On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration

Tarik Taleb, *Senior Member, IEEE*, Konstantinos Samdanis, Badr Mada,
Hannu Flinck, Sunny Dutta, and Dario Sabella

Abstract—Multi-access edge computing (MEC) is an emerging ecosystem, which aims at converging telecommunication and IT services, providing a cloud computing platform at the edge of the radio access network. MEC offers storage and computational resources at the edge, reducing latency for mobile end users and utilizing more efficiently the mobile backhaul and core networks. This paper introduces a survey on MEC and focuses on the fundamental key enabling technologies. It elaborates MEC orchestration considering both individual services and a network of MEC platforms supporting mobility, bringing light into the different orchestration deployment options. In addition, this paper analyzes the MEC reference architecture and main deployment scenarios, which offer multi-tenancy support for application developers, content providers, and third parties. Finally, this paper overviews the current standardization activities and elaborates further on open research challenges.

Index Terms—MEC, Multi-Access Edge Computing, Mobile Edge Computing, 5G Mobile Communications, Edge Computing, Application Virtualization, Network Function Virtualization, Software Defined Networking, Network Softwarization.

I. INTRODUCTION

IN THE recent years, the Telco world has undergone a tremendous growth in developing mobile communications technologies. With highly capable end-devices (e.g., smartphones and tablets) and new interactive mobile applications in place, we are in the midst of a data revolution. This is also reflected by the fact that 63% of the world population has currently a mobile subscription, whereas in the previous decade it was just 20% [1]. Services such as video, music, social networking, gaming and other interactive applications are gaining momentum. The usage of mobile applications (APPs) are expected to further contribute to this trend, on top of the fact that mobile networks would also provide a means of broadband connectivity to a number of emerging services such as augmented reality, resulting in a foreseen data traffic growth of 8-fold from 2015 to 2020 [2]. In addition, the emerging Internet of Things (IoT) and Machine-Type-Communications (MTC) are expected to introduce a huge number of machine connections, reaching 1 billion by 2020, defining promising business opportunities for mobile operators, yet impacting significantly the development of their mobile networks [1].

As such thriving range of new and diverse services is becoming an integral part of the mobile users entertainment and social life, the expectations towards immersive Quality of Experience (QoE) are also increasing. Currently 3GPP introduced a Stage 1 study, commonly known as SMARTER (New Services and Market Enablers), on the market segments and vertical services that should be addressed by the emerging 5G networks, identifying the main use cases to be considered [3]. The related performance requirements include the support of up to 1000 times higher data volumes, data rates up to 10 Gb/s, very low service level latency below 5ms, ubiquitous communicating things and mass connectivity supporting 300,000 devices within a single cell, ultra-high reliability of 99.999% (i.e., five nine availability) and reduced energy consumption by 90% [4], [5].

Such soaring demands for data services are now posing new challenges on the service providers as well as on mobile network operators. The vision of future 5G systems is to provide a customized and advanced user centric value at an affordable price, enabling context-aware and proximity services, service delivery in crowded areas and on the move, and advanced multimedia centric services. Such emerging 5G ecosystem will encompass a heterogeneous communication landscape integrating networking and Information Technology (IT) resources with cloud enabled services via the means of virtualization and softwarization [6]–[10]. Innovative service-oriented edge-cloud architectures will pave the way for enhanced QoE, while network virtualization will be the catalyst in supporting multi-service and multi-tenancy, enabling in this way efficient network operations and service provisioning.

ETSI has initiated Mobile Edge Computing (MEC) standardization to promote and accelerate the advancement of edge-cloud computing in mobile networks, by launching the MEC Industry Specification Group (ISG) in December 2014. The objective of ETSI MEC ISG is to create an open environment across multi-vendor cloud platforms located at the edge of the RAN, accessible by application/service providers and third parties in an effort to overcome the challenges related with centralized cloud computing environments and that is in terms of latency, and assurance of higher speeds [11]. By pushing data intensive tasks towards the edge and locally processing data in proximity to the users, mobile network operators can reduce traffic bottlenecks in the core and backhaul networks, while assisting in the offload of heavy computational tasks from power constrained User Equipment (UE) to the edge. In general, such decentralized cloud architecture constitutes a pillar technology for the emerging 5G systems transforming the legacy mobile base stations by offering cloud computing abilities and an IT service environment at the edge of the network. Since September 2016, ETSI ISG has dropped the 'Mobile' out of MEC and renamed it as Multi-access Edge Computing in order to broaden its applicability into heterogeneous networks including WiFi and fixed access technologies [12].

MEC enables new services and yields promising business models to the major stakeholders, namely Mobile Network Operators (MNOs), application developers, Over-The-Top (OTT) players, vertical business segments, Independent Software Vendors (ISVs), IT platform vendors and system integrators as elaborated in [13] and [14]. However, MEC is still in its early development phase and realizing its potential is significant to take full advantage of its benefits. For example, MNOs can increase revenues by providing open access of the MEC platforms (via specific Application Programming Interfaces-APIs) to service providers and imply usage-based charging for utilizing storage, network bandwidth, and computation resources (e.g., CPU). The most obvious advantage of edge computing, namely the reduction in delay, will motivate the deployment of diverse services (e.g., IoT/ M2M, 4K Ultra High Definition - UHD - video, and mobile serious gaming) [15], hence opening a wider market for application developers and content providers. MEC can also offer application providers local context awareness (e.g., Radio Access Network - RAN - analytics, traffic characteristics, and device location) to improve the operational efficiency of the offered services.

MEC has also gained momentum in the academia side, with a number of surveys published [15]–[17], focusing on the usage requirements and architectures, different deployment approaches, and applications that can benefit from edge cloud. However, to the best of the authors' knowledge, there is no survey that summarizes mobile edge-cloud as a 5G technology, above all considering the latest ETSI MEC ISG architecture, with respect to service orchestration and software defined control. This paper details the main enablers of MEC including virtualization technologies such as Virtual Machines (VMs) and containers, Software Defined Networking (SDN) and Network Function Virtualization (NFV) as well as network slicing. In addition, it provides an analysis of the MEC orchestration considering standalone services, service mobility, joint network and service optimization as well as a comprehensive study of current orchestrator deployment options. Finally, this paper provides an exhaustive state-of-the art study on related research in the area of edge-cloud computing, considering the impact on various applications and highlighting open research challenges.

The remainder of this paper is organized as follows. Section II presents the evolution of edge cloud considering cloudlets, fog computing and MEC. Section III provides an overview of the MEC and edge-cloud use cases, highlighting the main research fields and achievements. Section IV describes the MEC fundamental enablers and Section V elaborates the MEC framework and reference architecture considering also other standardization activities, particularly those relevant to OpenFog and 3GPP. Section VI analyzes the incentives for orchestration considering both individual services and an edge-cloud platform network. Section VII sheds light on the different MEC deployment scenarios analyzing also the orchestration deployment options. Section VIII presents a summary of the lessons learned and highlights research challenges. Finally, the paper concludes in Section IX.

II. EVOLUTION TOWARDS MOBILE EDGE COMPUTING

Since the beginning of the smart-phone era, managing efficiently the network resources has been a significant challenge for MNOs. With the rise of diverse and complex mobile applications, MNOs needed to additionally deal with cloud computing issues including data storage and processing. From the end-device's perspectives, certain constraints are raised in terms of battery lifetime, computational power and memory limitations, especially for complex processes. To cope with the relevant challenges, Mobile Cloud Computing (MCC) was introduced as a promising approach. MCC is envisioned as an expansion of cloud computing features, i.e., adaptability, scalability, availability and self-awareness, in a mobile computing environment [18], [19]. From the network architecture perspectives, MCC acts as an Infrastructure as a Service (IaaS) for data storage and processing, outside the mobile devices as considered by Dinh *et al.* [20], augmenting in this way the resource capabilities of mobile devices.

Generic cloud providers (e.g., Google, Amazon, and Salesforce) may facilitate third parties, e.g., OTT providers, with computing resources including IaaS (e.g., servers, networks, and storages), Platform as a Service (PaaS) (e.g., middle-ware services and operating systems) and Software as a Service (SaaS) (e.g., application programs) to develop applications with a lower cost. This paradigm facilitates easy development of new applications (i.e., APPS) and services, e.g., m-learning [21], m-gaming [22], and m-health care [23], which, in combination with the evolution of MCC, can transform the UE into a tool for learning, entertainment, social networking, updating news and businesses. Despite such potential usage, public cloud facilities failed in fulfilling latency requirements due to long response times, due, in turn, to the centralized cloud architecture model, affecting the QoE for end users.

Considering these weaknesses, the concept of edge-cloud was introduced, initially in the form of Cloudlets that assist

UEs to gain computational and storage performance benefits, which later paved the way towards the more sophisticated concepts of fog computing and MEC. Both fog computing and MEC envision an open platform offering similar type of features. Fog computing concentrates on applications, mainly IoT, that take advantage of a platform set that collectively assist UEs. MEC, on the other hand, focuses on application-related enhancements in terms of feedback mechanisms, information and content processing and storage, to mention a few, considering single platforms in the first phase.

A. Cloudlets

In [24] and [25], Satyanarayan *et al.* introduced the concept of cloudlet as the middle layer in a three-tier architecture consisting of an end-device, an edge cloud platform, i.e., cloudlet, and a centralized datacenter. The objective of cloudlet is to extend the remote datacenter cloud services in close proximity to the end users. Cloudlet is viewed as a trusted and resource-rich node with stable Internet connectivity, offering computing as well as access and storage resources to nearby mobile devices. It was proposed as an edge cloud node, which can reside in community places, e.g., coffee shops and shopping malls, and highly populated areas, e.g., train stations and exhibition halls. Mobile devices merely act as thin clients, i.e., lightweight devices which heavily depend on remote access to a cloud server for offloading resource-intense tasks in order to increase execution speed and save energy. Cloudlets are instantiated based on a soft state implementation that relies on Virtual Machines (VMs). Cloudlets merely act as micro datacenter in a box , offering access to end users over Wi-Fi for deploying and managing their own VMs [25].

Using a VM-based approach, cloudlets offer a transient guest environment for individual users, providing isolation from other host software. For transferring service requests from end devices to cloudlets, two distinct approaches were considered. The first one is based on VM migration that focuses on suspending and transferring an existing state of processor, disk and memory of the end device to the destination cloudlet, whereby execution is resumed from the exact state. The second approach considers the delivery of a small VM overlay image instead of the mentioned states. In this case, the cloudlet is ready with a VM base (OS). The overlay image is integrated on top and the execution continues seamlessly from that point. Depending on the application design, the clients are allowed to customize VMs to ensure that applications are executed smoothly; a process that helps to rapidly offload potential workload at run-time. The impact of cloudlets on interactive applications (e.g., video streaming, image processing, and file editing) is analyzed by Fesehaye *et al.* [26] in terms of caching and offloading, considering a content-centric approach for interconnecting cloudlets. Scalability issues relevant to cloudlets are studied in [27].

B. Fog Computing

Fog computing, alternatively known as fog networking or "fogging", is a term originally introduced by Cisco, proposed to enable a cloud computing architecture away from centralized cloud datacenters, considering a large number of geographically wide spread edge nodes as part of a distributed and collaborating cloud. Fog computing envisions cloud nodes residing directly at the edge of the network and being capable to deliver new applications and services, especially for the future Internet and wireless network-based IoT services [28], [29]. Fog computing offers storage in close proximity to the edge and helps to offload traffic load that would, otherwise, need to transverse the mobile backhaul and backbone. It also enables device management and network management such as control and configuration procedures at the edge of the network [30].

The notion of fog computing nodes is wide. Any equipment with processing power and storage, e.g., ranging from wireless access points, switches and routers to base stations and resource-rich datacenters or cloud platforms, can be qualified as a fog node [17]. Cisco introduced the first commercial fog device, IOx [31], capable of hosting applications in a guest operating system running on a hypervisor directly on routers. As a generic cloud platform to develop, compile and execute software, fog computing is considered as an open ecosystem for wearables/IoT, big-data analytics and for emerging services such as automotive, hostile and tactile applications.

C. Multi-Access Edge Computing

MEC enables IT and cloud computing capabilities at the RAN edge in a close proximity to end users, offering an ultra-low latency environment with high bandwidth and real-time access to radio and network analytics. The use of MEC can provide the potential for developing a wide range of new applications and services, bringing innovation and driving new businesses. In particular, contextual information and specific content, proximity and location awareness can create business value opportunities, offering a customized mobile broadband experience. Service providers may also benefit from MEC in collecting more information regarding customers in terms of content, location and interests, in order to differentiate their portfolio or introduce new services or simply use such information for commercial reasons.

MEC offers an open radio network edge platform, facilitating multi-service and multi-tenancy by allowing authorized third-parties to make use of the storage and processing capabilities, introducing new businesses on-demand and in a flexible manner. MNOs can then provide cloud services and additionally monetize the broadband experience providing an insight of RAN and network conditions to third parties to facilitate service enhancements. MEC can also benefit from the ubiquitous coverage of cellular networks to become the key enabler for supporting M2M and IoT services that have become sufficiently mature to shape vertical segments/services including energy utilities, automotive, and smart city services [11]. From the business perspective, MEC should enable a secure cloud platform architecture and should provide APIs to enable third players to share and use it dynamically, easily installing and modifying new services and efficiently interacting with the RAN, e.g., being able to retrieve RAN-relevant information (e.g., for transcoding videos at rates adequate to the current conditions of RAN [32]).

III. MEC Use Cases and Applications

MEC facilitates enhancements to the existing applications and offers tremendous potential for developing a wide range of new and innovative applications by enabling authorized third-parties to make use of local services and caching capabilities at the edge of the RAN. MEC can be seen as a natural evolution of legacy mobile base stations, enabling new business opportunities and revenue streams to operators, vendors and third-parties. This section provides the main MEC use cases and summarizes the related research efforts.

A. Computation Offloading

Computation offloading is a technique whereby a resource-constrained mobile device fully or partially offloads a computation-intensive task to a resource-sufficient cloud environment. Computation offloading is performed mostly to save energy, battery lifetime or due to inability of the end device to process a computation-heavy application. There have been numerous approaches on offloading methods in classical MCC environments. Cyber foraging [33] enables computation offloading, enhancing the capabilities of mobile devices, while notably improving energy efficiency. CloneCloud [34] performs automatic code partitioning at the thread level. Component-level partitioning was considered in Cuckoo [35] and a relevant system was implemented in MAUI [36]. However, due to unbound latency in computation offloading to a remote cloud, the session state may get prolonged. In certain cases, this might consume more power than local processing.

In comparison, energy consumption is less in MEC due to reduced latency and close proximity to the RAN edge. Hence, users experience faster execution and enhanced performance. A comprehensive overview on computation offloading considering small cell edge computing is provided in [37]. For video services, the most compute-intensive part is encoding. During video conferencing when a mobile device tries to make a video call, a negotiation phase is performed for selecting the type of encoding available before the video is encoded in the mobile device and then uploaded. Such a process consumes energy and may take time. As a solution, Beck et al. [38] have proposed a communication protocol to offload the encoding service to an edge server. The encoding part, being offloaded, shall save energy and decrease the latency in connectivity ensuring an optimal video quality.

In case of M2M, wearables and other IoT devices, compute-intensive applications can be offloaded to the edge. This can be performed by splitting the application, offloading only the data-intensive part to the edge. REPLISOM [39], proposed by Abdelwahab et al., introduces a computation offloading model for IoT applications, whereby the edge cloud periodically sends a request to the corresponding IoT devices acquiring information related to the associated service, i.e., typically a memory replica of the VM. REPLISOM is based on the Long Term Evolution (LTE) optimized memory replication protocol that relies on Device-to-Device (D2D) communication to create multiple memory replicas considering neighboring devices into a single compressed form, which is then pulled from the device. Compressed sampling construction algorithms are used to manage such memory replica, reducing the effort of the average devices to push, which saves energy and cost by reducing the amount of multiple requests at a time.

An edge cloud federation scheme focusing on IoT applications has been considered in [40] introducing the concept of hybrid cloud that splits applications seamlessly between core and edge cloud platforms, allowing delay sensitive and user-interactive components at the edge, while maintaining a scalable cloud solution. The objective is to facilitate efficiently two new classes of applications, a highly accurate indoor localization that relies on low latency, and a scalable and resilient video monitoring that saves bandwidth. Mobile gaming is another area that requires a significant computation power. By offloading the rendering part from mobile devices, mobile gaming can become more interactive due to shorter response times. In case of multi-player gaming, the trade-off between offloading and the performance of the cloud resource should be considered to avoid overloading a particular cloud platform. In this context, Chen et al. [41] have proposed a distributed computation offloading model using a game theoretic approach. The model considers a multi-channel wireless environment and multi-user offloading scheme, wherein the equilibrium state can be achieved considering the number of users that can benefit from the edge-cloud and the system-wide computation overhead.

A complementary service that leverages the benefits of MEC, named RAN-aware content optimization, can enhance the performance of computational offloading. Gaining information on the RAN quality and user context before performing offloading can assist both the device and the network to make the best out of the MEC platform services. Orsini et al. [42] proposed a cloud aware framework for application offloading based on network availability, radio signal quality and accessible surrogate computing resources. It breaks down applications into different components and creates an offloading online strategy based on the aforementioned optimization parameters.

