leng, J. Hu, X. Chen, and K. Yang, "Smart network slicing for vehicular fog-RANs," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3075–3085, Apr. 2019.
[226] Y. Cai, X. Lu, Y. Luo, K. Wang, D. Chen, and K. Yang, "A task allocation algorithm for profit maximization in NFC-RAN," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2019, pp. 203–207.
[227] P. Shantharama, A. S. Thyagaturu, N. Karakoc, L. Ferrari, M. Reisslein, and A. Scaglione, "LayBack: SDN management of multi-access edge computing (MEC) for network access services and radio resource sharing," *IEEE Access*, vol. 6, pp. 57545–57561, 2018.
[228] Q.-V. Pham, L. B. Le, S.-H. Chung, and W.-J. Hwang, "Mobile edge computing with wireless backhaul: Joint task offloading and resource allocation," *IEEE Access*, vol. 7, pp. 16444–16459, 2019.
[229] *Unified Architecture for Machine Learning in 5G and Future Networks*, document ITU-T FG-ML5G-ARC5G, Jan. 2019.
[230] O. Simeone, "A very brief introduction to machine learning with applications to communication systems," *IEEE Trans. Cognit. Commun. Netw.*, vol. 4, no. 4, pp. 648–664, Dec. 2018.
[231] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017.
[232] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3133–3174, 4th Quart., 2019.

[233] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3039–3071, 4th Quart., 2019.

[234] M. A. Alsheikh, D. Niyato, S. Lin, H.-P. Tan, and Z. Han, "Mobile big data analytics using deep learning and apache spark," *IEEE Netw.*, vol. 30, no. 3, pp. 22–29, May 2016.

[235] S. Deb and P. Monogioudis, "Learning-based uplink interference management in 4G LTE cellular systems," *IEEE/ACM Trans. Netw.*, vol. 23, no. 2, pp. 398–411, Apr. 2015.

[236] *Federated Learning: Collaborative Machine Learning Without Centralized Training Data*. Accessed: May 12, 2020. [Online]. Available: http://ai.googleblog.com/2017/04/federated-learning-collaborative.html

[237] H. Khelifi, S. Luo, B. Nour, A. Sellami, H. Moungla, S. H. Ahmed, and M. Guizani, "Bringing deep learning at the edge of information-centric Internet of Things," *IEEE Commun. Lett.*, vol. 23, no. 1, pp. 52–55, Jan. 2019.

[238] G. Zhu, D. Liu, Y. Du, C. You, J. Zhang, and K. Huang, "Toward an intelligent edge: Wireless communication meets machine learning," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 19–25, Jan. 2020.

[239] T. Yang, Y. Hu, M. C. Gursoy, A. Schmeink, and R. Mathar, "Deep reinforcement learning based resource allocation in low latency edge computing networks," in *Proc. 15th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2018, pp. 1–5.

[240] D. Van Le and C.-K. Tham, "A deep reinforcement learning based offloading scheme in ad-hoc mobile clouds," in *Proc. IEEE INFOCOM Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2018, pp. 760–765.

[241] S. Yu, X. Wang, and R. Langar, "Computation offloading for mobile edge computing: A deep learning approach," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Oct. 2017, pp. 1–6.

[242] X. Chen, H. Zhang, C. Wu, S. Mao, Y. Ji, and M. Bennis, "Optimized computation offloading performance in virtual edge computing systems via deep reinforcement learning," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4005–4018, Jun. 2019.

[243] Y. Wang, K. Wang, H. Huang, T. Miyazaki, and S. Guo, "Traffic and computation co-offloading with reinforcement learning in fog computing for industrial applications," *IEEE Trans. Ind. Informat.*, vol. 15, no. 2, pp. 976–986, Feb. 2019.

[244] H. Zhu, Y. Cao, W. Wang, T. Jiang, and S. Jin, "Deep reinforcement learning for mobile edge caching: Review, new features, and open issues," *IEEE Netw.*, vol. 32, no. 6, pp. 50–57, Nov. 2018.

