Article

Multi-Access Edge Computing Empowered Heterogeneous Networks: A Novel Architecture and Potential Works

June-Woo Ryu [1,†] [iD]**, Quoc-Viet Pham [2,*,†]** [iD]**, N. T. Huynh Luan[3], Won-Joo Hwang [4,*],
Jong-Deok Kim [1] and Jung-Tae Lee [1]**

[1]   School of Electrical and Computer Engineering, Pusan National University, 2, Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan 46241, Korea
[2]   High Safety Core Technology Research Center, Inje University, 197, Inje-ro, Gimhae-si, Gyeongsangnam-do 50834, Korea
[3]   Faculty of Technology Engineering, Thu Dau Mot University, 6, Tran Van On, Phu Hoa, Thu Dau Mot City, Binh Duong 820000, Vietnam
[4]   Department of Electronic, Telecommunications, Mechanical Automotive Engineering, Inje University, 197, Inje-ro, Gimhae-si, Gyeongsangnam-do 50834, Korea
*   Correspondence: vietpq90@gmail.com (Q.-V.P.); ichwang@inje.ac.kr (W.-J.H.)
†   These authors contributed equally to this work.

Abstract: One of the most promising approaches to address the mismatch between computation-intensive applications and computation-limited end devices is multi-access edge computing (MEC). To overcome the rapid increase in traffic volume and offload the traffic from macrocells, a massive number of small cells have been deployed, so-called heterogeneous networks (HetNets). Strongly motivated by the close integration of MEC and HetNets, in this paper, we propose an envisioned architecture of MEC-empowered HetNets, where both wireless and wired backhaul solutions are supported, flying base stations (BSs) can be equipped with MEC servers, and mobile users (MUs) need both communication and computation resources for their computationally heavy tasks. Subsequently, we provide the research progress summary of task offloading and resource allocation in the proposed MEC-empowered unmanned aerial vehicle (UAV)-assisted heterogeneous networks. We complete this article by spotlighting key challenges and open future directives for researches.

Keywords: computation offloading; Internet of Things (IoT); heterogeneous networks (HetNets); multi-access edge computing (MEC); non-orthogonal multiple access (NOMA); resource allocation; unmanned aerial vehicles (UAV)

1. Introduction

With the increasing popularity of smart devices, many sophisticated applications are emerging in the 5G network and beyond, for example, online gaming, augmented reality, and intelligent video acceleration [1]. These new applications often require intensive computation, high energy consumption, and high latency sensitivity. Unfortunately, end devices are usually constrained by finite battery capacity and low computation capability and are not able to handle the aforementioned applications. The contradiction between the low capabilities of end devices and high requirements of applications in 5G motivates the development of multi-access edge computing (MEC) [2]. Unlike conventional centralized clouds, MEC moves computing away from cloud nodes towards the network edges, thus enhancing the availability and reachability of cloud services at the edge of networks in close proximity to MUs. In light of this, MEC offers the cloud services and functions (e.g., storage, computation, and caching) at the edge of the network, thereby improving MU's quality of experience—for example,

extension of the battery lifetime, low latency, and relief of burden on the fronthaul links. Typically, MEC servers are smaller with moderate computing resources and deployed at aggregation points such as base stations (BSs). Moreover, MEC is highly scalable as the number of MEC servers is expected to increase significantly and is able to support low-latency applications, real-time mobility, and location awareness [3,4]. Among considered problems in MEC, computation offloading is of great importance and presents in many network use cases. The idea of computation offloading is to enable MUs to offload their intensive computations to the MEC servers, which are typically more powerful in computational capabilities and storage capacity, and are not constrained by energy consumption. Through computation offloading, many compute-intensive applications (e.g., 3D gaming, virtual reality, autonomous driving, and location-based service recommendation) can be deployed at the edge of the network so that MUs can make use of these services. Additionally, computation offloading is tightly coupled by radio resources—for example, transmit power and subchannel assignment, and computation resources such as computing frequency of MUs, computation scheduling at MEC servers, and MEC server selection.

To overcome the rapid increase in data traffic and rate requirements, one potential solution is the use of heterogeneous networks (HetNets), which is formed by the deployment of conventional macro base stations (MBSs) and small BSs (SBSs) such as pico-, femto-, and micro-BSs [5]. While HetNets bring substantial improvements to the network capacity and service coverage, the dense deployment in HetNets can result in severe interference, which in turn rapidly deteriorates into the network performance. Therefore, interference management is considered as the central premise of the practical deployment of HetNets. As small cells are being densely deployed and due to challenges of providing wired backhaul links for dense HetNets, how to forward and receive data in HetNets has been a crucial issue. In general, there are two main types of backhaul in HetNets, including wired and wireless backhauling solutions [6]. Many factors should be taken into consideration of choosing backhaul types—for example, implementation cost, traffic load, and quality of service (QoS) requirements. In fact, it is difficult to deploy all small cells with wireless backhaul links because of installation obstacles. Recently, wireless backhaul has been considered as a feasible and cost-effective solution for sending and receiving data in dense HetNets.

Since both MEC and HetNets could offer attractive features including low latency, real-time mobility, and high computation capabilities (MEC) and high spectral-energy efficiency and coverage extension (HetNets), which are exactly required by the 5G network, MEC and HetNets are recognized as two key enabling technologies in the emerging 5G network. Although there has been growing interest in MEC in recent years, most of the existing studies are dedicated to single-server MEC networks. In terms of computation offloading and resource allocation, previous works usually simplify the network model such as one MU in a small cell [7], one MEC server [8–10], and dedicated backhaul links with infinite capacity [11]. Existing works do not envision a holistic architecture for MEC in HetNets and consider heterogeneous MEC servers, i.e., MEC servers are distinctly different in sizes (computing units) and configurations (computational speeds). It is reasonable to believe that MEC servers will be densely equipped at various places such as cellular BSs, WiFi access points, business premises, and portable devices around our lives, e.g., mobile phones, vehicles, and wearable devices. Regarding the concept of small cell cloud, a general architecture was proposed in [12,13], where the small cell manager is introduced to manage cloud resources provided by edge servers collocated with small cell base stations. There were some approaches for computation migration in various edge cloud concepts—for example, small cell cloud in [14,15], follow-me cloud in [16], and mobile micro-cloud in [17]. The combination of genetic algorithm and particle swarm optimization in [18,19], game theory in [10,20,21], and machine learning in [22,23] were adopted to design algorithms for jointly optimizing task offloading and resource allocation in small cell networks. As security and privacy are important issues and edge devices are typically located at physical locations close to eavesdroppers and attackers [24,25], many recent studies have been devoted to privacy and security issues. For example, a survey of security threats and challenges for different edge computing paradigms

was provided in [25], the application of deep reinforcement learning for security in mobile edge caching was investigated in [26], and physical layer security for MEC systems with joint consideration of task offloading and resource allocation was studied in [27]. However, a common assumption in the existing works is that small cells are all equipped with wired backhaul.

It is well known that the implementation of HetNets faces the challenge of data transmission over limited-capacity backhaul links, especially when considering ultra-dense HetNets in the 5G network. Therefore, it is urgently required to integrate the two key concepts: MEC and HetNets, and propose a holistic architecture for the integration of MEC and HetNets (so-called MEC-empowered HetNets or MEC HetNets in short), which are the main objectives of this article. The integration of MEC and HetNets will provide a viable and realistic solution for the 5G network to take individual benefits of MEC and HetNets in order to meet critical requirements of applications and services in the 5G network. For instance, latency-sensitive Internet of Things (IoT) could rely on MEC for extremely-low computation latency requirements and on HetNets for massive connection over distributed geographic areas. In summary, our contributions are three-fold:

- Firstly, we provide an overview of computation offloading and resource allocation in single-server MEC networks.
- Secondly, we propose a novel framework of HetNets with MEC, where *both wireless and wired backhaul solutions are supported*, *flying BSs can be equipped with MEC servers*, and *MUs need both communication and computation resources for their computationally heavy tasks*.
- Finally, we discuss *five challenges* and *open issues* stemmed from our proposed architecture, including joint interference management and offloading decision, joint user association and offloading decision, hierarchical and collaborative computation offloading, joint user association and MEC server selection, and non-orthogonal multiple access (NOMA)-assisted MEC systems.

