Journal of Internet Services and Applications, 2023, 14:1, doi: 10.5753/jisa.2023.2913

Service Provisioning in Edge-Cloud Continuum: Emerging Applications for Mobile Devices

Diego O. Rodrigues [**University of Bern** | ***diego.oliveira@unibe.ch***]
Allan M. de Souza [**University of Campinas** | ***allanms@lrc.ic.unicamp.br***]
Torsten Braun [**University of Bern** | ***torsten.braun@unibe.ch***]
Guilherme Maia [**Federal University of Minas Gerais** | ***jgmm@dcc.ufmg.br***]
Antonio A. F. Loureiro [**Federal University of Minas Geraiss** | ***loureiro@dcc.ufmg.br***]
Leandro A. Villas [**University of Campinas** | ***lvillas@unicamp.br***]

University of Campinas - Institute of Computing, Av. Albert Einstein, 1251, Cidade Universitária, Campinas - SP. Cep: 13083-852

Received: 26 September 2022 • **Accepted:** 14 March 2023 • **Published:** 20 June 2023

Abstract Disruptive applications for mobile devices can be enhanced by Edge computing facilities. In this context, Edge Computing (EC) is a proposed architecture to meet the mobility requirements imposed by these applications in a wide range of domains, such as the Internet of Things, Immersive Media, and Connected and Autonomous Vehicles. EC architecture aims to introduce computing capabilities in the path between the user and the Cloud to execute tasks closer to where they are consumed, thus mitigating issues related to latency, context awareness, and mobility support. In this survey, we describe which are the leading technologies to support the deployment of EC infrastructure. Thereafter, we discuss the applications that can take advantage of EC and how they were proposed in the literature. Finally, after examining enabling technologies and related applications, we identify some open challenges to fully achieve the potential of EC, and also research opportunities on upcoming paradigms for service provisioning. This survey is a guide to comprehend the recent advances on the provisioning of mobile applications, as well as foresee the expected next stages of evolution for these applications.

Keywords: Fog Computing, Multi-access Edge Computing, Mobile Applications, Network Handover, Internet of Things, Smart Cities, Intelligent Transportation Systems, Unmanned Aerial Vehicles, Immersive Media, Software-Defined Networking, Future Internet.

1 Introduction

As we experience a fast growth in the number and types of devices connected to the Internet, new classes of applications emerge. These upcoming applications promise to change human life, from elementary tasks like turning on a lamp to complex activities such as running an entire industrial factory. Intelligent Transportation Systems (ITSs) optimize the way people and goods traverse the city towards their destinations. For instance, connected and autonomous vehicles (CAVs) are awaited to improve the quality of experience (QoE) for driving (Amadeo *et al.*, 2016). Still, to support these vehicles multiple sensors must be deployed in the vehicle or on its premises, leading to a high volume of data to be transmitted and processed. Unmanned Aerial Vehicles (UAVs) connected to the network will also perform tasks such as surveillance, search and rescue, and various monitoring tasks (Rahman *et al.*, 2018). To execute these tasks, UAVs will have to rely on computing-intensive navigation and cooperation support services provided continuously regardless of their mobility. Moreover, immersive media devices are awaited to disrupt scenarios related to entertainment, communication, and even more traditional ones such as medicine and education (You *et al.*, 2019). When real-time immersive media, such as augmented or virtual reality, is consumed inside vehicles, multiple tasks for data collection and analysis need to be simultaneously executed. Since immersive media devices (e.g., smart glasses) have limited capacity, technologies for task offloading and mobility support are necessary to achieve the expected levels of QoE (Wireless One, 2018). In addition, there is a vast amount of possibilities for monitoring and acting over real-world environments enabled by the Internet of Things (IoT) and its anticipation of 50 billions connected devices very soon (Salman *et al.*, 2018). It is worth noticing that the building blocks of these applications are multiple services that interact with each other. The execution of these services require the development of a complex service management solution that is aware of the characteristics and requirements of applications.

One factor that many of the above applications share is mobility. While some applications operate statically in the same place, mobile devices and wireless communication led to a culture shift. This turned concerns regarding user mobility and mobility-aware service provisioning into major research topics. Actually, mobility management is one of the driving forces of future applications pushing the development of EC architectures.

The emergence of such novel classes of applications comes together with ever-increasing demands for quality of service (QoS). Currently, the Cloud is the preferred approach for handling massive service provisioning. However, this paradigm cannot cope with some of the upcoming requirements of these applications (Roman *et al.*, 2018; Bi *et al.*, 2019), such as: (i) inability to maintain low levels of latency, due to the distance between the consumer and the host where applications run; (ii) difficulty in providing location and context awareness needed to enhance the QoE delivered from services to users; and (iii) lack of mobility support needed to keep service continuity in response to network transitions originated due to users' mobility patterns.

To address the issues faced by Cloud Computing and support modern applications, many architectures arise, such as Fog Computing (Naha *et al*., 2018), Multi-access Edge Computing (MEC) (Mao *et al*., 2017a), and other related computing technologies (Roman *et al*., 2018; Ning *et al*., 2019). These architectures propose the deployment of a computing infrastructure along the path from the Edge of the network to the Cloud. The communication infrastructure also has to be transformed to manage these resources since they add more complexity to the network in terms of management. To realise this communication transformation, Future Internet technologies, such as Software-Defined Networking (SDN) (Saraswat *et al*., 2019), Network Function Virtualization (NFV) (Nguyen *et al*., 2017b), and Information-Centric Networking (ICN) (Fang *et al*., 2018), emerge to empower wired and wireless communication channels.

To deploy EC, a series of technologies have been proposed and developed. For instance, software-defined networks are promising candidates to manage the communication between applications and users (Mao *et al*., 2017a). SDN decouples routing logic from forwarding in the network routers, turning them into programmable switches that can receive instructions from a centralized controller to only dispatch network packets. The existence of a centralized controller in SDN facilitates the management of the network since a general overview of the network can be used to make decisions. NFV is another important component to deploy a virtual network (Nguyen *et al*., 2017b). NFV supports the development of the underlying network node functionalities as virtual functions used to provide communication services. Through the virtualization of the network, a more adaptive network is expected, which facilitates the deployment of new network protocols and services, and also a faster response to dynamic changes in the network state.

1.1 Related Surveys

Table 1 shows a list of relevant recently published surveys in the literature that have similar goals to our study. For each reference, the year of publication, the application category, and the perspective of the survey are shown. The perspective column gives insights into the main subject of discussion of the survey. Thus, (i) Communication focus on the communication technology and channels used to implement the application, (ii) System describes the multiple component parts of the application and how they interact, and (iii) Application discusses the applications themselves, their relation to the real world, and the technologies involved in their implementation. The columns Cloud, Edge and Network virtualization inform whether the usage of these prominent technologies for Future Internet is considered on the survey when discussing the applications. It can be observed that multiple studies focus on discussing specific applications, while some are more general, providing a more broad view of the promising future applications. Also, it is possible to observe the increasing focus on applications for mobile devices over the last years due to the increasing popularization of such devices, not only smartphones, but also smart devices and vehicles, and wireless communication. Two studies in the list stand with closed objectives to our survey. Firstly, Santos *et al*. (2021) discusses a series of enabling technologies for the provisioning of services and applications with strict latency requirements. Our study differs from this study by focusing on a wider range of scenarios and also on the characteristics of the applications. Secondly, Khan *et al*. (2022) explore the studies in the literature that apply the Digital Twin abstraction when implementing and deploying applications. In our study we do not focus on a single abstraction, indeed we do not study implementation abstractions of applications. Our survey extensively explores service provisioning focusing on mobility challenges and computing and communication technologies, providing a broader view of trending applications use cases. It provides an entry point for researchers studying applications for mobile devices that run at the edge.

1.2 Objectives and Contributions

The objective of this survey is to provide an entry point for researchers studying applications for mobile devices and their supporting technologies. To achieve this aim we elaborated a review of the literature focusing on three main areas: (i) the leading technologies to support emerging application use cases, (ii) the mobility of devices that consume these applications, and (iii) the characteristics of the applications themselves. By "emerging" applications, we refer to applications that are coming to adoption with initial and, sometimes, incomplete implementations in the real world. For instance, currently there are self-driving vehicles out there, still these vehicles have not achieved their full potential. One of the main reasons is that these vehicles cannot freely communicate with each other and the infrastructure to aid the driving process. When discussing leading technologies, we aim to target the supporting technologies most adopted and/or discussed in the literature. There are a series of novelty application use cases, e.g., Tactile Internet (Promwongsa *et al*., 2021), and also technologies, e.g., re-configurable surfaces (Liu *et al*., 2021a), that were left out of the survey since it is not very clear when they will be implemented or adopted. The most important contributions of this survey are to:

- describe the history of the evolution of service provisioning over the internet highlighting the different paradigms adopted and the driving forces for their adoption;
- give an overview of technologies and architectures that enable service provisioning at the edge of the network;
- highlight most important features and characteristics of a variety of emerging applications;
- outline the main challenges faced from the communication perspective to the deployment of these applications.

Table 1. List of recently published related surveys in the literature.

Study	Year	Application	Cloud	Edge	Network Virtualization	Mobility	Perspective
Khan *et al.* (2022)	2022	General	✓	✓	✓	✓	Applications
Fizza *et al.* (2022)	2022	IoT	✓	✓		✓	Applications
Metzger *et al.* (2022)	2022	Gamming	✓	✓		✓	Systems/Communications
Arbabi *et al.* (2022)	2022	e-Health					Systems
Hafner *et al.* (2021)	2021	Autonomous Vehicles		✓		✓	Applications
Afrin *et al.* (2021)	2021	Robotics	✓	✓		✓	Systems
Santos *et al.* (2021)	2021	General	✓	✓	✓	✓	Communications
Navarro-Ortiz *et al.* (2020)	2020	General				✓	Communications
Qadri *et al.* (2020)	2020	IoT for e-Health	✓	✓	✓		Systems
Wang *et al.* (2020)	2020	UAV	✓	✓		✓	Systems
Barakabitze *et al.* (2020)	2020	Multimedia	✓	✓	✓		Communications
Aceto *et al.* (2019)	2019	IoT for Industry 4.0	✓	✓			Systems
This Survey		General	✓	✓	✓	✓	Applications

1.3 Organization

This study contributes to the current state of the literature by discussing several mobile applications that can take advantage of EC infrastructures. Mobile applications have a significant role in EC-enabled settings due to the importance of mobile devices in the envisioned applications for the Future Internet. In order to understand the different shifts of computing and networking paradigms, we briefly introduce a historical view of the evolution of the requirements for service provisioning over the Internet in Section 2. It is worth noticing that many studies focus on developing the underlying architectures and mobility management strategies to support the applications, while other studies focus on evolving the applications themselves. Thus, we first elaborate a classification of the main architectures in EC and enumerate studies about them in Section 3. Afterwards, we examine different mobility management strategies to seamless service provisioning in Section 4. Then, Section 5 presents different applications and discuss the community view on how to better explore EC infrastructures in their composition. Section 6 discusses challenges and opportunities for mobile service provisioning. Finally, Section 7 summarizes this study and presents some final remarks.

2 History of Service Provisioning Technologies

Society has evolved into a state that continuous information exchange is required to improve citizens' lives in large urban centers. For instance, the advances achieved due to smartphone popularization make it hard to imagine a non-connected future. Indeed, the requirements for connectivity tend to increase with the emergence of new technologies such as IoT, Autonomous Vehicles, Immersive Media, and others. We will survey the technologies developed to support these increasing requirements and also some envisioned applications that will take advantage of these technologies. Before, however, this section presents a brief historical overview of the computing and communication technologies for service provisioning. We organize this historical overview in four "stages" of evolution according to the placement of the computing resources: (1) Self-hosted Service Provisioning; (2) Cloud-based Service Provisioning; (3) Edge-based Service Provisioning; and (4) Ubiquitous Service Provisioning. Figure 1 illustrates these stages. The main cultural-technological shifts of each stage are discussed in Sections 2.1 to 2.4. As discussed later on, in the current state of evolution, technology and standards are being proposed and developed to fully achieve stage 3 of Edge-based service provisioning.

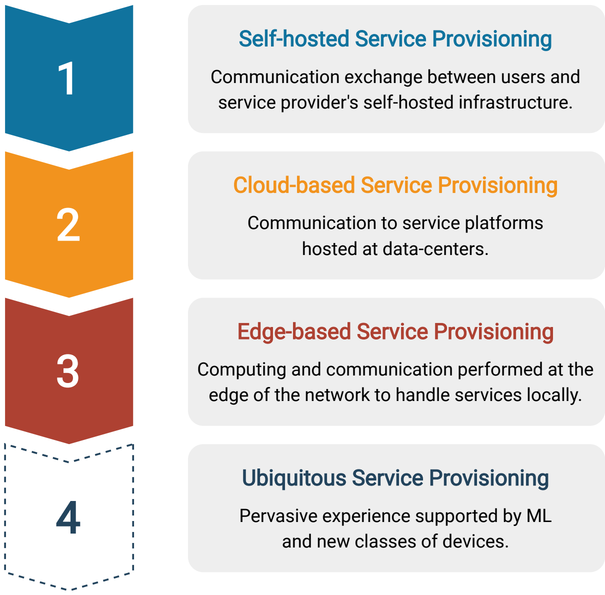

Figure 1. Evolution stages of service provisioning technologies.

2.1 Self-hosted Service Provisioning

The Internet design, made in the 1970s, was created with a very different objective than today's Internet usage. By then, service providers would have to deploy and maintain their own infrastructure to provide their services, which characterizes stage 1 of self-hosted service provisioning. Users consumed services through emerging technologies of the time, such as the TCP/IP protocol. The IP protocol was designed to handle the addressing of hosts in a topology-based network and remains the main addressing protocol nowadays. The initial applications evolving by that time were e-mails, chat rooms, and later shopping and banking. However, such applications became very popular and started to face scalability issues. These concerns became the main reasons to push forward the service provisioning paradigm to the next stage.

2.2 Cloud-based Service Provisioning

The term Cloud Computing was initially used in the late 1990s (Favaloro, 1996), becoming broadly adopted by the late 2000s (Weiss, 2007) when large companies start adopting it and pushing towards stage 2 of Cloud-based service provisioning. The idea was to offer companies the possibility to acquire computing resources that would scale on demand. Thus, virtually infinite computing resources were deployed to data centers strategically positioned around the globe. This paradigm is currently the main method for providing services on the Internet and allowed companies and their developers to overcome some of the scalability issues previously faced. Nevertheless, scalability issues once again became a big concern with the massive adoption of applications such as Video Streaming, Gaming, and Social Media, i.e., resource-consuming applications accessed by millions of users. Different technologies were developed to handle the emerging 1990s-2000s applications and meet business requirements of resource and energy saving, and massive provision to millions of users. For instance, the platforms sold to companies to run their services should be virtualized to facilitate portability and application deployment. Consequently, service virtualization became an important research topic, with virtual machines, and more recently containers, being the main approach to provide Platform as a Service (Dua *et al*., 2014). More complex applications are still under development. Even network functions are being virtualized to handle these applications inside datacenters, using technologies such as NFV (Jin and Wen, 2017) and SDN (Amin *et al*., 2018).

In stage 2 of the evolution, we also experienced the popularization of different types of devices. Mobile devices became mainstream, such as tablets and smartphones, and later a many other devices also started to be plugged into the Internet. Many envisioned applications are coming to reality due to the wide adoption of these devices. Such applications, however, are facing issues due to the 1970s topology-based design of the Internet. Therefore, we are again undergoing a paradigm shift in service provisioning.

2.3 Edge-based Service Provisioning

Multiple applications rely on low latency machine-type communication to run, thus, creating a need for faster communication and computation. Besides latency, the enormous amount of data generated by various devices connected to the Internet cannot constantly traverse significant distances, due to the risk of overloading the network infrastructure (Satyanarayanan, 2017). Furthermore, context-awareness became a requirement for the correct functioning of specific location-based applications. Therefore, the issues mentioned in Section 2.2, combined with the development of new applications supported by a wide range of devices, are driving service provisioning towards stage 3. In that stage, computing tasks run closer to end-users to maintain sustainable scalability. The main architectures of EC being studied to overcome those issues are Fog and Multi-access Edge Computing (Roman *et al*., 2018; Nguyen Gia *et al*., 2018). These computing architectures propose the placement of computing resources closer to users, allowing services to execute locally and mitigate issues of latency, network overloading, and context awareness.

2.3.1 Fog and Multi-Access Edge Computing

Fog Computing is an extension of the Cloud computing paradigm that expands the resource pool with resources from a plethora of devices, such as micro-datacenters (i.e., a smaller and self-contained category of a datacenter that comes in different sizes with cooling, security, and protection solutions out of the shelf) and also end-user devices. These computing facilities are deployed in spatially distributed Points of Presence (PoPs) to provide computational resources closer to users (Roman *et al*., 2018).

MEC (Mao *et al*., 2017a) is a similar paradigm where computing tasks may run at resources closer to users. Both architectures aim to provide computation closer to the Edge of the network by deploying resources or using idle resources from end-user devices, thus causing Fog Computing and MEC to share many features. The two main differences between Fog Computing and MEC are (Roman *et al*., 2018): (i) ownership: MEC is kept by telecommunications companies, while Fog Computing is typically maintained by private providers (e.g., AWS, Google); and (ii) deployment: MEC is only located at the edge of the network, while Fog Computing also uses resources placed strategically closer to end-users but not at the edge, for instance, data centers in neighbor cities/states or closer to the entrance of the core network (i.e., near-Edge resources).

With the current development of 5G networks, partnerships between Telecommunication companies and Cloud infrastructure providers have been created to allow service provisioning at the edge. For example, Amazon Web Services (AWS) has established partnerships with multiple companies in different countries to create AWS Wavelength, a publicly available Multi-access Edge Computing (MEC) Infrastructure as a Service (Vodafone, 2021). Still, EC-enabled applications have not yet been widely adopted. Emerging applications such as connected vehicles and Augmented Reality are not freely available to the public in cellular networks. Despite the more controlled settings of EC-enabled applications in real-world deployments, research on EC technologies has been widely performed. The EC paradigm comprises the two main Edge computing architectures, Fog Computing and MEC, and related architectures aimed to handle services closer to end-users.

Throughout the years, the definitions of Fog and MEC have evolved towards each other and many times, even in academic studies, these terms are used interchangeably. This evolution of the definitions is mostly due to two motivations. Firstly, a wider set of access technologies was envisioned for Multi-Access Edge Computing, which was previously called Mobile Edge Computing to highlight the usage of only mobile communication. And secondly, the expansion of the set of devices sometimes considered part of the Fog Computing infrastructure. These terms might be considered interchangeable in many aspects, being the most fundamental difference among them the focus of each architecture. Fog Computing studies target discussions about the infrastructure perspective and the placement of resources in the Cloud-Things continuum, whereas Edge Computing studies have their focus on the devices at the edge Shi *et al*. (2016). Devices discussed in Fog Computing may be placed not only at the edge, but also at the near-edge infrastructure, such as at the border to the core network or closer to the cloud. This distribution of resources often leads to the existence of a hierarchy of resources composing the Fog, with multiple layers closer or further to the end-user comprising more or less resources. Such hierarchy is not found when studying MEC, which is usually represented as a horizontal architecture.

In the EC stage of evolution, computing resources are deployed at the edge of the network, accessible through multiple access points and using different access technologies and channels. Wireless access points have limited coverage areas. Thus, when users move, they switch from one access point to another, causing a network handover, i.e., configuration updates in the network and user terminal to connect to the new access point. Current-state topology-based networking would struggle to handle multiple handovers while meeting the QoE requirements. Thus, research has been done to adapt networks for mobility-aware service provisioning. In this scenario, dynamically adaptable networks have an important role due to the ease of deploying new network algorithms and protocols to achieve better mobility management for connections.

2.3.2 Envisioned EC-enhanced Applications

Due to the possibilities envisioned by EC-enabled computing, developers started to foresee new classes of applications. The most discussed of these classes in the literature are: (i) Intelligent Transportation Systems (ITS), with applications for traffic management (Wang *et al*., 2018b; Ahmad *et al*., 2019), traffic lights control (Liu *et al*., 2018), and self-driving vehicles (Su *et al*., 2018b; Peng *et al*., 2019); (ii) Immersive Media (You *et al*., 2019), with applications in Augmented Reality or Virtual Reality; (iii) Unmanned Aerial Vehicles (UAVs) (Lei *et al*., 2019; Zhang *et al*., 2018c; Kalatzis *et al*., 2018) for monitoring or surveillance tasks; (iv) Smart Cities applications to handle public infrastructure (Katsaros *et al*., 2014; Li *et al*., 2017a); (v) other Internet of Things (IoT) related applications, such as healthcare (Liao *et al*., 2019), body area sensing with wearable and implantable devices (Lal and Kumar, 2017), and also Industry 4.0 (Hofer *et al*., 2019). These applications will be widely adopted in the future, thus generating more requirements for EC-enabled settings. One of these requirements is mobility awareness since users with different mobility patterns will use many of these applications.

Moreover, a composition of EC to run services and applications, and network management handled by SDN is envisioned as the Future Internet (Salman *et al*., 2018; Zhang *et al*., 2018b; Nobre *et al*., 2019). This vision is due to the possibility of using these technologies to deploy 5G networks. The challenges to fully achieve stage 3 of evolution are discussed in Section 6.

2.4 Ubiquitous Service Provisioning

Devices are evolving together with applications and adding a sense of pervasiveness to the Internet. Examples of these devices are smart glasses and head-mounted displays, some enabling movement-free access to immersive media virtually anywhere. These devices are key enablers on the information access revolution from desktop to smartphones and then toward freedom of form and location (Cuthbertson, 2019). Other technologies such as wearable devices, body area and ultra-dense networks also contribute to such increase in the sense of the pervasiveness of services. To deploy future applications envisioned for these devices, such as Tactile Internet and Internet of Skills (Antonakoglou *et al*., 2018), the architecture for service provisioning will have to evolve into an envisioned stage 4 of service provisioning. In this stage, EC, extended with pervasive devices, relies on Machine Learning (ML) to enhance its context awareness, latency reduction, mobility support, and other capabilities. There is still a big technological gap for achieving fully immersive and experiences and ubiquitous service provisioning, still the development of EC infrastructure and its architectures for seamless provisioning for mobile devices, as discussed in Section 3, are important steps towards reaching stage 4 of service provisioning.

3 Seamless Service Provisioning for Mobile Devices

This section describes the envisioned Future Internet for massive service provisioning in EC-enabled settings while dealing with mobility. Figure 2 shows a three layer architecture composed by: *(i)* the Edge; *(ii)* the core network; and *(iii)* the Cloud. Multiple technologies interact with each other to support service provisioning in each layer. The bottom layer is the Edge, where technologies such as Mobile Cloud Computing (MCC), Vehicular Cloud Computing (VCC), Vehicular Edge Computing (VEC), and Floating Content provide processing, storage and communication capabilities to run services. The core network layer in the middle manages networking and computing resources offered to users. This management is done by the SDN controller, which may use different abstractions and protocols for this task. Also, the core network layer enlarges the computing resource pool with near-Edge Fog nodes, nodes deployed with more computing power than observed at the edge but not as far as the Cloud. Finally, the highest computing power is provided at the cloud to handle eventual resource constraints of the lower layers.