[245] Y. He, F. R. Yu, N. Zhao, V. C. M. Leung, and H. Yin, "Software-defined networks with mobile edge computing and caching for smart cities: A big data deep reinforcement learning approach," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 31–37, Dec. 2017.

[246] L. Hou, L. Lei, K. Zheng, and X. Wang, "A *Q*-learning-based proactive caching strategy for non-safety related services in vehicular networks," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4512–4520, Jun. 2019.

[247] L. T. Tan and R. Q. Hu, "Mobility-aware edge caching and computing in vehicle networks: A deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10190–10203, Nov. 2018.

[248] J. Li, H. Gao, T. Lv, and Y. Lu, "Deep reinforcement learning based computation offloading and resource allocation for MEC," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1–6.

[249] L. Huang, S. Bi, and Y. J. Zhang, "Deep reinforcement learning for online computation offloading in wireless powered mobile-edge computing networks," *IEEE Trans. Mobile Comput.*, early access, Jul. 24, 2019, doi: 10.1109/TMC.2019.2928811.

[250] Y. Wei, F. R. Yu, M. Song, and Z. Han, "Joint optimization of caching, computing, and radio resources for fog-enabled IoT using natural actor–critic deep reinforcement learning," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2061–2073, Apr. 2019.

[251] Y. He, N. Zhao, and H. Yin, "Integrated networking, caching, and computing for connected vehicles: A deep reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 44–55, Jan. 2018.

[252] Y. He, C. Liang, F. R. Yu, and Z. Han, "Trust-based social networks with computing, caching and communications: A deep reinforcement learning approach," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 66–79, Jan. 2020.

[253] M. Gheisari, Q.-V. Pham, M. Alazab, X. Zhang, C. Fernandez-Campusano, and G. Srivastava, "ECA: An edge computing architecture for privacy-preserving in IoT-based smart city," *IEEE Access*, vol. 7, pp. 155779–155786, 2019.

[254] A. Abeshu and N. Chilamkurti, "Deep learning: The frontier for distributed attack detection in fog-to-things computing," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 169–175, Feb. 2018.

[255] L. Xiao, X. Wan, C. Dai, X. Du, X. Chen, and M. Guizani, "Security in mobile edge caching with reinforcement learning," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 116–122, Jun. 2018.

[256] S. Tu, M. Waqas, S. U. Rehman, M. Aamir, O. U. Rehman, Z. Jianbiao, and C.-C. Chang, "Security in fog computing: A novel technique to tackle an impersonation attack," *IEEE Access*, vol. 6, pp. 74993–75001, 2018.

[257] M. Du, K. Wang, Z. Xia, and Y. Zhang, "Differential privacy preserving of training model in wireless big data with edge computing," *IEEE Trans. Big Data*, vol. 6, no. 2, pp. 283–295, Jun. 2020.

[258] M. Yang, T. Zhu, B. Liu, Y. Xiang, and W. Zhou, "Machine learning differential privacy with multifunctional aggregation in a fog computing architecture," *IEEE Access*, vol. 6, pp. 17119–17129, 2018.

[259] M. Min, X. Wan, L. Xiao, Y. Chen, M. Xia, D. Wu, and H. Dai, "Learning-based privacy-aware offloading for healthcare IoT with energy harvesting," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4307–4316, Jun. 2019.

[260] Q. Meng, K. Wang, and X. He, "QoE-driven big data management in pervasive edge computing environment," *Big Data Mining Analytics*, vol. 1, no. 3, pp. 222–233, 2018.

[261] Z. Zhou, H. Liao, B. Gu, K. M. S. Huq, S. Mumtaz, and J. Rodriguez, "Robust mobile crowd sensing: When deep learning meets edge computing," *IEEE Netw.*, vol. 32, no. 4, pp. 54–60, Jul. 2018.