The article is organized as follows. In Sections 2 and 3, overviews of computation offloading and resource allocation in single-MEC networks and multi-server homogeneous networks are given. Then, the architecture of MEC HetNets is proposed and described. In Section 4, we discuss challenges and open issues of the proposed architecture. Finally, we conclude the article in Section 5.

2. Overview of Computation Offloading in Homogeneous Networks

In general, MEC use cases can be categorized into *consumer-oriented services*, *operator and third-party services*, and *network performance and quality-of-experience (QoE) improvements* [6,28,29]. Among them, computation offloading is regarded as the main use case for consumer-oriented services so that it enables end devices to run new resource-demanding applications, prolong the battery lives, and reduce the end-to-end latency. However, computation offloading presents several challenges—for example, efficient communication and computation resources allocation, appropriate selection of the MEC server among available ones, and efficient management of simultaneous offloading by multiple MUs [28]. Due to the profound importance of MEC, there has been a large amount of research advocated for technical issues in MEC, e.g., computation offloading, design of application and network architecture, resource allocation, and implementation and standardization.

Regarding computation offloading, a critical component is to decide whether or not to offload the tasks. Basically, the decision on offloading computation tasks can be categorized into three models: local execution, binary offloadable offloading, and partial offloadable offloading. When either the MU has a poor connection, no available MEC server exists, or the MU does not benefit from the computation offloading process, the MU will locally handle the whole computation tasks. Otherwise, the MU will offload their its computations to the remote MEC servers for computation handling. In binary offloading, i.e., fully offloading, an integrated task, which cannot be partitioned into sub-tasks, is performed using either local computing or MEC execution. In the meanwhile, in partial offloadable offloading, a computation task can be arbitrarily divided into at least two sub-tasks, some of which are processed locally while the others can be offloaded to and executed at multiple MEC servers

by considering a variety of factors, e.g., user and server capabilities, resource availabilities, and backhaul connection characteristics [1]. Two other important aspects of computation offloading are the determination of the application model and management of the offloading process in practice [28]. The former refers to the offloadability of application, determines the amount of processed data, and specifies the dependency among the offloadable parts, while the latter means how an MU determines what should be offloaded, checks external factors such as the channel connection, system bandwidth, and available computation resources at the MEC server, and makes the offloading decision.

The typical objective of computation offloading is to improve the user experience by providing lower execution latency, reduce the energy consumption by migrating computation tasks to the server, and optimize the trade-off between the energy consumption and execution latency. Actually, the objective can be designed from the system-wide level, e.g., minimization of the total execution latency, to the user-perceived performance, e.g., the minimum of individual user energy consumption and revenue. In addition, the objective should consider distinct requirements of the target applications, for instance, to exploit MEC in handling IoT applications, since the battery is limited at the IoT devices, the trade-off between computation and transmission energy consumption should be studied [30]. To achieve such the goals for MEC, resource (communication and computation/storage) allocation plays an important role. The resource may include the computational speed of the MUs, computation resource at the MEC servers, transmit power of MUs, and bandwidth and time allocated to offloading users. In fact, offloading decision and resource allocation are tightly coupled and jointly affect the achieved performance. Let us take an example of a two-task MEC setting, where the first latency-sensitive task is offloaded to the MEC server and requires a substantial portion of the MEC computation capability, which in turn affects the remaining computation resource allocated to the remaining task. Thus, the second task does not benefit from remote processing. A possible solution, in this case, is to let the first task and a part of the second task to be remotely accomplished. In Table 1, we summarize the existing research on MEC in terms of the offloading decision and resource allocation. In-depth surveys on the computation offloading, resource allocation, and other aspects of MEC systems can be referred to in [28,31].

Table 1. Summary of research into offloading decisions and resource allocation in MEC systems.

Theme	Resource	Study	Proposed Framework
Computation offloading	Local execution	[28,29,32]	If an MU does not get any benefit from remote execution, and the computation task is executed locally [1].
	Partial offloading	[28,31,33]	If a computation task can be divided into smaller parts, partial offloading is possible. Particularly, partial offloading is suitable for latency-critical applications due to the execution parallelism [33].
	Binary offloading	[1,10,19]	In binary offloading, a computation task can be processed using either local or remote computing. Thus, binary offloading is suitable for simple and highly integrated applications [1].
Computation resource	Computation speed	[34]	The CPU performance via adjusting the CPU frequency and/or supply voltage can be utilized to improve the performance of MEC systems in consumption and execution latency [33,34].
	Server computation resource	[1,30]	The amount of computing resources allocated by the MEC server for each offloading MU depends on many factors, e.g., input data size, task workload, server computation capability, and network settings (e.g., number of offloading MUs and MEC servers).
Radio resource	Transmit power	[10,32,35]	MUs can adjust their transmit power for different purposes, e.g., interference management and adaptation to channel conditions, to lower energy consumption and increase the offloading probability.
	Bandwidth	[1,36]	Since MUs may run applications with different task workloads and completion deadlines, designing efficient bandwidth allocation schemes can enhance performance of MEC systems.
	Time	[36,37]	MUs can share time resources to offload (accomplish) their computations to (at) the MEC server [36]. For example, in MEC systems with wireless power transfer [37], the time frame is slotted into different parts for wireless energy transfer and computation task offloading. The optimal allocation of time resources is highly coupled with the offloading decisions and task characteristics.
JCORA	Joint task offloading & resource allocation	[1,29,35]	With massive MUs, computation offloading would be a bottleneck. Moreover, the server's computation capability is typically finite, thus challenging the sharing of radio and computation resources for MUs. Jointly optimizing task offloading and resource allocation is highly needed to further improve the network performance.

3. Multi-Access Edge Computing with Multiple Homogeneous MEC Servers

The majority of the existing studies has focused on single-server MEC networks. Typical network settings are (*i*) single-cell networks, where an MEC server is integrated into the BS and (*ii*) single-server MEC HetNets, where multiple small cells are overlaid within a macrocell, the macro BS is equipped with an MEC server, and MUs offload tasks to the MEC server through the small BSs. Figure 1 provides the pictorial illustration of these two network settings. The similarities of the aforementioned literature are in the assumption of the single MEC server and/or the fixed user association even in the case where an MU is under the coverage of multiple BSs.
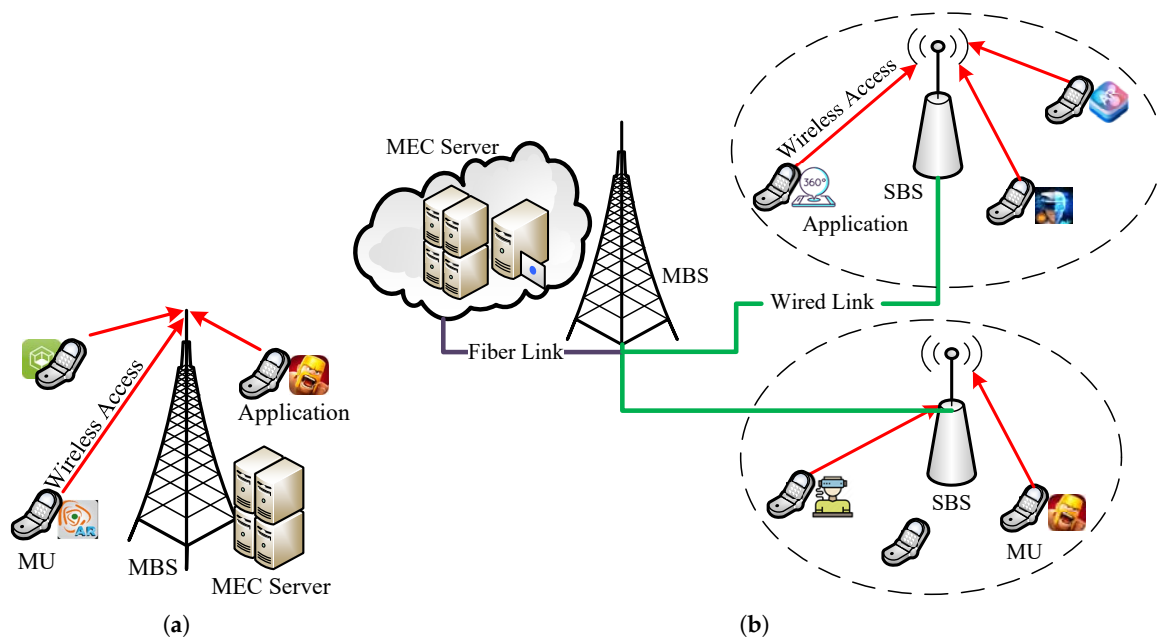

Figure 1. Two typical network settings of a single-server MEC system: (**a**) single-cell and single-server MEC system; (**b**) multi-cell and single-server MEC system.