B. Distributed Content Delivery and Caching

As content, and especially video, is the most popular service of mobile networks, thousands of video content items are uploaded daily to content providers' networks. Such content is stored in large volumes in the providers' centralized databases, and is then transcoded from source format to final delivery format, before distributed to multiple streaming servers within different network locations for further delivery. Despite content distribution efforts, several users may experience service interruptions because of buffering problems and jitter due to lack of content in the proximity of end-users, especially in mobile environments. Providing a distributed caching by extending Content Delivery Network (CDN) services towards the mobile edge can enhance the users' QoE, while reducing backhaul and core network usage. Such an outcome is confirmed through various research results [43], [44]. In particular, the research work in [26], [45], and [46] propose various architectural flavors for accommodating distributed parallel edges capable of performing video caching and streaming to increase QoE for content delivery. As for the impact to the

backhaul, caching content may result in reducing the capacity requirements by up to 35% [47].

Recently, a framework to efficiently distribute adaptive video streaming services has been proposed; referred to as Media Cloud [48], which proposes an elastic virtual content delivery network at the edge. By pre-caching at the edge popular content according to statistics and user/service forecasting intelligence, MNOs or OTT providers can fully take advantage of the MEC platform, increasing the users' QoE. A context-aware network with edge cloud caching capabilities based-on big data analytics is considered in [49], while a content placement and delivery approach based on the prediction of content popularity and the characteristics of the wireless environment is elaborated in [44] and [50].

For interconnected cloud edges, caching can be performed in a hybrid manner [51], wherein each edge shares the cached content information in a form of catalog. In a conventional cloud computing environment, when a user, connected to an edge platform, requests a content item for the first time, the request is redirected to remote cloud platform to fetch such desired content. Unlike this conventional cloud operation, hybrid caching enables the option of requesting content from other nearby edge platforms avoiding fetching content from remote clouds. Besides video, augmented reality is another type of content that highly depends on round trip time and network capacity. MEC can ensure an optimum QoS, especially for 3D image files and other heavy objects, by caching content (e.g., in popular areas and museums) locally without requiring centralized servers and core network resources [52].

C. Web Performance Enhancements

Web performance enhancements mainly focuses on reducing the access time, accelerating the loading of a Web-page by providing content optimization considering network, user and device conditions.

1) Content optimization: Content optimization is performed in the Web hosting sites based on cookies, surfing history, or by the users' provided geographical locations [53]. Content awareness can be performed dynamically by hosting a content optimizer at the MEC platform. Network conditions regarding, e.g., network load and network status, or users' geographical locations can be collected from RAN and can be used for rapidly improving QoE.

2) Accelerated browsing: When a Web-page is requested by a mobile device, the edge cloud platform first processes its URL. The content is fetched from the remote cloud and then selected parts are filtered out in the response. Takahashi *et al.* [54] introduced an Edge Accelerated Web Browsing (EAB) solution, wherein an edge server is fetching Web contents before performing an evaluation and task rendering.

3) Web acceleration: A Web-page may contain large files in size, which may affect the loading time of the Web-page. A CDN-based edge can help reducing the loading time by caching particular content [55]. Once content is stored in the CDN edge, it can no longer change.

If the content is requested for the first time, then a store-and-forward process needs to download all files in the CDN, before the Web acceleration service becomes available. When a user is moving or the network conditions are fluctuating, rendering content-heavy Web-pages becomes difficult due to limited connectivity. MEC can further help adapting the content size dynamically based on the network status, user mobility and device computing load [56].

D. Internet of Things (IoT) and Big-data

The emerging IoT and big data services have changed the traditional networking paradigm from data consuming to data producing relying on cloud infrastructures. MEC can facilitate computational and storage resources in close proximity to data sources, which can be used for processing purposes, e.g., for providing network scalability reducing IoT data and signaling, ensuring a fast response to user requests or enabling new services. Mechanisms that ensure scalability in a group of low mobility MTC devices at the edge of the RAN by reducing the amount of signaling and their content through the use of a group profile is elaborated in [57] and [58]. Such set of mechanisms may easily comprise a MEC service, which can speed-up data delivery requiring less networking capacity. Similarly, a smart gateway solution, residing at an edge-cloud platform, can filter IoT communications performing data trimming before the corresponding information reach the associated cloud server [59].

A MEC platform can also encompass a local IoT gateway functionality capable of performing data aggregation and big data analytics for event reporting (e.g., temperature measurements), smart grid, e-health, or alarm notifications, considering certain measurements received from a set of IoT devices. The work presented in [60] describes an edge IoT architecture that considers a hierarchy of edge cloud platforms for carrying out data computation tasks responsible for collecting, analyzing and converting raw IoT data into meta-data, which are then delivered to the corresponding application server via SDN means. A study that introduces edge-cloud medical cyber-physical systems, investigating task distribution and VM placement in relation with the medical device and edge-cloud association, is analyzed in [61], and that is based on mixed integer programming and derived heuristics considering the users' QoE and system costs. A cloud-based open source controller and architecture with a corresponding API that enables developers to create scalable sensor-centric applications, known as IoTCloud, is elaborated in [62], focusing on distributed cloud infrastructures.

E. Smart City Services

MEC can also facilitate a number of smart city services, including video analytics, transport, location services, intelligent public spaces, energy saving, public safety and emergency services, to mention a few. These sources can easily produce a potential amount of data, with location specific and latency stringent requirements such as retail targeted video advertising, stressing the big data aggregation and analytics a step further

as elaborated in [63], which studies the challenges of programmability, data abstraction, service management, privacy and security. Smart city services may also rely on video data including closed-circuit television (CCTV) or police patrol cameras, with the objective to locate people or objects, e.g., a criminal, a missing child or a stolen car, or analyze a situation, e.g., damage or accident. MEC is an ideal platform to perform video analytics, such as face recognition or computer vision algorithm at the edge of the network avoiding overloading the mobile backhaul and core network, while assuring low latency [64]. Giga Sight [65] introduces a video analytics service executed on a MEC platform based-on computer vision, which runs in near real-time recognizing objects, which are reported along with metadata, e.g., time and location. The application is expected to be useful for safety in an automotive environment.

MEC may also have a significant role in car-to-car and car-to-infrastructure communications. Vehicles connected via the distributed edges may transfer or receive information such as traffic congestion and warnings from other vehicles or through the network almost in real-time, ensuring safety. An architecture study for cloud connectivity among vehicles and towards edge-cloud platforms enabling a number of on-board services is presented in [66] considering also resource management and VM mobility. Cameras mounted on cars combined with MEC video analytics can help improve road safety even further, e.g., road damage, and promptly alert other passing vehicles. Car manufacturers can benefit from MEC, offering damage detection services via an on-board sensor system, software upgrades to the car operation system as well as providing the opportunity to offer a range of other services, e.g., from entertainment to provisioning of information on fuel refill locations and parking. MEC will also be the critical component for the emerging self-driving and autonomous cars providing assisting information.

F. Application-Awareness and Content Optimization

As the MEC platform resides closer to the radio environment and the end user, it can provide information regarding the application performance. In particular, it can provide application-aware performance optimization in near real-time by altering the encoding of streaming content inside the edge [67], e.g., for reducing video stalls. Such an operation can also consider radio analytics regarding the link quality enabling dynamic content optimization for improving QoE and network resource efficiency, e.g., for TCP traffic. A high quality stream can also be re-encoded in the edge for receivers with small screens that cannot take advantage of high quality. Alternatively, a video may be paused when there are no changes in the frame, e.g., CCTV, saving network resources. For saving energy on radio access equipment, MEC can buffer data bursts related to non-critical applications and align their transmission with the waking-up times of access points.

An architecture that combines client driven Dynamic Adaptive Streaming over HTTP (DASH) with network assisted capabilities based on MEC is introduced in [68]. The objective is to use the MEC platform to provide radio analytics towards the DASH streaming application in order to influence the size of media segments assuring QoE. A similar approach is adopted in [32] whereby NFV is leveraged to instantiate suitable Virtual Network Functions (VNFs) to transcode videos at rates suitable for the current network conditions to maintain acceptable QoE. Liquid APPs [69] and more general applications at the edge are a set of industry solutions of MEC-based applications that take advantage of users' proximity, short latency and radio analytics for performance optimization [70], [71].

Another aspect of application-awareness and content optimization is the impact of social networks as valuable sensors of the human activity [72]. Information derived from social networks can be useful for optimizing the resource allocation for specific applications based on indicated popularity at a certain region or on the potential of forming social groups, e.g., at university campuses. Personal clouds, introduced in [73], create an instance at the hypervisor level that selects the appropriate resources across a federated cloud platform that suits best the specific environment and use. Local transportation information can also assist proactive caching or application availability at particular MEC platforms based on knowledge of users' density. At occasional events, for instance sport tournaments, certain applications and video content can be stored at a local MEC platform providing on-demand video replay services [74]. Video processing at the MEC platform can also assist in removing certain video content from data uploaded directly by users based on regulation policies [65].

Online gaming can also benefit from MEC assisted application-awareness to ensure high bandwidth and strict latency alongside an on-demand powerful multi-player gaming engine in terms of processing and storage. In particular, thin client gaming, e.g., Nvidia Shield and Valve Steam, wherein the game is processed and executed in remote cloud and a stream, i.e., audio and video, is sent to the client is gaining momentum. MEC can offer gaming-as-a-service by hosting a full-stack gaming service (OS, middle-ware, data and application), which can be enjoyed by users on subscription usage basis. This facilitates hardware and software device independence access providing the opportunity for a huge variety of online gaming pool with reduced cost. Such an approach is already considered in Gaming Anywhere [75].

IV. MEC ENABLING TECHNOLOGIES

MEC relies on cloud computing and virtualization technologies including VMs and containers. However, a complete cellular system that offers MEC services counts on NFV, SDN and network slicing attributes that allow flexibility and multi-tenancy support [10], [76].

A. Cloud Computing, VMs and Containers

Cloud computing reduces the cost of managing software and hardware, while allowing high computing power and high performance, convenient accessibility and constant availability. It provides a shared pool of resources, which can be scaled dynamically. Could computing is currently in a mature phase

providing four distinct technology models and three service models [77] including:

1) Technology Models: (i) Private cloud that is entirely owned and maintained by an enterprise within the corporate firewall, ensuring security by providing dedicated access (e.g., RackSpace, Citrix, and Google), (ii) Public cloud - a publicly accessible pool of resources hosted by a cloud provider based on a pay-as-per-use basis (e.g., Amazon, Microsoft, and Dell), (iii) Hybrid Cloud - a combination of public and private cloud (e.g., VMWare, HP, and IBM) and (iv) Community cloud - a resource pool that consists of an aggregation of several providers, which can be shared by a certain group of users.

2) Service Models: (i) IaaS (Infrastructure-as-a-service) offers virtualized scalable computer infrastructure in terms of compute, storage and networking (e.g., Amazon EC2, and Azure VMs), (ii) PaaS (Platform-as-a-service) offers a platform for developing, managing, and running applications (e.g., Azure Websites and Amazon Beanstalk) and (iii) SaaS (Software- as-a-service) provides access to software hosted in the cloud (e.g., Dropbox and Google docs).

A cloud platform typically consists of a number of physical machines that comprise a single logical entity, which can be shared among different players carrying out distinct isolated tasks. One way to perform such operation is using a hypervisor, which can create and run VMs that can host separate tasks. The isolation property of VMs provides users an independent system, i.e., a fully functional machine, irrespective of the underlying hardware [78]. The technology of VMs offers a fine-grained control for instantiating and terminating tasks and processes anytime without affecting the underlying hardware, enabling in this way flexibility in resource provisioning. However, a VM is an abstraction of a physical hardware stack (i.e., virtualized BIOS, network adapters, storage, memory, and CPU) and requires a full guest OS image, additional binaries and libraries for hosting applications and services. Unless an application or service is in need for such an environment, a VM misuses a significant amount of resources, besides its slow startup process because of booting a complete OS [79], [80].

In contrast, the abstraction in a container takes place at the OS level, supporting programs and libraries, and system resources to run a specific application or service. Containers partition the resources of the physical machines, creating multiple isolated user-space instances of a much smaller size compared to VMs. This allows multiple containers to run inside a single OS enabling rapid deployment with near native performance in CPU, memory, disk and networking. A container typically executes an application or service, providing easy instantiation and fast migration benefits due to its lightweight nature, but is less secure compared to VMs, which could also run more efficiently multiple applications. In practice, a container can be instantiated in the scale of milliseconds, while the instantiation of a VM would require seconds or minutes, depending on the characteristics of the OS, the physical hardware and the system's load [80]. Table I elaborates the
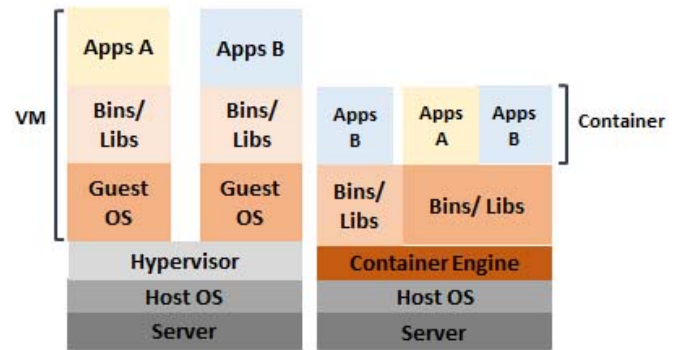

Fig. 1. Architectural representation of VMs vs Containers.

TABLE I
QUALITATIVE COMPARISON OF VMs AND CONTAINERS

VM	Container
Hypervisor-based virtualization works at hardware (HW) abstraction level	OS level virtualization; works at system call application binary interface (ABI) layer
Provides abstraction for full guest OSs	Provides abstractions directly for the guest processes
Requires a full virtualized hardware stack, hence heavy-weight	All virtual instances share a single operating system kernel, hence lightweight
Slow provisioning (instantiation, migration)	Fast scalable provisioning
High resource consumption, high overhead	Low resource consumption, more scalable
Provides full isolation, hence more secure	Process level isolation, hence less secure

differences between VMs and containers [81]. An architectural representation of VMs and containers is illustrated in Fig. 1.

The use of containers offers a light weight virtualization solution that allows a portable runtime of MEC services, which is particularly useful for mobile users. Additionally, MEC services can benefit from containers, because they offer mechanisms for fast packaging and deployment to a large number of interconnected MEC platforms. Containers provide a five-fold advantage over VMs in the context of MEC. Firstly, containers perform image layering and extension to build and create applications, which are then stored as images in a repository. This is later accessed by the container engine to run the application on the host. The engine helps in packaging, delivering and orchestrating which helps in fast deployment. Secondly, container API provides the support for life-cycle management, i.e., creating, defining, composing, distributing and running containers. Thirdly, storage facility is provided by attaching one or more data volume container with ongoing service. Fourthly, networking is performed by port mapping, which eases container linkage. Finally, support for microservice architecture (i.e., independent software components of a loosely coupled service, which can be rapidly mapped to form a business application based on requirements) facilitates container in paradigm of PaaS.

To this end, Docker is the most prominent container solution to facilitate an edge computing environment [82]. Containers can be also used as a replacement for VMs, assuring reduced

resource utilization in terms of storage and computation. Additionally, Google Kubernetes and Apache Mesos provide container cluster management within distributed nodes, which enables faster scaling [83]. Despite the benefits of containers, which are more evident in environments with user mobility, VMs can support heavier applications or a bundle of applications associated with a particular third party assuring a higher degree of security. The performance of VMs can be beneficial for such cases especially when service mobility is not required, for example considering applications such as IoT, augmented reality, enterprise and home networks. It shall be noted that containers, especially Docker containers [84], are inherently potable and can easily run inside a VM. These containers are portable from one VM to another or even to bare metal without requiring significant effort to facilitate the migration.

B. Network Function Virtualization (NFV)

NFV is a technology that allows operators to decouple network functions and services (e.g., Packet Data Network/Serving Gateways - PDN/S-GWs, firewall, Domain Name Service - DNS, and caching) from proprietary hardware, so they can run in software, accelerating in this way the speed of time to market [85]. This enables a new network architecture design and deployment by consolidating network functions and services into cloud platforms, enabling capital and operational expenditure savings. It equally allows flexibility in both data plane and control plane, by scaling up and down the allocated resources reflecting evolving service demands [86], [87]. In particular, NFV introduces Virtualized Network Functions (VNF) allowing the collocation of multiple instances in the same hardware or on top of the same virtualized environment, e.g., VMs. VNFs can reside within a single VM or may be distributed across multiple VMs over a cloud infrastructure.

The NFV architecture and orchestration framework [88], [89] defines the following three domains: (i) VNFs, which are software implementation versions of network functions, (ii) NFV Infrastructure (NFVI) that contains the hardware and software components, e.g., CPU, storage, virtualization layer, etc., that offer the network environment where VNFs are deployed and (iii) NFV Management and Orchestration (NFV MANO) that provides organization and management of the physical and virtual resources of the NFVI, being responsible for the lifecycle management of VNFs. In the context of MEC, NFV is a key technology for taking care of the virtualized instances related to certain applications and services. In particular, NFV can enable service flexibility, scalability and migration. For instance, increasing the resources for a popular application within a MEC platform can be easily achieved by adding another software instance or particular resources, e.g., CPU power or memory. The notion of virtual network embedding for efficient instantiation, placement and migration of virtual appliances in a dynamic edge network has been studied in [90] and [91], while the design of a middle-ware platform for increasing NFV's performance has been considered in [92].