The main difference between the layers in Figure 2 is the distance between users and resources. In Cloud Computing, a vast amount of resources is placed in data-centers usually positioned away from the users; while at the edge, the resources are distributed in smaller amounts closer to users. Furthermore, networking paradigms, such as SDN and ICN are expected to empower static networks and mobile networks such as Mobile Ad Hoc Networks (MANETs) and Vehicular Ad-hoc Networks (VANETs) and build the Future Internet. The computing architectures that compose EC are

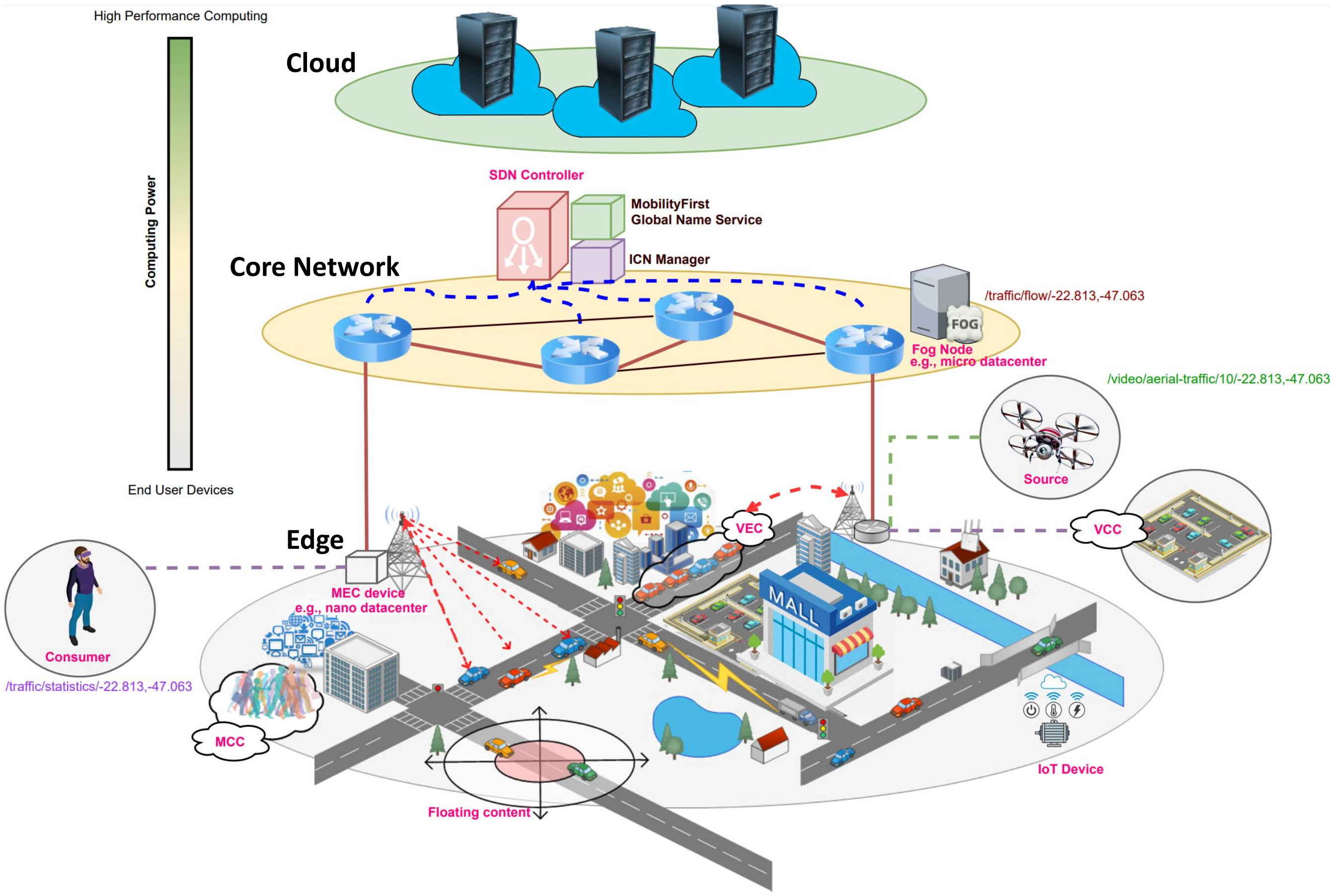

Figure 2. Landscape of Future Internet for massive service provisioning in EC-enabled settings.

discussed in Section 3.1, while networking architectures are discussed in Section 3.2.

3.1 Computing Architectures

As smart Edge devices become more popular, the IoT era emerges with the tendency of connecting these devices to the Internet to support different services and applications. Many of these devices are simple and resource constrained in terms of computing capabilities and power supply. Due to the observation that multiple services would demand more resources to execute, researchers developed offloading techniques to migrate their computation to the Cloud (Liu *et al.*, 2013). Yet, relying only on the Cloud has its drawbacks since datacenters are placed away from the Edge devices and end-users.

3.1.1 Fog Computing

Cisco introduced the idea of Fog Computing in 2012 (Bonomi *et al.*, 2012) with the objective of extending the Cloud paradigm closer to people (i.e., end-users). This new computing architecture was aimed to be placed between the traditional Cloud and the users in the form of datacenters positioned at the edge or near-Edge of the network, before the gateway to the core network. Yet lately, the concept evolved to contemplate even idle computer resources in end-users Edge devices to be added to the pool (Roman *et al.*, 2018). This framework is built to provide features such as low latency, geo-distribution, location awareness, and mobility support (Salman *et al.*, 2018). These features aim to fill the gaps of the traditional Cloud Computing paradigm, thus creating a complementary Cloud-Fog Computing architecture, where the resources are selected according to the volume and speed of the processing tasks.

Fog Computing was considered an enabler technology for new classes of applications (e.g., Immersive Media, ITS) because they can take advantage of its features to tackle their constraints. However, to deploy such services, there are some challenges to be overcome. For instance, different protocols and APIs need to be established for services to access information from the network and sensors (Roman *et al.*, 2018). Also, mobility creates issues related to network management at the edge.

User mobility, handovers, and intermittent communication channels in general may disrupt service provisioning because of the difficulty to keep reliable, low-latency, and high-throughput links to send messages to Fog nodes placed at the near-Edge of the network. One possible way of handling the mobility of the users is centralizing the handover control at the cloud (Bittencourt *et al.*, 2017). This solution faces issues when a connection to the Cloud data-centers fails. Thus, studies on handling mobility locally have emerged (Nguyen Gia *et al.*, 2018). Yet, to execute more complex algorithms and enhance the quality of service in network handovers, programmable networks have been considered (Salman *et al.*, 2018; Bi *et al.*, 2019). These networks create a large set of possibilities due to the flexibility in using different routing protocols (Nguyen *et al.*, 2017a) and employing ML approaches to perform data-driven decisions (Alawe *et al.*, 2018). Section 4 discusses the main possibilities found in the literature.

3.1.2 Multi-access Edge Computing

The MEC paradigm was proposed by ETSI in 2014 to use Edge devices to enhance mobile devices capabilities through mobile cellular networks (e.g., 3G, 4G/LTE, 5G) (Salman *et al.*, 2018; Mao *et al.*, 2017b). The concept of devices and connections was extended in MEC, which caused the terms Fog Computing and MEC computing to start to be used in an interchangeable fashion by the academic community (Mao *et al.*, 2017b). Initially, MEC would consider only datacenters at the edge of the network deployed by telecommunication companies to offload processing tasks, and these devices should be accessed via the cellular network. Currently, even devices from end-users can be added to the MEC resource pool in some collaborative approaches. Different connectivity technologies can also be used to communicate to these resources.

To widely deploy MEC for massive usage, there are still some open issues to be addressed. Researchers have studied how to better place servers spatially to ease the coverage of wide areas (Yang *et al.*, 2019). For instance, UAV-mounted micro-servers can be used to provide in-locus additional resource (Jeong *et al.*, 2018). Furthermore, to handle the vast amount of simultaneous users, techniques of MEC-enabled in-network caching have been proposed (Fang *et al.*, 2018; Nguyen *et al.*, 2017a). The use of caching aims to take advantage of the high popularity of content and services by storing them in servers closer to users to reduce the necessity to load them from the Cloud. User mobility causes several handover events in access networks, which is a complex task because of the many system configurations and policies to associate users and services (Mao *et al.*, 2017a). For instance, according to some mobility management protocols (Bi *et al.*, 2019; Perkins and Microsystems, 1997), the IP addresses of the users may have to change; also, a different host may be selected – according to a given policy – to execute services consumed by these users.

3.1.3 Mobile Cloud Computing

Another architecture aimed to enable the execution of computing-intensive tasks in resource-constrained mobile devices is Mobile Cloud Computing (MCC) (Akherfi *et al.*, 2018). This architecture advocates for offloading complete applications from Smart Mobile Devices to the Cloud infrastructure, thus integrating mobile computing and Cloud Computing. This architecture differs from ones described in Sections 3.1.1 and 3.1.2 because it does not use other Edge devices for offloading computing tasks.

3.1.4 Vehicular Edge Computing

An Edge resource that has gained recent attention for task offloading is the vehicle (Shah *et al.*, 2019; Ning *et al.*, 2019). This attention is due to the large number of vehicles (Albrahim *et al.*, 2019) and also the reasonable amount of computing resources on board both recently-released and upcoming vehicles (Wang *et al.*, 2019). These resources will be deployed in the form of On-Board Units (OBUs), which allow these vehicles to access network facilities. Due to the size of vehicles and its powerful batteries, when compared to other Edge devices, these OBUs can be deployed with a significant amount of processing power. Building infrastructure based on Road Side Units (RSUs) is costly and may require additional effort with maintenance (Shah *et al.*, 2019). These costs could be reduced by using idle resources of vehicles. This idea supports the emergence of the Vehicular Edge Computing (VEC) paradigm – this paradigm is also referred to as Vehicular Fog Computing (VFC).

The VEC paradigm adds to the EC architecture the possibility of collaboratively use vehicular idle resources. These vehicles have the ability to capture data from nearby or remote environments and use it to run diverse applications. The data of the environment can be captured using Intra-vehicle communication with its own sensors (V2X), Inter-vehicle communication (V2V) to collect data from neighboring vehicles, or even Extra-vehicle communication (V2I and/or V2X), where data can be collected from RSUs, remote EC-enabled sensors or the Cloud.

One way of building a VEC platform is by using VANETs. This approach is fully distributed, and nodes take decisions of sharing resources based on a limited view of the network status obtained from their neighborhoods. Studies considering using only VANETs focus on deploying vehicular-centric protocols to address communication among vehicles (Cao and Lee, 2018) or integrate this network to cellular networks (Khan *et al.*, 2019). To achieve a better resource sharing solution, SDN is proposed to centralize the VANET control (Abbas *et al.*, 2019). Studies on this field use RSUs-based VANETs (Kalogeiton and Braun, 2018), in which the SDN controller is placed in the RSUs. These controllers can also be hosted in some alternative infrastructure, such as UAVs (Seliem *et al.*, 2018; Sedjelmaci *et al.*, 2019).

3.1.5 Vehicular Cloud Computing

MCC faced some issues when applied to a vehicular scenario because of the strategy of sending all data to be processed at the cloud. Vehicular applications depend on a great variety of sensors that collect large amounts of data and have to be processed in real, or near-real, time. Offloading the entire application to the Cloud can create issues for the service provisioning because of the high volume of data that may be sent to the Cloud. Observing the reasonable amount of computing resources envisioned on-board of future vehicles, the Vehicular Cloud Computing architecture (VCC) (Ashok *et al.*, 2018) was proposed as an extension for MCC. In this architecture, only some parts of the application are offloaded to the Cloud, while others run in the vehicle itself. VCC is not fully able to cope with new application requirements in terms of delay and limited bandwidth as it relies on opportunistic communication with other vehicles, but part of this resources can be used in EC. To enable services to run in EC-enabled settings, approaches for network management, such as the ones presented in Section 3.2, have been proposed. Many of these techniques are expected to be used in conjunction to form a holistic platform and achieve the requirements established by the new generation of mobile networks (Zhang *et al*., 2016).

3.2 Networking Architectures

Different communication architectures emerged in the literature to enable computing technologies to cooperate and form an environment for service provisioning (Perkins and Microsystems, 1997; Das, 2018; Venkataramani *et al*., 2014; Zhang *et al*., 2018b). Many of these architectures envision a significant change on the basis of the Internet, such as virtualization of the network and a shift from a topology-based paradigm to new paradigms (e.g., information-centric). This section presents the main networking architectures proposed for the EC-enabled settings. To evolve the host-centric paradigm, researchers have proposed different paradigms, such as: (i) Geographical-based networking (Leontiadis and Mascolo, 2007; Di Maio *et al*., 2017; Hagihara *et al*., 2017; Tang *et al*., 2018a; Das, 2018) which routes content according to geographical locations; (ii) Mobility-centric networking (Venkataramani *et al*., 2014; Li *et al*., 2014), where mobile devices are addressed through unique identifiers generated to them; and (iii) Information-centric networking (Leontiadis and Mascolo, 2007; Fang *et al*., 2018; Zhang *et al*., 2018b; Nguyen *et al*., 2017a; Duarte *et al*., 2019) that uses interest names of contents or services for routing. This section presents some of the characteristics of these paradigms, which are summarized in Table 2. More flexible network control via SDN is a trend to implement these paradigms on EC-enabled scenarios and is discussed in Section 3.2.1. Opportunistic and geographical-based networking is discussed in Sectino 3.2.2. Finally, different architectures for Future Internet are discussed in Section 3.2.3.

3.2.1 Software-Defined Networking

SDN has emerged as a networking paradigm to make networks more flexible and ease the adoption of new protocols and algorithms for network routing and the implementation of other network functions (Rawat *et al*., 2017). This paradigm shifts the routing complexity from the network routers to a centralized instance called controller, where a complete network overview gives several benefits (Saraswat *et al*., 2019), such as: (i) granular control of policies that can be oriented to sessions, users, devices, or services; (ii) easy and on-demand adaption to changes; and (iii) cost savings due to better resource management.

The idea of a global network view present in SDN was adapted from the telephone network, where it was shown to be a secure and cost-efficient strategy (Rawat *et al*., 2017). Major SDN deployments were only observed after the evolution of the programmable router switches and the emergence of the OpenFlow protocol (McKeown *et al*., 2008). This protocol is based on a three-layered separation of network entities: (i) the application layer with services and end-users; (ii) the infrastructure layer with hardware to support storage, connectivity, and computation; and (iii) a control layer responsible for the virtualization of the infrastructure and enable its control by the applications. OpenFlow is a south-bound API to control programmable switches; the SDN controller also provides a north-bound API for management to be used by the application layer. For instance, Frentic (Gutz *et al*., 2012) is a north-bound API that abstracts the network management using the concept of slices. Pyretic (Monsanto *et al*., 2013) is another abstraction that uses a modular view of the network for management.

To handle all the expected traffic load exchanged at the edge of the network, different network traffic management tools were proposed to explore the centralized information maintained by SDN controllers (Fawcett *et al*., 2017; Priyadarsini and Bera, 2019). The programmability of the network enhances its flexibility because it allows not only adaptability and interoperability but it also opens space for innovation. This programmability also aids the process of intelligent management through the development of, for instance, efficient mobility management solutions (Bi *et al*., 2019), or the use of ML models to enhance networking (Alawe *et al*., 2018).

The flexibility and interoperability provided by SDN can be observed in 5G networks research, mainly to integrate new technologies and services in the networks (Le *et al*., 2017; Zaidi *et al*., 2018). SDN is also a key supporting technology to handle the scalability and complex management of IoT scenarios (Salman *et al*., 2018), where a vast number of heterogeneous devices need to be connected. The significant compatibility with other state-of-the-art technologies and applications makes SDN an important enabler for the next generation of networks.

3.2.2 Geo-Centric Networking

An intuitive way of managing networks in the presence of mobility is through geographical coordinates. This class of protocols uses geographical coordinates of the destination to support routing decisions. For instance, Geographical Routing using Partial Information (GRPI) (Jain *et al*., 2001) is an approach where each node in the network uses partial network information about its neighborhood to route packets to the closest neighbor from the destination. The route is not fixed, and, thus, if a packet reaches a node that "knows" a better route (i.e., based on distances of neighbors to the destination) the packet is sent through that route. This approach composes a distributed routing algorithm since no single node is required to have an overview of the entire network, but only knows information about its neighbors. GRPI is meant to operate in Wireless Ad-Hoc Networks. However, the problem of node sparsity is not studied. Network sparsity is a critical issue for routing protocols in wireless mobile environments. Consequently, this issue drives most studies on geographical routing for mobile entities to focus on opportunistic routing, which explores the links created opportunistically by the mobility of nodes. There are many applications for wireless mobile networks that aim at disseminating content to specific geographical areas, such as accident notifications or traffic flow conditions. Therefore, Geo-Centric Networking (GCN) protocols aim to allow nodes to address geographical areas and distribute network messages within them.

Table 2. Overview of main networking paradigms and protocols.

Paradigm	Addressing	Protocols
Host-centric	Numerical	IPv4
		IPv6
		MIPv6 (Perkins and Microsystems, 1997)
		PMIPv6 (Kellokoski *et al*., 2013)
		HMIPv6 (Castelluccia, 2000)
Geo-centric	Geographic Coordinates	GRPI (Jain *et al*., 2001)
		BLR (Heissenbüttel *et al*., 2004)
		GeOpps (Leontiadis and Mascolo, 2007)
		GeoNetworking (Tomatis *et al*., 2015)
		FloatingContent (Di Maio *et al*., 2017)
		PFCS (Hagihara *et al*., 2017)
		GSOR (Tang *et al*., 2018a)
		DGOR (Das, 2018)
		CBF (ETSI, 2021)
Future Internet	Content/Service Identifiers	HIP (Moskowitz *et al*., 2008, 2015)
		MobilityFirst (Venkataramani *et al*., 2014; Li *et al*., 2014)
		CCN (Jacobson *et al*., 2009)
		NDN (Zhang *et al*., 2014; Duarte *et al*., 2019)
		PUSUIT (Fotiou *et al*., 2012)
		SCN (Braun *et al*., 2013; Simoens *et al*., 2017)
		OON (Liu *et al*., 2017)
		NFN (Tschudin and Sifalakis, 2014)

Using opportunistic links to disseminate data, Geographical Opportunistic routing for vehicular networks (GeOpps) (Leontiadis and Mascolo, 2007) is a protocol that focuses on delay-tolerant networks. These networks are used by applications that can run without a continuous network connection. In particular, this protocol enables content distribution in target areas without fixed infrastructure. This approach relies on the *store-and-forward* strategy, where mobile nodes receive the content and carry it to later on forward it to the next node. Distributed Geographical Opportunistic Routing (DGOR) (Das, 2018) is a similar protocol that uses a different set of metrics to evaluate the link cost to select the forwarding path to send network packets.

It is worth noticing that these networks face scalability problems because most protocols rely on regularly sending messages to inform neighboring nodes of their existence and position in a process called beaconing. Therefore, scenarios with many nodes broadcasting beacon messages can create scalability issues, such as transmission interference. Beaconless routing algorithm (BLR) (Heissenbüttel *et al*., 2004) is a protocol designed to tackle such scalability issues in MANETs. In BLR, beacon messages are not used, since no information about the existence or position of neighbors is required. Instead, the protocol broadcasts data packets and uses a dynamic forwarding delay to ensure that only one node will forward the message. In this mechanism, every node computes a forwarding delay to send the data packet. The one that computes the shortest delay will send it first as a broadcast. Thus, the other nodes will receive it and cancel the forwarding of their own copies.

GeoNetworking (Tomatis *et al*., 2015) is a store-and-forward protocol standardized by ETSI aimed at vehicular communication that uses the location of OBUs and RSUs to disseminate data. This protocol has two main features: geographical addressing and geographical forwarding. This addressing allows unicast, where geographical positions are used together with node identifiers to aid the routing. It also allows broadcast and multicast, which may be performed by geographical or topological routing. According to the type of addressing being used, different methods can be used. For instance, in topological broadcasts, the forwarding process uses a simple flooding approach. Unicast, on the other hand, uses an approach called *line forwarding*, which applies different heuristics to create a forwarding path from source to destination.

Floating Content (Hagihara *et al*., 2017) uses an epidemic model for broadcasting content in an anchor zone (AZ) by keeping the content stored in the vehicles interested in it. In particular, vehicles inside the AZ can access the content via opportunistic links with other vehicles inside the zone that is carrying the content. Floating Content can also take advantage of a centralized SDN-based approach (Di Maio *et al*., 2017). SDN controllers, accessed via RSUs, can collect information from the moving vehicles and analyze this data to enhance Floating Content management.

One important concept of Geo-centric Networking standardized by ETSI is Contention-Based Forwarding (CBF) (ETSI, 2021; Füßler *et al*., 2003). This is a distributed and scalable forwarding strategy based on interest regions aimed for mobile ad hoc networks. This strategy does not rely on acquiring information of the neighborhood of a node via beacons, instead the message carries information of a contention window where it should be spread. Once receiving this packet, every node uses this information and its own position to decide whether it should become a forwarding hop or not. To avoid flooding, the re-transmission of the message is also conditioned to probabilities and/or timeouts that are evaluated by the node. For instance, a node will only re-transmit a message if it does not receive a transmission of the same message from another node in a given timeout. Multiple extensions of CBF exist in the literature that target, for instance, supporting the strategy using infrastructure (Bellache *et al*., 2017) or performing network congestion control (Meijerink and Heijenk, 2019).

It is worth noticing that while Geo-Centric Networking approaches address mobility issues in specific geographical zones, more complex and delay-critical applications require a higher level of quality of service to operate that sometimes cannot be obtained by opportunistic routing. Nevertheless, other networking paradigms can address these situations, such as Mobility-Centric and Information-Centric Networking, discussed hereafter.

3.2.3 Future Internet

Upon observing the mobility patterns of Internet users, the US National Science Foundation's Future Internet Architecture (NSF-FIA) project designed an architecture called MobilityFirst (Venkataramani *et al*., 2014) in 2010. This architecture aims to produce a network protocol for scalable service provisioning in mobility scenarios. This protocol is based on Globally Unique Identifiers (GUID) for in-network elements. These identifiers are used to separate names from addresses and locations, thus easing mobility management. This protocol relies on a distributed Global Name Service (GNS) that maps GUIDs to addresses. The strategy is similar to the one used nowadays on the Internet, where domains are translated to addresses via the Domain Names System (DNS).

The idea of using a global view of the network is shared with SDN. These technologies could in fact be used in collaboration by adding the GNS module to run within the SDN Controller. For content distribution, MobilityFirst relies on a in-network caching scheme (Zhang *et al*., 2012). In this strategy, storage-aware routers are used to cache content along the path it makes from its source to the consumer; this scheme facilitates dealing with intermittent connections due to mobility. Services are held in a similar fashion by mapping the service URI to a GUID and then using the GNS to resolve the GUID to an address. While handling services, MobilityFirst suggests the usage of in-network caching to store dynamic data (Li *et al*., 2012), which is unusual since dynamic data is supposed to change. Nevertheless, different caching strategies should be applied to handle services but not the same ones used for contents.