Very recently, few works have considered the existence of multiple servers in MEC networks for various problems, such as jointly adjusting the computational speed and offloading subtasks of a single user to multiple MEC servers [34], joint optimization of task offloading and resource allocation in multi-server MEC systems [1,3], and the minimum number of active MEC servers [38]. Compared with the single-server MEC network, MEC with multiple MEC servers could provide enormous benefits, as follows [1,3]:

- As previously mentioned, due to the economic and scalable deployments, the MEC server usually has limited computation resources. In addition, the mobile traffic is expected to increase 1000 times over the next decade and the number of connected users will reach 50 billion in 2020; the single MEC server will likely become seriously overloaded. In contrast, MUs can migrate their computation tasks to the best among several available MEC servers, thus reducing the burden of computation from the single MEC server.
- It is possible that an MEC is under the coverage of several cells (i.e., multiple MEC servers), thus the MU can offload its computations to the server with better channel condition and computation capability. Therefore, MUs can save the transmission energy consumption (thus the total energy consumption) and shorten the execution latency.
- In the case of partial offloading, a computation task can be divided into smaller parts and offloaded to different MEC servers, thereby reducing the execution latency. MEC servers can coordinate to exchange information to reduce inter-cell interference (ICI) among offloading users across

neighboring MEC servers and improve the resource allocation efficiency. Moreover, collaboration among different MEC servers can help potentially increase the cluster resources, thus affordable to support applications with higher computation requirements.

In addition to extra benefits, there exist some challenges in multi-server MEC networks that must be well addressed, as follows [1]:

- The availability of multiple servers makes the offloading decision problem particularly complex. Each MU must decide that (*i*) what is the optimal operation mode: offload or not offload and (*ii*) which MEC server should be selected such that the achieved performance is maximized and the resource and task constraints are satisfied.

- Due to the existence of ICI, the resource allocation problem in multi-server MEC networks is much more challenging than that in single-server MEC networks [3]. To mitigate this effect, the spectrum resource within each cell can be divided into orthogonal subchannels, which should be efficiently allocated to MUs (i.e., which subchannel an MU should use to offload its computation task to the MEC server). The network performance can be further improved by jointly optimizing the subchannel assignment and adjusting the transmit power at the MUs.

- On the one hand, it is foreseeable that a massive number of MEC servers will be deployed in distribution 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 is greatly dependent on the deployment locations of the MEC servers, e.g., a specific area may be densified by a large number of MEC servers while only one server is deployed in other areas. MUs may run applications and services with different priorities and security levels, e.g., some services can only run in private MEC servers due to the security consideration. Moreover, computation tasks may be different in input data size, computation workload, and execution deadline. As a result, the joint resource allocation problem should take into account the above properties and characteristics.

Considering the above advantages and challenges, we proposed a decentralized computation offloading scheme and resource allocation in multi-server MEC networks [1]. To enable decentralized computation offloading, matching theory, a powerful tool to design distributed algorithms for a large number of resource allocation problems in wireless communication, is adopted. Figure 2 illustrates the result comparison in offloading probability and computation overhead, among the schemes that only handle the computation tasks locally (local only) and remotely (offloading only), HODA algorithm (HODA) [35], and our proposed algorithm (JCORAMS). From the figure, the offloading probability decreases and the total computation overhead increases with the larger number of MUs. The explanation is obvious since, with more MUs, each MU has a lower probability to utilize its preferred MEC server and subchannel, and ICI among MUs becomes more severe. Another observation from Figure 2b is that offloading all the computation tasks may not a good solution. Actually, the performance of this scheme becomes worse than the local computing only method when the number of MUs is sufficiently large. This is because the limited amount of radio and computation resources needs to be shared among many MUs. Notably, the proposed algorithm outperforms these alternative schemes in terms of offloading probability and computation overhead.
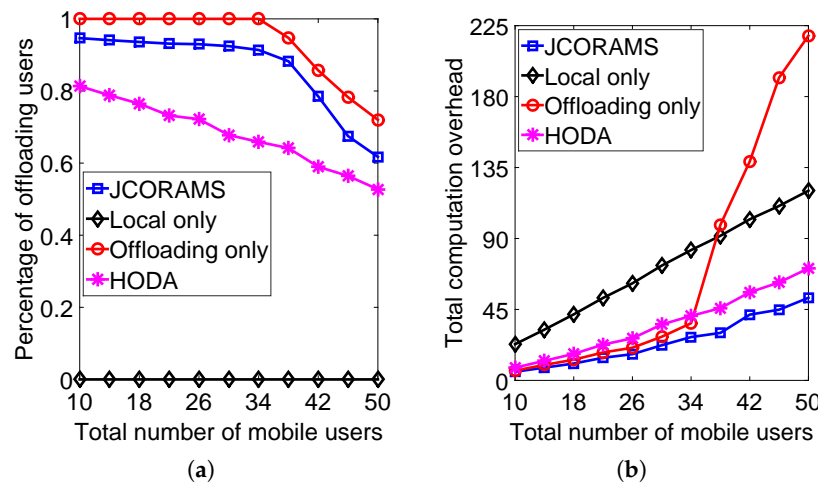
Figure 2. Comparison four offloading frameworks under different numbers of users: (**a**) offloading probability; (**b**) total computation overhead.

4. Multi-Access Edge Computing with Multiple Heterogeneous MEC Servers

The purpose of this section is to present an MEC-based HetNet architecture, where *the coexistence of wired and wireless backhaul links is supported* and *unmanned aerial vehicles (UAVs) can be deployed to serve as flying MEC servers*. We also present some future works that are stemmed from our proposed architecture.

4.1. Motivation

It is commonly assumed in the existing literature that SBSs connect with MBSs via dedicated links; however, providing all small cells with wired backhaul is not feasible in practice, which is further exacerbated by the densification of HetNets in the 5G network. In this context, wireless backhaul has been emerged as a viable solution. As opposed to wired backhaul, wireless backhaul is preferred due to the following reasons [6]:

- Despite many advantages, providing wireless backhaul connections for ultra-dense small cells is costly and time-consuming. Furthermore, there are many factors to be considered for the deployment of wired backhaul, such as locations of SBSs and mobile user QoS requirements.
- A number of frequency bands can be used for wireless backhaul, for example, microwave band between 6 GHz and 60 GHz, Millimeter-wave band including V-band (57–76 GHz), E-band (71–76, 81–86 and 92–95 GHz), W-band (92–94, 94.1–100, 102–109.5, and 111.8–114.25 GHz), and D-band (130–134, 141–148.5, 151.5–164, and 167–174.8 GHz), sub-6 GHz band, satellite frequency band, and TV white space band. Another easy-to-deploy and cost-effective approach that can provide high capacity wireless backhaul in hyper-dense HetNets is free-space optical communication [39]. It is therefore expected that the achievable capacity of wireless backhaul will reach beyond 10 Gbps, thus making wireless backhaul a practical solution for dense HetNets.
- In some cases like hard-to-reach and rural areas, the very high deployment cost is a great challenge of providing wireless access. Wireless backhaul is a vital solution for those areas, thus simplifying the deployment and reducing the maintenance cost.
- It is expected that the 5G network and beyond will support emergency services and disaster situations. Nonetheless, wired backhaul connections, which have been destroyed in the aftermath, may not be repaired immediately, thus severely affecting the network response and reliability. In such scenarios, temporarily deploying wireless backhaul links is a promising solution.