NFV MANO provides lifecycle and fault management for ongoing services, e.g., replacing a VM upon a failure, and can be also responsible for providing connectivity across MEC platforms that span over different locations, especially when the NFV is integrated with SDN that performs service chaining as documented in ETSI EVE [93]. For example, to cope with congestion in an ongoing MEC application due to a flash crowd event, NFV can allocate additional resources from another edge platform. This can be also a more systematic process for peak hours or at regular instances or on-demand [94]. NFV's dynamic aspects that can benefit MEC services include: (i) portability where independent blocks of all services can be moved to another cloud environment in a different network, (ii) federation support for deploying portable functions over inter-operable geographically distributed virtual networks, (iii) slicing through partitioning of virtual network resources for particular applications and (iv) sharing a pool of configurable resources for on-demand access.

C. Software Defined Networks (SDN)

SDN is a technology that allows a rapid deployment of innovative services, enabling network programmability and multi-tenancy support [95]–[97]. As mentioned before, MEC can introduce application-assisting services and processing capabilities at the edge of the mobile network. These operations and services can be offered to third parties and verticals, opening the mobile network edge. SDN can complement the use of MEC offering programmability capabilities to authorized tenants, while allowing a flexible and efficient service management as well as support for service trials. By decoupling the control plane from the data plane and through the use of common APIs, SDN introduces a logical centralized control, which can easily instantiate and offer virtual network instances, by abstracting the underlying network infrastructure. In the context of MEC, the SDN controller needs to handle MEC related VNFs, VMs and containers as another type of resource that can be dynamically allocated and re-located [98]. Hence, SDN can support flexible service chaining, offering a dynamic service provisioning by connecting VNFs and MEC services, which can also reside at the edge in a particular order that satisfies the required performance demands [99]–[101], while assisting service mobility by allowing application providers and third parties to guide the network infrastructure.

A survey that presents the main ideas behind SDN in mobile networks, focusing on network programmability, efficient resource sharing and near real-time network control, is elaborated in [102] and [103]. MEC can reveal the full potential when a network of platforms is considered, providing an edge service that stress to a greater coverage region. With the current heterogeneous hardware equipment and platforms, backhaul technologies and configuration interfaces, it is difficult to coordinate the dynamic provisioning of services especially in a distributed manner at the edge of the network. The SDN paradigm can enable a simple and agile solution for providing network connectivity and service management

towards a MEC platform or between MEC edges across a heterogeneous radio [104] and over edge-to-edge diverse transport networks [105], [106]. In particular, SDN can overcome the current routing challenges associated with IP address translation, large volume of control signaling, tunneling overhead and dynamic resource management, e.g., by re-routing or changing the codec scheme of wireless microwave links.

SDN transforms the proprietary firmware-based network switches and routers forming a simple data plane, which can be controlled at the ingress and egress points of the network. In mobile networks, the use of SDN brings the advantage of cross-layer operations between the mobile and transport system [107], e.g., updating flow tables in switches and routers without the need for redirecting traffic to a new mobility anchor point, avoiding IP address translation and tunneling [108]. This is particularly beneficial for the case of user mobility especially between MEC platforms where distributed mobility at the edge of the network can exploit RAN analytics, while reducing congestion in the core [109].

D. Network Slicing

Network slicing has emerged as a key concept for providing an agile networking platform to support emerging businesses with diverse service requirements in an efficient way [6], [110]. It consists in slicing one network into multiple instances, each architected and optimized for a specific requirement and/or specific application/service [10], [76]. Network slicing enables the deployment of multiple logical, self-contained networks on a common physical infrastructure, allowing resource isolation and a customized network operation as detailed in [111]; a key industry document that elaborates the network slicing concept. In other words, network slicing introduces a multi-tenant environment supporting flexible provisioning of network resources, as well as dynamic assignment of network functions, Radio Access Types (RATs) and applications, even with short lifecycles enabling new value creation opportunities.

Network slicing enables resource sharing among virtual MNOs, services and applications as elaborated in [112] considering 3GPP mobile networks, by introducing the notion of network slice broker that complements the network sharing management and the service exposure capability function. From the infrastructure perspective, network slicing allocates a set of dedicated or shared resources either of physical or virtual nature to particular tenants by introducing a network hypervisor. A survey that brings light into the different network hypervisor architectures is presented in [113] considering various deployment solution options. To accommodate the service requirements of incoming requests, network slices need to combine a set of network and cloud resources, including for example bandwidth, network functions as well as processing, storage and access to big data or RAN analytics, etc., which are common MEC utilities.

Fig. 2 depicts an example of network slicing on a common network infrastructure considering the potential role of MEC in mobile broadband, automotive and massive IoT services. Mobile broadband requires high capacity, ensuring that
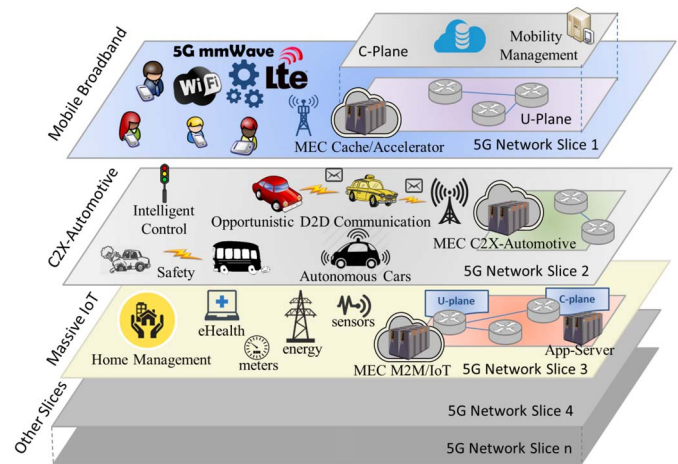

Fig. 2. Network slicing and the role of MEC.

an application receives the appropriate performance. The MEC platform can cache content at the edge increasing the capacity of the mobile backhaul and core network via traffic offloading to the local edge. MEC can also provide a number of services including, e.g., video accelerator or application aware performance optimization, to ensure the desired broadband experience. For the Car-to-X/automotive network slice that needs to accommodate strict latency and scalability with network functions instantiated at the edge, MEC is a catalyst element that shapes the capabilities of such an umbrella of services. Regarding massive IoT, scalability is essential for handling efficiently huge amounts of small data, hence MEC can provide processing and storage for performing signaling optimizations. To enable service customization in network slicing, a combination of NFV and SDN technologies is essential, supporting a tight coordination for VNF allocation and service provisioning at the edge-cloud, while allowing true and flexible service control in place [10].

V. MEC FRAMEWORK AND ARCHITECTURE

The introduction of MEC can be foreseen in parallel to the natural evolution of the legacy mobile base stations toward 5G systems, enabling software-based mobile edge applications and cloud computing services at the network edge. In particular, MEC enables modular, open solutions offering a programmable ecosystem, transforming the user experience, while at the same time, allowing application providers and third parties to get more information about their customers. The MEC platform can provide cloud storage, caching, computing, proximity benefits of resource provisioning, context and location awareness, agility and speed to the mobile applications. This section provides an overview on the current efforts of ETSI for the introduction of MEC technology, which has reached a stable state considering the MEC requirements, framework and the reference architecture [114]–[117]. In addition, it overviews the ongoing ETSI MEC efforts as well as the edge computing considerations in 3GPP and OpenFog.
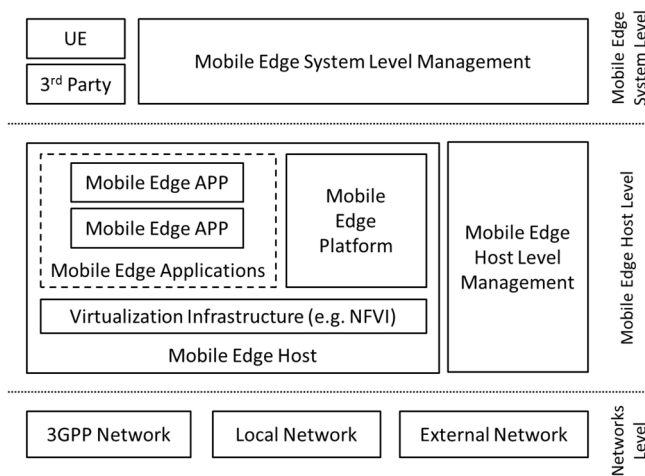
Fig. 3.   ETSI MEC framework [116].

A. MEC Framework

The MEC framework describes an ecosystem structure, including the entities and functions involved, which are categorized into mobile edge system level, mobile edge host level and networks level entities as illustrated in Fig. 3. The mobile edge host level is the fundamental part of the MEC framework. It consists of two main parts, namely the mobile edge host and the mobile edge host level management. The mobile edge host provides the virtualization infrastructure and the mobile edge platform, facilitating the execution of mobile edge applications. The underlying networks level offers connectivity to a variety of accesses including 3GPP mobile networks, local access network and other external ones such as the Internet. On top, the mobile edge system level management provides an abstraction of the underlying MEC system facilitating access for UEs and third parties.

B. MEC Reference Architecture

The MEC reference architecture, illustrated in Fig. 4, provides an insightful view of the functions and interfaces that compose the MEC system. The MEC reference architecture concentrates on the mobile edge system level and on the mobile edge host level excluding the networks level. The mobile edge host level consists of the mobile edge host, the mobile edge platform manager and the virtualization infrastructure manager. The mobile edge host facilitates mobile edge applications (APPs), offering a virtualization infrastructure that provides computation, storage and network resources, as well as a set of fundamental functionalities (mobile edge services) required to execute APPs, known as the mobile edge platform. The mobile edge platform enables applications to discover, advertise and consume edge services, and provides the virtualization infrastructure with a set of rules for the forwarding plane. Such forwarding rules are based on the policies received by the mobile edge platform manager and mobile edge applications. The mobile edge platform configures local DNS handling, which assists the user traffic to reach the desired mobile edge application and communicates with

other peer platforms via the Mp3 interface, which allows peer platform clustering.

Mobile edge applications (APPs) are executed as virtual instances on top of the virtualization infrastructure and interact with the mobile edge platform via the Mp1 interface in order to obtain the offered services, to indicate availability and to perform APP relocation in case of mobility. APPs can consume services from the platform and also provide services to other peer APPs, enhancing in this way the mobile edge platform. APPs typically indicate their service requirements in terms of resources, latency, etc., which are used for selecting the target mobile edge host(s). The mobile edge platform manager provides (i) the application lifecycle management, i.e., instantiation, termination, etc., interacting with the mobile edge orchestrator, (ii) the mobile edge platform element management, and (iii) the mobile edge application policy management functions, i.e., authorization, traffic rule, DNS configuration, etc. The mobile edge platform manager interacts with the mobile edge platform via the Mm5 interface to configure policies and traffic filtering rules, APP relocation and lifecycle procedures. The mobile edge platform manager associates with the Operations Support System (OSS) through Mm2 interface for fault, configuration and performance management purposes and with the mobile edge orchestrator via Mm3 for lifecycle management and policy provision.

The mobile edge orchestrator has the visibility over the resources and capabilities of the entire mobile edge network including a catalog of available APPs. The orchestrator is responsible for authenticating and managing the APPs, analyzing the service requirements to select the appropriate mobile edge host and to perform the APP relocation and policy provisioning. The orchestrator interacts with the Virtualized Infrastructure Manager (VIM) for managing the APPâĂŹs images and maintaining status information of the available resources. The OSS is an entity managing various services and subsystems within the mobile operatorâĂŹs network, receiving APP related requests, i.e., instantiate, terminates, etc., from the Customer Facing Service (CFS) portal and UEs. APP requests granted by the OSS are then forwarded to the orchestrator. The CFS portal allows access for third parties, providing APP developers with the opportunity of APP provision and management, while offering other customers the option to select an APP, providing Service Level Agreement (SLA) or billing related information. The user APP Lifecycle Management (LCM) proxy is a function that enables UEs to request APP related services, e.g., instantiation, termination, including relocation between MEC platforms or towards external cloud systems.

The VIM is responsible for managing the virtualized resources (i.e., storage, computing resources, etc.) for the residing APPs. It also provides software images for fast APP instantiation. It further facilitates fault and performance monitoring reporting information on virtualized resources to the orchestrator, which manages such APP images and virtualized resources. The Mm4 interface between the VIM and the orchestrator is used for monitoring and resource orchestration purposes, while the VIM also interacts with the virtualization infrastructure via Mm7 in order to manage the virtualization
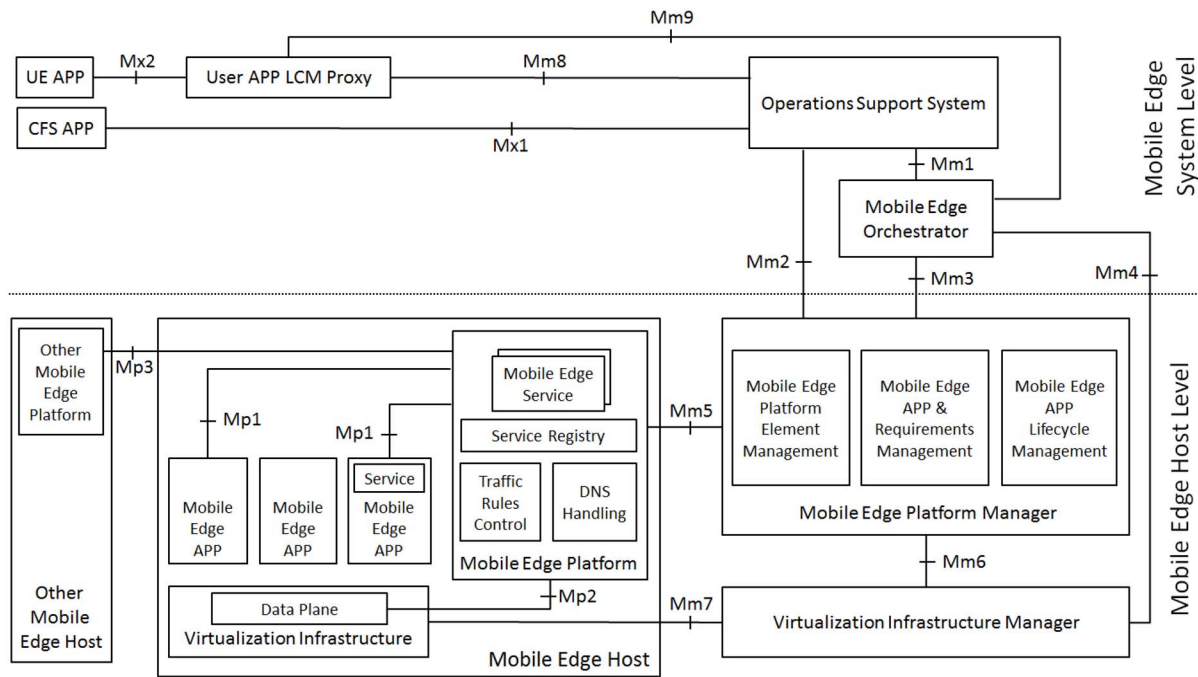
Fig. 4. MEC reference architecture [116].

resources. The VIM additionally interacts with the mobile edge platform manager enabling APP lifecycle management.

C. Ongoing MEC Activities and Other Related Standards

Following the MEC architecture and framework, ETSI MEC Industry Enabling Group (IEG) currently proceeds with the establishment of the application enablement platform, the mobile edge services and the corresponding APIs enabling innovative services from third parties. The standardization group (and in particular the IEG) is also working towards accelerating the adoption of MEC technologies by promoting the realization of proof-of-concepts (mecwiki.etsi.org) in accordance with the Proofs of Concept (PoC) MEC framework [118]. The following eight PoCs [119] are currently developed to demonstrate MEC as a viable technology with the results being fed back to the ETSI MEC Industry Specification Group (ISG):

1) Video user experience optimization via MEC: A service aware RAN PoC from Intel, China Mobile and iQiY, where the MEC application identifies paid video streams from designated content providers and assigns a higher priority ensuring an enhanced user experience.
2) Edge video orchestration and video clip replay via MEC: The edge video orchestration application from Nokia enables users to choose the camera angle and view video replays provided locally via the MEC host.
3) Radio-aware video optimization in a fully virtualized network: A MEC assisted video optimisation by Telecom Italia, Intel, Eurecom and Politecnico di Torino, via feeding radio conditions to the content provider, in order to accordingly adjust the video streams for improving the perceived QoS.

4) FLIPS - Flexible IP-based services: An operator-based MEC application from InterDigital, Bristol is Open, Intracom, CVTC and Essex University that transparently accelerates the delivery of IP-based content and streaming.
5) Enterprise Services: A MEC-based local breakout service for the enterprise users from Saguna, AVDA and Bezeq International.
6) Healthcare dynamic hospital user, IoT and alert status management: Developed by Quortus Ltd, Argela and Turk Telecom, to assign cellular network slicing and open edge access for local healthcare systems based on access rights, while also demonstrating dynamic network slicing considering Hospital 'alert' status.
7) Multi-service MEC platform for advanced service delivery: Developed by Brocade, Gigaspaces, Advantech, Saguna, Vasona and Vodafone to show an interoperable infrastructure controlled by a combined NFVO and cloud orchestration system that supports multiple MEC platforms offering distinct applications and flexible service chaining to enhance the user experience.
8) Video analytics: A MEC application by Nokia, Vodafone Hutchison Australia and SeeTec to provide video surveillance application over the LTE network, using MEC to analyze raw video streams and forwarding relevant incidents to local control room.