Another important protocol designed by the IETF is the Host Identity Protocol (HIP) (Moskowitz *et al*., 2008, 2015). This protocol allows hosts to share IP-level states to facilitate service provisioning continuity despite changes in IP addresses. By establishing Host Identities, HIP decouples transport-layer logic from network-layer logic. This separation creates many possibilities for network-layer mobility management. IETF has specified a basic network-level host mobility protocol (Henderson *et al*., 2017). This protocol defines how to create message flows and also other procedures to achieve host mobility. It is important to notice that CCN/NDN, PURSUIT, HIP, and other protocols can operate together, creating possibilities to apply them in the best suitable use cases.

To serve users with named content that can be stored anywhere in the network, ICN is a paradigm that has gained attention in both academy (Koponen *et al*., 2007; Yao *et al*., 2016; Tortonesi *et al*., 2019; Din *et al*., 2018; Zhang *et al*., 2018b; Nguyen *et al*., 2017a) and industry (NDN Project, 2010; GreenICN Project, 2013; ICN 2020 Project, 2016). Due to its lack of information about the content's location in its naming/addressing, ICN has the potential to solve many issues associated with the Host-centric paradigm, such as mobility (Fang *et al*., 2018). By replacing IP addresses by a naming-based scheme, ICN supports seamless mobility. A scalable content distribution in this scenario is achieved via the deployment of storage-aware routers throughout the network, augmenting the possibilities for caching and offloading the core network. Several architectures have been proposed, such as Content-Centric Networking (CCN) (Jacobson *et al*., 2009), Named Data Networking (NDN, an evolution of CCN) (Zhang *et al*., 2014), and Publish-Subscribe Internet Routing Paradigm (PURSUIT, earlier called PSIRP) (Fotiou *et al*., 2012). Although these three projects have similar objectives in terms of routing data based on its name, CCN and NDN advocate for hierarchical-based names to facilitate locating and sharing data. However, PURSUIT supports flat naming to allow a greater variety of naming approaches, in which names are organized in hierarchical scopes. This organization allows the constitution of information networks, similarly to IP topological sub-networks.

ICN is proposed to be a clean-slate paradigm, which means it demands infrastructure replacement. However, SDN is a promising future networking technology that can smooth this process of deployment of ICN (Zhang *et al*., 2018b). The protocol could be implemented over the virtual network controlled by the SDN controller. Furthermore, the benefits of SDN to the current network paradigm also apply to ICN. ICN can take advantage of the global view of the network and of actively controlling communication flows.

MobilityFirst and ICN are expected to run in EC-enabled Future Internet and deal with mobility-related issues. When comparing these two approaches they show similar performance to support scalability and mobility requirements for IoT applications. MobilityFirst outperforms ICN in terms of control overhead (Li *et al*., 2014). However, ICN strategies focused on VANETs have gained more attention. For instance, RSU-assisted NDN (RA-NDN) (Tiennoy and Saivichit, 2018) is a protocol that relies on RSUs to improve network connectivity in a VANET scenario. The protocol outperforms general ad-hoc communication in terms of data received, throughput, and reduction in total dissemination time and traffic load. Mobility in Vehicular NDN (MobiVNDN) (Duarte *et al*., 2019) is a protocol to mitigate issues related to vehicular communication, such as broadcast storms, message redundancy, network partitions, reverse path partitioning, and content source mobility. This protocol has good performance when sharing wireless medium with multiple applications. Cooperative Caching with Mobility Prediction (COMP) (Huang *et al*., 2019) is a caching strategy that focus on reducing the impact of mobility in VNDN. COMP reduces access delay and increases cache hit ratio by adding cooperative caching to VNDN, which usually is non-cooperative. The cooperative caching uses RSU resources to run caching decision algorithms that allow the vehicles to cache globally popular data instead of making caching decisions only based on local data. This strategy clusters vehicles with similar mobility patterns to store content on their OBUs. Later, these vehicles can share the content among themselves since their links are more stable.

The expected new classes of services, such as IoT, Immersive Media, and Autonomous Driving have drawn attention to mobile service provisioning in EC-enabled settings, thus reflecting the emergence of service-oriented ICN. One strategy to allow services to take advantage of the in-networking caching of the ICN paradigm is allowing cached content to be transformed and serve the requests (Braun *et al*., 2013) (e.g., transcoding a cached video). This approach is named Service-Centric Networking (SCN). Some other strategies use naming schemes of ICN to facilitate service consumption. For instance, Layered SCN (L-SCN) divides the network into inter-domain and intra-domain. It allows nodes within a domain to possess more information about available services in that domain and, thus, reduce overhead to share information about these services (Gasparyan *et al*., 2017a). Some other naming schemes have emerged, such as: Named-Function Networking (NFN) (Tschudin and Sifalakis, 2014), which describes chaining of named λ-expressions to compose in-network services; and Object-Oriented Networking (OON) (Liu *et al*., 2017) that proposes a programmable network using the same abstraction of object-oriented programming languages, where a set of specific functions can be accessed via named operable objects.

4 Mobility Management for Seamless Service Provisioning

Robust mobility management solutions have to be applied to achieve the expected levels of QoS and QoE at the edge. These solutions are needed to prevent communication and service disruptions for some network mobility events occur, such as users changing access points or services being reallocated to different hosts. Poor mobility management may cause service disruption when these mobility-related events occur. Therefore, maintaining service continuity in this environment is a key aspect for achieving the full potential of Edge-based service provisioning. This section explores technologies to support service continuity in the presence of user mobility.

When users are on the move they change from one access point to another. Thus, network configurations have to be updated to keep their connectivity. Such process is known as handover. There are different approaches to perform a handover, which will be discussed in Section 4.1. Also, the handover may result in other events to maintain the expected levels of QoS and QoE, such as service migration.

Figure 3 presents a classification of the main approaches used for the mobility management in EC environments. Network handover, and stateless and stateful service mobility definitions are present in the ETSI specification of end-to-end mobility aspects (ETSI, 2017, 2018). The present section divides mobility management into two parts as depicted in Figure 3. The first is network handover, which is a network operation to guarantee service and communication continuity when users change access points. Besides user mobility, some events in the network might trigger service mobility, which reallocates services in Edge nodes to (i) keep them near to consumers, to reduce latency and enhance bandwidth usage, or (ii) to better use of the resources (e.g., energy saving, load balancing). EC-enabled settings must support migration of two service types according to the presence of user-related state data (i.e., session): stateless and stateful services. Copies of stateless services can be deployed in different hosts, and the user can easily switch access points due to the absence of session data. Conversely, stateful services have session data to be migrated to keep service continuity without disruption, thus the mobility of stateful services is usually referred as service migration or service state transfer. Handover strategies are discussed in Section 4.1, while service mobility is solutions are shown in Section 4.2.

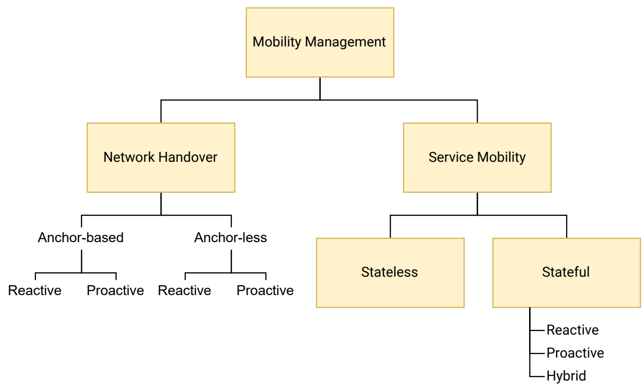

Figure 3. Overview of mobility management events and methodologies.

4.1 Network Handover

When users change access points, handover procedures to deal with the network transition process are used. Different approaches can be applied, in terms of using or not an anchor network or triggering or not the handover proactively, as depicted in the left branch of the diagram in Figure 3. The idea behind reactive and proactive handover approaches is straight forward. In reactive handover strategies, the process of migrating context occurs after the user connects to the new network. In proactive strategies, the migration process is anticipated by mobility prediction and can start before the user disconnects from the initial network. If the handover is executed without efficient mobility support, users will have to go over a set of repeated processes, such as service discovery and authentication, resulting in disruptions and reducing the QoE (Bi *et al.*, 2019). This section describes different approaches in the literature to deal with network handovers.

One approach is to perform both the network handover control logic and data forwarding procedure through mobility anchor networks. Despite the existence of multiple anchor networks, this approach is called Centralized Mobility Management (CMM) because it centralizes logic and forwarding. For instance, in Mobile IP (Perkins and Microsystems, 1997), a Internet Engineering Task Force (IETF) standard, creates a transparent interface for the TCP layer in which the IP address of a mobile user is kept constant after mobility events. To provide this feature, Mobile IP approach assigns internally two addresses to the mobile user: the Home Address (HoA), which is seen by the TCP layer and kept constant, and Care of Address (CoA), which is internally handled by the network and updated when the user moves. Each of these addresses have a respective entity associated with it. The first entity is the Home Agent (HA), which tracks the mobile users that belong to its network and forward packages to these users by using their CoA. The second entity is the Foreign Agent (FA), which advertises CoA for mobile users that visit its network so these users can be achieved. All the traffic sent to a mobile user is initially forwarded to its HoA, at its home network, and just after fowarded to the foreing network, using the CoA.

Anchor-based approaches lead to some drawbacks. For instance, since all the traffic is sent to the anchor network and only then to the access network of the user, the delay in communication increases. Anchor-less handover strategies can update communication paths within the network, for instance using SDN Bi *et al.* (2019), to carry packets directly to the current network of the mobile users, thus reducing the numnber of packets traveling in sunboptimal network paths.

Distributed Mobility Management (DMM) solutions have a similar objective as the SDN paradigm of separating the data forwarding from the control logic. According to the definitions of IETF (Liu *et al.*, 2015; Lee *et al.*, 2013), DMM solutions should not allow packets to be forwarded through anchor networks, thus resulting in a sub-optimal route. The IETF combines existing protocols, such as MIP (Perkins and Microsystems, 1997), PMIPv6 (Kellokoski *et al.*, 2013), and HMIPv6 (Castelluccia, 2000). The goal is to re-use the mobility management functions already deployed in these protocols, such as: (i) Anchoring: control of user original IP address; (ii) Localization: track of current access network where the user is connected; and (iii) Forwarding: receive and forward packets towards the user.

An SDN environment provides a series of advantages for mobility management. Since data forwarding and control logic are separated in the network, fewer configurations must be updated to perform the handover. The required update in the configurations can be achieved by updating SDN flow entries. For instance, user devices can request to the controller an address to use in the next access point (Bi *et al.*, 2019). Since the controller has a global view of the network, the switches in the network can be updated to forward data to users' new addresses without sending it through an anchor network. However, this protocol does not fully accomplish IETF requirements, since users have to change their IP addresses according to the networks they are currently connected, thus breaking the IP continuity.

The centralized view of the network of the SDN controller allows a better selection of the route to serve the users when considering their mobility. Furthermore, SDN already has the control logic and data forwarding separate from one another, one of the objectives to achieve optimal DMM according to IETF. Once users move to a new network, the attachment process is executed, and the SDN updates the flow rules in the forwarding switches. This is a reactive handover process; still, a proactive handover is also feasible. The mobile device gathers identifiers to connect to the new access point before leaving the previous one. Thus, all connections will be already set when the device arrives at the new network. Proactive handover increases the possibilities of enhancing QoS and QoE in the handover process since mobility prediction techniques can be used to estimate the users' positions in the future, allowing all the setup to run before the network shift happens. Mobility prediction to aid network management is one of the challenges to be addressed to achieve the full potential of EC. This challenge is discussed in more detail in Section 6.

In cellular networks, such as LTE and 5G, different types of handover, in different domains, can be triggered (Tayyab *et al.*, 2019). In the *frequency* domain, when the base stations involved in the handover operate with different frequencies and time multiplexing, user devices have to switch between different frequencies to perform measurements in both frequencies. Differently sized cells are deployed in the network to load balance users connected to a specific base station (e.g., macro, micro, or picocells). When a user device observes that a smaller cell has better QoS measurements, it offloads the bigger cell by migrating its connection to the smaller one. Handovers can also occur in the *radio access technology* (RAT) domain. In this case, a user device changes between different radio access technologies, such as 3G-LTE. In 5G networks, there are more access technology options, such as massive multiple-input multiple-output (MIMO) networks (i.e., networks implemented with arrays of multiple antennas), millimeter wave (mmWave) networks (i.e., network where carrier wavelength is between 1 and 10 millimeters), and also energy harvesting networks (i.e., network where user devices can obtain power, i.e., recharge) (Liu *et al.*, 2016a). The base station initiates the RAT handover process, which instructs the user device to change access points. Again, not only connectivity variables are considered in the process. For instance, load balancing and other factors may drive the decision for a base station to perform the handover. Finally, handovers can happen in the *operator's* domain. A common example of a handover between different operators is roaming, when users leave the area covered by their original operator and have to switch to another operator that covers that area.

Different metrics are explored in the literature to evaluate the handover in LTE and 5G scenarios. For instance, the number of handover failures, handover success rates, and handover frequency (Gelabert *et al*., 2013). Another metric is the number of ping-pong events, i.e., when a migration process from access points A to B is followed by another migration from B to A in a short period of time (Thakkar *et al*., 2017). Some metrics measure directly the impact on the final QoS delivered from services to users (Han *et al*., 2015), such as: (i) handover delay, the time between the user device receives the last packet at the original station and the reception of the first packet on the next base station; and (ii) handover interruption time, when user applications cannot send any packet. Many other metrics exist in other domains related to (Liu *et al*., 2016a): (i) spectrum efficiency, (ii) energy efficiency, and (iii) fairness. Spectrum and energy efficiency measure how well the resources in these domains are used by the connections, while fairness concerns the division of the communication resources fairly among users. The current state-of-the-art in handover management for cellular networks focuses on User Association in 5G networks. User Association creates policies to maintain acceptable levels of QoS and QoE in these networks. Different mature algorithms have been proposed in this scope (Zhang *et al*., 2019; Yazdinejad *et al*., 2019), and also ML models started to be considered using mobility-related and other data sources (Liu *et al*., 2016a).

Different approaches in the literature propose ways to improve the handover procedure and reduce its execution time (HET). The most relevant proposals include (i) schemes where configuration setups required to communicate to a target antenna are made before the disconnection with the currently used antenna, namely Make-before-break (MBB); (ii) schemes that do not use a Random Access to Channel (RACH) to perform timing alignment between the device and the antenna, also known as RACH-less; and (iii) schemes that are coordinated by SDN controllers, SDN-enabled handovers.

The MBB scheme consists of a straightforward idea where execution time is saved by preparing configuration setups before disconnecting from the current base station. This way, they are ready when needed to establish communication with the next base station. This strategy is included in 3GPP standards to be used for the next generations of cellular communication infrastructure (3GPP, 2014, 2016). In this case, the X2 interface for wired communication between the base stations is used to exchange information and prepare the configuration setups. RACH-less handovers consist of avoiding executing the RACH procedure during the handover, which is on average 10 12 ms when considering a total handover execution time of 40 50 ms (3GPP, 2016). RACH-less handovers were initially proposed for synchronized networks (Barbera *et al*., 2015). However, the exchange of internal clock references of current and target base stations on a handover can provide enough information for the user device to perform the timing alignment in a non-synchronized network without executing the RACH procedure (Choi and Shin, 2019). Still, users need to reach both base stations during the process, which makes it possible to receive the last message from the current base station and send the next message to the target base station. When using SDN controllers, the handover execution aims to take advantage of the global information of the network available for the controller. This information is used, for instance, to trigger handovers proactively and also evaluate the best antenna candidate considering the data plan, not only the signal quality (Bi *et al*., 2019).

4.2 Service Mobility

Service mobility (ETSI, 2017, 2018) can happen in the network triggered by different events, such as resource management, energy saving, or accompanying user's mobility. For instance, mobile devices may move away from the infrastructure hosting a service while still consuming it. As depicted in Figure 4, there are two ways to keep service continuity in this scenario. First, requests to the service can be forwarded to the original server. Still, problems may arise for maintaining low levels of latency to services that require high reliability, low jitter or also have high data transfer volumes. In this situation, one option is reallocating the service instance, thus requiring a migration of the service so it can run on an infrastructure closer to the device. This migration may add overhead in terms of service downtime, network traffic, and computing. Nevertheless, overall QoS should be increased to the final user allowing to meet application requirements. Studies on service migration focus on virtualization technologies, such as hypervisor-based (Zhang *et al*., 2018a; Baccarelli *et al*., 2018) and container-based (Tang *et al*., 2018b; Ma *et al*., 2018), and how to perform the migration procedure.

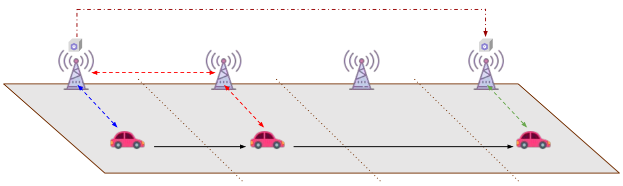

Figure 4. Service horizontal migration due to user mobility.

Service migration is divided into two broad categories according to the existence or absence of user session data. Services without sessions are called stateless services. In this case, the main data migrated is related to its running code. This code can be downloaded from the original host node, from other neighbor nodes, or replicated beforehand to enhance migration performance. On the other hand, stateful services hold sessions of users consuming them, thus all session data has to be migrated. This session data is stored in two forms: (i) main memory – i.e., stores data for immediate usage – and (ii) storage – also secondary memory, which stores persistent data. Main memory migration is the critical operation for migrating stateful services since this data is usually at constant usage. Storage migration is a bottleneck since it represents a large amount of traffic to traverse the network. In terms of storage migration, some distributed file systems have been proposed in the literature (Monga *et al*., 2019; Gupta and Ramachandran, 2018; Pamboris *et al*., 2019), which handle the responsibility of moving large chunks of data (this approach is the last topic discussed in Section 4). The migration process can be proactive or reactive for either main memory or storage. Thus, there is a great variety of combinations of proactive/reactive main memory migration with proactive/reactive storage migration and some hybrid approaches (Zhang *et al.*, 2018a).

As mentioned before, some studies focus on migrating VMs. For instance, to enhance mobile user experience Follow-Me Cloud (FMC) (Taleb *et al.*, 2019) explores the migration of services among different datacenters. A Markov Chain decision algorithm is used to decide whether or not to migrate VMs. Besides user experience, other factors can influence the decision to migrate VMs, such as, the trade-off between energy consumption in the migration process and delay. These variables can be measured using models (Baccarelli *et al.*, 2018) to allow the decision to migrate a machine and also select the most energy-saving migration strategy. To reduce the overhead of VM migration, identification of segments of in-memory data with imminent access can be used (Li and Gao, 2017). Thus, only these data segments can be migrated, reducing the traversed load among the hosts. This identification is made in a pre-deployment step where the application code is parsed, and metadata about context information used more often in the main memory stack is extracted and applied to support the decision of what to migrate.

Container-based migration is a recent research topic in the literature (Wang *et al.*, 2018a), but has gained attention in the industry (with Docker Swarm and Kubernetes) and academy (Tang *et al.*, 2018c; Kaur *et al.*, 2017; Ma *et al.*, 2018). Similar to FMC, a Markov Decision algorithm can be applied to container-based migration. For instance, this approach can be used to evaluate a trade-off between delay constraints and power consumption in the migration process to decide whether to migrate a container or not (Tang *et al.*, 2018c). Even mobility parameters are considered in this approach. CoESMS (Kaur *et al.*, 2017) is an EC migration framework that models power consumption and user QoE accessing a service in terms of utility functions in a cooperative game (i.e., game theory).

Docker containers are composed of a series of layers, each one of them added to the container as a result of one operation (e.g., move, copy or download files). This layered composition can be used to improve the migration process. For instance, multiple Docker containers share layers with the same content. These layers are mapped to unique identifiers. However, even containing the same content, they may be mapped differently. An algorithm that applies the same identifier to label the layers that have the same content were used to allow layer reuse. The goal is to reduce the amount of data downloaded from the Cloud to deploy a given service (Ma *et al.*, 2018).

To support services running and moving at the edge, most of the data persistence for EC-enabled applications is delegated to the Cloud. Yet, reliable storage and data management at the edge are necessary to support some classes of applications that perform frequent update to this data. Some initial studies proposed to deploy EC-hosted file systems. ElfStore (Monga *et al.*, 2019) is a methodology to store data blocks at selected locations to achieve data reliability. Stored data uses a block-level differential replication scheme to achieve a minimum reliability level. This replication scheme splits large chunks of data into blocks. These blocks can then be stored and replicated. Common segments between multiple blocks are then identified and some of the copies removed to reduce storage resource usage. The desired level of reliability can be maintained by replicating the different segments of the blocks and combining them with the common segments to obtain the complete block. Bloom filters are used to explore the hierarchical structure of EC and enhance data block retrieval. This filter is a data structure used to determine whether an element belongs to a set or not. Another data storage service that supports reliability is Fog Store (Gupta and Ramachandran, 2018), which proposes a solution for the placement of replicas of data blocks for Fog Computing. The proposed mechanism takes into account network topology and device heterogeneity to decide about data storage placement. Fog FileSystem (Pamboris *et al.*, 2019) is a solution to aid the process of migrating services in EC. Snapshotting and synchronization are applied to reduce the migration time of disk states between different nodes.

Besides data storage as a service, application state management services are important to support latency-critical applications. For instance, Do and Kim (2018) proposed a latency-aware placement of state management functions for 5G scenarios. This placement solution stores the state data at the cloud, which might create barriers to access it under certain latency thresholds, even more if considering the high update rates of such data. A more general mobility-aware state support was proposed for SDN by Peuster *et al.* (2018). The authors propose to rewrite communication flows to enable the user to consume the state from a static host. Yet, consuming from a static node may create bottlenecks when provisioning services. A similar solution for consuming state data from a constant node is proposed for SCN (Gasparyan *et al.*, 2017b, 2019). In their study, the authors also fix the path to consume the data. This strategy limits the application of standard ICN multi-path capabilities, which negatively impacts the scalability of the solution. Filho and Porter (2020) propose a transparent system to replicate state data in multiple hosts. The authors cover two main scenarios: (i) centralized state, in which the service should be stopped while the state is migrated; and (ii) distributed state, where a coherence mechanism is proposed that replicates all data update operations in all hosts.

5 EC-Enhanced Mobile Applications

The discussed set of EC architectures would support a plethora of modern applications in different domains discussed in the preset section, such as the Internet of Things (IoT) (Section 5.1), Immersive Media (AR/VR) (Section 5.4.1), Intelligent Transportation Systems (ITS) (Section 5.2), Unmanned Aerial Vehicles (UAVs) Services (Section 5.3), Smart Cities (Section 5.5), and Edge AI (Section 5.6). Some of these applications already exist today, while others are still being studied. Broad deployment of EC technologies is required to handle the massive adoption of such applications. This section highlights some instances of applications and the strategies they use. We present different classes of applications for each of the domains mentioned above. Table 3 shows an overview of all applications explored in this section.

Table 3. Taxonomy of main upcoming mobile applications and EC-enabling technologies used to deploy them.