Previous studies are typically based on the assumption of perfect backhaul links, i.e., no resource constraints on backhaul links. However, the realities are that such constraints exist in MEC-empowered

HetNets and the backhaul link qualities have a great impact on performance achievement. For instance, a small BS transfers the received data from a set of MUs to the MEC server through a wireless backhaul connection, but forwarding all the computations to the MEC server would be not possible due to the limited-capacity backhaul. How to design efficient wireless backhaul solutions for MEC empowered HetNets is still not investigated.

Historically and originally used in the military, unmanned aerial vehicles (UAVs) have recently used in a variety of applications, among which the UAV-aided wireless communication is an important use case [40]. Investigated as an alternative solution to meet temporary and/or unexpected traffic demands when terrestrial communication is not able to satisfy, UAV-aided wireless system offers several advantages [41,42]. First, thanks to the continued cost reduction in UAVs, the deployment of UAVs as flying BSs, which are used to provide communications for ground users and increase the network coverage and capacity, becomes more feasible than ever before. Next, UAVs can be flexibly and temporarily deployed in rapid response to either catastrophe and emergency scenarios or unexpected events. For instance, UAVs can be deployed to provide Internet connectivity to post-disaster areas, where the terrestrial network infrastructure fails to provide reliable communications and cannot be recovered soon. In addition, the UAV's status and communication can be jointly optimized such that the network performance is improved, e.g., the UAV's energy consumption can be improved by optimally adjusting its flying speed, direction, and acceleration, and end-user QoS requirements are satisfied, e.g., the UAV can adjust its altitude and mobility to establish reliable connections with potential users. Finally, signals from UAVs to ground users can be significantly improved through the line-of-sight communication links, thus reducing the effects of signal shadowing and blockage. However, due to the finite weight and size, the UAV's on-board energy is limited and the performance of UAV-aided wireless systems is highly affected. Accordingly, to take full advantage of UAVs in wireless communications, several challenges must be well addressed [43] such as air-to-ground channel modeling for UAVs-to-BSs communications and ground-to-air channel modeling for BSs-to-UAVs communications, optimal UAV placement, energy-aware operation of UAVs, interference-aware resource management in multi-UAV systems, and performance analysis of UAV-aided wireless networks.

4.2. State-Of-The-Art UAV-Assisted Heterogeneous Networks

In this subsection, we provide a summary of state-of-the-art research works that combine the advantages of UAV communications and heterogeneous networks. It is worth mentioning that 5G and beyond will support not only communication, but also computation, caching, and control (4C). This motivates us to propose a new architecture in the next subsection, which is able to integrate UAV communications, edge computing, and heterogeneous networks.

In [44], the authors integrated UAVs into HetNets for public safety communications. Due to the high mobility of UAVs and dense deployment of HetNets, designing effective interference management schemes faces several challenges: (1) outage problems due to damaged and dysfunctional BSs, (2) mobility and dynamic locations of UAVs, and (3) burst data transmission in the aftermath and (4) high-QoS and low-latency requirements for mission-critical applications. Unlike existing works, where UAVs are usually assumed to be deployed quickly, two deployment problems of UAV networks was studied in [45]. The first problem is to minimize the maximum deployment delay for fairness consideration and the other is to minimize the total deployment delay for efficiency consideration. The work in [46] first presented four representative scenarios to show potential applications of UAV-assisted HetNets, including UAV as flying BSs, UAV as mobile relays, UAV-supported energy transfer, and UAV-supported caching. The work also demonstrated the superior performance of UAV-assisted HetNets compared with static networks and highlighted some challenges and open issues—for example, coordination among UAVs, energy limitation of UAVs, flying ad hoc networks, and UAV deployment. A two-layer (ground and aerial layers) air-ground mobile edge network architecture (MEN) was proposed in [47], which combines three fundamentals components for an MET, including network densification, mobile edge caching, and MEC. Challenges of the proposed architectures were

also represented, which are network interworking, methods to evaluate performance of the proposed MEN, prediction and optimization of communication (i.e., access, fronthaul, and backhaul) links, and software-defined networking based control and communication schemes. Different from this work, where aerial and ground layers are connected with each other, the architecture in our work is proposed from the perspective of MEC and HetNets, where each MU can be served by either SBS, MBS, or flying UAV-supported BS, and the communication between any two SBSs and MBSs can be via either wireless backhaul or wired link. Moreover, many studies have been devoted to the optimization of UAV-HetNets, for example, capacity enhancement in [48], delay minimization in [49], and caching capability in [50].

4.3. Proposed Architecture: MEC-Empowered UAV-Assisted Heterogeneous Networks

Given the importance of MEC and HetNets, the 5G network will be highly integrative with a focus on the integration of MEC and HetNets, the coexistence of both wireless and wired backhaul for the deployment of MEC, and support of UAVs for catastrophe and emergency scenarios as well as enhancing the connection availability. We will envision a holistic architecture for MEC-empowered UAV-assisted HetNets.

Figure 3 demonstrates our envisioned architecture of MEC-empowered HetNets with the consideration of various communication and computation resources. It is worth noting that the MBS can be a wireless cellular tower, which is shared among multiple network operators to physically collocate their individual BSs. Therefore, the network operators of small cells may be different. To provide contents and services to MUs, the MBS is linked to one or a group of MEC servers through the dedicated fiber links. As aforementioned, the integration of MEC and HetNets supports the coexistence of both conventional wired and wireless backhaul in order to flexibly adapt to the actual network state for deploying small cells. For instance, due to the hard-to-reach deployment location, the small cells 1 and 3 connect to the macrocell by wireless backhaul links, whereas the small cell 2 can forward and receive data from the macrocell via a conventional wired backhaul connection. In this context, computations of MUs in the small cells 1 and 3 are firstly offloaded to the MBS via wireless backhaul links and then to the MEC server via fiber connections. As opposed, since the small cell 2 connects to the macrocell through the wired backhaul, MUs associated with the small cell 2 can ignore the offloading time over the wired backhaul. Similarly, MU 1 directly associates with the MBS, the computational time is therefore composed of the offloading time (the transmission time from itself to the MBS) and the computation time (due to the limited computing resources at the MEC server), i.e., the offloading time from the MBS to the MEC server over fiber links is entirely neglected. The majority of the existing literature on MEC networks are almost completely reliant on the assumption of infinite backhaul capacity that would, however, be improper in traditional multi-cell networks (due to the deployments cost of the backhaul links and heavy burden in the exchange for communication transportation), and that is further exacerbated in MEC empowered HetNets (due to big-data computation tasks and strict completion deadline). Because of the backhaul capacity constraints, the resource management problem should allocate the bandwidth for transmissions over access links and backhaul connections as well as considering other features such as task characteristics and heterogeneous computation capabilities.

The proposed architecture also supports the coexistence of conventional central clouds and distributed MEC to guarantee the availability of cloud services over the whole network. On the one hand, big-data and stateless applications would be better served by centralized clouds. As opposed to MEC servers, central clouds have much more powerful computation capabilities so that the massive number of huge signal tasks can be processed in a centralized fashion. However, traditional cloud computing centralization may lead to some limitations, e.g., high latency and burden on fronthaul links. On the other hand, distributed MEC is more suitable for novel applications and services like real-time video gaming, IoT, and mobility-related and location-aware applications. In the envisioned architecture, the macrocell is connected to centralized clouds through fiber-optic backbone networks [51]. In cloud