The ETSI MEC IEG also concentrates on describing key MEC metrics, best practices and guidelines on how to assess the MEC performance gains (with respect to a legacy system without MEC) [120]. This work will pave the way for a future activity on testing and performance measurements. The MEC platform is intended to be transparent to the 3GPP mobile network architecture, i.e., not affecting UEs, RAN and

core network functions or reducing the provisioning of lawful interception. In particular, the MEC platform is expected to have minimal impact on typical 3GPP network management processes and associated Key Performance Indicators (KPIs). The MEC management procedures should take into account and complement the 3GPP network management, ensuring application and service portability. From the regulation perspectives, MEC services should be transparent without discrimination for specialized services, i.e., being compliant with the net-neutrality framework, while allowing users to purchase specialized services subject to network neutrality.

Currently, 3GPP is exploring context-aware service delivery, considering MEC in a new work item with significant industry support, which is discussed in RAN 3 Working Group for assuring high-data rates and low latency [121]. The 3GPP work concentrates on the mechanisms, protocol and interface aspects between the radio access and the MEC platform. The focus of such a work item is on cross-layer optimization of the RAN with applications such as video-aware scheduling, proposing mobile video delivery optimization and local content caching. The rational is to explore use cases and derive requirements for efficient resource utilization and enhanced user experience, analyzing the potential impact on protocols and related signaling, considering also whether UE assistance has an impact.

Besides ETSI MEC, the OpenFog consortium is an industry body that drives the development of edge or fog computing architecture, testbed implementations and a variety of best practices and interoperability reports to enable end-to-end IoT services and deliver intelligence to customers and businesses. OpenFog was launched in November 2015 with the goal to "define a common framework for distributed computing based on open standard technologies" according to [122]. The objective of OpenFog is to enable an open computation, control and data storage across network edges leveraging the performance benefits of local, distributed cloud clusters. In particular, OpenFog uses multiple edge cloud platforms introducing a logically hierarchical architecture from the information processing perspective, considering a federation of cloud platforms with the objective to specify fog-to-fog and fog-to-cloud interfaces [123].

VI. MEC Service and Network Orchestration

The true impact of MEC relies on the service orchestration capabilities as well as on the interaction with the network architecture, aligning its operation with the VNFâăŹs orchestration. Since as the NFV framework, MEC is based on virtualized platform, the definition of the management and orchestration in MEC will reuse the infrastructure and infrastructure management of NFV to the largest extent possible, by hosting both VNFs and mobile edge applications on the same or similar infrastructure. Currently, ETSI MEC is studying how to implement MEC in an NFV environment and define (in collaboration with ETSI NFV) the management and orchestration framework, in the view of future 5G systems [124]. This section analyzes the notion of MEC service orchestration,

exploring MEC service mobility and the joint orchestration of VNFs and MEC.

A. MEC Service Orchestration

Integrating a MEC platform into a mobile network environment brings a number of challenges related to service orchestration, mainly due to the fluctuation of resources and the evolving radio and network conditions, caused by user mobility. A MEC system should support application lifecycle management, i.e., instantiating and terminating an application, on-demand or in response to a request by an authorized third-party. Orchestrating a MEC platform in terms of resource allocation and service placement is critical for assuring efficient network resource utilization, QoE and reliability. The following service-related attributes are particularly relevant for the efficient operation of a MEC service orchestration process:

1) Resource allocation: The flexible availability of resources plays a crucial role in the performance of a service. Various considerations around the resource allocation of VMs are analyzed in [125], taking into account the availability of CPU, memory, storage, network bandwidth along with resource contention [126]. The notion of application and its particular service characteristics when allocating resources in cloud environment are studied in [127]. Allocating resources in a multi-cloud deployment for offloading processing tasks considering a parallel execution among different platforms is analyzed in [128]. A multi-resource allocation strategy using semi-Markov decision processing is elaborated in [129], which determines the exact amount of resources, i.e., wireless bandwidth and computing resource, for achieving optimal system benefits. The capability to dynamically scale-up/down the allocated cloud resources for ensuring QoE is proposed in [130] based on the amount of users and active connections, whereas the work in [131] also considers SLAs and the notion of cloud-native elasticity in relation with QoE assurance [132].

2) Service placement: Placing MEC services over a number of edge cloud platforms can prove to be critical for the user QoE and should take into account gravity points, e.g., shopping malls, which attract a plethora of users. In addition, the placement of selected services need to consider potential mobility patterns, in order to ensure that associated users always receive the desired performance in terms of delay, capacity, etc. An optimization study for determining the edge cloud positions that best fit given mobility patterns and fluctuating resource demands is elaborated in [133] and [134]. Volley [135] is yet another solution, which concentrates on solving the VM placement problem over a number of given geographically distributed edge-cloud platforms, considering also service migration based on user location. Other solutions devised for VNF placements for the creation of virtual mobile networks can be also explored for MEC service placement [136], [137].

3) Edge selection: Typically allocating the nearest MEC platform that offers the requested service is desired from the performance perspectives, e.g., delay. However, such a strategy could prove to be inefficient creating performance bottlenecks, especially when the edge-cloud load is not considered. For mobile users, the selection of the edge cloud becomes crucial due to uncertainty of movement and wireless conditions.

4) Reliability: Another significant issue in deploying MEC services is reliability. A common technique to address fault tolerance in computing systems is dubbed checkpointing that maintains a regular snapshot of the application state, which can be used to resume an application upon failure. In mobile environments, dynamic network conditions may require frequent checkpointing raising scalability concerns. An alternative option to resolve scalability concerns is to replicate the MEC service instance, a process that also requires less time compared to checkpointing.

B. MEC Service Mobility

Ensuring service mobility is another aspect where optimal end-to-end session connectivity needs to be maintained for the entire course of service usage. In case of a mobile user, frequently changing the anchor points (e.g., from one edge node to another), ensuring optimum QoE for a delivered MEC service becomes challenging, especially for delay sensitive applications. Distributed Mobility Management (DMM) [138]–[140] is a notable solution towards managing the user mobility, overcoming also the scalability and reliability drawbacks of centralized mobility schemes. However, managing user mobility and redirecting user edge-cloud service requests to a distant edge hosting the service may not be an optimum solution. Considering MEC, which aims for a single hop connectivity to the service, there is a need to frequently migrate MEC services and place them close to the corresponding mobile users. Besides, in a federated cloud native environment (deployed across the edge), service migration requires virtual instances (VMs/containers) to be transferred to the target location.

Although there have been numerous studies on VM migration, such process imposes significant technical challenges when considered from service point of view [141]. The complexity in maintaining the service continuity during migration comes due to the time involved in preparing a VM for target node, transferring the same over the network and finally addressing the issue related with change of IP address after VM relocation. To perform migrations with IP continuity across a wide area network, IP mobility solutions have been proposed in [142] and [143]. However, these approaches involve the hypervisor in modifying the VM configuration to support mobility. Typically for an IP-based service, a change of the IP address would result in session breakdown, introducing the need to reestablish a new one, affecting in this way the QoE.

Technologies such as DNS and NAT have been introduced to cater with the breakdown of an IP session between two peers when the IP address of any of the two peers changes during mobility [144], but they do not natively support endpoint mobility or instant location update. To address this issue of IP-based service mobility and to bring the service close to the user, the follow-me-cloud concept is proposed in [145]–[147]. The authors introduced a framework, which ensures the cloud service to follow the user movement. The scheme enforces replacement of data anchor points with service anchoring and converts IP addressing by service/data identification. Moreover, by decoupling service mobility from layers 2 and 3, the framework ensures seamless migration and service continuity, elaborating the decision logic for migrating services. A study that enhances the follow-me-cloud concept considering mobility prediction models for supporting QoS/QoE is elaborated in [148], while a service migration policy focusing on the uncertainty of user mobility and the non-linearity of the migration and transmission costs is presented in [149].

Furthermore, to complement the follow-me-cloud solution, a SDN-based approach was introduced in [150] considering distributed elastic controllers, while the research work in [151]–[153] exploit the concept of VM migration along with user mobility introducing a Location/ID Separation Protocol (LISP) based approach to avoid triangular routing and reduce the downtime of VM migration. In general, the use of SDN ensures connectivity towards a cloud service over any radio technology (e.g., WiFi), while it also assures optimal routing and scalability [154]. For edge cloud platforms, the use of software defined control can assure a flexible connectivity and efficient QoE/QoS management via monitoring and closed-loop service adjustment means as presented in [155]. Service requirements, such as low delay, may vary depending on the nature of the application type. Hence, the controller or orchestrator should take this into account in case of service migration, e.g., for a delay-stringent application, bulk migrations would be advantageous over live migration [156]. A network planning and deployment framework, considering an optimization analysis for mobile edge computing and taking into account user and service mobility for assuring service level agreements, is elaborated in [157].

C. Joint Optimization of VNFs and MEC Services

With the launch of cloud-RAN, the network edge started containing commodity-of-the-shelf platforms that can host VNFs associated with legacy base stations allowing flexibility in deploying radio functions. Typical paradigms that introduce the notion of base station functional split in a cloud-RAN environment are presented in [158]–[160], allowing a number of different functional split options for a non-ideal fronthaul. A generic reconfigurable cloud platform considering both network functions and devices in terms of capacity, energy footprint, scalability and modularity is elaborated in [161]. An open environment for deploying such distinct C-RAN options is offered by the OpenAirInterface [162], which also enables a software defined control. An evolution of the functional split concept towards the network of functions that introduces new software based VNFs by de-composing a-priori monolithic 4G RAN and core functions into "atomic" sub-functions and
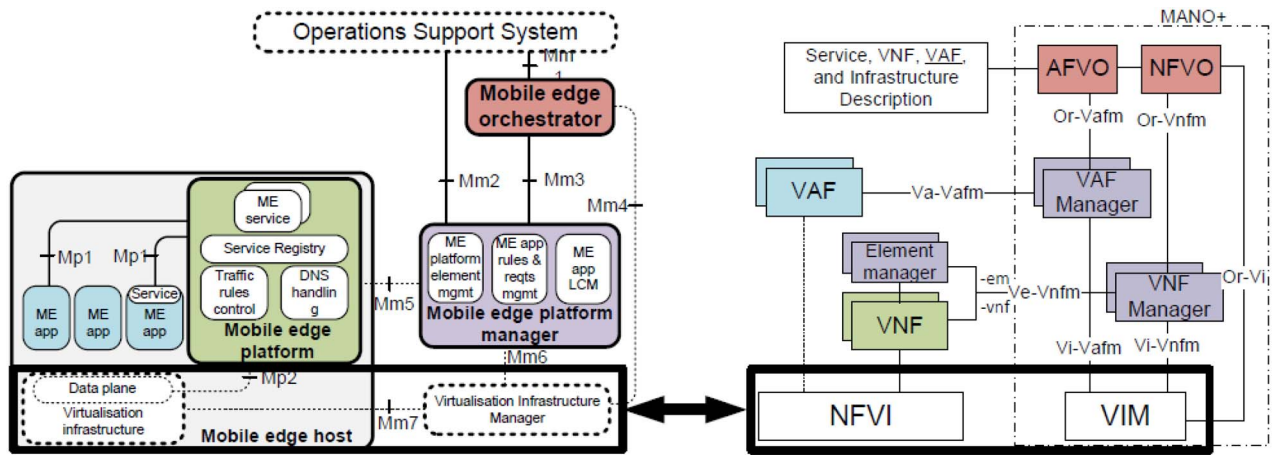
Fig. 5. Compound architectural evaluation of MEC and NFV structural blocks [167].

allocating them flexibly among edge cloud and core platforms is elaborated in [163]. MEC can take advantage of such edge platforms providing services and applications in a close proximity to the end user.

The MEC platform's lifecycle management can practically be supported by the current NFV MANO, i.e., from the infrastructure management perspective. However, the MEC orchestration should also consider service authentication, service availability announcement and support of service discovery. Such MEC orchestration can either be standalone in coordination with NFVO or a part of a common orchestrator, considering the alignment of both VNFs and MEC services for enabling a particular network utility. The suggested joint optimization can enhance the allocation and selection of VNFs considering the application type by taking into account additionally MEC related information, e.g., network conditions and big data. For mobile users, the common orchestrator can select the new attachment point considering the application type as elaborated in [164], transferring selected user related application content to the new location [165], while also having the capability to perform a VNF re-location, e.g., S-GW re-location in LTE. MEC can also provide feedback to the transport layer, e.g., making the TCP more radio aware [166], in order to optimize the network resource management.

A set of enhancements on the current ETSI MANO architecture for addressing such joint optimization are elaborated in [167]. In particular, two new functional entities are introduced as shown in Fig. 5, namely the Virtual Application Function (VAF) and VAF Manager (VAFM), which take care of the instantiation and life-cycle management of MEC services. The VAF and VAFM entities enable an enhanced orchestrator, referred to as NFVO+, to consider both VNF and MEC services jointly, providing an efficient network resource management and considering the application type. Such an enhanced architecture can enable for instance an enhanced gaming experience, e.g., avoiding the relocation of a MEC application during a crucial phase of the game. An alternative view for integrating MEC with NFV is provided in [168], wherein the mobile edge host, i.e., both mobile edge platform and APP, as well as the mobile edge platform manager

(of the MEC architecture in Fig. 4) interface with the ETSI MANO VNFM, which is responsible for their life-cycle management. The mobile edge orchestrator and the NFVO cooperate directly offering a logical enhanced orchestrator.

A converged edge cloud and cellular network resource orchestrator paradigm named CONCERT [169] abstracts cloud resources that can be flexibly utilized for both networking and computational services. CONCERT relies on the SDN paradigm adopting the concept of decoupling the control and data planes with the orchestrator coordinating data plane physical hardware such as servers, switches, and radio equipment in response to service requirements. A joint orchestration of radio and MEC resources is elaborated in [170] considering a multi-user computational offloading in a dense RAN deployment scenario. The objective is to minimize inter-cell interference optimizing radio resources, while minimizing the energy consumption of mobile devices and satisfying the corresponding service latency. A similar orchestration problem focusing on decentralized decisions is analyzed using game theory in [41] and [171], allowing devices to be self-organized adding autonomics into the mobile cloud computing paradigm.

VII. MEC Deployment Issues in Mobile Networks

This section studies the practical issues related with the adoption of MEC into nowadays mobile networks, considering a number of deployment scenarios and elaborating on the currently available service orchestration implementation options.

A. MEC Deployment Scenarios

MEC allows resources, typically residing in a centralized remote cloud, to be distributed among a set of multi-cloud platforms. A straightforward deployment of MEC is as an individual platform, i.e., running in isolation, allowing a MNO to integrate it into its RAN to provide local services without considering user mobility and service continuation. A more advanced MEC deployment scenario is to allow a network (i.e., federation) of MEC platforms that supports a number of services taking into account network and traffic characteristics as
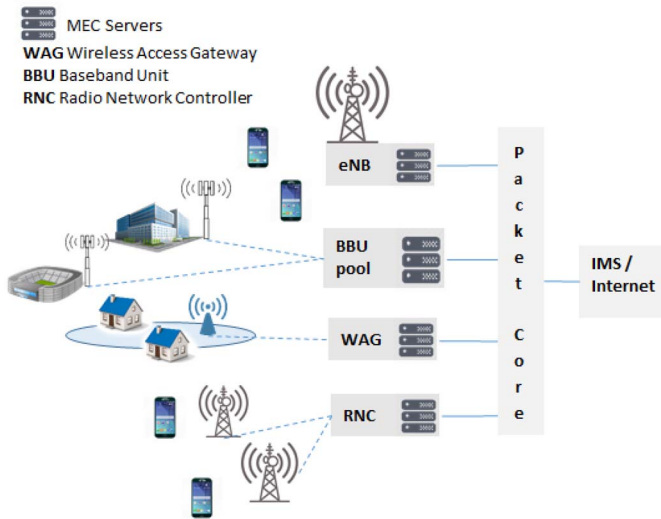
Fig. 6.   MEC deployment scenarios.

well as supporting service mobility. The deployment scenario of such network of MEC platforms can potentially introduce a heterogeneous computing environment for supporting a diverse range of applications. There are different commercial uses of MEC, considering the availability of the access technology and the deployment space [172]. An overview of the different MEC deployment scenarios is illustrated in Fig. 6.

For outdoor regions, MEC can be deployed either directly at the RAN or above in close proximity, enabling a close coordination of applications with the RAN, and a flexible service provisioning in key locations, understanding traffic characteristics and radio conditions, while enabling storage and processing at the RAN edge. MEC can also support and enable vertical segment services and offer emerging big data services, e.g., video analytics, to authorized third parties. The MEC platform can reside at macro-base station sites, e.g., eNB in LTE networks, or at the Radio Network Controller (RNC) of a 3G mobile system. Equally, the MEC platform can be located at an aggregation point such as the Baseband Unit (BBU) in a cloud-RAN deployment or directly at the mobile backhaul, e.g., at the small cell gateway. For indoor environments, MEC can act as a powerful on-premises gateway, enabling multiple services within a particular location, such as building management, retail services, augmented reality in museums, video streaming in sports and social events, security in public spaces and empower social network applications.