			Computation				Communication		Service Virtualization	
Domain	Class	Instance	Fog Computing	MEC	VEC/VFC	SDN/NFV	ICN	Geo-Centric Networking	Container-based	VM-based
IoT	Industry 4.0	Hofer *et al*. (2019)	✓	✓					✓	
		Kaur *et al*. (2018)	✓	✓						
		Li *et al*. (2018)	✓	✓						
	Cognitive IoT	Al-Turjman (2017)					✓			
		Zhao *et al*. (2019a)		✓	✓		✓			
		Chen *et al*. (2019b)	✓	✓		✓				
	Body Area Sensing	Li *et al*. (2017)				✓				
		Lal and Kumar (2017)	✓	✓		✓	✓			
		Quan *et al*. (2015)					✓			
	Healthcare	Liao *et al*. (2019)	✓	✓						
		Abdelmoneem *et al*. (2019a)	✓							
		Li *et al*. (2017)				✓				
ITS	Traffic Management Systems	Wang *et al*. (2018b)	✓	✓						
		Ahmad *et al*. (2019)	✓			✓				
		Ahmed *et al*. (2016)			✓		✓			
		Khaliq *et al*. (2019)	✓	✓	✓					
		Bhatia *et al*. (2019)		✓	✓	✓				
	CAVs	Su *et al*. (2018b)			✓	✓				
		Chekired *et al*. (2019)	✓	✓		✓				
		Peng *et al*. (2019)		✓		✓			✓	
		Zhao *et al*. (2019a)		✓	✓		✓			
	Internet of Vehicles	Yahiatene *et al*. (2019)	✓	✓	✓	✓				
		Ho *et al*. (2016)					✓			
		Liu *et al*. (2016b)					✓	✓		
		Venkatramana *et al*. (2017)				✓		✓		
UAV Services	Augmented Environment Information	Lei *et al*. (2019)					✓			
		Zhang *et al*. (2018c)		✓		✓	✓			
		Kalatzis *et al*. (2018)	✓	✓					✓	
	Navigation and Swarming	Rahman *et al*. (2018)				✓				
		Xiong *et al*. (2019)				✓				
		Zhao *et al*. (2019b)				✓				
Immersive Media	Augmented and Virtual Reality	Kim *et al*. (2018a)		✓		✓				
		Fraga-Lamas *et al*. (2018)	✓	✓						
	Teleoperation and Telepresence	Miao *et al*. (2018)	✓	✓		✓				
		Oteafy and Hassanein (2019)	✓	✓		✓	✓			
		Schneider *et al*. (2017)		✓						
	Gaming	Wireless One (2018)	✓							
		Hu *et al*. (2019b)	✓							
Smart Cities	Public Services	Katsaros *et al*. (2014)					✓			
		Vilalta *et al*. (2017)	✓	✓		✓			✓	✓
		Li *et al*. (2017a)			✓	✓			✓	
		Han *et al*. (2017)	✓			✓	✓			
	Location-Based Services	Tanaka *et al*. (2019)		✓			✓			
		Meneguette *et al*. (2019)	✓							✓
	Mobile Crowdsensing	Li *et al*. (2017b)		✓		✓				
		You *et al*. (2017)		✓			✓			
		Longo *et al*. (2018)	✓	✓					✓	
Edge AI	Infrastructure Management	Sami *et al*. (2020)	✓			✓			✓	
		Lan *et al*. (2020)	✓			✓			✓	
		Liu *et al*. (2021b)	✓	✓		✓				
		Shi *et al*. (2020)			✓					
		Ye *et al*. (2019)			✓					
	Support for Smart Services	Wan *et al*. (2022)	✓			✓			✓	
		Dalgkitsis *et al*. (2021)		✓	✓	✓			✓	✓
		Hu *et al*. (2019a)		✓						

For each domain/class in Table 3, some studies were explored in order to provide a view of the requirements in terms of delay and data rate, shown in Table 4. In each domain, applications may have different requirements, such as in immersive media, where requirements for data rate in AR/VR and Gaming are more strict than in teleoperation applications. In contrast, in other scenarios, requirements are constant for most of the applications, like UAVs. Smart Cities is a peculiar case where different applications, even in the same classes, have varied delay and data rate requirements. This happens because any application can use different types of data from nearby sensors and the Cloud together. The values present in Table 4 are the average values for those applications with references where more information about these requirements can be found. Still, there are exceptional cases where the requirements can be more strict. For instance, some control applications in smart factories have 10 μs of delay tolerance (Ma *et al*., 2019). In contrast, data collection for psychological applications in Body Area Networks (BANs) and healthcare require up to 10 Mb/s of data rate (Thotahewa *et al*., 2014).

5.1 Internet of Things

IoT is an infrastructure of physical and virtual connected devices with sensing and actuating capabilities. This infrastructure aims to create a collaborative environment between for many devices, augmenting the possibilities of monitoring and acting over a cyber-physical domain (Salman *et al*., 2018). Furthermore, these collaborative environments must simultaneously support millions of mobile users. To handle these users, issues related to management and scalability arise. There is a need to decentralize information and communication technologies and bring them closer to users through the enabling EC architectures discussed earlier to support a massive adoption of IoT (Salman *et al*., 2018; Wen *et al*., 2017; Dupont *et al*., 2017; Arshad *et al*., 2019). Once these enabling architectures become well studied and deployed, a variety of applications may emerge. This section discusses the main classes of IoT applications: Section

Table 4. Requirements for applications in different domain and classes.

	Domain/Class	Delay	Data rate
5.1	**IoT**		
5.1.1	Industry 4.0	1-10 ms (Schulz *et al*., 2017)	<1 kb/s (Schulz *et al*., 2017)
5.1.3, 5.1.4	Body Area Networks and e-Health	<250 ms (Akbar *et al*., 2017)	0,1-50 kb/s (Akbar *et al*., 2017)
5.2	**ITS**		
5.2.1	TMS	>1s (Boban *et al*., 2018)	<2 Mb/s (Boban *et al*., 2018)
5.2.2	CAVs	1-10 ms (Li *et al*., 2019)	>1 Gb/s (Li *et al*., 2019)
5.2.3	IoV	1-100 ms (Boban *et al*., 2018)	>25 Mb/s (Boban *et al*., 2018)
5.3	**UAV**	**5-50 ms** (Yuan *et al*., 2018)	**<1 Mb/s** (Zeng *et al*., 2016)
5.4	**Immersive Media**		
5.4.1,5.4.3	AR/VR and Gaming	5-30 ms (Han, 2019)	<10 Gb/s (Han, 2019)
5.4.2	Teleoperation	5-20 ms (Boban *et al*., 2018)	>25 Mb/s (Boban *et al*., 2018)
5.5	**Smart Cities**	**1 ms-1 s** (Ma *et al*., 2019)	**1 kb/s-1 Gb/s** (Kuzlu *et al*., 2014)

5.1.1 shows how the industry applies IoT applications in a manufacturing process; Section 5.1.2 describes how ML can be applied to the IoT domain creating new possibilities of applications; Section 5.1.3 describes applications that build networks around a human body, with devices such as wearables; finally, Section 5.1.4 discusses applications to facilitate health care of users. We also discuss how applications use EC technologies to enhance aspects of QoS, QoE, and business models in each section.

5.1.1 Industry 4.0

Cyber-Physical Systems (CPS) supported by IoT is also an interesting technology applied to control Smart Factories in Industry 4.0 scenarios. CPSs are systems that connect virtual and real environments and allow the interaction between them while being controlled – or monitored – by humans. There is still resistance to adopt virtualized solutions to keep manufacturing infrastructure. This resistance is mainly due to the short deadlines that machines must respond to real-world observed events, usually real or near-real time. Yet, requirements such as distributed sensing, data analytics, and enhanced network bandwidth usage are pushing forward this evolution. In these factories, services run on a great variety of devices and they usually migrate vertically (i.e., from the Edge towards the Cloud), seeking for more resources. To make these devices portable for multiple platforms and also to enable these services to run with a varied amount of resources, a lightweight virtualization technique is preferred, such as container-based service virtualization (Hofer *et al*., 2019). These vertical migrations increase network and service management intricacy to keep the levels of QoS. Therefore, SDN solutions to handle Cloud-Edge interplay have been proposed (Kaur *et al*., 2018; Li *et al*., 2018).

5.1.2 Cognitive IoT

Recent developments of IoT allied to advances in ML brought up the possibility of providing smart services. These services can collect and process information about the environment around them and make decisions to perform their tasks independently. This class of services is referred to as Cognitive IoT in the literature and can take advantage of EC technologies. Various network architectures exist to support these applications. Information-Centric Sensor Networks (ICSNs) are used to serve information based on user requirements, rather than providing an endpoint to read raw data (Al-Turjman, 2017). The authors evaluate the usage of different machine learning models to identify the best communication paths in the network to deliver the data to the consumer. A distributed map-reduce framework (Zhao *et al*., 2019a) was proposed to gather enormous amounts of vehicular and infrastructure sensor data and feed it to an architecture to apply ML and other analytical models and provide an intelligent route planning service. This framework uses ICN to allow vehicles to consume sensor data to evaluate traffic conditions and based on a . Data analysis tasks execute on MEC and VEC infrastructures to produce this information. Cognitive-LPWAN (Chen *et al*., 2019b) is a framework that uses SDN management of network traffic in Low-Power Wide Area Networks and some unlicensed spectrum technologies. This framework proposes the usage of a cognitive engine to create a smart orchestration of wireless communication technologies including 4G, 5G, LoRa, and SigFox. Using the cognitive engine and the combination of these wireless technologies the authors could achieve a sustainable trade-off between transmission delay and energy consumption compared to the technologies individually.

5.1.3 Body Area Sensing

Sensing devices have been spread in urban environments to facilitate the task of monitoring fast-changing city dynamics. Most commonly, these devices use Wireless Sensors Networks (WSNs) to connect and cover wide areas for different applications such as fire detection and building monitoring (Khan *et al*., 2016). Studies point out that these networks have been brought closer to users with wearable (and implantable (Santagati and Melodia, 2017; Jiang *et al*., 2018)) devices, forming Wireless Body Area Networks (WBANs) (Li *et al*., 2017). Wearables in WBANs are constrained devices that still have to run multiple tasks and report data to other wider-area networks. Since these devices are attached to users, they are subject to the same mobility patterns as them. These characteristics create the need for solutions to handle communications within these networks and bridge their interaction with other networks. Some of the enabling EC technologies are expected to make this level of interaction of WBANs and other networks achievable. For instance, some applications for reporting users' vital signs use SDN-based solutions to handle network issues (Li *et al.*, 2017). Some other studies have applied ICN-based solutions to improve efficiency in WBANs. By using ICN and exploring in-network caching thus reducing the amount of redundant sensors (Quan *et al.*, 2015), or minimize traffic load when connecting to external networks (Lal and Kumar, 2017).

5.1.4 Healthcare

Wearable and implantable devices gained wide attention due to their use in healthcare systems. In this scenario, mission-critical applications and mobility increase even more the complexity involved in deploying systems. EC-enabling technologies have a fundamental contribution in implementing such systems. Multiple studies discuss the application of these technologies individually. For instance, Fog Computing was used to deploy a task scheduling and offloading platform for healthcare with native support for patient mobility (Abdelmoneem *et al.*, 2019b; Liao *et al.*, 2019). SDN was applied to reduce in-network traffic load due to the vast amount of monitoring devices that need to access real-time information (Li *et al.*, 2017). The combination of Fog Computing and ICN was studied together to reduce latency and allocate safer storage for privacy matters (Guibert *et al.*, 2017).

5.2 Intelligent Transportation Systems

The increasing number of vehicles has forced the deployment of complex transportation infrastructure in large urban centers, yet in many cases, this infrastructure is inefficient, which results in a waste of valuable time for the citizens. Due to this inefficiency, multiple studies have explored strategies to create a more intelligent transportation infrastructure (Su *et al.*, 2018b; Ahmed *et al.*, 2015; Yahiatene *et al.*, 2019; Rodrigues *et al.*, 2018). These efforts explore classes of applications such as Traffic Management Systems, Connected and Autonomous Vehicles (CAVs), and Internet of Vehicles (IoV). These specific classes of applications have high mobility requirements. This section focus on these classes and how they use EC technologies to run their services.

5.2.1 Traffic Management Systems

To enhance road network usage by vehicles, Traffic Management Systems (TMS) emerged as part of ITS. Studies in this class of applications vary from road accident detection based on vehicle to vehicle and vehicle to infrastructure communication (Ahmed *et al.*, 2016), to issuing violation tickets in vehicular named data networks (Khaliq *et al.*, 2019). The distributed nature of EC brings advantages to collect and process localized data to produce real-time traffic information and reduce unnecessary movement of data, alleviating bandwidth of the core network and mitigating privacy issues. For instance, to obtain an overview of a road state, SDN-based crowdsensing (Wang *et al.*, 2018b) can be applied to collect and provide data to support context awareness. Also, SDN and VANETs are used to identify congestion-sensitive spots using GPS data collected from vehicles (Bhatia *et al.*, 2019). The SDN controller global view of the VANET is explored, centralizing the data and applying recurrent neural networks to forecast traffic behavior.

5.2.2 Connected and Autonomous Vehicles (CAVs)

The need to deploy fast-moving vehicles that could operate without human intervention on urban roads and solve traffic congestion issues raised multiple research projects. One challenge to achieve the full potential of CAVs is related to the computing-intensive services they have to run, such as trajectory and route planning, object detection and tracking, and even behavioral reasoning on proceeding in an intersection or overtaking. Besides that, vehicle-to-vehicle communication can enable a better performance of the entire ITS by allowing more cooperative decisions to be taken. The amount of data exchange to support CAVs is expected to surpass 1 Gb/s (Li *et al.*, 2019) for every vehicle with use cases in which the maximum tolerable end-to-end latency is in the range from 1 to 10 ms (Boban *et al.*, 2018), which challenges even emerging 5G networks. EC will be present inside the vehicles, in the form of On-board Units (OBUs), or attached to nearby infrastructure, in the form of Road-side Units (RSUs) that can be used to meet these requirements. SDN-based VANETs architectures have been studied to allow offloading of tasks to nearby vehicles (Su *et al.*, 2018b) or infrastructure (Chekired *et al.*, 2019; Peng *et al.*, 2019) to address the problem of limited resources to run computing-intensive tasks for autonomous driving.

A market mechanism is used to motivate users to share the idle resources of their vehicles (Su *et al.*, 2018b). This mechanism creates an ad-hoc marketplace where the prices of the resources are settled according to the number of idle resources available in the seller vehicle. SDN/NFV network slicing is used to isolate certain driving functionalities in service slices (Chekired *et al.*, 2019) to attend to ultra-low latency requirements of autonomous driving services. AVNET (Peng *et al.*, 2019) is an architecture to address issues related to the amount of data transmitted and the number of processing tasks in CAVs scenarios. This architecture proposes the usage of: (i) NFV to implement multiple network functions and allow more diverse CAVs services to be deployed; (ii) MEC to offload tasks of these services to infrastructure and also nearby vehicles; and (iii) SDN to maintain a global view of the network to achieve efficient resource management. Routing protocols to better manage task offloading are also studied to operate using the ICN principle (Zhao *et al.*, 2019a).

5.2.3 Internet of Vehicles

Constituted by distributed transport communication networks, IoV (Lee *et al.*, 2016) allows ITS applications to make decisions based on data collected from other vehicles and sensors, which aid the process of driving people and goods towards their destinations. The communication features provided by IoV are important for applications, such as TMS and CAVs, and also applications related to smart-parking and virtual traffic lights. EC, in the form of VEC/VFC, is essential to this class of applications since communication and processing facilities are deployed in the vehicles. These vehicles use OBUs to perform communication and run tasks to support services. These OBUs use dedicated short-range devices and enable the formation of VANETs. VANETs do not require any infrastructure to be formed, yet RSUs can be used to improve the network QoS and overall capacity. One of the duties of vehicular communication is to handle emergency communication, such as car accident notifications or traffic flow reports. In this context, eVNDN (Ho *et al.*, 2016) applies ICN to broadcast emergency-related messages in vehicular networks, exploring the facility of communicating to fast-moving nearby nodes and also determining their interest in a given message. An emerging class of applications for vehicular networks is Vehicular Social Networking (VSN), in which vehicle riders share spatio-temporal data with other vehicles in similar conditions. SDN and Blockchain technologies can help to certify data exchange transactions in a distributed fashion while ensuring data source anonymity (Yahiatene *et al.*, 2019). Different networking technologies have been combined to enhance VANETs and cope with applications that rely on vehicular communication, such as GOFP (Liu *et al.*, 2016b), which supports geographically tagged information retrieval in VNDN, or SCGRP (Venkatramana *et al.*, 2017), an SDN-enabled geographic routing protocol.

One of the main standards to realise IoV is the Cellular Vehicle-to-Everything (C-V2X) developed by 3GPP (Chen *et al.*, 2020). It envisions the evolution of LTE-V2X (4G) to NR-V2X (5G) to enable highly-reliable low-latency service provisioning for vehicular applications. This evolution will allow the support of advanced vehicular applications with stringent requirements. For instance, the most stringent application for LTE-V2X (i.e., pre-crashing sensing and warning) requires 20 ms latency and 95% reliability (ETSI, 2015), while use cases for NR-V2X (e.g., emergency trajectory alignment) require latency levels to reduce to 3 ms with 99.999% reliability (ETSI, 2020). Finally, besides LTE and NR, more technologies can be combined and complement each other to meet application requirements in Heterogeneous Vehicular NETworks (HetVNETs) (Zheng *et al.*, 2015) framework (e.g., IEEE 802.11p).

5.3 UAV-Enabled Services

UAV-based platforms become an infrastructure alternative to network management and sensing for multiple applications. The advantage of such platforms is related to their aerial characteristics, which facilitate deployment almost anywhere. This possibility of easy deployment brought attention from the government and industry to adopt UAVs (Rao *et al.*, 2016). Communication among UAVs usually is supported by satellites, cellular networks, or Unmanned Aerial Vehicles Ad-hoc Networks (UAVANETs). This section highlights some UAV classes of applications and discusses how networking technologies are applied to support them.

5.3.1 Augmented Environment Information

Diverse applications for UAVs obtain information about a given variable of the environment to feed this information to other applications or systems. In these applications, UAVANETs have to handle a significant amount of data collected by themselves or terrestrial nodes that report to them. EC infrastructure can be used to aid the data collection by running tasks related to data aggregation and fusion and also by coordinating how data should be reported to its consumers. For instance, an early fire detection system that uses the sensing capabilities of UAVs to produce short videos that are sent for analysis at EC infrastructure is proposed (Kalatzis *et al.*, 2018). A container orchestrator handles the processing tasks in EC infrastructure. This orchestrator can create, run, scale, and stop services. A different use case of UAVs to aid monitoring is to replace network infrastructure. Data collection can be executed in areas with no wired infrastructure by deploying network backbones with UAVs (Zhang *et al.*, 2018c). A load-balancing algorithm operated by an SDN controller manages the data traffic in this backbone. ICN in-network caching is used to mitigate the issue of content dissemination in UAVANETs (Lei *et al.*, 2019). A blockchain-based strategy to handle content poisoning that may contaminate cache and prevent the fetching of valid content is used to enhance the security of the UAVANET.

5.3.2 Navigation and Swarming

Due to the flying capabilities of UAVs, UAVANETs can quickly adapt their topology to respond to network events. EC infrastructure can aid the process of coordination of the drones by offloading tasks or providing a wider view of the system. One approach that can be used to handle dynamic changes to the network topology is by using SDN in UAVANETs. SDN controllers can be used to send control packets to UAVs, demanding that they move to different positions (Rahman *et al.*, 2018). To allow this control, the controllers use a search procedure that looks at the rate demands and paths of each communication flow and changes the topology to maximize throughput. A more complex collaboration scenario for UAVs is the formation of drone swarms, which are open networks that can organize themselves. SD-UAVNet (Zhao *et al.*, 2019b) is another architecture for UAV placement to optimize UAVANETs. SDN can control operational parameters of UAVs and mitigate the impact of the mobility of UAVANETs for streamed video transmissions by positioning relay UAV nodes. SDN is used with the MQTT[1] protocol to enable flexible swarms formation while allowing the control of topology and bandwidth (Xiong *et al.*, 2019). A multi-path routing scheme is proposed, in which drones move to produce multiple communication paths. These multiple paths are used to increase the bandwidth to meet the desired QoS.

5.4 Immersive Interactive Media

Augmented Reality (AR), Virtual Reality (VR), and other mixed reality experiences have recently been applied to pro-

[1] http://mqtt.org/

duce immersive media applications. Such applications extend reality by emulating it on a device and adding new layers of information. The increase in such applications is due mainly to the recent popularization of head-mounted devices and also smartphone capabilities to support immersive media (e.g., smartphone-enabled cardboard headsets). Applications for immersive media are resource-consuming, which reduces the user experience of such applications because of the necessity of the headset being wired to a powerful computer – or at least usage with limited mobility when in wireless scenarios. However, EC technologies can support applications consumed in (mobility-free) wireless headsets/devices in the near future by offloading resource-consuming tasks. In contrast, there are immersive-only classes of applications, such as 360º Videos (Zink *et al.*, 2019). This section focus on Augmented and Virtual Reality and Teleoperation and Telepresence.

5.4.1 Augmented and Virtual Reality

A large amount of data is being gathered from IoT devices and other remote sources that can be accessed through the Internet. AR is an interactive experience with real-world mediated by human-machine interfaces. These interfaces allow a better visualization of this data collected from different sources. Such data needs to be organized and aligned to coordinate systems on top of the real-world coordinates to allow this visualization. Since AR/VR terminals have limited resources, these tasks can be offloaded to EC infrastructure. Such a setup has been applied, for instance, to build Industry 4.0. Navantia's Industrial AR (Fernández-Caramés *et al.*, 2018; Fraga-Lamas *et al.*, 2018) is commercially used to facilitate the execution of certain tasks in shipyards. This system uses EC infrastructure to reduce the response delay to handle real or near-real-time communication wirelessly – wireless technology is required to allow the desired level of mobility inside the shipyard industry. VR-CPES (Kim *et al.*, 2018a) is an education system that also offloads tasks to EC. This system uses an SDN-enabled Time-Sensitive Networking (TSN) framework to address issues of QoE of users due to delay and packet loss in real-time network communication.

5.4.2 Teleoperation and Telepresence

Immersive media are also expanding the horizons of people with multiple interactive applications. For instance, different applications involve the remote operation of devices or even the telepresence of people; such applications allow to save time and reduce costs. Such applications require a reasonable amount of data to be sent and processed (e.g., videos and metadata about the environment around both ends of the communication), which can be handled by EC infrastructure. A use case of immersive media in the industry, for example, is remote live support (Schneider *et al.*, 2017). In this application, a machine operator can receive help to fix an issue from an expert. The system uses an approach to offload AR tasks to EC infrastructure to deploy a real-time live support system. Thus, a remote expert can make annotations to a video stream recorded by the machine operator. The operator also visualizes these annotations to facilitate the process of fixing the issue. In this use case, an important concept is applied, transferable skills. Immersive media is one of the key enablers for the anticipated Internet of Skills and Tactile Internet (Antonakoglou *et al.*, 2018). The concept of Tactile Internet is to reproduce touch-based human communication to the network – the subject will be discussed later in Section 6.9. In the field of Tactile Internet, some applications are already being proposed, such as telesurgery. SDN-enabled EC infrastructure is used to reduce latency dramatically and allow surgeons to remote control a surgery robot (Miao *et al.*, 2018). To realize Tactile Internet, EC architectures (e.g., Fog Computing, MEC, SDN, and ICN) are being studied to work together while also using robust ML models to predict movement and actions of users, thus reducing latency even further (Oteafy and Hassanein, 2019).