radio access networks (C-RANs), the BS is composed of distributed remote radio heads (RRHs) and the centralized baseband unit (BBU), which is responsible for the baseband signal processing functionalities. Multi-tier computing architecture can be designed in C-RANs, where the cloud centers are collocated at the central BBUs and the edge servers are deployed at the distributed RRHs (e.g., macro-RRHs and small-RRHs). With this hierarchical architecture, MUs can exploit various computing servers with heterogeneous computation capabilities. Moreover, computing servers at the same and even different tiers can collaborate to execute enormous computations so as to improve the computing performance. Note that central servers can be integrated into the cellular networks; they, however, can be standalone like Amazon Web Service, Microsoft Azure, and Google Cloud Platform.
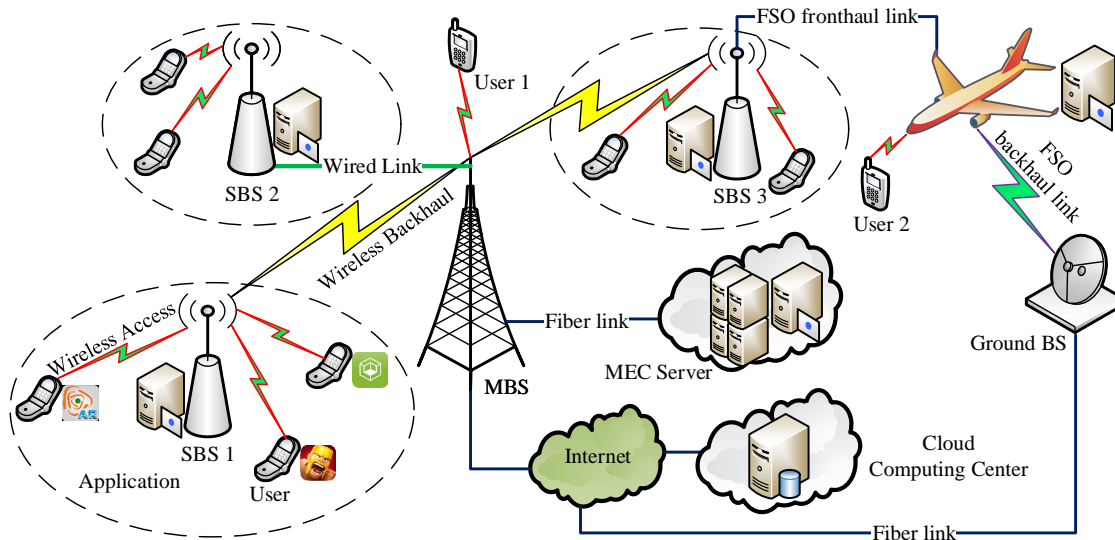

Figure 3. Proposed architecture of MEC-empowered HetNets.

To augment the computation capabilities of the MEC-empowered HetNet and guarantee the coverage and services for natural disaster and emergency situations, UAV-mounted BSs with computation resources are considered as a part of the envisioned architecture. Two potential use scenarios are considered in the following as demonstrations of *UAV-enabled MEC-empowered HetNets*. First, since after a great earthquake, the traditional terrestrial communication infrastructure can be either damaged or partially dysfunctional, UAVs could be used (*i*) to collect information, e.g., pictures of buildings and factories, for remote assessments of site damage; (*ii*) to find people in the aftermath of the earthquake for emergent rescue operations; and (*iii*) to provide fundamental services to MUs. Another use case of UAVs is for crowd events such as a soccer match. During the game, the existing MEC servers may not be able to satisfy massive computations and to provide additional services; therefore, more UAV-mounted MEC servers could be used (*i*) to carry out the surveillance of illegal activities and (*ii*) to generate 360-degree football match in virtual reality.

Similar to the task offloading from MUs to the MEC server, MUs can, either directly or indirectly, offload to the UAV-mounted MEC server. For instance, since there does not exist any SBS/MBS for MU 1 to associate, the UAV adjusts its altitude and location close to and handle MU 1's computation tasks, MUs in the small cell 3 can offload their computation tasks to the UAV-mounted MEC server through the SBS 3 in the meanwhile. Recently, free-space optical (FSO) communication is very attractive for a wide range of applications including the backhauling/fronthauling for wireless networks [52]. Major advantages of FSO communication come from ease of deployment, license-exempt long-range operation (up to several kilometers), high data rate (beyond 10 Gbps), robustness to electromagnetic interference, increased security, and full-duplex operation [39,52]. We propose utilizing FSO links for the fronthaul between the UAV and SBS. It is stressed again that the UAV cannot endure over a long time due to its finite on-board energy, thus being not able to process energy-hungry tasks. In this scenario, the UAV can act as a cellular-connected user and offload its computationally heavy

tasks to the ground MEC server for computation execution. Furthermore, the UAV can establish the FSO backhaul link with the ground control station (GCS) to receive control signals and feedback collected information and statistic data. Similar to multi-tier computing, UAVs can be orchestrated to multi-coverage scales when flying vehicles are positioned at different altitudes (low-altitude as a UAV, mid-altitude as a balloon, and high-altitude as a satellite).

4.4. Open Challenges and Potential Works

Potential capabilities of MEC and HetNets converge on the proposed architecture of MEC-empowered HetNets with the integration of both wired and wireless backhaul and the UAV consideration, and, at the same time, the envisioned architecture introduces new challenges and open research issues.

4.4.1. Challenge 1: Joint Interference Management and Offloading Decision

In HetNets, the wireless backhaul capacity could be seriously degraded when ICI becomes significant. On the one hand, the wireless backhaul is able to support rates of several Gbps that is, however, much more stringent compared to that of the traditional wired backhaul, which can provide rates of hundreds of Gbps. On the other hand, while the wired backhaul capacity is fixed, it is highly flexible to allocate the bandwidth resource for the wireless backhaul. For instance, the macrocell provides wireless backhaul for multiple small cells and simultaneously allocates resources to different wireless backhaul connections, e.g., bandwidth partitioning for wireless and access transmissions and time division for different small cells sharing the same backhaul with the macrocell. In MEC-empowered HetNets, the transmit power of SBSs and bandwidth partitioning factor jointly affect the wireless backhaul capacity; the offloading decision should be determined according to, for example, computation task characteristics, the transmit power of SBSs, the bandwidth allocation factor, and the maximum computational capability at the MEC server. In other words, a joint problem, where interference management, offloading decision, and resource (bandwidth and computation) allocation problem are jointly optimized, is needed to improve the network performance.

As ICI is the major obstacle to gain high performance in conventional multi-cell networks as well as MEC-empowered HetNets, this calls for efficient schemes to mitigate ICI among adjacent MEC servers. Several techniques can be utilized to avoid ICI in MEC-empowered HetNets, as described below.

Fractional Frequency Reuse (FFR): The basic mechanism of FFR is to partition the cell space into multiple regions and assigns each region a distinct frequency band. In such the way, the cell-edge users do not interfere with cell-center users and may not interfere with cell-edge users of adjacent cells, thus improving the quality of receive signals. Through the optimal FFR, inter-cell and intra-cell interference can be significantly reduced.

Power Control: To harness ICI, adaptively adjusting the transmit power of MUs is usually used in the literature. This is simply because increasing the transmit power of an offloading MU can increase the transmission rate and the offloading time as well. Hence, the task's completion deadline is not violated and the MU can benefit from computation offloading. Nevertheless, increasing the transmit power would seriously interfere with MUs in other cells and, as a consequence, MUs are not profitable from remote execution while the computation tasks cannot be handled locally due to the limitation on device capabilities. Therefore, compared to single-server MEC networks, power control in multi-server MEC networks are more difficult to resolve because of (*i*) the coupling with the offloading decisions of MUs in adjacent MEC servers and (*ii*) the computation task requirements.

3D Beamforming: Most research on MEC simply assumed that the output result size is small and the downlink transmission from the MEC servers is with the high transmit power, so the downlink latency and energy consumption for the return of computational results back to the MU are negligible [1]. This would not be correct all the time, where the computational result is comparable and the downlink transmission must be considered. On the one hand, a large number of MUs can instantly request computation offloading and the output can be multicast from the MEC server for a group of

MUs, e.g., in augmented reality applications [53]. On the other hand, MUs can be distributed over not only the horizontal space but also the vertical dimension, such as in urban areas with many skyscrapers. In such a case, 3D beamforming is particularly suitable due to the capability to dynamically control the radio patterns on both vertical and horizontal dimensions (i.e., adapting different downtilts for center-zone and edge-zone MUs). One interesting idea is to consider the joint optimization of downlink and uplink with the optimal control of the transmit power of the MEC servers and the antenna downtilt.

Optimization of UAV Deployment: UAVs can be deployed in cellular networks as either flying MUs in the cellular-connected communication mode or UAV-mounted MEC servers in the UAV-assisted wireless communication mode. However, the UAV transmission may cause ICI to ground MUs in other cells or the transmission for UAV can interfere with other MUs in the uplink. Successful implementations of UAVs could be possible if the challenges of interference management are addressed. An effective solution is to optimize the UAV trajectory so that UAVs can establish the line-of-sight connectivities to the associated terrestrial BS (in the cellular-connected mode) or the ground MUs (in the UAV-assisted mode), and concurrently avoid interference to ground MUs. This could provide substantial gains such as increasing the transmission rate and benefits for the ground MUs from computation offloading. An alternative approach is to enable UAVs to associate with the best ground server thanks to the dense deployment of MEC servers. Transmit power of UAVs can be also optimized, thus alleviating ICI, and boosting the UAV endurance. While the high controllable mobility of UAVs offers many opportunities to optimize the network, it poses the challenges of distributed, online, and autonomous algorithms for resource allocation in UAV-enabled MEC networks.