Considering IoT applications, the research work in [29] investigates the location of edge-cloud platforms at a cell aggregation site and even to a WiFi access point. A seamless integration of edge-cloud platforms in a small cell 5G deployment without any impact on the operations of 3GPP LTE networks is elaborated in [173] and [174], forming a cluster of interconnected computing resources. In these solutions, mobile users transmit cloud and conventional data over the radio bearers, with the separation of cloud data carried out at the access point via the use of a local gateway, e.g., Local IP Access (LIPA), which is responsible for routing and encapsulation between the mobile users and the edge cloud platform.

For facilitating dynamic resource management, a logically centralized orchestration entity is introduced taking into account a cluster of small cell-based edge-clouds via a new interface.

As mentioned earlier, a set of MEC platforms can be tightly coupled or highly dispersed. A typical tightly coupled scenario considers a central cloud orchestrator controlling multiple edge platforms deployed at different locations under the same administrator domain, i.e., MNO. In an alternative deployment scenario, MEC platforms may also work independently under the control of individual owners based on "on-site" policies and commercial usage. A homogeneous, i.e., single vendor, deployment scenario considers a set of MEC platforms, as part of the infrastructure of the same cloud service provider, whereby each MEC platform uses either its own individual orchestrator to manage the corresponding services or a common orchestrator. Typically, homogeneous MEC deployments are applied within a single MNO. In a heterogeneous deployment scenario, MEC services are hosted on a multi-vendor cloud service provider infrastructure, with each cloud service provider adopting a different orchestration arrangement. Heterogeneous MEC deployments can even stress between different administrative domains.

In handling an edge-architecture across a federation of edge and centralized cloud platforms, the OpenFog work in [123] considers a hierarchical cloud arrangement, wherein operations with a local scope are handled by edge platforms, while broader decisions are centralized. Such architecture can be seen as an extension of the traditional cloud, allowing flexibility in service deployment and mobility, by enabling an elastic combination of different resources across separate platforms for particular application types. The deployment of this arrangement requires an orchestration system to manage, control and configure the corresponding services across the set of cloud platforms. A similar layered fog to cloud architecture is also considered in [175], which also highlights the need for coordinated network management.

B. MEC Orchestration Deployment Options

As 5G mobile networks progressively incorporate different technologies and cloud infrastructures, becoming more heterogeneous in nature, the resource allocation and management processes turn to be more complex. On top of such a heterogeneous environment, emerging requirements on distributed service provisioning, programmability, open networking and multi-tenancy support drive the network control approaches towards a unified networking and cloud orchestration. Such an orchestration should take into account networking, cloud and service requirements. Currently, a number of different orchestration deployment options have emerged from the industry and standardization, with the most significant ones elaborated below.

  1) OpenNFV [176]: Developed by HP, it is an open source platform for orchestrating an end-to-end NFV and SDN infrastructure. OpenNFV is based on the ETSI NFV reference architecture and consists of three parts, namely NFV director, NFV manager and Openstack (HPE

TABLE II
QUALITATIVE COMPARISON OF DIFFERENT ORCHESTRATOR DEPLOYMENT OPTIONS

	Project Type/Driver	Orchestrator Scope	Operational Objectives	Technologies Features	Service/Legacy Management Features
OpenNFV	HP	Cloud networks and services	Deploy ETSI MANO	Helion OpenStack	–
CloudNFV	Collaborative	SDN/NFV enabled cloud services	Unifies management and orchestration in a data model	OpenStack, TM Forum SID	Service chaining and OSS/BSS
OpenBaton	TU Berlin, Fraunhofer FOKUS	Heterogeneous virtual infrastructures	Enables virtual network services on a modular architecture	OpenStack, TOSCA, REST-API, Zabbix	Event management and auto scaling
Open MANO	Telefonica	SDN/NFV infrastructure	Deploy ETSI MANO	OpenMANO, OpenVIM, REST-API	–
Cloudify	GigaSpaces Technologies	Cloud platforms	Multi-cloud solution for automating APP deployment	OpenStack, TOSCA, Docker, Kubernetes	Service chaining, OSS/BSS
T-NOVA	European Union	Network service and virtual resource	Network Function as a Service	OpenStack, OpenDaylight	Service lifecycle, OSS/BSS
OSM	Collaborative	Cloud networks and services	MANO with SDN control, multi-site/multi-VIM capability	OpenMANO, JuJu, OpenVIM, OpenStack	–
OPNFV	Linux Foundation	NFV platforms	Validate multi-vendor NFV solutions and advance NFV	OpenStack, OpenDaylight	–
ExperiaSphere	CIMI Corporation	Cloud, SDN/NFV and legacy networks	Flexible service model	USDL, TOSCA	Service events, derived operations
M-CORD	ON.Lab and partners	NFV/SDN - edge cloud for mobile networks	Anything as a Service, Micro-services architecture	ONOS, OpenStack, XOS	Real-time resource management, monitoring/analytics, service chaining
ZOOM	TM Forum	Cloud services, NFV	Enable personalized services, develop OSS/BSS for VNFs	–	Improved OSS/BSS
Open-O	Linux Foundation	Cloud, SDN/NFV and legacy networks	End-to-end composite services	TOSCA, YANG, REST-API, OpenStack, Multi-VNFM/VIM, ONOS, OpenDaylight	Service chain, improved OSS/BSS, policy management
ECOMP	AT&T	Cloud, SDN/NFV and legacy networks	Software-centric network capabilities and automated end-to-end services	TOSCA, YANG, REST-API, OpenStack	Service chain, improved OSS/BSS, policy management
NGSON	IEEE	Cloud services	Context-aware services	–	Service chain, improved OSS/BSS

Helion). The NFV director acts as an NFV orchestrator performing an automatic deployment and monitoring of the VNF ecosystem. It supports heterogeneous hardware platforms enabling virtualization environments for efficiently deploying VNF instances. The NFV manager is responsible for the life cycle of the VNF instances enabling accordingly scale-up or scale-down. The Helion Openstack offers an open source platform supporting VNFs.

2) CloudNFV [177]: CloudNFV is another open source orchestration platform for integrating NFV-based cloud computing with the SDN functionality. In addition to ETSI NFV MANO, CloudNFV considers management alongside orchestration in a unified data model. CloudNFV's functionality is divided amongst three elements: active virtualization, orchestrator and manager. Active virtualization is a data model representing resources, functions and services using the active resource and active contract sub-elements. The active resource represents the status of infrastructure resources, whereas the active contract defines service templates according to the characteristics of the available network functions. Depending on the allocated policy rules, resource status and the requested service order, the orchestrator addresses the VNF location for a particular service, including also the connectivity between

them. The manager operates the active resources and maintains an information base of the running services.

3) OpenBaton [178]: OpenBaton, developed by Fraunhofer FOKUS and TU Berlin, ensures the development of virtual network infrastructures by porting and further adapting network functions to the specific cloud environment. It focuses on improving the performance and grants security of the overall infrastructure by merging the underlying infrastructures, software architectures, networking, management, and orchestration. The framework considers a generic VNFM for the life cycle management of the VNFs based on the corresponding descriptors and a Juju VNFM Adapter in order to deploy Juju Charms. OpenBaton integrates two different engines: i) event management engine for dispatching and ii) auto scaling engine for managing scaling operations. A fault management system is also included for automatic run-time management where monitoring information are gathered using Zabbix. Finally, it provides plugin for addition and deletion inside the orchestration logic.

4) OpenMANO [179]: Based on Enhanced Platform Awareness principle, Telefonica's OpenMANO framework addresses aspects related to performance and portability. The architecture consists of three

components: OpenMANO, OpenVIM and a graphical user interface (GUI). OpenVIM is responsible for creating the infrastructure network topology and applying the Enhanced Platform Awareness principles. It is a lightweight VIM, directly interfacing with compute and storage resources in NFVI and with an OpenFlow controller. It offers a REST-API based interface towards OpenMANO. OpenMANO offers cloud services including flavors, instances and life cycle management of images. OpenMANO is the first of its kind to have a northbound interface (i.e., OpenVIM API extended with OpenStack API) for the creation and deletion of instances and templates of both network function and services.

5) Cloudify [180]: Cloudify is another open source framework based on TOSCA (Topology and Orchestration Specification for Cloud Applications), which acts as a cloud platform orchestrator. It provides a complete solution for automating and managing application deployment and DevOps processes on top of a multi-cloud environment. Cloudify eliminates the boundaries between orchestration and monitoring, assuring an automatic reaction to pre-defined events with the appropriate corrective measures. It organizes workflow for environment setup, application installation, infrastructure management, scaling and fault recovery. It provides compute, network, storage for application provisioning and deployment. Cloudify offers interoperability among diverse cloud platforms (e.g., VMware, Cloudstack, Amazon, and Azure) and reduces multi-vendor lock-in. Command Line Interface (CLI) based client is used to perform the different operations. Providers rely on the IaaS API to setup the required environment and to also install the manager, where the manager takes care of deployment, management and monitoring of the running applications.

6) T-NOVA [181]: T-NOVA is a management and orchestration platform for automated provisioning of Network Function-as-a-Service (NFaaS) on top of virtualized network infrastructure. It leverages the benefits of SDN and cloud management architectures to enable automated provisioning, configuration, monitoring and efficient operations of VNFs. T-NOVA differs from the other frameworks in terms of an additional marketplace layer, which allows operators to offer their infrastructures as a value added service. This layer is placed on top of the orchestrator and contains a customer facing module for implementing business related functionalities in a multi-user setting, employing the paradigm of "APP-store". T-NOVA follows the ETSI NFV architecture separating the VIM and NFVI, which are based on OpenStack and OpenDaylight. The orchestrator divides its functionality into two modules, namely Network Service Orchestrator (NSO) and Virtualized Resource Orchestrator (VRO). NSO maintains the lifecycle of the network services focusing on connectivity. On the other hand, VRO manages compute, storage and network resources.

7) OpenSource MANO [182]: This is an ETSI hosted open source project for automating end-to-end service deployment and orchestration, formed in collaboration with Telefonica, British Telecom, Telenor, Telecom Austria Group, Intel, Canonical, RIFT.io and Mirantis. The framework offers SDN underlay control (integrating multiple SDN controllers), multi-site capability, multi-VIM capability with enhanced performance awareness and support for physical functions. The architectural components contain: resource orchestrator (Telefonica OpenMANO), VNF configuration component (Canonical Juju), network service orchestrator (RIFT.io), GUI (RIFT.io), virtualized infrastructure based on Intel architecture, virtual infrastructure manager (OpenVIM and Openstack) and finally service VNFs (Metaswitch and 6wind).

8) OPNFV [183]: An open source project hosted by the Linux Foundation. OPNFV establishes a reference NFV platform for facilitating the development and evolution of multi-vendor NFV components. It conducts performance and use case-based testing on current standards specifications and work from open source communities for particular NFV use cases. OPNFV work concentrates on NFVI and VIM, considering components from OpenDaylight, ONOS, OpenStack, Ceph Storage, KVM, Open vSwitch, DPDK and Linux. OPNFV accelerates the development of emerging NFV products and services, ensuring certain performance targets and interoperability.

9) ExperiaSphere [184]: ExperiaSphere by CIMI Corporation is a combination of open source components that forms a universal service-layer approach for management and orchestration of the cloud, SDN, NFV, and even legacy networks. It introduces the concept of flexible service models by enabling abstractions of the underlying infrastructure resources into functional objects that expose service features. Functional objects are then integrated vertically to facilitate a service instance. A broker, e.g., accessed via a user portal, can select the appropriate set of service models creating a service instance, which is tracked during its entire lifecycle. An orchestration process is responsible for the structured commitment of resources, while management means take care of service and resource changes. ExperiaSphere is based on two principles named the structured intelligence that links data models to service events and the derived operations of virtual elements, while its deployment relies on the Universal Service Definition Language (USDL) and TOSCA.

10) M-CORD [185]: M-CORD from ON.Lab and partners, combines NFV/SDN and cloud solutions for mobile networks leveraging the benefits of mobile edge computing offering scalable customized services with enhanced QoE, e.g., higher throughput and lower delay. M-CORD enables virtualization of the RAN and core network functions, while separating the control functions from the data plane enabling a unified network orchestration and management. In addition, it allows third parties to

build mobile edge services facilitating localized applications. M-CORD offers a single SDN control plane following ONOS [186] to control the virtual network infrastructure, SDN/NFV resources based on OpenStack and TOSCA that facilitate VNFs and network slices, providing mobile services with the desired performance, orchestrated by XOS [187].

11) ZOOM [188]: ZOOM is a TM Forum project that enables the delivery and management of virtualized networks and services, facilitating the development of virtualization and NFV/SDN best practices and standards. ZOOM identifies an operations environment considering physical and virtual components, which can easily and dynamically be assembled into personalized services considering also new security processes to protect NFVI and NVFs. To achieve the aforementioned goals, ZOOM regularly provides demos supported by operators and vendors within the context of the so called catalyst project, establishing DevOps, NEtOps and ServOps user scenarios.

12) NGSON [189]: NGSON is an IEEE standardization effort that specifies an open service ecosystem that allows the establishment of dynamic services across different service providers and network domains considering also a seamless mobility environment. The main feature of NGSON is context awareness in the form of service, user, network and device information, which enable an intermediate layer between the network and application. The NGSON architecture defines service, transport and management functions for establishing context-aware service chains, enhancing QoE, while efficiently utilizing the network and cloud resources. The process of context-aware service composition and delivery consists of: (i) service composition that defines a flow of functional capabilities, (ii) service selection that specifies the service points among a set of candidates and (iii) service delivery considering the underlying transport network. NGSON can easily complement SDN/NFV control and orchestration taking advantage of the network virtualization properties and programmable traffic steering capabilities as elaborated in [190].

13) Open-O [191]: An open source project supported by the Linux Foundation that establishes a carrier grade orchestration platform to deliver end-to-end composite services across a virtualized SDN/NFV infrastructure and legacy networks. OPEN-O enables flexibility, supporting multi-domain and multi-location, and enhances the service lifecycle via automation. It also accelerates innovation by shortening the time to market through a hierarchy of three orchestration modules. Such an orchestration hierarchy consists of (i) the Global Service-Orchestrator that enables end-to-end service composition and delivery, (ii) the NFV-O responsible for NFV orchestration, considering diverse VNFs across a wide range of VNFMs and VIMs and (iii) the SDN-Orchestrator that provides network connectivity and traffic steering via the means of different SDN controllers (e.g., OpenDaylight and ONOS), and/or the conventional element management system. Open-O adopts TOSCA, YANG data models, REST APIs, OpenStack and supports resource abstraction over diverse SDN, NFV, and legacy networks, allowing a set of common services including policy management, security and other management capabilities.

14) ECOMP [192]: A project originally initiated by AT&T and later turned over to the Linux Foundation. It focuses on VNF management and the provisioning of software-centric network capabilities, leveraging the benefits of cloud technologies and network virtualization to offer automated services. ECOMP defines a master service orchestrator responsible for automating end-to-end service instances, which interacts with three distinct Controller types, namely the infrastructure controller - typically within the cloud layer, the network controller, and the application controller. The orchestration automates configuration processes, programmability rules and a policy-driven operational management, facilitating flexibility in instantiating, modifying and terminating network, application or infrastructure services and resources, considering network data and service-oriented analytics. ECOMP expands the scope of ETSI MANO, introducing the notion of the resource controller and policy component as well as the concept of resource description, i.e., meta-data, for lifecycle management of the virtual environment enabling network agility and elasticity, while improving the time-to-market. To this end, ECOMP supports open cloud standards (e.g., OpenStack, OPNFV, and TOSCA) and follows Netconf, Yang configuration and management models, and REST-APIs.

It shall be noted that recently, Open-O and ECOMP have merged into a new open source orchestrator project called Open Network Automation Platform (ONAP) [193], which is carried out within the Linux Foundation with the support of AT&T, China Mobile and many other leading industry partners. ONAP provides a unified architecture adopting the best components from both previous projects, offering a policy-driven software automation of VNFs and network capabilities that allows software, network and cloud providers to rapidly create and efficiently orchestrate new services.

A summary of the different orchestrators is provided in Table II, considering the project type and the main driving organization, the orchestrator scope and its objectives as well as PoC related technology features. In addition, service management and legacy management aspects are included to show specific capabilities that can complement the corresponding MEC services. An analysis that elaborates a number of different orchestrators, including also a broader range of solutions is provided in [86] and [194].

VIII. Lessons Learned and Research Challenges

The concept of edge-cloud computing has evolved from a simple content provisioning, e.g., CDN, and computation offloading service to an open platform that can accommodate third party services and can provide a means of interaction

between networking and applications. This creates a number of new service opportunities and brings new technical and business challenges. This section discusses some potential enhancements to existing solutions and highlights some future research directions.

A. MEC Service Orchestration and Programmability

Service orchestration and programmability with respect to different layers of the MEC platform (i.e., infrastructure, application platform and services hosted on the platform) are still open issues, which introduce significant challenges. Service orchestration should be performed in coordination with the network resource programmability, considering capacity and VNF allocation, especially the stretch of services across a set of edge-cloud platforms. Edge-cloud platforms located across different administrative domains raise even further challenges for the service orchestration considering federated resources, where more research is required for the processes of resource aggregation and service mapping, besides the definition of the corresponding APIs.

This article sheds light on the different service aspects of edge-cloud orchestration and programmability including: (i) service operations such as resource allocation, service placement, platform selection and reliability, (ii) service continuity and mobility within a set of edge platforms and (iii) joint optimization of VNFs and MEC services on common edge-cloud platforms to achieve efficient resource utilization and cross-layer optimization among edge-cloud services and network resources. In addition, it overviews the different MEC network architecture scenarios and elaborates the potential MEC orchestrator deployment options considering various SDN/NFV integration opportunities that enable distinct edge-cloud resource control flavors.