5.4.3 Gaming

Among the top 10 highest downloaded games for mobile devices nowadays, Niantic's Pokemon GO[2], an AR game, can enhance the gaming experience provided to users using EC infrastructure. Games that immerse players in the real world have high QoE requirements and challenging mobility characteristics to be handled by communication facilities, which may require EC-enabling technologies. Indeed, in 2018 Deutsche Telekom[3] started placing decentralized micro-servers to leverage EC infrastructure deployment, and Pokemon GO was one of the first AR applications to use this platform (Wireless One, 2018). Such infrastructure, and also 5G networks, will augment the possibilities for game development for this and other AR and VR games where mobility is a critical factor. Another type of EC approach to enhance mobile gaming experience is UAV-assisted EC (Kim *et al.*, 2018b; Hu *et al.*, 2019b). In such a scenario, mobility awareness is an even more critical factor, since both consumers and producers will be mobile entities. This setup supports AR and VR games to be played inside vehicles (Hu *et al.*, 2019b). Multiple UAVs are clustered to offload tasks related to computing, caching, communication, and AI-based decision-making.

5.5 Smart Cities

The application of information and communication technology to enhance the performance of services in large urban centers added a lot of attention to smart cities. The idea behind smart cities is to build infrastructure for monitoring of several city dynamics and act according to insights obtained with this data to serve citizens better. We discuss Public Services (Section 5.5.1) and Location-based Services (Section 5.5.2) that can be enhanced by EC infrastructure.

5.5.1 Public Services

Various services in urban centers can take advantage of information and communication technology, such as power, water, environment monitoring and waste management. Often,

[2]https://www.pokemongo.com/
[3]https://www.telekom.com/

sensor networks are deployed to collect data over large regions. A4-Mesh Jamakovic *et al*. (2012) is a wireless mesh sensor network deployed to collect weather data in near-real-time. This network produces a large amount of data that needs to be sent to a central remote processing station. Wireless sensor networks are convenient since they can be easily deployed without a big effort on underlying infrastructure. In order to make better usage of the communication channels, multiple sensors in an urban center can use Narrowband communication technology for IoT applications (NB-IoT). Although this type of communication technology has a reduced bandwidth, it uses fewer frequencies of the wireless spectrum and has low power consumption. Such technology may be adequate in scenarios where a large number of sensors can be spread in the urban perimeter to increase data collection coverage. In environmental monitoring applications, for instance, each sensor does not transmit a large volume of data and usually has a limited battery, making it an interesting use case for exploring NB-IoT (Shen *et al*., 2022; Cheng *et al*., 2019).

Many applications for Smart Grid aimed to allow power generation and electricity transmission are delay-critical. ICN is used to enhance the communication needed to manage such an infrastructure (Katsaros *et al*., 2014). ICN allows the reduction of delay to obtain data about the current state of the grid. Battery Status Sensing Software-Defined Multicast (BSS-SDM) (Li *et al*., 2017a) is a battery status sensing scheme based on SDN to reduce the latency of the communication between electric vehicles and the power grid. An SDN controller keeps the status of the batteries in vehicles. This information is used to schedule vehicle recharges. The vehicles are notified via messages transmitted by multicast. Another important issue pursued in smart cities is security. For instance, different technologies have been applied to analyze surveillance videos and identify events. One way to manage all the surveillance application and video analysis services – while also reducing the traffic load sent to the core network – is by orchestrating containerized services over the EC infrastructure (Vilalta *et al*., 2017). SDN-enabled containers allow the SDN controller to orchestrate better and save EC resources. AODV-SPEED (Ahmed and Rani, 2018) is a communication protocol to enable smart street highlights. This protocol combines SDN and ICN to enhance network QoS for service provisioning.

5.5.2 Location-based Services

Location-based services consume strategic spatio-temporal information to deliver value to their users. In general, many services running on EC infrastructure can take advantage of location awareness. For instance, the position information of mobile users can be used to improve service resource scheduling and deployment in virtualized platforms (Meneguette *et al*., 2019). Tracking moving objects is an important source of information when studying more reliable and predictive services. ICN and in-network service-provisioning functions were used to develop a moving object tracker application (Tanaka *et al*., 2019). This application coordinates a distributed video service that produces a video stream of a single moving entity (e.g., vehicle) using a system of multiple cameras. The video consumer sends an interest containing the vehicle ID to the network, and, later on, each camera receives this vehicle ID and transmits the video only when the vehicle is in its capture area. The vehicle sends its position to the network to verify in which camera capture area it is at a given moment.

5.5.3 Mobile Crowdsensing

Due to the large urban perimeter in some cities, deploying infrastructure to sense entire urban areas may become challenging. One solution to tackle this issue is by applying mobile crowdsensing (You *et al*., 2017; Li *et al*., 2017b; Longo *et al*., 2018). Mobile crowdsensing advocates for the sharing of spatiotemporal annotated data collected from mobile devices (e.g., smartphones and tablets) about different urban phenomena. To manage the formation of opportunistic networks in mobile crowdsensing, Software Defined Opportunistic Networks (SDON) (Li *et al*., 2017b) uses the centralized control of SDN. Statistical data stored at the SDN controller allow the creation of an incentive mechanism for users' participation in the sensing process. A different approach to manage opportunistic networks for mobile crowdsensing is by using ICN (You *et al*., 2017). An urban pollution monitoring system orchestrates container-based microservices to integrate data from multiple heterogeneous data sources (Longo *et al*., 2018). These services compose a layer where data streams from different sources (e.g., mobile phones, IoT sensors) are integrated.

5.6 Edge AI

Edge Computing enables a series of interesting use cases for applications to be explored, among them Edge AI aims at allowing the usage of Machine Learning models to enhance the applications already running or envisioned for the Edge. Thus, in the present section, we discuss two main aspects of intelligent service provisioning at the edge: (i) how to manage the Cloud-Edge infrastructure in Section 5.6.1, and (ii) technologies to support smart services at the edge in Section 5.6.2. The usage of ML models raises challenges on data privacy that can be overcome by using Edge architectures. Data privacy issue and related challenges are discussed in Section 6.7.

5.6.1 Infrastructure Management

One application of intelligence at the edge is the orchestration of the envisioned multiple services provisioned in Fog Computing and MEC infrastructure. Fog Scaler (Sami *et al*., 2020) is a service orchestrator that targets horizontally scaling services running at the edge. The authors use a Reinforcement Learning algorithm and model different cost functions to allow their solution to take decisions on the placement of containerized service instances. OctoFog (Lan *et al*., 2020) is another service orchestrator solution focusing on optimizing the migration of services at the edge. The authors minimize two cost functions that model latency and energy consumption of the migration procedure. This minimization is achieved by applying a Deep Reinforcement Learning algorithm that is divided into two layers, one hosted in the Cloud and the other in the Fog. The Cloud layer hosts the main control of the resources, while local decisions are taken at the Fog layer with reduced latency. Liu *et al*. (2021b) divide IoT services into a collection of chained service functions. They proposed a VNF placement and service path routing framework that minimizes the end-to-end delay observed when consuming the services. Their solution uses a Deep Reinforcement Learning approach to achieve this minimization in real-time. The solution observes the IoT network state and the number of requests to the IoT services and outputs orchestration strategies composed by VNF placement and network routing paths.

Besides orchestrating services, allocating resources is an important aspect of provisioning services at the edge that can take advantage of ML models to be enhanced. Shi *et al*. (2020) propose a mechanism to match idle vehicular computing resources to tasks that have to be processed for delay-sensitive applications. The authors propose a reinforcement learning algorithm that evaluates wireless channel state and idle resource pool in vehicles and outputs efficient task allocation strategies. Since the resource of vehicles may not be voluntarily shared, the authors propose a pricing scheme to stimulate this sharing. In a similar setting, Ye *et al*. (2019) propose an alternative reinforcement learning solution for communication resource allocation in vehicular computing resources. One advantage of this method is the possibility of independent vehicles taking decentralized decisions to satisfy desired latency constraints.

5.6.2 Support for Smart Services

Kubernetes-Based Fog Computing IoT Platform for Online Machine Learning (KFIML) (Wan *et al*., 2022) is a platform for service orchestration at the edge developed on top of Kubernetes. This platform has the potential to facilitate the deployment and management of different service stacks at the edge, including mainstream data processing frameworks. The authors use their proposal to manage a LSTM-based real-time data stream processing applied in an IoT scenario. Dalgkitsis *et al*. (2021) propose a service orchestration platform for Vehicular-to-Everything scenarios that aims at predicting the next access point of vehicles in the cellular network and migrating services consumed by these vehicles proactively. The authors use a Convolutional Neural Network to predict the next access point of the vehicles, then a Genetic Algorithm is used to search for a service allocation strategy to place services closer to their users while considering user priorities and resource utilization.

When dealing with IoT applications, one common limitation is the reduced processing capacity and energy available in the end devices. Sometimes, although a ML model that performs well is trained, it cannot run on the device and therefore cannot be used. One solution in the literature for these scenarios is the partitioning of the model for inference acceleration. Partitioned models are composed of many layers that can run at different distances from the source node consuming the model predictions. Dynamic Adaptive DNN Surgery (DADS) (Hu *et al*., 2019a) is a framework that allows the partitioning of DNN models to adapt its usage according to the status of the network. This model can dynamically adapt the model partitioning to maximize the throughput or minimize the delay of predictions. A similar strategy of partitioning models can be applied during the training phase in order to save the resources of devices. Adaptive REsource-aware Split-learning (ARES) (Samikwa *et al*., 2022) is a solution that accelerates the training phase of a global model by selecting split points for every device involved in the training considering network and computing resource variation over time. The approach also reduces the impact of slower devices in the time taken for training the model, which is important in highly heterogeneous scenarios.

6 Challenges and Opportunities for Mobility-Aware Service Provisioning

EC architectures for service provisioning in urban environments have been extensively studied (Naha *et al*., 2018; Tayyab *et al*., 2019; Rejiba *et al*., 2019; Mao *et al*., 2017a; Laghrissi and Taleb, 2019; Zhang *et al*., 2018a; Roman *et al*., 2018). Academia has put much work into proposing EC technologies, as shown throughout this survey, and also some successful industry use cases can be observed[4], mass adoption of EC has not happened yet. However, some issues related to its practical deployment still need to be addressed to fully achieve the stage 3 depicted in Figure 2. Besides the costs involved in deploying such an infrastructure, it is unclear how to overcome many obstacles.

While running services at the edge facilitates the handling of some of these obstacles, it also brings new challenges to networking. In the present section, we discuss the new challenges that emerge together when computation is performed at the edge. Section 6.1 discusses how mobility predictors may be used to avoid communication disruption caused by the mobility of the users. Section 6.2 discusses issues related to the migration of service instances running at the edge. Section 6.3 outlines how caching strategies can be used to support service provisioning. Section 6.4 discusses the usage of distributed authorization to secure data access hosted in multiple nodes at the edge. Section 6.5 shows challenges when allocating tasks to run at Edge infrastructure. Section 6.6 discusses the implementation of distributed file systems over Edge networks. Section 6.7 highlights the importance of data privacy solutions when ML models are trained at the edge of the network. Section 6.8 describes scalability-related challenges of services. Finally, Section 6.9 discusses the challenges to be overcome when moving towards the next stage of service provisioning over the Internet.

6.1 Mobility Prediction

Short-range coverage of EC access points will lead to multiple handovers due to users' mobility. These multiple han-

[4] AWS Lambda@Edge (https://aws.amazon.com/en/lambda/Edge/) and Green Grass (https://aws.amazon.com/en/greengrass/), and Google Serverless with KNative (https://Cloud.google.com/serverless/).

dovers will turn mobility management into an essential aspect of service provisioning. These handovers may add significant overhead to use communication and computation infrastructure, depending on their implementation. One way to mitigate this issue is by exploring historical data about users' trajectories to enable proactive handover mechanisms. For instance, for a communication handover in an EC setting with SDN, the flow tables of the forwarding switches can be updated before the users even enter a specific access point. In terms of service migration, data prefetching can start loading service dependencies (e.g., software libraries) and also deploying services, thus making them ready for users and reducing migration downtime. Recent studies have shown a significant impact of ML models in mobility management (Sun *et al*., 2019), causing different methodologies for predicting users' mobility to appear in the literature using, for example: Markov Chain Models (Qiao *et al*., 2018), Reinforcement Learning (Zhang and Zheng, 2019), and Deep Learning (Jiang *et al*., 2018).

6.2 Service Migration

EC-enabled services are pushed forward because of their closeness to users. However, when users move, the host where services run may become far from them, which might be critical for some applications (ETSI, 2018; Campolo *et al*., 2019). For instance, optical X2 links for backhaul in mobile networks are expected to have latency of ≈0.3-0.5 ms when operating between 40%-70% of traffic load (Li *et al*., 2017). In this scenario, round trips with three or four hops adds a few milliseconds to respond a request, when considering also other delays in processing and in the wireless channel it might be challenging to meet the expectations of applications that demand (ultra-)low latency combined with high throughput, high reliability, or low tolerance to latency jitter. VMs and containers are technologies expected to handle the fast deployment of services to allow live migration to keep them running near users. Each methodology for virtualization has its advantages. VMs provide a more isolated and secure environment for services (Manco *et al*., 2017), while containers are more lightweight and have an overall better performance (Chae *et al*., 2018). However, most of the available studies to compare these technologies do not profile them thoroughly. These studies lack awareness of the possibilities in terms of virtualization architectures. Different architectures may have an impact on the performance of the services. Also, according to the virtualization technology, various possibilities of migration strategies have been proposed and adopted, but comparative studies on the migration feature are still needed.

6.3 Service Caching

Mobile services may require different types of resources and have different requirement levels of QoE, which make this problem different from content caching (e.g., Content Distribution Networks (CDN)) (Yang *et al*., 2016). For instance, VR applications may require a reasonable amount of processing power but a reduced amount of memory; compared to data collection tasks that may need more storage and not so much CPU and memory. One approach to enhance physical resource allocation for service caching is to perform *spatio-temporal popularity-driven service caching* (Mao *et al*., 2017a). Content and service popularity are modeled using the Zipf model (Zipf, 1950; Mehrabi *et al*., 2019) to when evaluate caching solutions. Popularity-driven caching is used to predict which types of services are more used in a given location at a certain time. Thus, service shutdowns (i.e., stop a virtual instance of a service and clean unnecessary files from the host) may be prevented when predicting upcoming usage. This type of approach may also aid in the process of service migration combined with mobility prediction. Different service components can be stored for a more extended period in case a user is migrated to a specific host; popular services can be kept running in this scenario, reducing time with re-deployment. Besides controlling the life cycle of the service instances, in order to run stateful services in geographically distributed nodes, state data replication may be required (Filho and Porter, 2020). Different from regular contents, state data is often more volatile, which may demand complex solutions to ensure availability and coherence.

6.4 Service Authorization and Session Support

Due to mobility, while changing access points, and performing handovers, applications have to perform multiple authorizations and exchange access keys to grant access and keep user sessions. This process may add significant overhead to consume services at the edge, which leads to a necessity of shared trust domains or alternative security protocols for accessing distributed services. Ideas to embed authorization in networking are present in Data-Oriented Network Architecture (DONA) (Koponen *et al*., 2007), yet there is a lack for support of distributed and federated sessions. CCNx Key Exchange Protocol (CCNxKE) (Mosko *et al*., 2017) proposes mobile sessions in CCNx. However, the mobility in this protocol is handled by exchanging keys (i.e., migration token) which still results in network overhead in this process. Session support has also been studied in SCN (Gasparyan *et al*., 2019), yet it relies on consuming the service from the same host infrastructure and does not support the mobility of the sessions.

6.5 Service Load Balancing

EC offloads computing-intensive tasks from simple devices to idle resources in its resource pool. Resources in this pool might be in mobile network base stations or Edge devices. In this second group, these devices have a limited communication radius that, given user mobility, may reduce the available time to use that resource. This problem becomes even more complicated when both consumers and resources are mobile entities. In this scenario, the communication links lifespan can become longer or shorter depending on their mobility patterns. Evaluation, clustering, and classification of these mobility patterns, which use this information for task offloading, are challenges for EC deployment. Research has been done in this direction for task offloading in the presence of mobility (Tran and Pompili, 2019; Liu and Zhang, 2019).

However, most of these studies address the problem in single networks. EC will be achieved by the usage of a series of heterogeneous networks, which increase the complexity of the task offloading process. Models to understand the relevant metrics to decide when to migrate service instances have to be developed, for instance by considering ML models (Urgaonkar *et al*., 2015; Chen *et al*., 2019a).

6.6 Distributed File Systems

In Cloud Computing, different distributed storage mechanisms have emerged to address scalability, performance, and reliability issues and provide virtually infinite storage services in distributed resources. Ceph (Weil *et al*., 2006) is a commercially adopted solution to create such a storage service that relies on a pseudo-random function to distribute data to resources. Using this function, users can evaluate the location of the data, which saves resources to perform searches. Although Ceph is a well-known solution, it was designed to work in a scenario with plenty of resources, which is not a reality in EC cases. Limited bandwidth at the edge allied to massive access by users and connected devices makes the process of creating a fully-distributed EC-oriented file system difficult (Ma *et al*., 2018). In the literature, some strategies for these file systems are already emerging, for instance snapshot and synchronization techniques to transfer disk state in the network (Pamboris *et al*., 2019). Also, FogStore (Gupta and Ramachandran, 2018) is an EC-oriented file system that uses key-value storage and a relevance system to guarantee low latency in file access. Reliable mechanisms to provide file systems at the edge are still under study. One studied issue is how to perform redundant storage to ensure reliability while also not overusing resources. Duplicating contents to selected locations may be a solution to this issue. Another solution is to apply data deduplication (Zhang and Ansari, 2014), which identify intra- and inter-document duplicated data blocks and store these blocks only once. Fewer blocks (i.e., less data) to be stored can facilitate the process of deploying a distributed file system with reliability assured by copying blocks at selected locations.

6.7 Data Privacy and Federated Learning

Data leakage points are reduced when processing data at the edge since this processing happens closer to the data generation and fewer data transfers are needed. Still, similarly to what happens in the Cloud, processing user data in public servers at the edge introduces issues on data privacy. Such problem becomes more apparent when using this data to train different ML models, which sometimes requires this user data to be shared across multiple servers. One possible solution that emerges to tackle this issue is Federated Learning (FL) (Lim *et al*., 2020), in which ML models can be trained locally in devices and only share its learning updates (i.e., model weights or gradients) with centralized servers, instead of sharing the raw data. These centralized servers can then collect multiple updates from several devices and aggregate them to produce a global model. By training the model locally and avoiding data sharing, FL reduces the number of points where user data may leak. However, personal data sometimes may still be extracted from the trained model with model-inversion attacks (He *et al*., 2019). Besides not fully resolving the data leakage problem, FL still lacks maturity in other aspects. For instance, models trained with FL tend to have convergence problems resulting from the non-identical distribution of the data produced by each device (Zhu *et al*., 2021). On the communication perspective, when dealing with mobile devices some problems may arise. For example, when training an online model with FL with a large user pool, it is a common practice to select some users to be the ones feeding the global model with the learning updates. However, in mobile scenarios users may disconnect from the network or not be available for a certain period of time. Reducing the pool of selected devices for training may impact the training quality of the model. Furthermore, even when these users can return to the network, most FL approaches are synchronous (Bonawitz *et al*., 2019), thus training has to be halted to wait for the updates of these users.

6.8 Scalability

Most of the issues for EC infrastructure arise from the existence of many connected devices creating tremendous amounts of data to be consumed. However, creating such a platform to support all this load is not an easy task. First, a plethora of heterogeneous devices will be integrated into the pool of resources. In this scenario, an interoperable architecture to control these devices is a requirement to allow the construction of EC. Second, multiple technologies to run EC are being proposed in the literature, as shown in Section 3. While most of them will not make it to the real world, many will need to be integrated, thus raising questions about the compatibility of devices, algorithms, and protocols. Even the integration between SDN, NFV, and ICN paradigms is not well defined in the literature, although these technologies are expected to run together in the future. Finally, the Cloud computing paradigm has become mainstream due to the offer of IaaS solutions by big IT companies, such as Amazon, IBM, and Google. Market driving forces to produce a similar platform based on EC are studied (Frank *et al*., 2013). However, a complete architecture to elastically serve this infrastructure for companies to deploy their service is unknown to the best of our knowlEdge.

6.9 Ubiquitous Service Provisioning

Applications and services consumed over the Internet are used as facilitators to perform activities that were previously only possible using physical means such as sending a letter, buying products, and attending a theatre play. As Internet technologies evolve, more activities are performed online. At stage 1 of service provisioning depicted in Figure 1, humans had to formalize instructions via messages through mostly textual interfaces. Later, these text interfaces evolved to graphical browser and phone-based interfaces and also virtual environments in video-games in stage 2 to facilitate the way humans interact with the applications. Recently, the advances in IoT and EC started to allow the direct control of devices to act over the real-world in stage 3. This last approach uses a virtual world as a middleware. Humans input instructions to a device, which modifies a virtual world, and the changes made to it are replicated by other devices (i.e., actuators) in the real world using the Cybertwin paradigm (Yu *et al*., 2019).

The bridge towards stage 4 of ubiquitous service provisioning will be built with the development of technologies that blend virtual and physical worlds. Important enablers for ubiquitous service provisioning are the Mixed Reality (XR) technologies (Kim *et al*., 2018b; Chakareski, 2017) such as Augmented Reality, Virtual Reality, and 360° videos. While multiple of these devices can already be commercially acquired, there is still a significant barrier to be surpassed in terms of user experience and freedom of movement. Most of these interfaces are developed in head-mounted, glasses, and other wearables devices with limited capabilities and sometimes mobility. EC and related communication infrastructure are expected to allow applications to meet the increasing QoE requirements from users by providing a reasonable resource pool accessible with very-low end-to-end latency, while also dealing with mobility-related issues. With immersive experiences applications such as the Metaverse (Dionisio *et al*., 2013) gain attention, which foresees the development of multiple 3D immersive virtual worlds. The popularization of immersive virtual worlds also brings interest on moving entities, objects and sensations in and out of the virtual reality with neural interfaces (Sadeghi *et al*., 2016; Martins *et al*., 2019; Grau *et al*., 2014), immersive projections (Takeda *et al*., 2016; Hoggenmueller and Tomitsch, 2019), holographs (Matsushima and Sonobe, 2018; Su *et al*., 2018a), and haptic communications (Kataoka *et al*., 2019; Schneider and Blikstein, 2018; Bong *et al*., 2018). New technologies easing the process of moving entities between different realities will then allow the emergence of novel and complex applications, such as the Tactile Internet (Antonakoglou *et al*., 2018) and the Internet of Skills (Oteafy and Hassanein, 2019; Antonakoglou *et al*., 2018). These applications are awaited to disrupt the way humans interact by allowing the transference of abilities, skills, and knowlEdge over the Internet.