4.4.2. Challenge 2: Joint User Association and Offloading Decision

In practice, small cells connect to the macrocell through distinct backhaul solutions. For small cells with wireless backhaul, besides the bandwidth partitioning factor, the transmit power is of critical importance in determining the backhaul capacity [1]. Additionally, small cells are densely deployed in the 5G network and an MU may be under the coverage of multiple small cells. Therefore, each offloading MU needs to select its preferred SBS with regard to the quotas of different cells, channel conditions, and backhaul qualities. To find the offloading decision of MUs in a cell, the corresponding BS can optimally partition the bandwidth such that the backhaul capacity constraint is not violated and the computation resource at the MEC server is fairly shared among offloading users.

Since the centralized optimization of offloading decisions requires that all MUs report their local information to the central entity, which may result in the huge network-wide computation overhead. In addition, heuristic schemes are often adopted to solve the centralized computation offloading and the central entity is not always feasible in wireless networks. Therefore, distributed approaches are highly desirable, where each MU can independently decide the offloading decision with local information only and/or limited information exchange. In such contexts, matching theory, game theory, and machine learning are promising tools [1], where the individual objective captures the channel qualities, task characteristics, and the cost of backhaul links. One must consider that the design of computation offloading policies should take account of the network dynamics such as the highly time-varying channel qualities, the dynamic task arrival rate, the number of users requesting for task offloading, and the number of MEC servers. In addition, the integration of communication, computation, caching, and control is vital for the successful implementation of many applications and services in the emerging 5G network. Therefore, one difficulty facing in dynamic MEC-empowered HetNets is the long-term optimization of a huge amount of variables. Accordingly, conventional techniques, e.g., one-shot optimization, become unsuitable, but long-term approaches, e.g., Markov decision process and machine learning, are particularly suitable. For example, MUs determine the offloading decisions and/or the server selection at the current network state by learning from the interaction with other MUs and their past experience, machine learning can be used to enable MEC servers to predict the MUs' behaviors in order to efficiently allocate computation resource, and finally

UAVs can use machine learning to optimize their trajectory and optimization variables such as the transmit power and allocated computation resource.

4.4.3. Challenge 3: Hierarchical and Collaborative Computation Offloading

The development of 5G is mainly driven by three categories: enhanced mobile broadband, massive IoT, and mission-critical services [29,54]. The first class targets at an extremely high data rate and extreme capacity, the second one is to serve a massive number of IoT devices with requirements of ultra-high density, ultra-low energy, and ultra-high scalability, and the last class aims for critical applications with high requirements on strong security, ultra-high reliability and ultra-low latency (ultra-reliable and low-latency communication uRLLC). In addition, massive and heterogeneous MUs in the emerging 5G network are distinctly different in, e.g., battery size, computation capability, and targeted applications. Consider a network with two IoT devices as the illustration; the former runs an application with a requirement of ultra-low 1 ms end-to-end latency and the latter continuously observes environmental conditions such as temperature, humidity, and pictures, and deliver to the aggregator for processing the raw data. As opposed to the first IoT device, the second device can generate big data with huge computation requirement for data processing but can be tolerant of latency. In such a case, IoT devices (second one in the above example) with computation-intensive and latency-tolerant applications prefer to offload their computations to the high-tier servers due to their huge computation capabilities, e.g., central clouds and macro-MEC servers. In contrast, IoT devices with latency-critical applications highly prefer to utilize the low-tier MEC servers due to the close proximity, e.g., macro- and small-MEC servers.

Moreover, the MEC servers are likely to have comparable size and computation capabilities as a desktop computer; therefore, some computation tasks cannot be strictly handled by only one MEC server. This motivates different servers to collaborate in executing the same enormous computations in order to improve computing performance. Hierarchical and collaborative computation will become much more complicated when jointly optimizing with user association and resource allocation for wireless backhaul. Furthermore, compared to MUs with latency-tolerant applications, those with latency-sensitive applications get higher priorities to access the MEC server with higher computation resources and better connection qualities.

4.4.4. Challenge 4: Joint User Association and MEC Server Selection

When the MBS is a multi-operator wireless tower, it is possible that multiple MEC servers are physically collocated in the same site. Besides the user association problem due to the various backhaul solutions, an efficient scheme to select the suitable MEC server for offloading users is highly necessary to investigate. For single-user networks, MUs strongly prefer to connect with the small cell with wired backhaul and the best channel quality and then offload to the MEC server with the highest computation capability. However, the practical network scenario is multi-user and small cells normally have the limited quota of few users [1]. Consequently, one can design that each MU acts according to its own interest and makes the offloading decision independently as in non-cooperative game, where each MU always tries to maximize its own benefit by selecting the best small cell and MEC server, with the objective of, for example, minimizing (1) the energy consumption subject the maximum latency and (2) the computation time with the power budget constraints.

4.4.5. Challenge 5: NOMA-Assisted MEC Networks

MEC was originally proposed to improve the computation capabilities of computation- and battery-limited end devices so that various computation-intensive and latency-critical applications can be deployed. By 2020, the number of devices connected to the Internet will exceed 50 billion and there will be more than several hundred millions of sensors deployed in the world. One can foresee that a large number of devices simultaneously integrate their computation-intensive tasks into a single MEC server. To serve multiple MUs, bandwidth and time resources can be divided into orthogonal parts

and assigned to different MUs for computation offloading. Recently, some studies on non-orthogonal multiple access (NOMA) enabled MEC networks have reported that the combination of NOMA and MEC can reduce the energy consumption and latency [55,56]. However, existing works are limited to specific scenarios of single-server MEC networks. For example, considering the set of offloading users and no ICI, the energy consumption of offloading users is minimized through jointly optimizing user clustering, resource allocation, and transmit power control per cluster [55].

As more and more MEC servers are deployed, ICI becomes the major obstacle since ICI strictly limits the performance of cell-edge users. To achieve the full benefits of NOMA in MEC-empowered HetNets, user clustering, power allocation, and user scheduling (UCPAUS) are important issues, which are however coupled with each other, thus solving the problem really challenging [57]. One solution is to consider only one cluster in each cell and then allocate the transmit power of MUs [58]. However, this solution has an immense complexity because signals of all the same-cell MUs are superimposed at the transmitter side and largely-superimposed signals are decoded at the receiver side. An alternative solution is to have multiple clusters in each cell, assign each cluster to a subchannel, and adapt power allocation in each cluster. This could reduce the complexity and hardware requirements compared to the single-cluster approach, but sub-optimal in the performance. The joint UCPAUS problem becomes highly complicated in MEC-empowered HetNets, as explained below. First, each MU must decide to offload or not in the case of binary offloading or determine the fraction of offloadable computation in the case of partial offloading. Second, determining the optimal number of clusters and how to select MUs to form a cluster is non-trivial. For instance, a clustering scheme can assign two latency-critical MUs to a cluster and two other latency-tolerant MUs to another cluster; this could improve the network performance, but the completion deadlines of latency-critical MUs may not be met [56]. Therefore, one suitable solution is to group one latency-critical MU with one latency-tolerant MU together. Lastly, the performance of MEC networks is highly affected by the computation resource allocation at MEC servers. As a result, studies on offloading decision, user clustering, and resource allocation are strongly desired to further improve the performance of NOMA-enabled MEC networks, which require efficient but close-optimal schemes for the successful deployment of MEC and NOMA in the emerging 5G network and beyond.