[262] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, "Cache in the air: Exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, Feb. 2014.

[263] D. Liu, B. Chen, C. Yang, and A. F. Molisch, "Caching at the wireless edge: Design aspects, challenges, and future directions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 22–28, Sep. 2016.

[264] Z. Chang, L. Lei, Z. Zhou, S. Mao, and T. Ristaniemi, "Learn to cache: Machine learning for network edge caching in the big data era," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 28–35, Jun. 2018.

[265] K. Thar, N. H. Tran, T. Z. Oo, and C. S. Hong, "DeepMEC: Mobile edge caching using deep learning," *IEEE Access*, vol. 6, pp. 78260–78275, 2018.

[266] Q.-V. Pham, H.-L. To, and W.-J. Hwang, "A multi-timescale cross-layer approach for wireless ad hoc networks," *Comput. Netw.*, vol. 91, pp. 471–482, 2015.

[267] Q.-V. Pham and W.-J. Hwang, "Network utility maximization-based congestion control over wireless networks: A survey and potential directives," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1173–1200, 2nd Quart., 2017.

[268] M. Nduwayezu, Q.-V. Pham, and W.-J. Hwang, "Online computation offloading in NOMA-based multi-access edge computing: A deep reinforcement learning approach," *IEEE Access*, vol. 8, pp. 99098–99109, 2020.

[269] P. Patel, M. Intizar Ali, and A. Sheth, "On using the intelligent edge for IoT analytics," *IEEE Intell. Syst.*, vol. 32, no. 5, pp. 64–69, Sep. 2017.

[270] A. Ndikumana, N. H. Tran, T. M. Ho, Z. Han, W. Saad, D. Niyato, and C. S. Hong, "Joint communication, computation, caching, and control in big data multi-access edge computing," *IEEE Trans. Mobile Comput.*, vol. 19, no. 6, pp. 1359–1374, Jun. 2020.

[271] M. Marjanovic, A. Antonic, and I. P. Zarko, "Edge computing architecture for mobile crowdsensing," *IEEE Access*, vol. 6, pp. 10662–10674, 2018.

[272] P. Zhou, W. Chen, S. Ji, H. Jiang, L. Yu, and D. Wu, "Privacy-preserving online task allocation in Edge-Computing-Enabled massive crowdsensing," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7773–7787, Oct. 2019.

[273] H. Li, K. Ota, and M. Dong, "Learning IoT in edge: Deep learning for the Internet of Things with edge computing," *IEEE Netw.*, vol. 32, no. 1, pp. 96–101, Jan. 2018.

[274] J. Kone ný, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," 2016, *arXiv:1610.02527*. [Online]. Available: http://arxiv.org/abs/1610.02527

[275] S. Samarakoon, M. Bennis, W. Saad, and M. Debbah, "Distributed federated learning for ultra-reliable low-latency vehicular communications," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 1146–1159, Feb. 2020.