7 Lessons Learned and Final Remarks

Networks and the Internet are constantly evolving, and for every stage of evolution, different challenges have to be overcome. In the present stage, service provisioning is expanding from the centralized Cloud to the Edge of the network, which raises various questions as presented in this survey. Also, in the future, new paradigm shifts will raise different questions about how to better provide Internet services. This survey shows an overview of the main paradigm shifts experienced in the Internet and the application evolution that have pushed their changes. Also, we discussed the current state-of-the-art for the provisioning of in-network mobile services. We discussed many recently-proposed enabling technologies and also highlighted some applications that use these technologies. Furthermore, we presented incoming challenges for broad adoption of EC technologies in the near future and even upcoming opportunities to researchers looking for the next stages of evolution in service provisioning. We believe technologies such as SDN, NFV, and ICN have a significant role in deploying EC and future service provisioning paradigms.

Different technologies and architectures have been developed to fullfil the requirements of emerging and upcoming services and applications. The resource pool includes computing and communication infrastructure deployed in the Cloud-Edge continuum, and also user equipment such as the vehicles. All these technologies are expected to coexist and to be accessible to be used by different service providers in a similar fashion of Cloud solutions nowadays. The interoperability, control and management of such a complex platform will rely on high levels of virtualization and orchestration of services, containers, VMs and communication infrastructure. These orchestration platforms will make use of multiple Machine Learning solutions to automate decision taking and automation of tasks.

The idea of EC emerged to bring computing power closer to users in a context where accessing Cloud resources could take up to a few hundreds of milliseconds. Nowadays, latency to access Cloud Computing has reduced, still only EC will allow the support of emerging applications with requirements of ultra-reliable very-low latency with low jitters and high data throughput. For instance, a reliable and near-deterministic support of very-low latency for wireless communications can enable the provisioning of real-time services in which task deadline meeting is critical. EC must be able to operate isolated from the Cloud in case of link failure and maintain services running. Furthermore, computing tasks running closer to data generation will reduce the amount of data sent to the core of the network. Reducing this data movement improves communication resource allocation and reduce the number of points for data leak, thus reducing concerns about data security and privacy. Finally, EC architectures will allow a better exploration of the spatio-temporal locality of data generation and service consumption. Innovative networking and computing solutions are emerging that exploit this locality in order to better provision services, such as the new networking paradigms for Future Internet (e.g., Geo-centric and Information-centric networking).

While helping to conceive, design, evaluate, and test EC infrastructure, academia has another important role on understanding the impacts caused by these technologies on society and envisioning what will be the next applications for service provisioning over the Internet. The popularization of immersive devices has created a trend for immersive experiences and put a lot of attention in applications such as the Metaverse and other possibilities of immersive-first applications such as the Internet of Skills. While difficult to predict whether these applications will penetrate society as the Internet and mobile devices, they have possibilities of highly impacting the traditional ways of human interactions. Possible use-cases for such applications are still to be unveiled and can lead to interesting research topics.

Acknowledgment

The authors would like to thank the São Paulo Research Foundation (FAPESP), grants #2020/11259-9, #2018/12447-3, #2018/19639-5, #2015/24494-8, #2018/23064-8, the FAPEMIG, grants #APQ-02510-18, and #APQ-00426-22.

Declarations

Authors' Contributions

All authors contributed to the writing of this article, read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Data can be made available upon request.

References

3GPP (2014). TS 136 133 - V12.5.0 - LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Requirements for support of radio resource management (3GPP TS 36.133 version 12.5.0 Release 12). Available at: `https://www.etsi.org/deliver/etsi_ts/136100_136199/136133/12.05.00_60/ts_136133v120500p.pdf`.

3GPP (2016). TR 36.881 V14.0.0 - Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Study on latency reduction techniques for LTE (Release 14). Technical Report Release 14, 3GPP. Available at: `https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2901`.

Abbas, M. T., Muhammad, A., and Song, W.-C. (2019). SD-IoV: SDN enabled routing for internet of vehicles in road-aware approach. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–16. DOI: 10.1007/s12652-019-01319-w.

Abdelmoneem, R. M., Benslimane, A., Shaaban, E., Abdelhamid, S., and Ghoneim, S. (2019a). A Cloud-Fog Based Architecture for IoT Applications Dedicated to Healthcare. In *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE. DOI: 10.1109/ICC.2019.8761092.

Abdelmoneem, R. M., Benslimane, A., Shaaban, E., Abdelhamid, S., and Ghoneim, S. (2019b). A Cloud-Fog Based Architecture for IoT Applications Dedicated to Healthcare. In *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE. DOI: 10.1109/ICC.2019.8761092.

Aceto, G., Persico, V., Pescapé, A., and Member, S. (2019). A survey on information and communication technologies for industry 4.0: State-of-the-art, taxonomies , perspectives , and challenges. *IEEE Communications Surveys and Tutorials*, 21:3467–3501. DOI: 10.1109/COMST.2019.2938259.

Afrin, M., Jin, J., Rahman, A., Rahman, A., Wan, J., and Hossain, E. (2021). Resource allocation and service provisioning in multi-agent cloud robotics: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 23:842–870. DOI: 10.1109/COMST.2021.3061435.

Ahmad, E., Alaslani, M., Dogar, F. R., and Shihada, B. (2019). Location-Aware, Context-Driven QoS for IoT Applications. *IEEE Systems Journal*, pages 1–12. DOI: 10.1109/JSYST.2019.2893913.

Ahmed, S. H. and Rani, S. (2018). A hybrid approach, Smart Street use case and future aspects for Internet of Things in smart cities. *Future Generation Computer Systems*, 79:941–951. DOI: 10.1016/j.future.2017.08.054.

Ahmed, S. H., Yaqub, M., Bouk, S., and Kim, D. (2015). Towards content-centric traffic ticketing in VANETs: An application perspective. In *2015 Seventh International Conference on Ubiquitous and Future Networks*, pages 237–239. IEEE. DOI: 10.1109/ICUFN.2015.7182541.

Ahmed, S. H., Yaqub, M. A., Bouk, S. H., and Kim, D. (2016). SmartCop: Enabling Smart Traffic Violations Ticketing in Vehicular Named Data Networks. *Mobile Information Systems*, 2016:1–12. DOI: 10.1155/2016/1353290.

Akbar, M. S., Yu, H., and Cang, S. (2017). Tmp: Tele-medicine protocol for slotted 802.15.4 with duty-cycle optimization in wireless body area sensor networks. *IEEE Sensors Journal*, 17(6):1925–1936. DOI: 10.1109/JSEN.2016.2645612.

Akherfi, K., Gerndt, M., and Harroud, H. (2018). Mobile cloud computing for computation offloading: Issues and challenges. *Applied Computing and Informatics*, 14(1):1–16. DOI: 10.1016/j.aci.2016.11.002.

Al-Turjman, F. M. (2017). Information-centric sensor networks for cognitive IoT: an overview. *Annals of Telecommunications*, 72(1):3–18. DOI: 10.1007/s12243-016-0533-8.

Alawe, I., Ksentini, A., Hadjadj-Aoul, Y., and Bertin, P. (2018). Improving Traffic Forecasting for 5G Core Network Scalability: A Machine Learning Approach. *IEEE Network*, 32(6):42–49. DOI: 10.1109/MNET.2018.1800104.

Albrahim, M., Zahrani, A. A., Arora, A., Dua, R., Fattouh, B., and Sieminski, A. (2019). An overview of key evolutions in the light-duty vehicle sector and their impact on oil demand. *Energy Transitions*, 3(1):81–103. DOI: 10.1007/s41825-019-00017-7.

Amadeo, M., Campolo, C., and Molinaro, A. (2016). Information-centric networking for connected vehicles: A survey and future perspectives. *IEEE Communications Magazine*, 54(2):98–104. DOI: 10.1109/MCOM.2016.7402268.

Amin, R., Reisslein, M., and Shah, N. (2018). Hybrid SDN networks: A survey of existing approaches. *IEEE Communications Surveys and Tutorials*, 20(4):3259–3306. DOI: 10.1109/COMST.2018.2837161.

Antonakoglou, K., Xu, X., Steinbach, E., Mahmoodi, T., and Dohler, M. (2018). Toward haptic communications over the 5g tactile internet. *IEEE Communications Surveys Tutorials*, 20(4):3034–3059. DOI: 10.1109/COMST.2018.2851452.

Arbabi, M. S., Lal, C., Veeraragavan, N. R., Marijan, D., Nygard, J. F., and Vitenberg, R. (2022). A survey on blockchain for healthcare: Challenges, benefits, and future directions. *IEEE Communications Surveys and Tutorials*. DOI: 10.1109/COMST.2022.3224644.

Arshad, S., Azam, M. A., Rehmani, M. H., Member, S., and Loo, J. (2019). Recent Advances in Information-Centric Networking-Based Internet of Things (ICN-IoT). *IEEE Internet of Things Journal*, 6(2):2128–2158. DOI: 10.1109/JIOT.2018.2873343.

Ashok, A., Steenkiste, P., and Bai, F. (2018). Vehicular Cloud Computing through Dynamic Computation Offloading. *Computer Communications*, 120:125–137. DOI: 10.1016/J.COMCOM.2017.12.011.

Baccarelli, E., Scarpiniti, M., and Momenzadeh, A. (2018). Fog-Supported Delay-Constrained Energy-Saving Live Migration of VMs Over MultiPath TCP/IP 5G Connections. *IEEE Access*, 6:42327–42354. DOI: 10.1109/ACCESS.2018.2860249.

Barakabitze, A. A., Barman, N., Ahmad, A., Zadtootaghaj, S., Sun, L., Martini, M. G., and Atzori, L. (2020). Qoe management of multimedia streaming services in future networks: A tutorial and survey. *IEEE Communications Surveys and Tutorials*, 22:526–565. DOI: 10.1109/COMST.2019.2958784.

Barbera, S., Pedersen, K. I., Rosa, C., Michaelsen, P. H., Frederiksen, F., Shah, E., and Baumgartner, A. (2015). Synchronized RACH-less handover solution for LTE heterogeneous networks. In *Proceedings of the International Symposium on Wireless Communication Systems*, volume 2016-April, pages 755–759. IEEE. DOI: 10.1109/ISWCS.2015.7454451.

Bellache, T., Shagdar, O., and Tohme, S. (2017). Dcc-enabled contention based forwarding scheme for vanets. In *2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 1–8. DOI: 10.1109/WiMOB.2017.8115767.

Bhatia, J., Dave, R., Bhayani, H., Tanwar, S., and Nayyar, A. (2019). SDN-based real-time urban traffic analysis in VANET environment. *Computer Communications*, 149:162–175. DOI: 10.1016/j.comcom.2019.10.011.

Bi, Y., Han, G., Lin, C., Guizani, M., and Wang, X. (2019). Mobility Management for Intro/Inter Domain Handover in Software-Defined Networks. *IEEE Journal on Selected Areas in Communications*, 37(8):1739–1754. DOI: 10.1109/JSAC.2019.2927097.

Bittencourt, L. F., Diaz-Montes, J., Buyya, R., Rana, O. F., and Parashar, M. (2017). Mobility-Aware Application Scheduling in Fog Computing. *IEEE Cloud Computing*, 4(2):26–35. DOI: 10.1109/MCC.2017.27.

Boban, M., Kousaridas, A., Manolakis, K., Eichinger, J., and Xu, W. (2018). Connected roads of the future: Use cases, requirements, and design considerations for vehicle-To-everything communications. *IEEE Vehicular Technology Magazine*, 13(3):110–123. DOI: 10.1109/MVT.2017.2777259.

Bonawitz, K. A., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C. M., Konečný, J., Mazzocchi, S., McMahan, B., Overveldt, T. V., Petrou, D., Ramage, D., and Roselander, J. (2019). Towards federated learning at scale: System design. In *SysML 2019*. Available at:`https://arxiv.org/abs/1902.01046` note: To appear.

Bong, W. K., Chen, W., and Bergland, A. (2018). Tangible User Interface for Social Interactions for the Elderly: A Review of Literature. *Adv. Human-Computer Interaction*, 2018:7249378:1–7249378:15. DOI: 10.1155/2018/7249378.

Bonomi, F., Milito, R., Zhu, J., and Addepalli, S. (2012). Fog computing and its role in the internet of things. In *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing*, MCC '12, page 13–16, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/2342509.2342513.

Braun, T., Mauthe, A., and Siris, V. (2013). Service-centric Networking Extensions. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, SAC '13, pages 583–590, New York, NY, USA. ACM.

Campolo, C., Iera, A., Molinaro, A., and Ruggeri, G. (2019). MEC Support for 5G-V2X Use Cases through Docker Containers. *IEEE Wireless Communications and Networking Conference, WCNC*, 2019-April:0–5. DOI: 10.1109/WCNC.2019.8885515.

Cao, S. and Lee, V. C. (2018). A novel adaptive TDMA-based MAC protocol for VANETs. *IEEE Communications Letters*, 22(3):614–617. DOI: 10.1109/LCOMM.2017.2785378.

Castelluccia, C. (2000). HMIPv6: A Hierarchical Mobile IPv6 Proposal. *SIGMOBILE Mob. Comput. Commun. Rev.*, 4(1):48–59. DOI: 10.1145/360449.360474.

Chae, M., Lee, H., and Lee, K. (2018). A performance comparison of linux containers and virtual machines using Docker and KVM. *Cluster Computing*, 22(s1):1765–1775. DOI: 10.1007/s10586-017-1511-2.

Chakareski, J. (2017). Vr/ar immersive communication: Caching, edge computing, and transmission trade-offs. In *Proceedings of the Workshop on Virtual Reality and Augmented Reality Network*, VR/AR Network '17, page 36–41, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3097895.3097902.

Chekired, D. A., Togou, M. A., Khoukhi, L., and Ksentini, A. (2019). 5G-Slicing-Enabled Scalable SDN Core Network: Toward an Ultra-Low Latency of Autonomous Driving Service. *IEEE Journal on Selected Areas in Communications*, 37(8):1769–1782. DOI: 10.1109/JSAC.2019.2927065.

Chen, M., Li, W., Fortino, G., Hao, Y., Hu, L., and Humar, I. (2019a). A dynamic service migration mechanism in edge cognitive computing. *ACM Trans. Internet Technol.*, 19(2). DOI: 10.1145/3239565.

Chen, M., Miao, Y., Jian, X., Wang, X., and Humar, I. (2019b). Cognitive-LPWAN: Towards Intelligent Wireless Services in Hybrid Low Power Wide Area

Networks. *IEEE Transactions on Green Communications and Networking*, 3(2):409–417. DOI: 10.1109/TGCN.2018.2873783.

Chen, S., Hu, J., Shi, Y., Zhao, L., and Li, W. (2020). A Vision of C-V2X: Technologies, Field Testing, and Challenges With Chinese Development. *IEEE Internet of Things Journal*, 7(5):3872–3881. DOI: 10.1109/JIOT.2020.2974823.

Cheng, Y., Xu, X., Du, Y., Guan, P., Liu, S., and Zhao, L. (2019). Design of air quality monitoring system based on nb-iot. In *2019 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*, pages 385–388. DOI: 10.1109/ICPICS47731.2019.8942533.

Choi, J.-H. and Shin, D.-J. (2019). Generalized RACH-Less Handover for Seamless Mobility in 5G and Beyond Mobile Networks. *IEEE Wireless Communications Letters*, 8(4):1264–1267. DOI: 10.1109/LWC.2019.2914435.

Cuthbertson, A. (2019). Galaxy Fold: Inside Samsung's struggle to deliver a foldable phone – and why the future of smartphones hinges on it. Available at: `https://www.independent.co.uk/life-style/gadgets-and-tech/features/samsung-galaxy-fold-foldable-phone-release-date-when-explained-a8980056.html`.

Dalgkitsis, A., Mekikis, P.-V., Antonopoulos, A., and Verikoukis, C. (2021). Data driven service orchestration for vehicular networks. *IEEE Transactions on Intelligent Transportation Systems*, 22(7):4100–4109. DOI: 10.1109/TITS.2020.3011264.

Das, D. (2018). Distributed algorithm for geographic opportunistic routing in VANETs at road intersection. *Proceedings - 2017 IEEE 15th International Conference on Dependable, Autonomic and Secure Computing, 2017 IEEE 15th International Conference on Pervasive Intelligence and Computing, 2017 IEEE 3rd International Conference on Big Data Intelligence and Computing and 2017 IEEE Cyber Science and Technology Congress, DASC-PICom-DataCom-CyberSciTec 2017*, 2018-January:1202–1209. DOI: 10.1109/DASC-PICom-DataCom-CyberSciTec.2017.192.

Di Maio, A., Soua, R., Palattella, M. R., Engel, T., and Rizzo, G. A. (2017). A centralized approach for setting floating content parameters in VANETs. In *2017 14th IEEE Annual Consumer Communications and Networking Conference, CCNC 2017*, pages 712–715. IEEE. DOI: 10.1109/CCNC.2017.7983220.

Din, I. U., Hassan, S., Khan, M. K., Guizani, M., Ghazali, O., and Habbal, A. (2018). Caching in Information-Centric Networking: Strategies, Challenges, and Future Research Directions. *IEEE Communications Surveys & Tutorials*, pages 1–1. DOI: 10.1109/COMST.2017.2787609.

Dionisio, J. D. N., III, W. G. B., and Gilbert, R. (2013). 3d virtual worlds and the metaverse: Current status and future possibilities. *ACM Comput. Surv.*, 45(3). DOI: 10.1145/2480741.2480751.

Do, T. X. and Kim, Y. (2018). Latency-aware placement for state management functions in service-based 5G mobile core network. In *2018 IEEE 7th International Conference on Communications and Electronics, ICCE 2018*, pages 102–106. Institute of Electrical and Electronics Engineers Inc.. DOI: 10.1109/CCE.2018.8465746.

Dua, R., Raja, A. R., and Kakadia, D. (2014). Virtualization vs containerization to support PaaS. *Proceedings - 2014 IEEE International Conference on Cloud Engineering, IC2E 2014*, pages 610–614. DOI: 10.1109/IC2E.2014.41.

Duarte, J. M., Braun, T., and Villas, L. A. (2019). MobiVNDN: A distributed framework to support mobility in vehicular named-data networking. *Ad Hoc Networks*, 82:77–90. DOI: 10.1016/J.ADHOC.2018.08.008.

Dupont, C., Giaffreda, R., and Capra, L. (2017). Edge computing in IoT context: Horizontal and vertical Linux container migration. *GIoTS 2017 - Global Internet of Things Summit, Proceedings*, pages 2–5. DOI: 10.1109/GIOTS.2017.8016218.

ETSI (2015). TR 22.885 - V14.0.0 - Technical Specification Group Services and System Aspects; Study on LTE support for Vehicle to Everything (V2X) services (Release 14). Technical Report Release 14, European Telecommunications Standards Institute. Available at: `https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2898`.

ETSI (2017). GR MEC 018 - V1.1.1 - Mobile edge computing: End to End Mobility Aspects. Technical report, European Telecommunications Standards Institute. Available at: `https://www.etsi.org/deliver/etsi_gr/mec/001_099/018/01.01.01_60/gr_mec018v010101p.pdf`.

ETSI (2018). Study on MEC Support for V2X Use Cases. Technical report, European Telecommunications Standards Institute. Available at: `https://www.etsi.org/deliver/etsi_gr/mec/001_099/022/02.01.01_60/gr_mec022v020101p.pdf`.

ETSI (2020). TS 123 287 - V16.3.0 - 5G; Architecture enhancements for 5G System (5GS) to support Vehicle-to-Everything (V2X) services (3GPP TS 23.287 version 16.3.0 Release 16). Technical report, European Telecommunications Standards Institute. Available at: `https://www.etsi.org/deliver/etsi_ts/123200_123299/123287/16.03.00_60/ts_123287v160300p.pdf`.

ETSI (2021). TS 102 636-4-2 - V1.4.1 - Intelligent Transport Systems (ITS); Vehicular Communications; GeoNetworking; Part 4: Geographical addressing and forwarding for point-to-point and point-to-multipoint communications; Sub-part 2: Media-dependent functionalities F. Technical report, European Telecommunications Standards Institute. Available at: `https://www.etsi.org/deliver/etsi_ts/102600_102699/1026360402/01.04.01_60/ts_1026360402v010401p.pdf`.

Fang, C., Yao, H., Wang, Z., Wu, W., Jin, X., and Yu, F. R. (2018). A Survey of Mobile Information-Centric Networking: Research Issues and Challenges. *IEEE Communications Surveys Tutorials*, 20(3):2353–2371. DOI: 10.1109/COMST.2018.2809670.

Favaloro, G. (1996). Internet Solutions Division Strategy for Cloud Computing.

Fawcett, L., Mu, M., Hareng, B., and Race, N. (2017). REF:

Enabling Rapid Experimentation of Contextual Network Traffic Management Using Software Defined Networking. *IEEE Communications Magazine*, 55(7):144–150. DOI: 10.1109/MCOM.2017.1600507.

Fernández-Caramés, T. M., Fraga-Lamas, P., Suárez-Albela, M., and Vilar-Montesinos, M. (2018). A Fog Computing and Cloudlet Based Augmented Reality System for the Industry 4.0 Shipyard. *Sensors*, 18(6):1–18. DOI: 10.3390/s18061798.

Filho, R. R. and Porter, B. (2020). Autonomous State-Management Support in Distributed Self-adaptive Systems. In *2020 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion*, pages 176–181. Institute of Electrical and Electronics Engineers (IEEE). DOI: 10.1109/acsos-c51401.2020.00052.

Fizza, K., Banerjee, A., Jayaraman, P. P., Auluck, N., Ranjan, R., Mitra, K., and Georgakopoulos, D. (2022). A survey on evaluating the quality of autonomic internet of things applications. *IEEE Communications Surveys and Tutorials*. DOI: 10.1109/COMST.2022.3205377.

Fotiou, N., Trossen, D., and Polyzos, G. C. (2012). Illustrating a publish-subscribe Internet architecture. *Telecommunication Systems*, 51(4):233–245. DOI: 10.1007/s11235-011-9432-5.

Fraga-Lamas, P., Fernández-Caramés, T. M., Blanco-Novoa, Ó., and Vilar-Montesinos, M. A. (2018). A Review on Industrial Augmented Reality Systems for the Industry 4.0 Shipyard. *IEEE Access*, 6:13358–13375. DOI: 10.1109/ACCESS.2018.2808326.

Frank, B., Poese, I., Lin, Y., Smaragdakis, G., Feldmann, A., Maggs, B., Rake, J., Uhlig, S., and Weber, R. (2013). Pushing CDN-ISP collaboration to the limit. *ACM SIGCOMM Computer Communication Review*, 43(3):34. DOI: 10.1145/2500098.2500103.

Füßler, H., Widmer, J., Käsemann, M., Mauve, M., and Hartenstein, H. (2003). Contention-based forwarding for mobile ad hoc networks. *Ad Hoc Networks*, 1(4):351–369. DOI: 10.1016/S1570-8705(03)00038-6.

Gasparyan, M., Braun, T., and Schiller, E. (2017a). L-SCN: Layered SCN architecture with supernodes and Bloom filters. In *2017 14th IEEE Annual Consumer Communications Networking Conference (CCNC)*, pages 899–904. DOI: 10.1109/CCNC.2017.7983252.

Gasparyan, M., Corsini, G., Braun, T., Schiller, E., and Saltarin, J. (2017b). Session support for SCN. In *2017 IFIP Networking Conference (IFIP Networking) and Workshops*, pages 1–6. IEEE. DOI: 10.23919/IFIPNetworking.2017.8264879.