In summary, in MEC-empowered UAV-assisted HetNets, user association is of central importance since MUs can choose their preferred small cells to associate with and offload to. This is motivated by heterogeneous wireless backhaul solutions among different small cells. In addition, computation resource allocation should receive enough attention since the task requirement and computation overhead are highly affected by the assigned resource. We must stress that our main focus in this paper is on an envisioned architecture of MEC-empowered UAV-assisted HetNets. However, many unexplored issues need to be thoroughly considered and discussed, e.g., content caching and delivery, implementation of real applications such as augmented and virtual reality, optimal deployment of MEC servers, and planning MEC networks with UAVs (i.e., how many UAV-mounted MEC servers are needed to provide satisfactory computation to MUs and how to place UAV-mounted MEC servers in a geographical area). This calls for further efforts from the research community for issues and challenges in MEC-empowered HetNets.

5. Conclusions

In this article, we have described the evolution of computation offloading and resource allocation in single-server and multi-server MEC networks. This has been followed by our envisioned architecture of MEC-empowered UAV-assisted HetNets, which is able to support the integration of both wired and wireless backhaul, the coexistence of conventional central clouds and distributed MEC, and the deployment of UAVs as flying-MEC servers. Finally, we have highlighted several challenges and open issues stemmed from our proposed architecture.

References

1. Pham, Q.V.; Anh, T.L.; Tran, N.H.; Park, B.J.; Hong, C.S. Decentralized Computation Offloading and Resource Allocation for Mobile-Edge Computing: A Matching Game Approach. *IEEE Access* **2018**, *6*, 75868–75885. [CrossRef]

2. Nguyen, T.D.T.; Nguyen, T.D.; Nguyen, V.D.; Pham, X.Q.; Huh, E.N. Cost-Effective Resource Sharing in an Internet of Vehicles-Employed Mobile Edge Computing Environment. *Symmetry* **2018**, *10*, 594. [CrossRef]

3. Tran, T.X.; Hajisami, A.; Pandey, P.; Pompili, D. Collaborative Mobile Edge Computing in 5G Networks: New Paradigms, Scenarios, and Challenges. *IEEE Commun. Mag.* **2017**, *55*, 54–61. [CrossRef]

4. Pham, X.Q.; Nguyen, T.D.; Nguyen, V.; Huh, E.N. Joint Node Selection and Resource Allocation for Task Offloading in Scalable Vehicle-Assisted Multi-Access Edge Computing. *Symmetry* **2019**, *11*, 58. [CrossRef]

5. Pham, Q.V.; Hwang, W.J. Energy-efficient power control for uplink spectrum-sharing heterogeneous networks. *Int. J. Commun. Syst.* **2018**, *31*, e3717. [CrossRef]

6. Pham, Q.V.; Le, L.B.; Chung, S.H.; Hwang, W.J. Mobile Edge Computing With Wireless Backhaul: Joint Task Offloading and Resource Allocation. *IEEE Access* **2019**, *7*, 16444–16459. [CrossRef]

7. Wang, C.; Yu, F.R.; Liang, C.; Chen, Q.; Tang, L. Joint Computation Offloading and Interference Management in Wireless Cellular Networks with Mobile Edge Computing. *IEEE Trans. Veh. Technol.* **2017**, *66*, 7432–7445. [CrossRef]

8. Sardellitti, S.; Scutari, G.; Barbarossa, S. Joint Optimization of Radio and Computational Resources for Multicell Mobile-Edge Computing. *IEEE Trans. Signal Inf. Process. Networks* **2015**, *1*, 89–103. [CrossRef]

9. Zhao, M.; Zhou, K. Selective Offloading by Exploiting ARIMA-BP for Energy Optimization in Mobile Edge Computing Networks. *Algorithms* **2019**, *12*, 48. [CrossRef]

10. Liu, B.; Xu, H.; Zhou, X. Resource Allocation in Wireless-Powered Mobile Edge Computing Systems for Internet of Things Applications. *Electronics* **2019**, *8*, 206. [CrossRef]

11. Zhang, J.; Xia, W.; Yan, F.; Shen, L. Joint Computation Offloading and Resource Allocation Optimization in Heterogeneous Networks with Mobile Edge Computing. *IEEE Access* **2018**, *6*, 19324–19337. [CrossRef]

12. Lobillo, F.; Becvar, Z.; Puente, M.A.; Mach, P.; Lo Presti, F.; Gambetti, F.; Goldhamer, M.; Vidal, J.; Widiawan, A.K.; Calvanesse, E. An architecture for mobile computation offloading on cloud-enabled LTE small cells. In Proceedings of the IEEE Wireless Communications and Networking Conference Workshops (WCNCW), Istanbul, Turkey, 6–9 April 2014; pp. 1–6.

13. Puente, M.A.; Becvar, Z.; Rohlik, M.; Lobillo, F.; Strinati, E.C. A Seamless Integration of Computationally-Enhanced Base Stations into Mobile Networks towards 5G. In Proceedings of the 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), Glasgow, UK, 11–14 May 2015; pp. 1–5.

14. Plachy, J.; Becvar, Z.; Strinati, E.C. Dynamic resource allocation exploiting mobility prediction in mobile edge computing. In Proceedings of the 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, Spain, 4–8 September 2016; pp. 1–6.

15. Plachy, J.; Becvar, Z.; Mach, P. Path selection enabling user mobility and efficient distribution of data for computation at the edge of mobile network. *Comput. Netw.* **2016**, *108*, 357–370. [CrossRef]

16. Ksentini, A.; Taleb, T.; Chen, M. A Markov Decision Process-based service migration procedure for follow me cloud. In Proceedings of the 2014 IEEE International Conference on Communications (ICC), Sydney, Australia, 10–14 June 2014; pp. 1350–1354.

17. Wang, S.; Urgaonkar, R.; He, T.; Chan, K.; Zafer, M.; Leung, K.K. Dynamic Service Placement for Mobile Micro-Clouds with Predicted Future Costs. *IEEE Trans. Parallel Distrib. Syst.* **2017**, *28*, 1002–1016. [CrossRef]

18. Guo, F.; Zhang, H.; Ji, H.; Li, X.; Leung, V.C.M. An Efficient Computation Offloading Management Scheme in the Densely Deployed Small Cell Networks With Mobile Edge Computing. *IEEE/ACM Trans. Netw.* **2018**, *26*, 2651–2664. [CrossRef]

19. Huynh, L.N.T.; Pham, Q.V.; Nguyen, Q.D.; Pham, X.Q.; Nguyen, V.; Huh, E.N. Energy-Efficient Computation Offloading with Multi-MEC Servers in 5G Two-Tier Heterogeneous Networks. In *Proceedings of the 13th International Conference on Ubiquitous Information Management and Communication (IMCOM) 2019*; Lee, S., Ismail, R., Choo, H., Eds.; Springer International Publishing: Phuket, Thailand, 2019; pp. 120–129.

20. Yang, L.; Zhang, H.; Li, X.; Ji, H.; Leung, V.C.M. A Distributed Computation Offloading Strategy in Small-Cell Networks Integrated With Mobile Edge Computing. *IEEE/ACM Trans. Netw.* **2018**, *26*, 2762–2773. [CrossRef]

21. Dong, C.; Wen, W. Joint Optimization for Task Offloading in Edge Computing: An Evolutionary Game Approach. *Sensors* **2019**, *19*, 740. [CrossRef] [PubMed]

22. Huang, L.; Feng, X.; Zhang, C.; Qian, L.; Wu, Y. Deep reinforcement learning-based joint task offloading and bandwidth allocation for multi-user mobile edge computing. *Digit. Commun. Netw.* **2019**, *5*, 10–17. [CrossRef]

23. Huang, L.; Feng, X.; Zhang, L.; Qian, L.; Wu, Y. Multi-Server Multi-User Multi-Task Computation Offloading for Mobile Edge Computing Networks. *Sensors* **2019**, *19*, 1446. [CrossRef]

24. Albanese, M.; Erbacher, R.F.; Jajodia, S.; Molinaro, C.; Persia, F.; Picariello, A.; Sperlì, G.; Subrahmanian, V.S., Recognizing Unexplained Behavior in Network Traffic. In *Network Science and Cybersecurity*; Pino, R.E., Ed.; Springer: New York, NY, USA, 2014; pp. 39–62.