[276] S. Wang, T. Tuor, T. Salonidis, K. K. Leung, C. Makaya, T. He, and K. Chan, "Adaptive federated learning in resource constrained edge computing systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1205–1221, Jun. 2019.
[277] S. Johnston and S. Cox, "The raspberry pi: A technology disrupter, and the enabler of dreams," *Electronics*, vol. 6, no. 3, p. 51, Jul. 2017.
[278] A. van Kempen, T. Crivat, B. Trubert, D. Roy, and G. Pierre, "MEC-ConPaaS: An experimental single-board based mobile edge cloud," in *Proc. 5th IEEE Int. Conf. Mobile Cloud Comput., Services, Eng. (Mobile-Cloud)*, Apr. 2017, pp. 17–24.
[279] Y. Elkhatib, B. Porter, H. B. Ribeiro, M. F. Zhani, J. Qadir, and E. Riviere, "On using micro-clouds to deliver the fog," *IEEE Internet Comput.*, vol. 21, no. 2, pp. 8–15, Mar. 2017.
[280] K. Jaiswal, S. Sobhanayak, A. K. Turuk, S. L. Bibhudatta, B. K. Mohanta, and D. Jena, "An IoT-cloud based smart healthcare monitoring system using container based virtual environment in edge device," in *Proc. Int. Conf. Emerg. Trends Innov. Eng. Technol. Res. (ICETIETR)*, Jul. 2018, pp. 1–7.
[281] S.-C. Huang, Y.-C. Luo, B.-L. Chen, Y.-C. Chung, and J. Chou, "Application-aware traffic redirection: A mobile edge computing implementation toward future 5G networks," in *Proc. IEEE 7th Int. Symp. Cloud Service Comput. (SC2)*, Nov. 2017, pp. 17–23.
[282] Q. Liu, S. Huang, and T. Han, "Demo: Fast and accurate object analysis at the edge for mobile augmented reality," in *Proc. 2nd ACM/IEEE Symp. Edge Comput.*, San Jose, CA, USA, Oct. 2017, pp. 1–2.
[283] Z. Zhao, Z. Jiang, N. Ling, X. Shuai, and G. Xing, "ECRT: An edge computing system for real-time image-based object tracking," in *Proc. 16th ACM Conf. Embedded Netw. Sensor Syst.*, Nov. 2018, pp. 394–395.
[284] D. Zhang, Y. Ma, Y. Zhang, S. Lin, X. S. Hu, and D. Wang, "A real-time and non-cooperative task allocation framework for social sensing applications in edge computing systems," in *Proc. IEEE Real-Time Embedded Technol. Appl. Symp. (RTAS)*, Apr. 2018, pp. 316–326.
[285] S. Yi, Z. Hao, Q. Zhang, Q. Zhang, W. Shi, and Q. Li, "LAVEA: latency-aware video analytics on edge computing platform," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun. 2017, pp. 1–13.
[286] F. Boabang, H.-H. Nguyen, Q.-V. Pham, and W.-J. Hwang, "Network-assisted distributed fairness-aware interference coordination for device-to-device communication underlaid cellular networks," *Mobile Inf. Syst.*, vol. 2017, pp. 1–11, Jan. 2017.
[287] G. G. Girmay, Q.-V. Pham, and W.-J. Hwang, "Joint channel and power allocation for Device-to-Device communication on licensed and unlicensed band," *IEEE Access*, vol. 7, pp. 22196–22205, 2019.
[288] S. Singh, Y.-C. Chiu, Y.-H. Tsai, and J.-S. Yang, "Mobile edge fog computing in 5G era: Architecture and implementation," in *Proc. Int. Comput. Symp. (ICS)*, Dec. 2016, pp. 731–735.
[289] J. Wen, C. Ren, and A. K. Sangaiah, "Energy-efficient device-to-device edge computing network: An approach offloading both traffic and computation," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 96–102, Sep. 2018.
[290] J. Wang, Y. Hu, H. Li, and G. Shou, "A lightweight edge computing platform integration video services," in *Proc. Int. Conf. Netw. Infrastruct. Digit. Content (IC-NIDC)*, Aug. 2018, pp. 183–187.
[291] A. Lertsinsrubtavee, A. Ali, C. Molina-Jimenez, A. Sathiaseelan, and J. Crowcroft, "PiCasso: A lightweight edge computing platform," in *Proc. IEEE 6th Int. Conf. Cloud Netw. (CloudNet)*, Sep. 2017, pp. 1–7.
[292] M. Körner, T. M. Runge, A. Panda, S. Ratnasamy, and S. Shenker, "Open carrier interface: An open source edge computing framework," in *Proc. Workshop Netw. for Emerg. Appl. Technol. NEAT*, 2018, pp. 27–32.
[293] G. Orsini, D. Bade, and W. Lamersdorf, "CloudAware: A context-adaptive middleware for mobile edge and cloud computing applications," in *Proc. IEEE 1st Int. Workshops Found. Appl. Self* Syst. (FAS*W)*, Sep. 2016, pp. 216–221.
[294] T. Rausch, S. Nastic, and S. Dustdar, "EMMA: Distributed QoS-aware MQTT middleware for edge computing applications," in *Proc. IEEE Int. Conf. Cloud Eng. (IC2E)*, Apr. 2018, pp. 191–197.
[295] K. E. Benson, G. Wang, N. Venkatasubramanian, and Y.-J. Kim, "Ride: A resilient IoT data exchange middleware leveraging SDN and edge cloud resources," in *Proc. IEEE/ACM 3rd Int. Conf. Internet-of-Things Design Implement. (IoTDI)*, Apr. 2018, pp. 72–83.
[296] A. Carrega, M. Repetto, P. Gouvas, and A. Zafeiropoulos, "A middleware for mobile edge computing," *IEEE Cloud Comput.*, vol. 4, no. 4, pp. 26–37, Jul. 2017.
[297] C.-Y. Li, H.-Y. Liu, P.-H. Huang, H.-T. Chien, G.-H. Tu, P.-Y. Hong, and Y.-D. Lin, "Mobile edge computing platform deployment in 4G LTE networks: A middlebox approach," in *Proc. USENIX Workshop Hot Topics Edge Comput. (HotEdge)*, Boston, MA, USA, Jul. 2018, pp. 1–6.
[298] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May/Jun. 2019.
[299] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, early access, Apr. 8, 2020, doi: 10.1109/COMST.2020.2986024.