The research and standardization efforts for enabling efficient MEC services are still ongoing with various challenges yet to be addressed [195]. One of the key areas is the development of advanced APIs that will enable third parties to acquire and simply and efficiently manage resources on MEC platforms, including also the corresponding data models. Currently, various APIs and data models are still under discussion in ETSI MEC considering emerging applications, e.g., using the MEC platform for assisting the discovery process of millimeter Wave (mmWave). Such APIs should be also enhanced to provide RAN and network related information, making the network visible to the application.

B. MEC Service Continuity and Mobility

Mobility is an essential challenging feature for MEC services. Hence, more research is encouraged considering the user's activity and application-specific information, in order to assure service continuity for high speed mobile users [196]. For vehicular communications and automotive, integrating MEC with mobile cloud computing [197], [198] or vehicular cloud [199], wherein mobile or vehicle resources are utilized for communication and computation services, is a highly challenging issue considering the service orchestration perspective. Since different MEC services may be allocated to particular platforms for service optimization purposes, users that consume multiple applications may be associated with distinct MEC platforms at the same time.

Supporting multiple paths and potentially streaming among different MEC platforms in a mobile network is challenging from the perspective of both network scalability and user performance, especially for mobile users. Such an operation can become even more complex for applications that benefit from multi-connectivity, e.g., applications that decouple the uplink and downlink or use multiple communication paths splitting a communication session among different access points. Mechanisms such as control/data-plane separation [200] or mechanisms analogous to phantom cells [201] can help in managing user connectivity, while the MEC platform itself should support operations to handle traffic re-direction based on application policies provided by the operator.

C. Service Enhancements: QoE and Resiliency

The use of MEC systems can also inspire further research for developing new services and applications, in order to assure network efficiency and enhanced user experience. MEC can be used to maintain network or service states for emerging applications, e.g., for resiliency reasons by keeping a state of a critical parameter and providing backup features or can help analyzing QoE or other performance related states, for instance in [202] considering the performance of Time Division-LTE. MEC can further provide proxying functionality on behalf of end users, e.g., for energy saving purposes, and can enhance the network scalability by reducing the signaling overhead related to network processes, through the aggregation of signaling messages or via mechanisms that can enable connectivity sharing, e.g., sharing an established bearer among a set of MTC devices with the same service requirements [203]. MEC can be used to enhance the experience of location services, making Proximity Services [204] more intelligent and personalized, by combining user-context information, big data and social applications.

From the resource optimization perspective, it is worth exploring the problem of network functional (de)composition and allocation, enabling a flexible "network of capabilities" by combining network and cloud resources with the different types of services considering the offered QoE. New KPIs that reflect QoE, coupling both network and cloud resources, should be studied to simplify the performance measurements and resource optimization process. For enhancing the efficiency of the MEC system, the concept of HAEC (Highly Adaptive Energy-efficient Computing) [205], which flexibly reflects energy considering the corresponding needs of a computational problem, should be further explored considering resource and service abstraction models.

D. MEC Security and Privacy

In contrast to traditional cloud computing, MEC imposes significant security risks especially when it is located at the base station or at areas where it is relatively vulnerable to physical attacks. Hence, MEC deployments raise additional

security measures against on site attacks. MEC also requires more stringent security policies as third party stakeholders can gain access to the platform and derive information regarding user proximity and radio analytics. Authentication based on privileges of the third parties accessing the platform should be considered. One such option is to ensure Public Key Infrastructure (PKI) based techniques.

Isolation between different parties, i.e., in-between the hosted applications, is another critical issue. In particular, a security attack on a particular application should not affect other running applications, while isolation should also provide privacy assuring a bidirectional trust among cooperating parties. Fine-grained access control needs to be investigated with appropriate encryption to ensure a secure collaboration and interoperability between heterogeneous resources and different operational parties. A state-of-the-art analysis on security and privacy issues associated with edge cloud is performed in [206], while a preliminary study on MEC security is presented in [207]–[209]. Various intrusion detection techniques are in place for cloud computing, but for large-scale geo-distributed environments, this is still a challenge.

E. MEC Service Monetization

Besides the technical challenges, MEC also brings new business related opportunities allowing mobile operators to monetize combined cloud and network resources as well as particular services to third parties. Resource brokering solutions are one approach that requires further investigation in these types of networking environments. Tariff planning for service usage, e.g., radio analytics or video optimization, is an open issue that depends on the particular demand. In addition, more dynamic pricing models are needed, requiring advanced accounting and monitoring. Sufficient attention needs to be paid to the economic incentives for providing MEC solutions and its architecture for assuring a sustainable competition environment for various participants. Methods for resolving potential economic conflicts among different participants should be also considered to assure a smooth MEC system operation.

IX. CONCLUSION

MEC is an emerging technology that brings forward the technical benefits of edge-cloud computing with networking and support multi-tenancy allowing third parties to provision applications and services on-demand through standardized APIs. MEC makes the radio and network layer visible to the application providers/developers, providing a range of new advancements on the QoE. MEC is recognized as one of the key emerging technologies for 5G systems, thanks to its significant contribution to low latency assurance and capacity enhancements in the backhaul and core networks. The success of MEC fundamentally hinges on the alignment of the technology with ETSI NFV ISG for the proper definition of management and orchestration system with respect to the service elasticity and life-cycle management, service mobility as well as regarding joint optimization with the network resources. Currently, MEC brings forward a range of different

challenges that are yet to be solved. However, considering its potential, it is obvious that MEC will significantly uplift the shape and experience of mobile communications.

REFERENCES

[1] *GSMA Mobile Economy*, GSMA Alliance, London, U.K., 2016.

[2] "Cisco visual networking index: Global mobile data traffic forecast update, 2015–2020," White Paper, Feb. 2016.

[3] "Feasibility study on new services and markets technology enablers, stage 1, rel. 14," 3GPP, Tech. Rep. 22.891, Jun. 2016.

[4] A. Osseiran *et al.*, "Scenarios for 5G mobile and wireless communications: The vision of the METIS project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, May 2014.

[5] 5GPPP, "5G vision: The 5G infrastructure public private partnership: The next generation of communication networks and services," in *Proc. EUCNC*, Athens, Greece, Jun. 2016. [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-5G-Architecture-WP-For-public-consultation.pdf

[6] *NGMN 5G White Paper*, NGMN Alliance, San Diego, CA, USA, Feb. 2015.

[7] P. K. Aapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, "Design considerations for a 5G network architecture," *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 65–75, Nov. 2014.

[8] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[9] T. Taleb *et al.*, "EASE: EPC as a service to ease mobile core network deployment over cloud," *IEEE Netw. Mag.*, vol. 29, no. 2, pp. 78–88, Mar. 2015.

[10] T. Taleb, B. Mada, M.-I. Corici, A. Nakao, and H. Flinck, "PERMIT: Network slicing for personalized 5G mobile telecommunications," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 88–93, May 2017.

[11] "Mobile edge computing: A key technology towards 5G," White Paper, ETSI, Sophia Antipolis, France, Sep. 2015.

[12] I. Morris, *ETSI Drops "Mobile" From MEC*, Light Reading, New York, NY, USA, Sep. 2016.

[13] "Stakeholder analysis and feasibility study for MEC," White Paper, version 1.0, GSMA, London, U.K., Jun. 2016.

[14] "Business opportunities and monetisation options for MEC," White Paper, version 1.0, GSMA, London, U.K., Jun. 2016.

[15] A. Ahmed and E. Ahmed, "A survey on mobile edge computing," in *Proc. IEEE 10th Int. Conf. Intell. Syst. Control (ISCO)*, Salerno, Italy, May 2016, pp. 1–8.

[16] M. T. Beck, M. Werner, S. Feld, and S. Schimper, "Mobile edge computing: A taxonomy," in *Proc. 6th Int. Conf. Adv. Future Internet*, Lisbon, Portugal, Nov. 2014, pp. 48–54.

[17] S. Yi, C. Li, and Q. Li, "A survey of fog computing: Concepts, applications and issues," in *Proc. Workshop Mobile Big Data*, Hangzhou, China, Jun. 2015, pp. 37–42.

[18] N. Fernando, S. W. Loke, and W. Rahayu, "Mobile cloud computing: A survey," *Future Gener. Comput. Syst.*, vol. 29, no. 1, pp. 84–106, Jan. 2013.

[19] H. Qi and A. Gani, "Research on mobile cloud computing: Review, trend and perspectives," in *Proc. IEEE 2nd Int. Conf. Digit. Inf. Commun. Technol. Appl. (DICTAP)*, Bangkok, Thailand, May 2012, pp. 195–202.

[20] 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.

[21] G. Sun and J. Shen, "Facilitating social collaboration in mobile cloud-based learning: A teamwork as a service (TaaS) approach," *IEEE Trans. Learn. Technol.*, vol. 7, no. 3, pp. 207–220, Jul./Sep. 2014.

[22] J. Wu, C. Yuen, N.-M. Cheung, J. Chen, and C. W. Chen, "Enabling adaptive high-frame-rate video streaming in mobile cloud gaming applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 12, pp. 1988–2001, Dec. 2015.

[23] D. B. Hoang and L. Chen, "Mobile cloud for assistive healthcare (MoCAsH)," in *Proc. IEEE Services Comput. Conf. (APSCC)*, Hangzhou, China, Dec. 2010, pp. 325–332.

[24] M. Satyanarayanan, "Fundamental challenges in mobile computing," in *Proc. 15th Annu. ACM Symp. Principles Distrib. Comput.*, Philadelphia, PA, USA, May 1996, pp. 1–7.

[25] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, "The case for VM-based cloudlets in mobile computing," *IEEE Pervasive Comput.*, vol. 8, no. 4, pp. 14–23, Oct./Dec. 2009.

[26] D. Fesehaye, Y. Gao, K. Nahrstedt, and G. Wang, "Impact of cloudlets on interactive mobile cloud applications," in *Proc. IEEE 16th Int. Conf. Enterprise Distrib. Object Comput. Conf. (EDOC)*, Beijing, China, Sep. 2012, pp. 123–132.

[27] Y. Jararweh, L. Tawalbeh, F. Ababneh, A. Khreishah, and F. Dosari, "Scalable cloudlet-based mobile computing model," *Proc. Comput. Sci.*, vol. 34, pp. 434–441, Aug. 2014.

[28] "Fog computing and the Internet of Things: Extend the cloud to where the things are," White Paper, Cisco, San Jose, CA, USA, 2015.

[29] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proc. ACM SIGCOMM Workshop Mobile Cloud Comput.*, Helsinki, Finland, Aug. 2012, pp. 13–16.

[30] L. M. Vaquero and L. Rodero-Merino, "Finding your way in the fog: Towards a comprehensive definition of fog computing," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 5, pp. 27–32, Oct. 2014.

[31] *Cisco Iox Overview*, Cisco, San Jose, CA, USA, 2014. [Online]. Available: http://goo.gl/n2mfiw

[32] S. Dutta, T. Taleb, P. A. Frangoudis, and A. Ksentini, "On-the-fly QoE-aware transcoding in the mobile edge," in *Proc. IEEE Globecom*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[33] R. Balan, J. Flinn, M. Satyanarayanan, S. Sinnamohideen, and H.-I. Yang, "The case for cyber foraging," in *Proc. 10th Workshop ACM SIGOPS Eur. Workshop*, Saint-Émilion, France, Sep. 2002, pp. 87–92.

[34] B.-G. Chun, S. Ihm, P. Maniatis, M. Naik, and A. Patti, "CloneCloud: Elastic execution between mobile device and cloud," in *Proc. 6th ACM Eur. Conf. Comput. Syst.*, Salzburg, Austria, Apr. 2011, pp. 301–314.

[35] R. Kemp, N. Palmer, T. Kielmann, and H. Bal, "Cuckoo: A computation offloading framework for smartphones," in *Mobile Computing, Applications and Services* (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering), vol. 76, M. Gris and G. Yang Eds. Heidelberg, Germany: Springer, 2012, pp. 59–79.

[36] E. Cuervo *et al.*, "MAUI: Making smartphones last longer with code offload," in *Proc. ACM MobiSys*, San Francisco, CA, USA, Jun. 2010, pp. 49–62.

[37] 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.

[38] M. T. Beck, S. Feld, A. Fichtner, C. Linnhoff-Popien, and T. Schimper, "ME-volte: Network functions for energy-efficient video transcoding at the mobile edge," in *Proc. IEEE 18th Int. Conf. Intell. Next Gener. Netw. (ICIN)*, Paris, France, Feb. 2015, pp. 38–44.

[39] S. Abdelwahab, B. Hamdaoui, M. Guizani, and T. Znati, "REPLISOM: Disciplined tiny memory replication for massive IoT devices in LTE edge cloud," *IEEE Internet Things J.*, vol. 3, no. 3, pp. 327–338, Jun. 2016.

[40] H. Chang, A. Hari, S. Mukherjee, and T. V. Lakshman, "Bringing the cloud to the edge," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Toronto, ON, Canada, May 2014, pp. 346–351.

[41] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2795–2808, Oct. 2015.

[42] G. Orsini, D. Bade, and W. Lamersdorf, "Computing at the mobile edge: Designing elastic android applications for computation offloading," in *Proc. IEEE 8th IFIP Wireless Mobile Netw. Conf. (WMNC)*, Munich, Germany, Oct. 2015, pp. 112–119.

[43] P. A. Frangoudis, L. Yala, A. Ksentini, and T. Taleb, "An architecture for on-demand service deployment over a telco CDN," in *Proc. IEEE ICC*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[44] S. Retal, M. Bagaa, T. Taleb, and H. Flinck, "Content delivery network slicing: QoE and cost awareness," in *Proc. IEEE ICC*, Paris, France, May 2017.

[45] W. Zhu, C. Luo, J. Wang, and S. Li, "Multimedia cloud computing," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 59–69, May 2011.

[46] Y. Jararweh, L. Tawalbeh, F. Ababneh, and F. Dosari, "Resource efficient mobile computing using cloudlet infrastructure," in *Proc. IEEE 9th Int. Conf. Mobile Ad-Hoc Sensor Netw. (MSN)*, Dalian, China, Dec. 2013, pp. 373–377.

[47] "Small cell revolutionize service delivery," White Paper, Intel Corporat., Santa Clara, CA, USA, 2013.

[48] Y. Jin, Y. Wen, and C. Westphal, "Optimal transcoding and caching for adaptive streaming in media cloud: An analytical approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 12, pp. 1914–1925, Dec. 2015.

[49] E. Zeydan *et al.*, "Big data caching for networking: Moving from cloud to edge," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 36–42, Sep. 2016.

[50] 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.

[51] H. Routaib, E. Badidi, M. Elmachkour, E. Sabir, and M. Elkoutbi, "Modeling and evaluating a cloudlet-based architecture for mobile cloud computing," in *Proc. 9th IEEE Int. Conf. Intell. Syst. Theories Appl. (SITA)*, Rabat, Morocco, May 2014, pp. 1–7.

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

[53] G. Simmons, G. A. Armstrong, and M. G. Durkin, "An exploration of small business website optimization: Enablers, influencers and an assessment approach," *Int. Small Bus. J.*, vol. 29, no. 5, pp. 534–561, Feb. 2011.

[54] N. Takahashi, H. Tanaka, and R. Kawamura, "Analysis of process assignment in multi-tier mobile cloud computing and application to edge accelerated web browsing," in *Proc. 3rd IEEE Int. Conf. Mobile Cloud Comput. Services Eng. (MobileCloud)*, San Francisco, CA, USA, Mar. 2015, pp. 233–234.

[55] D. O. Mau, T. Taleb, and M. Chen, "MM3C: Multi-source mobile streaming in cache-enabled content-centric networks," in *Proc. IEEE Globecom*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[56] J. Zhu *et al.*, "Improving web sites performance using edge servers in fog computing architecture," in *Proc. IEEE 7th Int. Symp. Service Oriented Syst. Eng. (SOSE)*, Redwood City, CA, USA, Mar. 2013, pp. 320–323.

[57] 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.

[58] T. Taleb, A. Ksentini, and A. Kobbane, "Lightweight mobile core networks for machine type communications," *IEEE Access Mag.*, vol. 2, pp. 1128–1137, Oct. 2014.

[59] M. Aazam and E.-N. Huh, "Fog computing and smart gateway based communication for cloud of things," in *Proc. IEEE FICloud*, Barcelona, Spain, Aug. 2014, pp. 464–470.

[60] X. Sun and N. Ansari, "EdgeIoT: Mobile edge computing for the Internet of Things," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 22–29, Dec. 2016.

[61] L. Gu, D. Zeng, S. Guo, A. Barnawi, and Y. Xiang, "Cost efficient resource management in fog computing supported medical cyber-physical system," *IEEE Trans. Emerg. Topics Comput.*, vol. 5, no. 1, pp. 108–119, Jan./Mar. 2017.

[62] G. C. Fox, S. Kamburugamuve, and R. D. Hartman, "Architecture and measured characteristics of a cloud based Internet of Things," in *Proc. IEEE Int. Conf. CTS*, Denver, CO, USA, May 2012, pp. 6–12.

[63] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Jun. 2016.