25. Roman, R.; Lopez, J.; Mambo, M. Mobile edge computing, Fog et al.: A survey and analysis of security threats and challenges. *Future Gener. Comput. Syst.* **2018**, *78*, 680 – 698. [CrossRef]

26. Xiao, L.; Wan, X.; Dai, C.; Du, X.; Chen, X.; Guizani, M. Security in Mobile Edge Caching with Reinforcement Learning. *IEEE Wirel. Commun.* **2018**, *25*, 116–122. [CrossRef]

27. Xu, J.; Yao, J. Exploiting Physical-Layer Security for Multiuser Multicarrier Computation Offloading. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 9–12. [CrossRef]

28. Mach, P.; Becvar, Z. Mobile Edge Computing: A Survey on Architecture and Computation Offloading. *IEEE Commun. Surv. Tutorials* **2017**, *19*, 1628–1656. [CrossRef]

29. Pham, Q.V.; Fang, F.; Ha, V.N.; Le, M.; Ding, Z.; Le, L.B.; Hwang, W.J. A Survey of Multi-Access Edge Computing in 5G and Beyond: Fundamentals, Technology Integration, and State-of-the-Art. *arXiv* **2019**, arXiv:1906.08452.

30. Porambage, P.; Okwuibe, J.; Liyanage, M.; Ylianttila, M.; Taleb, T. Survey on Multi-Access Edge Computing for Internet of Things Realization. *IEEE Commun. Surv. Tutorials* **2018**, *20*, 2961–2991. [CrossRef]

31. Mao, Y.; You, C.; Zhang, J.; Huang, K.; Letaief, K.B. A Survey on Mobile Edge Computing: The Communication Perspective. *IEEE Commun. Surv. Tutorials* **2017**, *19*, 2322–2358. [CrossRef]

32. Zhang, J.; Hu, X.; Ning, Z.; Ngai, E.C..; Zhou, L.; Wei, J.; Cheng, J.; Hu, B. Energy-Latency Tradeoff for Energy-Aware Offloading in Mobile Edge Computing Networks. *IEEE Internet Things J.* **2018**, *5*, 2633–2645. [CrossRef]

33. Wang, Y.; Sheng, M.; Wang, X.; Wang, L.; Li, J. Mobile-Edge Computing: Partial Computation Offloading Using Dynamic Voltage Scaling. *IEEE Trans. Commun.* **2016**, *64*, 4268–4282. [CrossRef]

34. Dinh, T.Q.; Tang, J.; La, Q.D.; Quek, T.Q.S. Offloading in Mobile Edge Computing: Task Allocation and Computational Frequency Scaling. *IEEE Trans. Commun.* **2017**, *65*, 3571–3584.

35. Lyu, X.; Tian, H.; Sengul, C.; Zhang, P. Multiuser joint task offloading and resource optimization in proximate clouds. *IEEE Trans. Veh. Technol.* **2017**, *66*, 3435–3447. [CrossRef]

36. You, C.; Huang, K.; Chae, H.; Kim, B.H. Energy-Efficient Resource Allocation for Mobile-Edge Computation Offloading. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 1397–1411. [CrossRef]

37. Bi, S.; Zhang, Y.J.A. Computation Rate Maximization for Wireless Powered Mobile-Edge Computing with Binary Computation Offloading. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 4177–4190. [CrossRef]

38. Ranadheera, S.; Maghsudi, S.; Hossain, E. Computation Offloading and Activation of Mobile Edge Computing Servers: A Minority Game. *IEEE Wirel. Commun. Lett.* **2018**, *7*, 688–691. [CrossRef]

39. Khalighi, M.A.; Uysal, M. Survey on Free Space Optical Communication: A Communication Theory Perspective. *IEEE Commun. Surv. Tutorials* **2014**, *16*, 2231–2258. [CrossRef]

40. Zeng, Y.; Zhang, R.; Lim, T.J. Wireless communications with unmanned aerial vehicles: Opportunities and challenges. *IEEE Commun. Mag.* **2016**, *54*, 36–42. [CrossRef]

41. Li, B.; Fei, Z.; Zhang, Y. UAV Communications for 5G and Beyond: Recent Advances and Future Trends. *IEEE Internet Things J.* **2019**, *6*, 2241–2263. [CrossRef]

42. Ullah, H.; Gopalakrishnan Nair, N.; Moore, A.; Nugent, C.; Muschamp, P.; Cuevas, M. 5G Communication: An Overview of Vehicle-to-Everything, Drones, and Healthcare Use-Cases. *IEEE Access* **2019**, *7*, 37251–37268. [CrossRef]

43. Mozaffari, M.; Saad, W.; Bennis, M.; Nam, Y.; Debbah, M. A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems. *IEEE Commun. Surv. Tutor. arXiv* **2019**. arXiv:1803.00680.

44. Merwaday, A.; Guvenc, I. UAV assisted heterogeneous networks for public safety communications. In Proceedings of the 2015 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), New Orleans, LA, USA, 9–12 March 2015; pp. 329–334.

45. Zhang, X.; Duan, L. Fast Deployment of UAV Networks for Optimal Wireless Coverage. *IEEE Trans. Mob. Comput.* **2019**, *18*, 588–601. [CrossRef]

46. Wang, H.; Ding, G.; Gao, F.; Chen, J.; Wang, J.; Wang, L. Power Control in UAV-Supported Ultra Dense Networks: Communications, Caching, and Energy Transfer. *IEEE Commun. Mag.* **2018**, *56*, 28–34. [CrossRef]

47. Cheng, N.; Xu, W.; Shi, W.; Zhou, Y.; Lu, N.; Zhou, H.; Shen, X. Air-Ground Integrated Mobile Edge Networks: Architecture, Challenges, and Opportunities. *IEEE Commun. Mag.* **2018**, *56*, 26–32. [CrossRef]

48. Sharma, V.; Bennis, M.; Kumar, R. UAV-Assisted Heterogeneous Networks for Capacity Enhancement. *IEEE Commun. Lett.* **2016**, *20*, 1207–1210. [CrossRef]

49. Sharma, V.; Sabatini, R.; Ramasamy, S. UAVs Assisted Delay Optimization in Heterogeneous Wireless Networks. *IEEE Commun. Lett.* **2016**, *20*, 2526–2529. [CrossRef]

50. Lin, X.; Xia, J.; Wang, Z. Probabilistic caching placement in UAV-assisted heterogeneous wireless networks. *Phys. Commun.* **2019**, *33*, 54–61. [CrossRef]

51. Guo, H.; Liu, J. Collaborative Computation Offloading for Multiaccess Edge Computing Over Fiber-Wireless Networks. *IEEE Trans. Veh. Technol.* **2018**, *67*, 4514–4526. [CrossRef]

52. Alzenad, M.; Shakir, M.Z.; Yanikomeroglu, H.; Alouini, M. FSO-Based Vertical Backhaul/Fronthaul Framework for 5G+ Wireless Networks. *IEEE Commun. Mag.* **2018**, *56*, 218–224. [CrossRef]

53. Al-Shuwaili, A.; Simeone, O. Energy-Efficient Resource Allocation for Mobile Edge Computing-Based Augmented Reality Applications. *IEEE Wirel. Commun. Lett.* **2017**, *6*, 398–401. [CrossRef]

54. Agiwal, M.; Roy, A.; Saxena, N. Next Generation 5G Wireless Networks: A Comprehensive Survey. *IEEE Commun. Surv. Tutorials* **2016**, *18*, 1617–1655. [CrossRef]

55. Kiani, A.; Ansari, N. Edge Computing Aware NOMA for 5G Networks. *IEEE Internet Things J.* **2018**, *5*, 1299–1306. [CrossRef]

56. Ding, Z.; Fan, P.; Poor, H.V. Impact of Non-Orthogonal Multiple Access on the Offloading of Mobile Edge Computing. *IEEE Trans. Commun.* **2019**, *67*, 375–390. [CrossRef]

57. Shin, W.; Vaezi, M.; Lee, B.; Love, D.J.; Lee, J.; Poor, H.V. Non-Orthogonal Multiple Access in Multi-Cell Networks: Theory, Performance, and Practical Challenges. *IEEE Commun. Mag.* **2017**, *55*, 176–183. [CrossRef]

58. Pham, Q.V.; Hwang, W.J. α-Fair resource allocation in non-orthogonal multiple access systems. *IET Commun.* **2018**, *12*, 179–183. [CrossRef]