QUOC-VIET PHAM (Member, IEEE) received the B.S. degree in electronics and telecommunications engineering from the Hanoi University of Science and Technology, Vietnam, in 2013, and the Ph.D. degree in telecommunications engineering from Inje University, South Korea, in 2017. From September 2017 to December 2019, he was with Kyung Hee University, Changwon National University, and Inje University on various academic positions. He is currently a Research Professor with the Research Institute of Computer, Information, and Communication, Pusan National University, South Korea. His research interests include convex optimization, game theory, machine learning to analyze and optimize edge/cloud computing, and 5G and beyond networks. He received the Best Ph.D. Dissertation Award in Engineering from Inje University, in 2017.

FANG FANG (Member, IEEE) received the B.A.Sc. and M.A.Sc. degrees in electronics engineering from Lanzhou University, in 2010 and 2013, respectively, and the Ph.D. degree in electrical engineering from The University of British Columbia (UBC), Kelowna, BC, Canada, in 2018. She is currently a Research Associate with the Department of Electrical and Electronics Engineering, The University of Manchester, U.K., and an Assistant Professor with the Department of Engineering, Durham University, Durham, U.K. Her current research interests include 5G and beyond wireless networks, NOMA, IRS, and mobile edge computing. She has served as a TPC Member of the IEEE conferences, such as the GLOBECOM and ICC. She received the Exemplary Reviewer Certificate of the IEEE TRANSACTIONS ON COMMUNICATIONS, in 2017. She is currently an Associate Editor of the IEEE Open Journal of the Communications Society.

VU NGUYEN HA (Member, IEEE) received the B.Eng. degree from the Ho Chi Minh City University of Technology (HCMUT), Vietnam, through the French Training Program for Excellent Engineers in Vietnam (PFIEV), the Addendum degree from École Nationale Supérieure des Télécommunications de Bretagne-Groupe des École des Télécommunications, Bretagne, France, in 2007, and the Ph.D. degree from the Institut National de la Recherche Scientifique-Énergie, Matériaux et Télécommunications (INRS-EMT), Université du Québec, Montréal, Québec, Canada, in 2017. From 2008 to 2011, he was a Research Assistant with the School of Electrical Engineering, University of Ulsan, Ulsan, South Korea. He is currently a Postdoctoral Fellow of the École Polytechnique de Montréal, Montréal. His research interests include radio resource management and emerging enabling technologies for 5G wireless systems with a special emphasis on heterogeneous small-cell networks, cloud RAN, massive MIMO communications, mmWave, and mobile edge computing. He is currently a recipient of the FRQNT Postdoctoral Fellowship for International Researchers (PBEEE).