[64] N. H. Motlagh, M. Bagaa, and T. Taleb, "UAV-based IoT platform: A crowd surveillance use case," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 128–134, Feb. 2017.

[65] M. Satyanarayannan *et al.*, "Edge analytics in the Internet of Things," *IEEE Pervasive Comput.*, vol. 14, no. 2, pp. 24–31, Apr./Jun. 2015.

[66] R. Yu, Y. Zhang, S. Gjessing, W. Xia, and K. Yang, "Toward cloud-based vehicular networks with efficient resource management," *IEEE Netw.*, vol. 27, no. 5, pp. 48–55, Sep./Oct. 2013.

[67] M. A. Hoque, M. Siekkinen, and J. K. Nurminen, "Energy efficient multimedia streaming to mobile devices—A survey," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 579–597, 1st Quart., 2014.

[68] J. O. Fajardo, I. Taboada, and F. Liberal, "Improving content delivery efficiency through multi-layer mobile edge adaptation," *IEEE Netw.*, vol. 29, no. 6, pp. 40–46, Nov./Dec. 2015.

[69] "Nokia liquid applications intelligent base station," White Paper, Nokia Netw., Espoo, Finland, 2014.

[70] H. Flinck, *Applications at the Edge*, Nokia Bell Labs Presentation, Murray Hill, NJ, USA, Jul. 2016.

[71] "Saguna and intel—Using mobile edge computing to improve mobile network performance and profitability," White Paper, Intel, Santa Clara, CA, USA, 2016.

[72] P. G. Lopez *et al.*, "Edge-centric computing: Vision and challenges," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 5, pp. 37–42, Oct. 2015.

[73] M. Jang, K. Schwan, K. Bhardwaj, A. Gavrilovska, and A. Avasthi, "Personal clouds: Sharing and integrating networked resources to enhance end user experiences," in *Proc. IEEE INFOCOM*, Toronto, ON, Canada, Apr. 2014, pp. 2220–2228.

[74] S. Davy *et al.*, "Challenges to support edge-as-a-service," *IEEE Commun. Mag.*, vol. 52, no. 1, pp. 132–139, Jan. 2014.

[75] C. Y. Huang, C. H. Hsu, Y. C. Chang, and K. T. Chen, "GamingAnywhere: An open cloud gaming system," in *Proc. ACM Multimedia Syst.*, Oslo, Norway, Feb. 2013. [Online]. Available: http://gaminganywhere.org/

[76] I. Afolabi *et al.*, "Towards 5G network slicing over multiple-domains," *IEICE Trans. Commun.* May 2017. [Online]. Available: https://www.jstage.jst.go.jp/article/transcom/advpub/0/advpub_2016NNI0002/_article

[77] S. K. Garg, S. Versteeg, and R. Buyya, "A framework for ranking of cloud computing services," *Future Gener. Comput. Syst.*, vol. 29, no. 4, pp. 1012–1023, Jun. 2013.

[78] J. Kołodziej, S. U. Khan, E. Gelenbe, and E.-G. Talbi, "Scalable optimization in grid, cloud, and intelligent network computing—Foreword," *Concurrency Comput. Pract. Exp.*, vol. 25, no. 12, pp. 1719–1721, Aug. 2013.

[79] S. J. Vaughan-Nichols, *Containers vs. Virtual Machines: How to Tell Which Is the Right Choice for Your Enterprise*, IT-World, Geneva, Switzerland, Apr. 2015.

[80] K. Townsend, *Containers: The Pros and the Cons of These VM Alternatives*, TechRepublic, San Francisco, CA, USA, Feb. 2015.

[81] M. G. Xavier *et al.*, "Performance evaluation of container-based virtualization for high performance computing environments," in *Proc. 21st IEEE Euromicro Int. Conf. Parallel Distrib. Netw. Based Process. (PDP)*, Belfast, U.K., Feb. 2013, pp. 233–240.

[82] B. I. Ismail *et al.*, "Evaluation of docker as edge computing platform," in *Proc. IEEE Conf. Open Syst. (ICOS)*, Malacca, Malaysia, Aug. 2015, pp. 130–135.

[83] C. Pahl and B. Lee, "Containers and clusters for edge cloud architectures—A technology review," in *Proc. 3rd IEEE Int. Conf. Future Internet Things Cloud (FiCloud)*, Rome, Italy, Aug. 2015, pp. 379–386.

[84] (2016). *Docker: Build, Ship, Run Any App, Anywhere*. [Online]. Available: https://www.docker.com/

[85] ETSI NFV, "Network function virtualisation: An introduction, benefits, enablers, challenges & call for action," Introductory White Paper, Issue 1, SDN & OpenFlow World Congr., Darmstadt, Germany, Oct. 2012.

[86] R. Mijumbi *et al.*, "Network function virtualization: State-of-the-art and research challenges," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 236–262, 1st Quart., 2016.

[87] J. G. Herrera and J. F. Botero, "Resource allocation in NFV: A comprehensive survey," *IEEE Trans. Netw. Service Manag.*, vol. 13, no. 3, pp. 518–532, Sep. 2016.

[88] *Network Functions Virtualisation (NFV); Architectural Framework, GS NFV 002*, ETSI NFV, Sophia Antipolis, France, Oct. 2013.

[89] *Network Functions Virtualisation (NFV); Management and Orchestration, GS NFV-MAN 001*, ETSI NFV, Sophia Antipolis, France, Dec. 2014.

[90] M. T. Beck and M. Maier, "Mobile edge computing: Challenges for future virtual network embedding algorithms," in *Proc. 8th Int. Conf. Adv. Eng. Comput. Appl. Sci. (ADVCOMP)*, Rome, Italy, Aug. 2014.

[91] A. Fischer, J. F. Botero, M. T. Beck, H. de Meer, and X. Hesselbach, "Virtual network embedding: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1888–1906, 4th Quart., 2013.

[92] J. Hwang, K. K. Ramakrishnan, and T. Wood, "NetVM: High performance and flexible networking using virtualization on commodity platforms," *IEEE Trans. Netw. Service Manag.*, vol. 12, no. 1, pp. 34–47, Mar. 2015.

[93] "Network function virtualization (NFV); ecosystem; report on SDN usage in NFV architecture framework," ETSI, Sophia Antipolis, France, Tech. Rep. ETSI GS NFV-EVE 005, v1.1.1, Dec. 2015.

[94] T. Taleb, "Toward carrier cloud: Potential, challenges, and solutions," *IEEE Wireless Commun. Mag.*, vol. 21, no. 3, pp. 80–91, Jun. 2014.

[95] "SDN architecture," ONF, Montreal, QC, Canada, Tech. Rep. TR-502, Jun. 2014.

[96] "SDN Architecture," ONF, Montreal, QC, Canada, Tech. Rep. TR-521, Feb. 2016.

[97] W. Xia, Y. Wen, C. H. Foh, D. Niyato, and H. Xie, "A survey on software-defined networking," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 27–51, 1st Quart., 2015.

[98] "Relationship of SDN and NFV," ONF, Montreal, QC, Canada, Tech. Rep. TR-518, Oct. 2015.

[99] J. Costa-Requena *et al.*, "SDN and NFV integration in generalized mobile network architecture," in *Proc. IEEE Eur. Conf. Netw. Commun. (EuCNC)*, Paris, France, Jun. 2015, pp. 154–158.

[100] Y. Li and M. Chen, "Software-defined network function virtualization: A survey," *IEEE Access*, vol. 3, pp. 2542–2553, Dec. 2015.

[101] A. M. Medhat *et al.*, "Service function chaining in next generation networks: State of the art and research challenges," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 216–223, Feb. 2017.

[102] M. Yang *et al.*, "Software-defined and virtualized future mobile and wireless networks: A survey," *Mobile Netw. Appl.*, vol. 20, no. 1, pp. 4–18, Feb. 2015.

[103] X. Jin, L. E. Li, L. Vanbever, and J. Rexford, "SoftCell: Scalable and flexible cellular core network architecture," in *Proc. ACM CoNEXT*, Santa Barbara, CA, USA, Dec. 2013, pp. 163–174.

[104] C. J. Bernardos *et al.*, "An architecture for software defined wireless networking," *IEEE Wireless Commun. Mag.*, vol. 21, no. 3, pp. 52–61, Jun. 2014.

[105] D. Siracusa, E. Salvadori, and T. Rasheed, "Edge-to-edge virtualization and orchestration in heterogeneous transport networks," in *Proc. IEEE Future Netw. Services (SDN4FNS)*, Trento, Italy, Nov. 2013, pp. 1–6.

[106] A. S. Thyagaturu, A. Mercian, M. P. McGarry, M. Reisslein, and W. Kellerer, "Software defined optical networks (SDONs): A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2738–2786, 4th Quart., 2016.

[107] K. Pentikousis, Y. Wang, and W. Hu, "MobileFlow: Toward software-defined mobile networks," *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 44–53, Jul. 2013.

[108] L. Valtulina, M. Karimzadeh, G. Karagiannis, G. Heijenk, and A. Pras, "Performance evaluation of a SDN/OpenFlow-based distributed mobility management (DMM) approach in virtualized LTE systems," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Austin, TX, USA, Dec. 2014, pp. 18–23.

[109] S. Jeon, C. Guimarães, and R. L. Aguiar, "SDN-based mobile networking for cellular operators," in *Proc. 9th ACM Workshop Mobility Evolving Internet Archit. (MobiArch)*, Maui, HI, USA, Sep. 2014, pp. 13–18.

[110] "Study on architecture for next generation system Rel. 14," 3GPP, Tech. Rep. 23.799, Dec. 2016.

[111] *Description of Network Slicing Concept, P1 WS1 E2E Architecture Team*, NGMN Alliance, San Diego, CA, USA, Jan. 2016.

[112] K. Samdanis, X. Costa-Perez, and V. Sciancalepore, "From network sharing to multi-tenancy: The 5G network slice broker," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 32–39, Jul. 2016.

[113] A. Blenk, A. Basta, M. Reisslein, and W. Kellerer, "Survey on network virtualization hypervisors for software defined networking," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 655–685, 1st Quart., 2016.

[114] *Mobile Edge Computing (MEC); Terminology, v1.1.1*, ETSI GS MEC Standard 001, Mar. 2016.

[115] *Mobile Edge Computing (MEC); Technical Requirements, v1.1.1*, ETSI GS MEC Standard 002, Mar. 2016.

[116] *Mobile Edge Computing (MEC); Framework and Reference Architecture, v1.1.1*, ETSI GS MEC Standard 003, Mar. 2016.

[117] D. Sabella, A. Vaillant, P. Kuure, U. Rauschenbach, and F. Giust, "Mobile-edge computing architecture: The role of MEC in the Internet of Things," *IEEE Consum. Electron. Mag.*, vol. 5, no. 4, pp. 84–91, Oct. 2016.

[118] *Mobile Edge Computing (MEC); Mobile-Edge Computing Proof-of-Concept Framework, V1.1.1*, ETSI GS MEC-IEG Standard 005, Aug. 2015.

[119] *MEC Proofs of Concept*, ETSI, Sophia Antipolis, France, 2016. [Online]. Available: http://www.etsi.org/technologies-clusters/technologies/mobile-edge-computing/mec-poc

[120] *Mobile Edge Computing (MEC); MEC Metrics Best Practice and Guidelines, V0.1.0*, ETSI GS MEC-IEG Standard 006, Jun. 2016.

[121] "Study on context aware service delivery for LTE, 3GPP TSG RAN meeting#71," 3GPP, Gothenburg, Sweden, Tech. Rep. RP-160633, Mar. 2016.

[122] "Communications of the ACM: Fog computing harnesses personal devices to speed wireless networks," Dept. Office Eng. Commun., ACM TechNews, New York, NY, USA, Tech. Rep., Dec. 2015.

[123] OpenFog Consortium, "OpenFog architecture overview," White Paper, Archit. Working Group, Feb. 2016.

[124] *Deployment of Mobile Edge Computing in an NFV Environment*, ETSI MEC work item DGS/MEC-0017MECinNFV, Sophia Antipolis, France, Dec. 2017.

[125] G. Somani, P. Khandelwal, and K. Phatnani, "VUPIC: Virtual machine usage based placement in IaaS cloud," *arXiv preprint arXiv:1212.0085*, Dec. 2012. [Online]. Available: https://arxiv.org/abs/1212.0085

[126] F. Z. Yousaf and T. Taleb, "Fine-grained resource-aware virtual network function management for 5G carrier cloud," *IEEE Netw. Mag.*, vol. 30, no. 2, pp. 110–115, Mar./Apr. 2016.

[127] H. Liang, L. X. Cai, D. Huang, X. Shen, and D. Peng, "An SMDP-based service model for interdomain resource allocation in mobile cloud networks," *IEEE Trans. Veh. Technol.*, vol. 61, no. 5, pp. 2222–2232, Jun. 2012.

[128] S. Kosta, A. Aucinas, P. Hui, R. Mortier, and X. Zhang, "ThinkAir: Dynamic resource allocation and parallel execution in the cloud for mobile code offloading," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012, pp. 945–953.

[129] Y. Liu, M. J. Lee, and Y. Zheng, "Adaptive multi-resource allocation for cloudlet-based mobile cloud computing system," *IEEE Trans. Mobile Comput.*, vol. 15, no. 10, pp. 2398–2410, Oct. 2016.

[130] T. C. Chieu, A. Mohindra, A. A. Karve, and A. Segal, "Dynamic scaling of web applications in a virtualized cloud computing environment," in *Proc. IEEE Int. Conf. e-Bus. Eng. (ICEBE)*, Oct. 2009, pp. 281–286.

[131] B. B. Nandi, A. Banerjee, S. C. Ghosh, and N. Banerjee, "Dynamic SLA based elastic cloud service management: A SaaS perspective," in *Proc. IFIP/IEEE Int. Symp. Integr. Netw. Manag. (IM)*, Ghent, Belgium, May 2013, pp. 60–67.

[132] S. Dutta, T. Taleb, and A. Ksentini, "QoE-aware elasticity support in cloud-native 5G systems," in *Proc. IEEE ICC*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[133] M. Jia, J. Cao, and W. Liang, "Optimal cloudlet placement and user to cloudlet allocation in wireless metropolitan area networks," *IEEE Trans. Cloud Comput.*, to be published, doi: 10.1109/TCC.2015.2449834,

[134] M. Bagaa, T. Taleb, and A. Ksentini, "Service-aware network function placement for efficient traffic handling in carrier cloud," in *Proc. IEEE WCNC*, Istanbul, Turkey, Apr. 2014, pp. 2402–2407.

[135] S. Agarwal *et al.*, "Volley: Automated data placement for geo-distributed cloud services," in *Proc. 7th Symp. Netw. Syst. Design Implement.*, San Jose, CA, USA, Apr. 2010, p. 2.

[136] A. Ksentini, M. Bagaa, T. Taleb, and I. Balasingham, "On using bargaining game for optimal placement of SDN controllers," in *Proc. IEEE ICC*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[137] T. Taleb, M. Bagaa, and A. Ksentini, "User mobility-aware virtual network function placement for virtual 5G network infrastructure," in *Proc. IEEE ICC*, London, U.K., Jun. 2015, pp. 3879–3884.

[138] H. Chan, D. Liu, P. Seite, H. Yokota, and J. Korhonen, "Requirements for distributed mobility management," Internet Eng. Task Force, Fremont, CA, USA, RFC 7333, Aug. 2014.

[139] D. Liu, J. C. Zuniga, P. Seite, H. Chan, and C. J. Bernardos, "Distributed mobility management: Current practices and gap analysis," Internet Eng. Task Force, Fremont, CA, USA, RFC 7429, Jan. 2015.

[140] F. Giust, L. Cominardi, and C. J. 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.

[141] L. F. Bittencourt, M. M. Lopes, I. Petri, and O. F. Rana, "Towards virtual machine migration in fog computing," in *Proc. IEEE 10th Int. Conf. P2P Parallel Grid Cloud Internet Comput. (3PGCIC)*, Kraków, Poland, Nov. 2015, pp. 1–8.

[142] H. Watanabe, T. Ohigashi, T. Kondo, K. Nishimura, and R. Aibara, "A performance improvement method for the global live migration of virtual machine with IP mobility," in *Proc. 5th Int. Conf. Mobile Comput. Ubiquitous Netw. (ICMU)*, Seattle, WA, USA, Apr. 2010.

[143] E. Harney, S. Goasguen, J. Martin, M. Murphy, and M. Westall, "The efficacy of live virtual machine migrations over the Internet," in *Proc. Int. Workshop Virtualization Technol. Distrib. Comput.*, Reno, NV, USA, Nov. 2007, Art. no. 8.

[144] K. Samdanis, T. Taleb, and S. Schmid, "Traffic offload enhancements for eUTRAN," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 3, pp. 884–896, 3rd Quart., 2012.

[145] T. Taleb and A. Ksentini, "Follow me cloud: Interworking federated clouds and distributed mobile networks," *IEEE Netw.*, vol. 27, no. 5, pp. 12–19, Sep./Oct. 2013.

[146] A. Ksentini, T. Taleb, and M. Chen, "A Markov decision process-based service migration procedure for follow me cloud," in *Proc. IEEE ICC*, Sydney, NSW, Australia, Jun. 2014, pp. 1350–1354.

[147] T. Taleb and A. Ksentini, "An analytical model for follow me cloud," in *Proc. IEEE Globecom*, Atlanta, GA, USA, Dec. 2013, pp. 1291–1296.