Gasparyan, M., Marandi, A., Schiller, E., and Braun, T. (2019). Fault-Tolerant Session Support for Service-Centric Networking. In *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, pages 312–320.

Gelabert, X., Zhou, G., and Legg, P. (2013). Mobility performance and suitability of macro cell power-off in lte dense small cell hetnets. In *2013 IEEE 18th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pages 99–103. DOI: 10.1109/CAMAD.2013.6708097.

Grau, C., Ginhoux, R., Riera, A., Nguyen, T. L., Chauvat, H., Berg, M., Amengual, J. L., Pascual-Leone, A., and Ruffini, G. (2014). Conscious Brain-to-Brain Communication in Humans Using Non-Invasive Technologies. *PLOS ONE*, 9(8):1–6. DOI: 10.1371/journal.pone.0105225.

GreenICN Project (2013). GreenICN: Architecture and Applications of Green Information Centric Networking. Available at: `http://www.greenicn.org/`.

Guibert, D., Wu, J., He, S., Wang, M., and Li, J. (2017). CC-fog: Toward content-centric fog networks for E-health. In *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pages 1–5. IEEE. DOI: 10.1109/HealthCom.2017.8210830.

Gupta, H. and Ramachandran, U. (2018). Fogstore: A geo-distributed key-value store guaranteeing low latency for strongly consistent access. In *Proceedings of the 12th ACM International Conference on Distributed and Event-Based Systems*, DEBS '18, page 148–159, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3210284.3210297.

Gutz, S., Story, A., Schlesinger, C., and Foster, N. (2012). Splendid Isolation: A Slice Abstraction for Software-defined Networks. In *Proceedings of the First Workshop on Hot Topics in Software Defined Networks*, HotSDN '12, pages 79–84, New York, NY, USA. ACM. DOI: 10.1145/2342441.2342458.

Hafner, B., Bajpai, V., Ott, J., and Schmitt, G. A. (2021). A survey on cooperative architectures and maneuvers for connected and automated vehicles. *IEEE Communications Surveys and Tutorials*, 24:380–403. DOI: 10.1109/COMST.2021.3138275.

Hagihara, R., Yamasaki, Y., and Ohsaki, H. (2017). On delivery control for floating contents sharing with epidemic broadcasting. *2017 14th IEEE Annual Consumer Communications and Networking Conference, CCNC 2017*, 1(2):353–356. DOI: 10.1109/CCNC.2017.7983133.

Han, B. (2019). Mobile Immersive Computing: Research Challenges and the Road Ahead. *IEEE Communications Magazine*, 57(10):112–118. DOI: 10.1109/MCOM.001.1800876.

Han, D., Shin, S., Cho, H., Chung, J., Ok, D., and Hwang, I. (2015). Measurement and stochastic modeling of handover delay and interruption time of smartphone real-time applications on lte networks. *IEEE Communications Magazine*, 53(3):173–181. DOI: 10.1109/MCOM.2015.7060501.

Han, W., Dong, M., Ota, K., Wu, J., Li, J., and Li, G. (2017). SD-OPTS: Software-Defined On-Path Time Synchronization for Information-Centric Smart Grid. In *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, pages 1–6. DOI: 10.1109/GLOCOM.2017.8254517.

He, Z., Zhang, T., and Lee, R. B. (2019). Model inversion attacks against collaborative inference. In *Proceedings of the 35th Annual Computer Security Applications Conference*, ACSAC '19, page 148–162, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3359789.3359824.

Heissenbüttel, M., Braun, T., Bernoulli, T., and Wälchli, M. (2004). BLR: Beacon-less routing algorithm for mobile ad hoc networks. *Computer Communications*, 27(11):1076–1086. DOI: 10.1016/j.comcom.2004.01.012.

Henderson, T., Vogt, C., and Arkko, J. (2017). Host mobility with the host identity protocol. Technical report, Internet Engineering Task Force. Available at:`https://www.rfc-editor.org/rfc/rfc8046`.

Ho, Y., Chun-Han Lin, Shih-Jie Sun, and Ling-Jyh Chen (2016). Emergency application for Vehicle-to-Vehicle Communication using Named Data Networking (eVNDN). In *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*, pages 611–616. DOI: 10.1109/ICUFN.2016.7537105.

Hofer, F., Sehr, M. A., Iannopollo, A., Ugalde, I., Sangiovanni-Vincentelli, A. L., and Russo, B. (2019). Industrial Control via Application Containers: Migrating from Bare-Metal to {IAAS}. *arXiv*, abs/1908.0. DOI: 10.48550/arXiv.1908.04465.

Hoggenmueller, M. and Tomitsch, M. (2019). Enhancing Pedestrian Safety Through In-situ Projections: A Hyperreal Design Approach. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays*, PerDis '19, pages 29:1—-29:2, New York, NY, USA. ACM. DOI: 10.1145/3321335.3329682.

Hu, C., Bao, W., Wang, D., and Liu, F. (2019a). Dynamic Adaptive DNN Surgery for Inference Acceleration on the Edge. In *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pages 1423–1431. DOI: 10.1109/INFOCOM.2019.8737614.

Hu, L., Tian, Y., Yang, J., Taleb, T., Xiang, L., and Hao, Y. (2019b). Ready Player One: UAV-Clustering-Based Multi-Task Offloading for Vehicular VR/AR Gaming. *IEEE Network*, 33(3):42–48. DOI: 10.1109/MNET.2019.1800357.

Huang, W., Song, T., Yang, Y., and Zhang, Y. (2019). Cluster-Based Cooperative Caching With Mobility Prediction in Vehicular Named Data Networking. *IEEE Access*, 7:23442–23458. DOI: 10.1109/ACCESS.2019.2897747.

ICN 2020 Project (2016). ICN2020: Advancing ICN towards real-world deployment through research, innovative applications, and global scale experimentation. Available at:`http://www.icn2020.org/`.

Jacobson, V., Smetters, D. K., Thornton, J. D., Plass, M. F., Briggs, N. H., and Braynard, R. L. (2009). Networking named content. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies - CoNEXT '09*, page 1, New York, New York, USA. ACM Press. DOI: 10.1145/1658939.1658941.

Jain, R., Puri, A., and Sengupta, R. (2001). Geographical Routing Using Partial Information for Wireless Ad Hoc Networks. *IEEE Personal Communications*, pages 48–57. DOI: 10.1109/98.904899.

Jamakovic, A., Anwander, M., Braun, T., Kropf, P., Schiller, E., Schwanbeck, J., and Staub, T. (2012). A4-mesh: Connecting remote sites. *SWITCH Journal*, pages 15–17. Available at: `https://libra.unine.ch/handle/123456789/22847`.

Jeong, S., Simeone, O., and Kang, J. (2018). Mobile Edge Computing via a UAV-Mounted Cloudlet: Optimization of Bit Allocation and Path Planning. *IEEE Transactions on Vehicular Technology*, 67(3):2049–2063. DOI: 10.1109/TVT.2017.2706308.

Jiang, H., Guo, Y., Wu, Z., Zhang, C., Jia, W., and Wang, Z. (2018). Implantable wireless intracranial pressure monitoring based on air pressure sensing. *IEEE Transactions on Biomedical Circuits and Systems*, 12(5):1076–1087. DOI: 10.1109/TBCAS.2018.2845462.

Jiang, R., Song, X., Fan, Z., Xia, T., Chen, Q., Miyazawa, S., and Shibasaki, R. (2018). DeepUrbanMomentum: An Online Deep-Learning System for Short-Term Urban Mobility Prediction. In *Procedings of AAAI Conference on Artificial Intelligence*, pages 784–791. DOI: 10.1609/aaai.v32i1.11338.

Jin, Y. and Wen, Y. (2017). When Cloud Media Meet Network Function Virtualization: Challenges and Applications. *IEEE Multimedia*, 24(3):72–82. DOI: 10.1109/MMUL.2017.3051519.

Kalatzis, N., Avgeris, M., Dechouniotis, D., Papadakis-Vlachopapadopoulos, K., Roussaki, I., and Papavassiliou, S. (2018). Edge Computing in IoT Ecosystems for UAV-Enabled Early Fire Detection. In *2018 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 106–114. DOI: 10.1109/SMARTCOMP.2018.00080.

Kalogeiton, E. and Braun, T. (2018). Infrastructure-Assisted Communication for NDN-VANETs. *19th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WoWMoM 2018*. DOI: 10.1109/WoWMoM.2018.8449740.

Kataoka, K., Yamamoto, T., Otsuki, M., Shibata, F., and Kimura, A. (2019). A new interactive haptic device for getting physical contact feeling of virtual objects. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1323–1324. DOI: 10.1109/VR.2019.8797762.

Katsaros, K. V., Chai, W. K., Wang, N., Pavlou, G., Bontius, H., and Paolone, M. (2014). Information-centric networking for machine-to-machine data delivery: a case study in smart grid applications. *IEEE Network*, 28(3):58–64. DOI: 10.1109/MNET.2014.6843233.

Kaur, K., Dhand, T., Kumar, N., and Zeadally, S. (2017). Container-as-a-Service at the Edge: Trade-off between Energy Efficiency and Service Availability at Fog Nano Data Centers. *IEEE Wireless Communications*, 24(3):48–56. DOI: 10.1109/MWC.2017.1600427.

Kaur, K., Garg, S., Aujla, G. S., Kumar, N., Rodrigues, J. J., and Guizani, M. (2018). Edge Computing in the Industrial Internet of Things Environment: Software-Defined-Networks-Based Edge-Cloud Interplay. *IEEE Communications Magazine*, 56(2):44–51. DOI: 10.1109/MCOM.2018.1700622.

Kellokoski, J., Koskinen, J., Rusanen, T., Kalliolahti, P., and Hämäläinen, T. (2013). Proxy Mobile IPv6-Based Seamless Handover. In Balandin, S., Andreev, S., and Koucheryavy, Y., editors, *Internet of Things, Smart Spaces, and Next Generation Networking*, pages 214–223, Berlin, Heidelberg. Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-40316-3_19.

Khaliq, K. A., Chughtai, O., Shahwani, A., Qayyum, A., and Pannek, J. (2019). Road Accidents Detection, Data Collection and Data Analysis Using V2X Communication and Edge/Cloud Computing. *Electronics*, 8(8). DOI: 10.3390/electronics8080896.

Khan, A., Abolhasan, M., Ni, W., Lipman, J., and Jamalipour, A. (2019). A Hybrid-Fuzzy Logic Guided Genetic Algorithm (H-FLGA) Approach for Resource Optimization in 5G VANETs. *IEEE Transactions on Vehicular Technology*, 68(7):1–1. DOI: 10.1109/tvt.2019.2915194.

Khan, I., Belqasmi, F., Glitho, R., Crespi, N., Morrow, M., and Polakos, P. (2016). Wireless sensor network virtualization: A survey. *IEEE Communications Surveys Tutorials*, 18(1):553–576. DOI: 10.1109/COMST.2015.2412971.

Khan, L. U., Han, Z., Saad, W., Hossain, E., Guizani, M., and Hong, C. S. (2022). Digital twin of wireless systems: Overview, taxonomy, challenges, and opportunities. *IEEE Communications Surveys and Tutorials*, 24:2230–2254. DOI: 10.1109/COMST.2022.3198273.

Kim, H., Shin, H., Kim, H. S., and Kim, W. T. (2018a). VR-CPES: A novel cyber-physical education systems for interactive VR services based on a mobile platform. *Mobile Information Systems*, 2018:10. DOI: 10.1155/2018/8941241.

Kim, S. J., Jeong, Y., Park, S., Ryu, K., and Oh, G. (2018b). A Survey of Drone use for Entertainment and AVR (Augmented and Virtual Reality). In Jung, T. and tom Dieck, M. C., editors, *Augmented Reality and Virtual Reality: Empowering Human, Place and Business*, pages 339–352. Springer International Publishing. DOI: 10.1007/978-3-319-64027-$3_2$3.

Koponen, T., Kim, K. H., Shenker, S., Chawla, M., Chun, B.-G., Ermolinskiy, A., Kim, K. H., Shenker, S., and Stoica, I. (2007). A Data-oriented (and Beyond) Network Architecture. *SIGCOMM Comput. Commun. Rev.*, 37(4):181–192. DOI: 10.1145/1282427.1282402.

Kuzlu, M., Pipattanasomporn, M., and Rahman, S. (2014). Communication network requirements for major smart grid applications in HAN, NAN and WAN. *Computer Networks*, 67:74–88. DOI: 10.1016/j.comnet.2014.03.029.

Laghrissi, A. and Taleb, T. (2019). A Survey on the Placement of Virtual Resources and Virtual Network Functions. *IEEE Communications Surveys & Tutorials*, 21(2):1409–1434. DOI: 10.1109/COMST.2018.2884835.

Lal, K. N. and Kumar, A. (2017). E-health application using network coding based caching for information-centric networking (icn). In *2017 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pages 427–431. DOI: 10.1109/ICRITO.2017.8342464.

Lan, D., Taherkordi, A., Eliassen, F., Chen, Z., and Liu, L. (2020). Deep reinforcement learning for intelligent migration of fog services in smart cities. In Qiu, M., editor, *Algorithms and Architectures for Parallel Processing*, pages 230–244, Cham. Springer International Publishing. DOI: 10.1007/978-3-030-60239-$0_1$6.

Le, N. T., Hossain, M. A., Islam, A., Kim, D.-Y., Choi, Y.-J., and Jang, Y. M. (2017). Survey of Promising Technologies for 5G Networks. *Mobile Information Systems*, 2016:25. DOI: 10.48550/arXiv.1701.05234.

Lee, E. K., Gerla, M., Pau, G., Lee, U., and Lim, J. H. (2016). Internet of Vehicles: From intelligent grid to autonomous cars and vehicular fogs. *International Journal of Distributed Sensor Networks*, 12(9):0–14. DOI: 10.1177/1550147716665500.

Lee, J., Bonnin, J., Seite, P., and Chan, H. A. (2013). Distributed IP mobility management from the perspective of the IETF: motivations, requirements, approaches, comparison, and challenges. *IEEE Wireless Communications*, 20(5):159–168. DOI: 10.1109/MWC.2013.6664487.

Lei, K., Zhang, Q., Lou, J., Bai, B., and Xu, K. (2019). Securing ICN-Based UAV Ad Hoc Networks with Blockchain. *IEEE Communications Magazine*, 57(6):26–32. DOI: 10.1109/MCOM.2019.1800722.

Leontiadis, I. and Mascolo, C. (2007). GeOpps: Geographical opportunistic routing for vehicular networks. In *2007 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WOWMOM*. DOI: 10.1109/WOWMOM.2007.4351688.

Li, C., Luo, Q., Mao, G., Sheng, M., and Li, J. (2019). Vehicle-Mounted Base Station for Connected and Autonomous Vehicles: Opportunities and Challenges. *IEEE Wireless Communications*, 26(4):30–36. DOI: 10.1109/MWC.2019.1800541.

Li, G., Wu, J., Li, J., Ye, T., Morello, R., Member, S., Wu, J., and Li, J. (2017a). Battery Status Sensing Software-Defined Multicast for V2G Regulation in Smart Grid. *IEEE Sensors Journal*, 17(23):7838–7848. DOI: 10.1109/JSEN.2017.2731971.

Li, H., Ota, K., Dong, M., and Guo, M. (2017b). Mobile Crowdsensing in Software Defined Opportunistic Networks. *IEEE Communications Magazine*, 55(6):140–145. DOI: 10.1109/MCOM.2017.1600719.

Li, J., Shvartzshnaider, Y., Francisco, J.-a., Martin, R. P., Nagaraj, K., and Raychaudhuri, D. (2012). Delivering Internet-of-Things Services in MobilityFirst Future Internet Architecture. In *2012 3rd IEEE International Conference on the Internet of Things*, pages 31–38. DOI: 10.1109/IOT.2012.6402301.

Li, R., Lai, D. T. H., and Lee, W. (2017). A survey on biofeedback and actuation in wireless body area networks (wbans). *IEEE Reviews in Biomedical Engineering*, 10:162–173. DOI: 10.1109/RBME.2017.2738009.

Li, S., Zhang, Y., Raychaudhuri, D., and Ravindran, R. (2014). A comparative study of MobilityFirst and NDN based ICN-IoT architectures. *Proceedings of the 2014 10th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, QSHINE 2014*, 1:158–163. DOI: 10.1109/QSHINE.2014.6928680.

Li, T. M., Liao, C. C., Cho, H. H., Chien, W. C., Lai, C. F., and Chao, H. C. (2017). An e-healthcare sensor network load-balancing scheme using SDN-SFC. In *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services, Healthcom 2017*, volume 2017-December, pages 1–4. Institute of Electrical and Electronics Engineers Inc.. DOI: 10.1109/Health-

Com.2017.8210833.

Li, X., Li, D., Wan, J., Liu, C., and Imran, M. (2018). Adaptive transmission optimization in sdn-based industrial internet of things with edge computing. *IEEE Internet of Things Journal*, 5(3):1351–1360. DOI: 10.1109/JIOT.2018.2797187.

Li, Y. and Gao, W. (2017). Minimizing Context Migration in Mobile Code Offload. *IEEE Transactions on Mobile Computing*, 16(4):1005–1018. DOI: 10.1109/TMC.2016.2586056.

Liao, Y., Qiao, X., Shou, L., Zhai, X., Ai, Q., Yu, Q., and Liu, Q. (2019). Caching-Aided Task Offloading Scheme for Wireless Body Area Networks with MEC. In *2019 NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*, pages 49–54. IEEE. DOI: 10.1109/AHS.2019.000-1.

Lim, W. Y. B., Luong, N. C., Hoang, D. T., Jiao, Y., Liang, Y.-C., Yang, Q., Niyato, D., and Miao, C. (2020). Federated learning in mobile edge networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 22(3):2031–2063. DOI: 10.1109/COMST.2020.2986024.

Liu, B., Jiang, T., Wang, Z., and Cao, Y. (2017). Object-Oriented Network: A Named-Data Architecture Toward the Future Internet. *IEEE Internet of Things Journal*, 4(4):957–967. DOI: 10.1109/JIOT.2017.2714658.

Liu, D., Wang, L., Chen, Y., Elkashlan, M., Wong, K. K., Schober, R., and Hanzo, L. (2016a). User Association in 5G Networks: A Survey and an Outlook. *IEEE Communications Surveys and Tutorials*, 18(2):1018–1044. DOI: 10.1109/COMST.2016.2516538.

Liu, D., Zúñiga, J. C., Seite, P., Chan, H. A., and Bernardos, C. J. (2015). Distributed mobility management: Current practices and gap analysis. Technical report, Internet Engineering Task Force. Available at: `https://www.rfc-editor.org/rfc/rfc7429`.

Liu, F., Shu, P., Jin, H., Ding, L., Yu, J., Niu, D., and Li, B. (2013). Gearing resource-poor mobile devices with powerful clouds: architectures, challenges, and applications. *IEEE Wireless Communications*, 20(3):14–22. DOI: 10.1109/MWC.2013.6549279.

Liu, J., Li, J., Zhang, L., Dai, F., Zhang, Y., Meng, X., and Shen, J. (2018). Secure intelligent traffic light control using fog computing. *Future Generation Computer Systems*, 78:817–824. DOI: 10.1016/J.FUTURE.2017.02.017.

Liu, J. and Zhang, Q. (2019). Code-partitioning offloading schemes in mobile edge computing for augmented reality. *IEEE Access*, 7:11222–11236. DOI: 10.1109/ACCESS.2019.2891113.

Liu, X., João Nicolau, M., Costa, A., Macedo, J., and Santos, A. (2016b). A Geographic Opportunistic Forwarding Strategy for Vehicular Named Data Networking. In Novais, P., Camacho, D., Analide, C., El Fallah Seghrouchni, A., and Badica, C., editors, *Intelligent Distributed Computing IX*, pages 509–521, Cham. Springer International Publishing. DOI: 10.1007/978-3-319-25017-$5_4$8.

Liu, Y., Liu, X., Mu, X., Hou, T., Xu, J., Di Renzo, M., and Al-Dhahir, N. (2021a). Reconfigurable intelligent surfaces: Principles and opportunities. *IEEE Communications Surveys & Tutorials*, 23(3):1546–1577. DOI: 10.1109/COMST.2021.3077737.

Liu, Y., Lu, H., Li, X., Zhang, Y., Xi, L., and Zhao, D. (2021b). Dynamic service function chain orchestration for nfv/mec-enabled iot networks: A deep reinforcement learning approach. *IEEE Internet of Things Journal*, 8(9):7450–7465. DOI: 10.1109/JIOT.2020.3038793.

Longo, A., De Matteis, A., and Zappatore, M. (2018). Urban Pollution Monitoring Based on Mobile Crowd Sensing: An Osmotic Computing Approach. In *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*, pages 380–387. DOI: 10.1109/CIC.2018.00057.

Ma, L., Yi, S., Carter, N., and Li, Q. (2018). Efficient Live Migration of Edge Services Leveraging Container Layered Storage. *IEEE Transactions on Mobile Computing*, 18(9):2020–2033. DOI: 10.1109/TMC.2018.2871842.

Ma, Z., Xiao, M., Xiao, Y., Pang, Z., Poor, H. V., and Vucetic, B. (2019). High-Reliability and Low-Latency Wireless Communication for Internet of Things: Challenges, Fundamentals, and Enabling Technologies. *IEEE Internet of Things Journal*, 6(5):7946–7970. DOI: 10.1109/JIOT.2019.2907245.

Manco, F., Lupu, C., Schmidt, F., Mendes, J., Kuenzer, S., Sati, S., Yasukata, K., Raiciu, C., and Huici, F. (2017). My vm is lighter (and safer) than your container. In *Proceedings of the 26th Symposium on Operating Systems Principles*, SOSP '17, pages 218–233, New York, NY, USA. ACM. DOI: 10.1145/3132747.3132763.

Mao, Y., You, C., Zhang, J., Huang, K., and Letaief, K. B. (2017a). A Survey on Mobile Edge Computing: The Communication Perspective. *IEEE Communications Surveys and Tutorials*, 19(4):2322–2358. DOI: 10.1109/COMST.2017.2745201.

Mao, Y., You, C., Zhang, J., Huang, K., and Letaief, K. B. (2017b). A Survey on Mobile Edge Computing: The Communication Perspective. *IEEE Communications Surveys & Tutorials*, 19(4):2322–2358. DOI: 10.1109/COMST.2017.2745201.

Martins, N. R. B., Angelica, A., Chakravarthy, K., Svidinenko, Y., Boehm, F. J., Opris, I., Lebedev, M. A., Swan, M., Garan, S. A., Rosenfeld, J. V., Hogg, T., and Freitas, R. A. (2019). Human Brain/Cloud Interface. *Frontiers in Neuroscience*, 13:112. DOI: 10.3389/fnins.2019.00112.

Matsushima, K. and Sonobe, N. (2018). Full-color digitized holography for large-scale holographic 3D imaging of physical and nonphysical objects. *Appl. Opt.*, 57(1):A150—-A156. DOI: 10.1364/AO.57.00A150.

McKeown, N., Anderson, T., Balakrishnan, H., Parulkar, G., Peterson, L., Rexford, J., Shenker, S., and Turner, J. (2008). Openflow: Enabling innovation in campus networks. *SIGCOMM Comput. Commun. Rev.*, 38(2):69–74. DOI: 10.1145/1355734.1355746.