MD. JALIL PIRAN (Member, IEEE) received the Ph.D. degree in electronics and radio engineering from Kyung Hee University, South Korea, in 2016. He is currently a Professor with the Department of Computer Science and Engineering, Sejong University, Seoul, South Korea. His research fields include resource allocation and management in 5G mobile and wireless communication, the Internet of Things (IoT), multimedia communication, cognitive radio networks, and machine learning. In the worldwide communities, he has been an active Delegate of the Moving Picture Experts Group (MPEG), South Korea, since 2013, and an active Member of the International Association of Advanced Materials (IAAM), since 2017.

MAI LE received the B.S. degree in electronics and telecommunications engineering from the Hanoi University of Science and Technology, Vietnam, in 2014. She is currently pursuing the M.S. degree with the Department of Information and Communications Systems, Inje University, South Korea. Her research interests include device-to-device (D2D) communications, edge computing, and computation intelligence.

LONG BAO LE (Senior Member, IEEE) received the B.Eng. degree in electrical engineering from the Ho Chi Minh City University of Technology, Vietnam, in 1999, the M.Eng. degree in telecommunications from the Asian Institute of Technology, Thailand, in 2002, and the Ph.D. degree in electrical engineering from the University of Manitoba, Canada, in 2007. He was a Postdoctoral Researcher at the Massachusetts Institute of Technology, from 2008 to 2010, and the University of Waterloo, from 2007 to 2008. Since 2010, he has been with the Institut National de la Recherche Scientifique (INRS), Université du Québec, Montréal, QC, Canada, where he is currently an Associate Professor. He has coauthored the book *Radio Resource Management in Multi-Tier Cellular Wireless Networks* (Wiley, 2013) and *Radio Resource Management in Wireless Networks: An Engineering Approach* (Cambridge University Press, 2017). His current research interests include smart grids, cognitive radios, radio resource management, network control and optimization, and emerging enabling technologies for 5G wireless systems. He is currently a member of the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS.

WON-JOO HWANG (Senior Member, IEEE) received the B.S. and M.S. degrees in computer engineering from Pusan National University, Busan, South Korea, in 1998 and 2000, respectively, and the Ph.D. degree in information systems engineering from Osaka University, Osaka, Japan, in 2002. From 2002 to 2019, he was a Full Professor at Inje University, Gimhae, South Korea. He is currently a Full Professor with the Department of Biomedical Convergence Engineering, Pusan National University. His research interests include optimization theory, game theory, machine learning, and data science for wireless communications and networking.

ZHIGUO DING (Fellow, IEEE) received the B.Eng. degree in electrical engineering from the Beijing University of Posts and Telecommunications, in 2000, and the Ph.D. degree in electrical engineering from Imperial College London, in 2005.

From July 2005 to April 2018, he was with Queen's University Belfast, Imperial College London, Newcastle University, and Lancaster University. From October 2012 to September 2018, he was an Academic Visitor with Princeton University. Since April 2018, he has been with The University of Manchester as a Professor of communications. His research interests are 5G networks, game theory, cooperative and energy harvesting networks, and statistical signal processing. He was a recipient of the Best Paper Award from IET ICWMC 2009 and the IEEE WCSP 2014, the EU Marie Curie Fellowship, from 2012 to 2014, the Top IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY Editor, in 2017, the IEEE Heinrich Hertz Award, in 2018, the IEEE Jack Neubauer Memorial Award, in 2018, and the Best IEEE SIGNAL PROCESSING LETTER Award, in 2018. He was an Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS and the IEEE COMMUNICATIONS LETTERS, from 2013 to 2016. He is also serving as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and the journal of *Wireless Communications and Mobile Computing*.