[148] A. Nadembega, A. S. Hafid, and R. Brisebois, "Mobility prediction model-based service migration procedure for follow me cloud to support QoS and QoE," in *Proc. IEEE ICC*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[149] S. Wang *et al.*, "Dynamic service migration in mobile edge-clouds," in *Proc. IEEE/IFIP Netw. Conf.*, Toulouse, France, May 2015, pp. 1–9.

[150] A. Aissioui, A. Ksentini, A. M. Gueroui, and T. Taleb, "Toward elastic distributed SDN/NFV controller for 5G mobile cloud management systems," *IEEE Access Mag.*, vol. 3, pp. 2055–2064, Nov. 2015.

[151] M. Portoles-Comeras, J. Mangues-Bafalluy, A. Krendzel, M. Requena-Esteso, and A. Cabellos-Aparicio, "An evolutionary path for the evolved packet system," *IEEE Commun. Mag.*, vol. 53, no. 7, pp. 184–191, Jul. 2015.

[152] S. Secci, P. Raad, and P. Gallard, "Linking virtual machine mobility to user mobility," *IEEE Trans. Netw. Service Manag.*, vol. 13, no. 4, pp. 927–940, Dec. 2016.

[153] A. Ksentini, T. Taleb, and F. Messaoudi, "A LISP-based implementation of follow me cloud," *IEEE Access*, vol. 2, pp. 1340–1347, Sep. 2014.

[154] H. Yang and Y. Kim, "SDN-based distributed mobility management," in *Proc. IEEE Int. Conf. Inf. Netw. (ICOIN)*, Kota Kinabalu, Malaysia, Jan. 2016, pp. 337–342.

[155] M. Gramaglia *et al.*, "Flexible connectivity and QoE/QoS management for 5G networks: The 5G NORMA view," in *Proc. IEEE ICC Workshops*, Kuala Lumpur, Malaysia, May 2016, pp. 373–379.

[156] A. Ceselli, M. Premoli, and S. Secci, "Cloudlet network design optimization," in *Proc. IEEE/IFIP Netw. Conf.*, Toulouse, France, May 2015, pp. 1–9.

[157] A. Ceselli, M. Premoli, and S. Secci, "Mobile edge cloud network design optimization," *IEEE/ACM Trans. Netw.*, to be published, doi: 10.1109/TNET.2017.2652850.

[158] P. Rost *et al.*, "Cloud technologies for flexible 5G radio access networks," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 68–76, May 2014.

[159] I. Chih-Lin *et al.*, "New paradigm of 5G wireless Internet," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 474–482, Mar. 2016.

[160] K. Sundaresan, M. Y. Arslan, S. Singh, S. Rangarajan, and S. V. Krishnamurthy, "FluidNet: A flexible cloud-based radio access network for small cells," in *Proc. 19th ACM MobiCom*, Miami, FL, USA, Sep. 2013, pp. 99–110.

[161] M. Payaro *et al.*, "Flex5Gware: Flexible and efficient hardware and software platforms for 5G network elements and devices," *Wiley Trans. Emerg. Telecommun. Technol.*, vol. 27, no. 9, pp. 1242–1249, Sep. 2016.

[162] N. Nikaein *et al.*, "OpenAirInterface: A Flexible platform for 5G research," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 5, pp. 33–38, Oct. 2014.

[163] P. Rost *et al.*, "Mobile network architecture evolution towards 5G," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 84–91, May 2016.

[164] T. Taleb, Y. Hadjadj-Aoul, and K. Samdanis, "Efficient solutions for enhancing data traffic management in 3GPP networks," *IEEE Syst. J.*, vol. 9, no. 2, pp. 519–528, Jun. 2015.

[165] D. Munaretto, F. Giust, G. Kunzmann, and M. Zorzi, "Performance analysis of dynamic adaptive video streaming over mobile content delivery networks," in *Proc. IEEE ICC*, Sydney, NSW, Australia, Jun. 2014, pp. 1053–1058.

[166] Y. Chen and L. Ruckenbusch, "Mobile edge computing: Bring the values back to networks," *IEEE Softwarization eNewsletter*, pp. 1–42, Mar. 2016.

[167] V. Sciancalepore, F. Giust, K. Samdanis, and Z. Yousaf, "A double-tier MEC-NFV architecture: Design and optimization," in *Proc. IEEE CSCN*, Berlin, Germany, Nov. 2016, pp. 1–6.

[168] Y. C. Hu, "MEC integration with NFV," in *Proc. Layer123 SDN OpenFlow World Congr.*, The Hague, The Netherlands, Oct. 2016.

[169] J. Liu, T. Zhao, S. Zhou, Y. Cheng, and Z. Niu, "CONCERT: A cloud-based architecture for next-generation cellular systems," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 14–22, Dec. 2014.

[170] 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.

[171] X. Chen, "Decentralized computation offloading game for mobile cloud computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 4, pp. 974–983, Apr. 2015.

[172] M. Patel *et al.*, "Mobile-edge computing introductory technical white paper," ETSI, Sophia Antipolis, France, and MEC, London, U.K., Tech. Rep., Sep. 2014.

[173] F. Lobillo *et al.*, "An architecture for mobile computation offloading on cloud-enabled LTE small cells," in *Proc. IEEE WCNC Workshops*, Istanbul, Turkey, Apr. 2014, pp. 1–6.

[174] M. A. Peunte, Z. Becvar, M. Rohlik, F. Lobillo, and E. C. Strinati, "A seamless integration of computationally-enhanced base stations into mobile networks towards 5G," in *Proc. IEEE VTC Spring*, Glasgow, U.K., May 2015, pp. 1–5.

[175] X. Masip-Bruin, E. Marin-Tordera, G. Tashakor, A. Jukan, and G.-J. Ren, "Foggy clouds and cloudy fogs: A real need for coordinated management of fog-to-cloud computing systems," *IEEE Wireless Commun.*, vol. 23, no. 5, pp. 120–128, Oct. 2016.

[176] (Mar. 2017). *OpenNFV*. [Online]. Available: http://www8.hp.com/us/en/cloud/nfv-architecture.html

[177] (Mar. 2017). *CloudNFV*. [Online]. Available: www.cloudnfv.com/WhitePaper.Pdf

[178] (Mar. 2017). *OpenBaton*. [Online]. Available: http://openbaton-docs.readthedocs.io/en/stable/

[179] (Mar. 2017). *OpenMANO*. [Online]. Available: https://www.tmforum.org/collaboration/catalyst-program/current-catalysts/

[180] (Mar. 2017). *Cloudify*. [Online]. Available: http://getcloudify.org/

[181] (Mar. 2017). *T-NOVA*. [Online]. Available: http://www.t-nova.eu

[182] (Mar. 2017). *OpenSourceMANO*. [Online]. Available: https://insights.ubuntu.com/wp-content/uploads/0e9c/E2E_OSM_Services_Instantiation_FINAL.pdf

[183] (Mar. 2017). *Open Platform for NFV (OPNFV)*. [Online]. Available: https://www.opnfv.org/

[184] (Mar. 2017). *ExperiaSphere*. [Online]. Available: http://www.experiasphere.com/

[185] CORD. (Mar. 2017). *M-CORD: Mobile CORD, Enable 5G on CORD*. [Online]. Available: http://www.opencord.org/

[186] (Mar. 2017). *ONOS*. [Online]. Available: http://onosproject.org/

[187] ONOS, "XOS: Service orchestration for CORD," White Paper, Jun. 2015.

[188] "Zero-time orchestration, operations and management (ZOOM)," TMF, Amsterdam, The Netherlands, Tech. Rep., Aug. 2014.

[189] *The Functional Architecture of Next Generation Service Overlay Networks*, IEEE Standard 1903-2011, Oct. 2011.

[190] F. Paganelli, M. Ulema, and B. Martini, "Context-aware service composition and delivery in NGSONs over SDN," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 97–105, Aug. 2014.

[191] (Mar. 2017). *Open-O*. [Online]. Available: https://www.open-o.org

[192] *ECOMP (Enhanced Control, Orchestration, Management & Policy) Architecture* White Paper, AT&T, Dallas, TX, USA, 2016.

[193] (Mar. 2017). *Open Network Automation Platform (ONAP) Project*. [Online]. Available: https://www.onap.org/

[194] R. Mijumbi *et al.*, "Management and orchestration challenges in network functions virtualization," *IEEE Commun. Mag.*, vol. 54, no. 1, pp. 98–105, Jan. 2016.

[195] "FG IMT-2020: Report on standards gap analysis," ITU, Geneva, Switzerland, Tech. Rep. TD 208 (PLEN/13), 2015.

[196] T. Taleb, K. Samdanis, and A. Ksentini, "Supporting highly mobile users in cost-effective decentralized mobile operator networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 7, pp. 3381–3396, Sep. 2014.

[197] D. Huang, T. Xing, and H. Wu, "Mobile cloud computing service models: A user-centric approach," *IEEE Netw.*, vol. 17, no. 6, pp. 6–11, Sep./Oct. 2013.

[198] X. Hou *et al.*, "Vehicular fog computing: A viewpoint of vehicles as the infrastructure," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 3860–3873, Jun. 2016.

[199] M. Whaiduzzaman, M. Sookhak, A. Gani, and R. Buyya, "A survey on vehicular cloud computing," *J. Netw. Comput. Appl.*, vol. 40, pp. 325–344, Apr. 2014.

[200] "Architecture enhancements for control and user plane separation of EPC nodes," 3GPP, Tech. Rep. 23.873, Dec. 2016.

[201] H. Ishii, Y. Kishihisa, and H. Takahashi, "A novel architecture for LTE-B: C-plane/U-plane split and phantom cell concept," in *Proc. IEEE Globecom Workshops*, Anaheim, CA, USA, Dec. 2012, pp. 624–630.

[202] S. Costanzo *et al.*, "An SDN-based virtual cell framework for enhancing the QoE in TD-LTE pico cells," *IEEE QoMEX*, Pilos, Greece, May 2015, pp. 1–6.

[203] K. Samdanis, K. Kunz, M. I. Hossain, and T. Taleb, "Virtual bearer management for efficient MTC radio and backhaul sharing in LTE networks," in *Proc. IEEE PIMRC*, London, U.K., Sep. 2013, pp. 2780–2785.

[204] "Proximity-based services (ProSe); stage 2, Rel.14," 3GPP, Tech. Rep. TS 23.303, Dec. 2016.

[205] G. Fettweise, W. Nagel, and W. Lehner, "Pathways to servers of the future: Highly adaptive energy efficient computing (HAEC)," in *Proc. Design Autom. Test Europe Conf. Exhibit. (DATE)*, Dresden, Germany, Mar. 2012, pp. 1161–1166.

[206] I. Stojmenovic and S. Wen, "The fog computing paradigm: Scenarios and security issues," in *Proc. IEEE FedCSIS*, Warsaw, Poland, Sep. 2014, pp. 1–8.

[207] A. Mtibaa, K. A. Harras, and H. A. Inuweiri, "Friend or foe? Detecting and isolating malicious nodes in mobile edge computing," in *Proc. IEEE CloudCom*, Vancouver, BC, Canada, Nov./Dec. 2015, pp. 42–49.

[208] S. Lal, T. Taleb, and A. Dutta, "NFV: Security threats and best practices," *IEEE Commun. Mag.*, to be published, doi: 10.1109/MCOM.2017.1600899.

[209] S. Lal, I. Oliver, S. Ravidas, and T. Taleb, "Assuring virtual network function image integrity and host sealing in telco cloud," in *Proc. IEEE ICC*, Paris, France, May 2017.

Tarik Taleb (M'05–SM'10) received the B.E. (with Distinction) degree in information engineering and the M.Sc. and Ph.D. degrees in information sciences from Tohoku University, in 2001, 2003, and 2005, respectively. He is currently a Professor with the School of Electrical Engineering, Aalto University, Finland. In 2009, he was an Assistant Professor with the Graduate School of Information Sciences, Tohoku University, Japan, in a lab fully funded by KDDI. He was a Senior Researcher and 3GPP Standards Expert with NEC Europe Ltd., Heidelberg, Germany. He was then leading the NEC Europe Labs Team working on Research and Development projects on carrier cloud platforms, an important vision of 5G systems. From 2005 to 2006, he was a Research Fellow with Intelligent Cosmos Research Institute, Sendai, Japan.

His research interests lie in the field of architectural enhancements to mobile core networks (particularly 3GPP's), mobile cloud networking, network function virtualization, software defined networking, mobile multimedia streaming, intervehicular communications, and social media networking. He has been also been directly engaged in the development and standardization of the Evolved Packet System as a member of 3GPP's System Architecture working group.

Prof. Taleb is a member of the IEEE Communications Society Standardization Program Development Board. As an attempt to bridge the gap between academia and industry, he founded the IEEE Workshop on Telecommunications Standards: from Research to Standards, a successful event that received the Best Workshop Award from the IEEE Communication Society (ComSoC). Based on the success of this workshop, he has also founded and has been the Steering Committee Chair of the IEEE Conference on Standards for Communications and Networking.

He is the General Chair of the 2019 edition of the IEEE Wireless Communications and Networking Conference (WCNC'19) to be held in Marrakech, Morocco. He is/was on the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the *IEEE Wireless Communications Magazine*, the IEEE JOURNAL ON INTERNET OF THINGS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE COMMUNICATIONS SURVEYS & TUTORIALS, and a number of Wiley journals. In 2016, he served as the Chair of the Wireless Communications Technical Committee, the largest in IEEE ComSoC. He also served as the Vice Chair of the Satellite and Space Communications Technical Committee of IEEE ComSoc from 2006 to 2010. He has been on the Technical Program Committee of different IEEE conferences, including Globecom, ICC, and WCNC, and chaired some of their symposia.

He was a (co)-recipient of the 2017 IEEE Communications Society Fred W. Ellersick Prize in 2017, the 2009 IEEE ComSoc Asia–Pacific Best Young Researcher Award in 2009, the 2008 TELECOM System Technology Award from the Telecommunications Advancement Foundation in 2008, the 2007 Funai Foundation Science Promotion Award in 2007, the 2006 IEEE Computer Society Japan Chapter Young Author Award in 2006, the Niwa Yasujirou Memorial Award in 2005, and the Young Researcher's Encouragement Award from the Japan Chapter of the IEEE Vehicular Technology Society in 2003. He was a recipient of the Best Paper Awards at prestigious conferences. He is an IEEE ComSoc Distinguished Lecturer.

Konstantinos Samdanis received the M.S. degree from Kings College London. He is a Principal Researcher with Huawei for 5G carrier networks. He is involved in strategy on 5G network slicing for transport and mobile networks, and in research for 5G SDN/NFV architectures and network slice OS. He is also active at the 5GPPP Architecture Working Group and at Broadband Forum in Wireless-Wired Converged Networks Working Group, where he has initiated a study item on Network Slicing. He has worked for NEC Europe, Germany, as a Senior Researcher and a Standardization Specialist, involved in numerous EU projects such as 5G-NORMA, iJOIN, BeFemto, CROSSFIRE, and standardization activities related to 3GPP SA2 and SA5 as well as at the Broadband Forum in the field of Mobile Backhaul and IP/MPLS.

Sunny Dutta received the bachelor's degree from the West Bengal University of Technology, India, in 2006, and the M.Sc. degree from the School of Electrical Engineering, Aalto University, Finland, in 2016. He was an Engineer assuming different roles in network administration, energy automation, and smart grid communication network infrastructure. His present research focus includes MEC, NFV, SDN, and multimedia content delivery.

Badr Mada received the bachelor's degree in mathematical and computer science and the master's degree in software engineering from University Mohammed V, Rabat, Morocco, in 2014 and 2017, respectively. He is currently pursuing the Doctoral degree with Aalto University. His research focuses on mobile edge computing and open source networking.

Dario Sabella received the degree in electronic engineering and the master's degree in telecommunications in 2004. In 2002, he was a System Specialist and Embedded Applications Developer with Telecom Italia Group. In 2006, he entered the Wireless Innovation Department where he built a deep experience on communication systems (WiFi, WiMAX/WiBro, UMTS/HSPA, LTE, LTE-A), with special knowledge of wireless radio access technologies and system engineering for mobile networks toward 5G. He has recently specialized his technical skills in the field of energy efficiency for mobile networks (also as delegate in ETSI TC EE) and mobile edge computing (since 2015, he has been serving as the Vice Chairman of ETSI MEC IEG). Since 2009, he has been involved in European projects (FP7, EIT Digital, Horizon 2020), often in leadership roles. He has collected a wide set of over 18 international patents and over 40 publications. He has organized international conferences and workshops and has experience with IPR management and project management. He is a Co-Founder and the Chairman of the CLEEN Workshops Series. In 2017, he joined Intel as a Senior Standards and a Research Engineer. He is currently serving ETSI MEC as an ISG Secretary and Lead of Industry Group Relationships. He is also active delegate in 5GAA. He has been a member of the AEIT since 2012.

Hannu Flinck received the M.Sc. and Lic.Tech. degrees in computer science and communication systems from Aalto University (formerly, Helsinki University of Technology) in 1986 and 1993, respectively. He was with Nokia Research Center and the Technology and Innovation Unit of Nokia Networks in various positions. He is a Research Manager with Nokia Bell Labs, Espoo, Finland. He has been actively participating in a number of EU research projects. His current research interests include mobile edge computing, SDN, and content delivery in mobile networks, particularly in 5G networks.