Mehrabi, M., You, D., Latzko, V., Salah, H., Reisslein, M., and Fitzek, F. H. (2019). Device-Enhanced MEC: Multi-Access Edge Computing (MEC) Aided by End Device Computation and Caching: A Survey. *IEEE Access*, 7:166079–166108. DOI: 10.1109/ACCESS.2019.2953172.

Meijerink, B. and Heijenk, G. (2019). Infrastructure support for contention-based forwarding. In *2019 IEEE Vehicular Networking Conference (VNC)*, pages 1–8. DOI: 10.1109/VNC48660.2019.9062795.

Meneguette, R. I., Rodrigues, D. O., Costa, J. B. D., Rosário, D., and Villas, L. A. (2019). A Virtual Machine Migration Policy Based on Multiple Attribute Decision in Vehicular Cloud Scenario. In *Proceeddings of The IEEE International Conference on Communications*. DOI: 10.1109/ICC.2019.8761248.

Metzger, F., Geisler, S., Grigorjew, A., Loh, F., Moldovan, C., Seufert, M., and Hosfeld, T. (2022). An introduction to online video game qos and qoe influencing factors. *IEEE Communications Surveys and Tutorials*, 24:1894–1925. DOI: 10.1109/COMST.2022.3177251.

Miao, Y., Jiang, Y., Peng, L., Hossain, M. S., and Muhammad, G. (2018). Telesurgery Robot Based on 5G Tactile Internet. *Mobile Networks and Applications*, 23(6):1645–1654. DOI: 10.1007/s11036-018-1110-3.

Monga, S. K., R, S. K., and Simmhan, Y. (2019). Elfstore: A resilient data storage service for federated edge and fog resources. *CoRR*, abs/1905.08932. DOI: 10.48550/arXiv.1905.08932.

Monsanto, C., Reich, J., Foster, N., Rexford, J., and Walker, D. (2013). Composing Software-defined Networks. In *Proceedings of the 10th USENIX Conference on Networked Systems Design and Implementation*, nsdi'13, pages 1–14, Berkeley, CA, USA. USENIX Association. DOI: 10.5555/2482626.2482629.

Mosko, M., Uzun, E., and Wood, C. A. (2017). Mobile sessions in content-centric networks. *2017 IFIP Networking Conference, IFIP Networking 2017 and Workshops*, 2018-January(2):1–9. DOI: 10.23919/IFIPNetworking.2017.8264861.

Moskowitz, R., Heer, T., Jokela, P., and Henderson, T. (2015). Host identity protocol version 2 (hipv2). Technical report, Internet Engineering Task Force. Available at:`https://www.rfc-editor.org/rfc/rfc7401.html`.

Moskowitz, R., Nikander, P., Jokela, P., and Henderson, T. (2008). Host identity protocol. Technical report, Internet Engineering Task Force. Available at: `https://www.rfc-editor.org/rfc/rfc5201`.

Naha, R. K., Garg, S., Georgakopoulos, D., Jayaraman, P. P., Gao, L., Xiang, Y., and Ranjan, R. (2018). Fog Computing: Survey of Trends, Architectures, Requirements, and Research Directions. *IEEE Access*, 6:47980–48009. DOI: 10.1109/ACCESS.2018.2866491.

Navarro-Ortiz, J., Romero-Diaz, P., Sendra, S., Ameigeiras, P., Ramos-Munoz, J. J., and Lopez-Soler, J. M. (2020). A survey on 5g usage scenarios and traffic models. *IEEE Communications Surveys & Tutorials*, 22:905–929. DOI: 10.1109/COMST.2020.2971781.

NDN Project (2010). Named Data Networking (NDN) - A Future Internet Architecture. Available at: `https://named-data.net/`.

Nguyen, D., Shen, Z., Jin, J., and Tagami, A. (2017a). ICN-Fog: An Information-Centric Fog-to-Fog Architecture for Data Communications. In *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, pages 1–6. IEEE. DOI: 10.1109/GLOCOM.2017.8254724.

Nguyen, V. G., Brunstrom, A., Grinnemo, K. J., and Taheri, J. (2017b). SDN/NFV-Based Mobile Packet Core Network Architectures: A Survey. *IEEE Communications Surveys and Tutorials*, 19(3):1567–1602. DOI: 10.1109/COMST.2017.2690823.

Nguyen Gia, T., Rahmani, A. M., Westerlund, T., Liljeberg, P., and Tenhunen, H. (2018). Fog computing approach for mobility support in internet-of-things systems. *IEEE Access*, 6:36064–36082. DOI: 10.1109/ACCESS.2018.2848119.

Ning, Z., Huang, J., and Wang, X. (2019). Vehicular Fog Computing: Enabling Real-Time Traffic Management for Smart Cities. *IEEE Wireless Communications*, 26(1):87–93. DOI: 10.1109/MWC.2019.1700441.

Nobre, J. C., de Souza, A. M., Rosário, D., Both, C., Villas, L. A., Cerqueira, E., Braun, T., and Gerla, M. (2019). Vehicular software-defined networking and fog computing: Integration and design principles. *Ad Hoc Networks*, 82:172–181. DOI: 10.1016/J.ADHOC.2018.07.016.

Oteafy, S. M. and Hassanein, H. S. (2019). Leveraging tactile internet cognizance and operation via IoT and edge technologies. *Proceedings of the IEEE*, 107(2):364–375. DOI: 10.1109/JPROC.2018.2873577.

Pamboris, A., Andreou, P., Polycarpou, I., and Samaras, G. (2019). FogFS: A Fog File System For Hyper-Responsive Mobile Applications. In *2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pages 1–6. IEEE. DOI: 10.1109/CCNC.2019.8651807.

Peng, H., Ye, Q., and Shen, X. (2019). SDN-Based Resource Management for Autonomous Vehicular Networks: A Multi-Access Edge Computing Approach. *IEEE Wireless Communications*, pages 1–7. DOI: 10.1109/MWC.2019.1800371.

Perkins, C. E. and Microsystems, S. (1997). Mobile IP. *IEEE Communications Magazine*, 35(5):84–99. DOI: 10.1109/35.592101.

Peuster, M., Kuttner, H., and Karl, H. (2018). Let the state follow its flows: An SDN-based flow handover protocol to support state migration. In *2018 4th IEEE Conference on Network Softwarization and Workshops, NetSoft 2018*, pages 406–414. Institute of Electrical and Electronics Engineers Inc.. DOI: 10.1109/NETSOFT.2018.8460007.

Priyadarsini, M. and Bera, P. (2019). A Secure Virtual Controller for Traffic Management in SDN. *IEEE Letters of the Computer Society*, page 1. DOI: 10.1109/LOCS.2019.2933418.

Promwongsa, N., Ebrahimzadeh, A., Naboulsi, D., Kianpisheh, S., Belqasmi, F., Glitho, R., Crespi, N., and Alfandi, O. (2021). A comprehensive survey of the tactile internet: State-of-the-art and research directions. *IEEE Communications Surveys & Tutorials*, 23(1):472–523. DOI: 10.1109/COMST.2020.3025995.

Qadri, Y. A., Nauman, A., Zikria, Y. B., Vasilakos, A. V., and Kim, S. W. (2020). The future of healthcare internet of things: A survey of emerging technologies. *IEEE Communications Surveys and Tutorials*, 22:1121–1167. DOI: 10.1109/COMST.2020.2973314.

Qiao, Y., Si, Z., Zhang, Y., Abdesslem, F. B., Zhang, X., and Yang, J. (2018). A hybrid Markov-based model for human mobility prediction. *Neurocomputing*, 278:99–109. DOI: 10.1016/j.neucom.2017.05.101.

Quan, W., Guan, J., Jiang, Z., and Zhang, H. (2015). I-wban: Enabling information-centric data retrieval in heterogeneous wban. In *2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 138–143. DOI: 10.1109/INFCOMW.2015.7179374.

Rahman, S., Kim, G., Cho, Y., and Khan, A. (2018). Positioning of UAVs for throughput maximization in software-defined disaster area UAV communication networks. *Journal of Communications and Networks*, 20(5):452–463. DOI: 10.1109/JCN.2018.000070.

Rao, B., Gopi, A. G., and Maione, R. (2016). The societal impact of commercial drones. *Technology in Society*, 45:83 – 90. DOI: 10.1016/j.techsoc.2016.02.009.

Rawat, D. B., Member, S., and Reddy, S. R. (2017). Software Defined Networking Architecture , Security and Energy Efficiency : A Survey. *IEEE Communications Surveys & Tutorials*, 19(1):325–346. DOI: 10.1109/COMST.2016.2618874.

Rejiba, Z., Masip-Bruin, X., and Marín-Tordera, E. (2019). A survey on mobility-induced service migration in the fog, edge, and related computing paradigms. *ACM Computing Surveys*, 52(5). DOI: 10.1145/3326540.

Rodrigues, D. O., Boukerche, A., Silva, T. H., Loureiro, A. A., and Villas, L. A. (2018). Combining taxi and social media data to explore urban mobility issues. *Computer Communications*. DOI: 10.1016/J.COMCOM.2018.10.004.

Roman, R., Lopez, J., and Mambo, M. (2018). Mobile edge computing, Fog et al.: A survey and analysis of security threats and challenges. *Future Generation Computer Systems*, 78:680–698. DOI: 10.1016/J.FUTURE.2016.11.009.

Sadeghi, K., Banerjee, A., Sohankar, J., and Gupta, S. K. S. (2016). Optimization of Brain Mobile Interface Applications Using IoT. In *2016 IEEE 23rd International Conference on High Performance Computing (HiPC)*, pages 32–41. DOI: 10.1109/HiPC.2016.014.

Salman, O., Elhajj, I., Chehab, A., and Kayssi, A. (2018). IoT survey: An SDN and fog computing perspective. *Computer Networks*, 143:221–246. DOI: 10.1016/J.COMNET.2018.07.020.

Sami, H., Mourad, A., Otrok, H., and Bentahar, J. (2020). Fscaler: Automatic resource scaling of containers in fog clusters using reinforcement learning. In *2020 International Wireless Communications and Mobile Computing (IWCMC)*, pages 1824–1829. DOI: 10.1109/IWCMC48107.2020.9148401.

Samikwa, E., Maio, A. D., and Braun, T. (2022). Ares: Adaptive resource-aware split learning for internet of things. *Computer Networks*, 218:109380. DOI: 10.1016/j.comnet.2022.109380.

Santagati, G. E. and Melodia, T. (2017). Experimental evaluation of impulsive ultrasonic intra-body communications for implantable biomedical devices. *IEEE Transactions on Mobile Computing*, 16(2):367–380. DOI: 10.1109/TMC.2016.2561277.

Santos, J., Wauters, T., Volckaert, B., and Turck, F. D. (2021). Towards low-latency service delivery in a continuum of virtual resources: State-of-the-art and research directions. *IEEE Communications Surveys and Tutorials*, 23:2557–2589. DOI: 10.1109/COMST.2021.3095358.

Saraswat, S., Agarwal, V., Gupta, H. P., Mishra, R., and Gupta, A. (2019). Journal of Network and Computer Applications Challenges and solutions in Software Defined Networking : A survey. *Journal of Network and Computer Applications*, 141(April):23–58. DOI: 10.1016/j.jnca.2019.04.020.

Satyanarayanan, M. (2017). The emergence of edge computing. *Computer*, 50(1):30–39. DOI: 10.1109/MC.2017.9.

Schneider, B. and Blikstein, P. (2018). Tangible User Interfaces and Contrasting Cases as a Preparation for Future Learning. *Journal of Science Education and Technology*, 27(4):369–384. DOI: 10.1007/s10956-018-9730-8.

Schneider, M., Rambach, J., and Stricker, D. (2017). Augmented reality based on edge computing using the example of remote live support. In *Proceedings of the IEEE International Conference on Industrial Technology*, pages 1277–1282. IEEE. DOI: 10.1109/ICIT.2017.7915547.

Schulz, P., Matthe, M., Klessig, H., Simsek, M., Fettweis, G., Ansari, J., Ashraf, S. A., Almeroth, B., Voigt, J., Riedel, I., Puschmann, A., Mitschele-Thiel, A., Muller, M., Elste, T., and Windisch, M. (2017). Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture. *IEEE Communications Magazine*, 55(2):70–78. DOI: 10.1109/MCOM.2017.1600435CM.

Sedjelmaci, H., Messous, M. A., Senouci, S. M., and Brahmi, I. H. (2019). Toward a lightweight and efficient UAV-aided VANET. *Transactions on Emerging Telecommunications Technologies*, 30(8):e3520. DOI: 10.1002/ett.3520.

Seliem, H., Shahidi, R., Ahmed, M. H., and Shehata, M. S. (2018). Drone-based highway-VANET and das service. *IEEE Access*, 6:20125–20137. DOI: 10.1109/ACCESS.2018.2824839.

Shah, S. S., Ali, M., Malik, A. W., Khan, M. A., and Ravana, S. D. (2019). VFog: A Vehicle-Assisted Computing Framework for Delay-Sensitive Applications in Smart Cities. *IEEE Access*, 7:34900–34909. DOI: 10.1109/ACCESS.2019.2903302.

Shen, D., Chen, W., Liu, P., and Wang, Z. (2022). Nb-iot-based environmental monitoring system for peach orchard. In *2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, pages 1–4. DOI: 10.1109/ICBAIE56435.2022.9985875.

Shi, J., Du, J., Wang, J., and Yuan, J. (2020). Distributed v2v computation offloading based on dynamic pricing using deep reinforcement learning. In *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. DOI: 10.1109/WCNC45663.2020.9120816.

Shi, W., Cao, J., Zhang, Q., Li, Y., and Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5):637–646. DOI:

10.1109/JIOT.2016.2579198.

Simoens, P., Griffin, D., Maini, E., Phan, T. K., Rio, M., Vermoesen, L., Vandeputte, F., Schamel, F., and Burstzynowski, D. (2017). Service-Centric Networking for Distributed Heterogeneous Clouds. *IEEE Communications Magazine*, 55(7):208–215. DOI: 10.1109/MCOM.2017.1600412.

Su, Y., Cai, Z., Wu, K., Shi, L., Zhou, F., Chen, H., and Wu, J. (2018a). Projection-Type Multiview Holographic Three-Dimensional Display Using a Single Spatial Light Modulator and a Directional Diffractive Device. *IEEE Photonics Journal*, 10(5):1–12. DOI: 10.1109/JPHOT.2018.2871936.

Su, Z., Hui, Y., and Luan, T. H. (2018b). Distributed Task Allocation to Enable Collaborative Autonomous Driving With Network Softwarization. *IEEE Journal on Selected Areas in Communications*, 36(10):2175–2189. DOI: 10.1109/JSAC.2018.2869948.

Sun, Y., Peng, M., Zhou, Y., Huang, Y., and Mao, S. (2019). Application of Machine Learning in Wireless Networks: Key Techniques and Open Issues. *IEEE Communications Surveys & Tutorials*, 21(4):1–1. DOI: 10.1109/comst.2019.2924243.

Takeda, S., Iwai, D., and Sato, K. (2016). Inter-reflection Compensation of Immersive Projection Display by Spatio-Temporal Screen Reflectance Modulation. *IEEE Transactions on Visualization and Computer Graphics*, 22(4):1424–1431. DOI: 10.1109/TVCG.2016.2518136.

Taleb, T., Ksentini, A., and Frangoudis, P. A. (2019). Follow-Me Cloud : When Cloud Services Follow Mobile Users. *IEEE Transactions on Cloud Computing*, 7(2):369–382. DOI: 10.1109/TCC.2016.2525987.

Tanaka, T., Eum, S., Ata, S., and Murata, M. (2019). Design and implementation of tracking system for moving objects in information-centric networking. In *2019 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*, pages 302–306. DOI: 10.1109/ICIN.2019.8685886.

Tang, X., Zhou, J., Xiong, S., Wang, J., and Zhou, K. (2018a). Geographic Segmented Opportunistic Routing in Cognitive Radio Ad Hoc Networks Using Network Coding. *IEEE Access*, 6:62766–62783. DOI: 10.1109/ACCESS.2018.2875566.

Tang, Z., Zhou, X., Zhang, F., Jia, W., and Zhao, W. (2018b). Migration Modeling and Learning Algorithms for Containers in Fog Computing. *IEEE Transactions on Services Computing*, page 1. DOI: 10.1109/TSC.2018.2827070.

Tang, Z., Zhou, X., Zhang, F., Jia, W., and Zhao, W. (2018c). Migration Modeling and Learning Algorithms for Containers in Fog Computing. *IEEE Transactions on Services Computing*, 14(8). DOI: 10.1109/TSC.2018.2827070.

Tayyab, M., Gelabert, X., and Jantti, R. (2019). A Survey on Handover Management: From LTE to NR. *IEEE Access*, 7(1):118907–118930. DOI: 10.1109/access.2019.2937405.

Thakkar, M. K., Agrawal, L., Rangisetti, A. K., and Tamma, B. R. (2017). Reducing ping-pong handovers in lte by using a1-based measurements. In *2017 Twenty-third National Conference on Communications (NCC)*, pages 1–6. DOI: 10.1109/NCC.2017.8077083.

Thotahewa, K. M., Khan, J. Y., and Yuce, M. R. (2014). Power efficient ultra wide band based wireless body area networks with narrowband feedback path. *IEEE Transactions on Mobile Computing*, 13(8):1829–1842. DOI: 10.1109/TMC.2013.120.

Tiennoy, S. and Saivichit, C. (2018). Using a Distributed Roadside Unit for the Data Dissemination Protocol in VANET With the Named Data Architecture. *IEEE Access*, 6:32612–32623. DOI: 10.1109/ACCESS.2018.2840088.

Tomatis, A., Menouar, H., and Roscher, K. (2015). *Forwarding in VANETs: GeoNetworking*, pages 221–251. Springer International Publishing, Cham. DOI: 10.1007/978-3-319-15497-8_8.

Tortonesi, M., Govoni, M., Morelli, A., Riberto, G., Stefanelli, C., and Suri, N. (2019). Taming the IoT data deluge: An innovative information-centric service model for fog computing applications. *Future Generation Computer Systems*, 93:888–902. DOI: 10.1016/J.FUTURE.2018.06.009.

Tran, T. X. and Pompili, D. (2019). Joint task offloading and resource allocation for multi-server mobile-edge computing networks. *IEEE Transactions on Vehicular Technology*, 68(1):856–868. DOI: 10.1109/TVT.2018.2881191.

Tschudin, C. and Sifalakis, M. (2014). Named functions and cached computations. *2014 IEEE 11th Consumer Communications and Networking Conference (CCNC)*, pages 851–857. DOI: 10.1109/ccnc.2014.6994417.

Urgaonkar, R., Wang, S., He, T., Zafer, M., Chan, K., and Leung, K. K. (2015). Dynamic service migration and workload scheduling in edge-clouds. *Perform. Eval.*, 91(C):205–228. DOI: 10.1016/j.peva.2015.06.013.

Venkataramani, A., Kurose, J., Raychaudhuri, D., Nagaraja, K., Mao, M., and Banerjee, S. (2014). MobilityFirst: a mobility-centric and trustworthy internet architecture. *ACM SIGCOMM Computer Communication Review*, 44(3):74–80. DOI: 10.1145/2656877.2656888.

Venkatramana, D. K. N., Srikantaiah, S. B., and Moodabidri, J. (2017). SCGRP: SDN-enabled connectivity-aware geographical routing protocol of VANETs for urban environment. *IET Networks*, 6(5):102–111. DOI: 10.1049/iet-net.2016.0117.

Vilalta, R., Popescu, I., Mayoral, A., Cao, X., Casellas, R., Yoshikane, N., Martínez, R., Tsuritani, T., Morita, I., and Muñoz, R. (2017). End-to-end SDN/NFV orchestration of video analytics using edge and cloud computing over programmable optical networks. In *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3. DOI: 10.1364/OFC.2017.W3I.1.

Vodafone (2021). White paper: Multi-access edge computing. Technical report, Vodafone GmbH. Available at: `https://www-vodafone-de.translate.goog/business/loesungen/whitepaper-mec/?_x_tr_sl=auto&_x_tr_tl=en&_x_tr_hl=en&_x_tr_pto=wapp`.

Wan, Z., Zhang, Z., Yin, R., and Yu, G. (2022). Kfiml: Kubernetes-based fog computing iot platform for online machine learning. *IEEE Internet of Things Journal*, 9(19):19463–19476. DOI: 10.1109/JIOT.2022.3168085.

Wang, H., Zhao, H., Zhang, J., Ma, D., Li, J., and Wei, J. (2020). Survey on unmanned aerial vehicle networks: A cyber physical system perspective. *IEEE Communications Surveys and Tutorials*, 22:1027–1070. DOI: 10.1109/COMST.2019.2962207.

Wang, J., Liu, J., and Kato, N. (2019). Networking and Communications in Autonomous Driving: A Survey. *IEEE Communications Surveys & Tutorials*, 21(2):1243–1274. DOI: 10.1109/COMST.2018.2888904.

Wang, S., Member, S., Xu, J., Zhang, N., Member, S., and Liu, Y. (2018a). A Survey on Service Migration in Mobile Edge Computing. *IEEE Access*, 6:23511–23528. DOI: 10.1109/ACCESS.2018.2828102.

Wang, X., Ning, Z., Hu, X., Ngai, E. C., Wang, L., Hu, B., and Kwok, R. Y. K. (2018b). A City-Wide Real-Time Traffic Management System : Enabling Crowdsensing in Social Internet of Vehicles. *IEEE Communications Magazine*, 56(September):19–25. DOI: 10.1109/MCOM.2018.1701065.

Weil, S. A., Brandt, S. A., Miller, E. L., Long, D. D. E., and Maltzahn, C. (2006). Ceph: A Scalable, High-Performance Distributed File System. In Bershad, B. N. and Mogul, J. C., editors, *7th Symposium on Operating Systems Design and Implementation (OSDI '06), November 6-8, Seattle, WA, USA*, pages 307–320. USENIX Association. Available at: `https://www.usenix.org/legacy/events/osdi06/tech/full_papers/weil/weil_html/`.

Weiss, A. (2007). Computing in the clouds. *NetWorker*, 11(4):16–25. DOI: 10.1145/1327512.1327513.

Wen, Z., Rovatsos, M., Yang, R., Xu, J., Garraghan, P., and Lin, T. (2017). Fog Orchestration for Internet of Things Services. *IEEE Internet Computing*, 21(2):16–24. DOI: 10.1109/mic.2017.36.

Wireless One (2018). Pokémon GO first to DT "Edge network in the core". Available at: `http://wirelessone.news/24-edge/1215-edge-networks-in-the-core-jumbo-shrimp`.

Xiong, F., Li, A., Wang, H., and Tang, L. (2019). An SDN-MQTT Based Communication System for Battlefield UAV Swarms. *IEEE Communications Magazine*, 57(8):41–47. DOI: 10.1109/MCOM.2019.1900291.

Yahiatene, Y., Rachedi, A., Riahla, M. A., Menacer, D. E., and Nait-Abdesselam, F. (2019). A blockchain-based framework to secure vehicular social networks. *Transactions on Emerging Telecommunications Technologies*, 30(8):e3650. DOI: 10.1002/ett.3650.

Yang, L., Cao, J., Liang, G., and Han, X. (2016). Cost Aware Service Placement and Load Dispatching in Mobile Cloud Systems. *IEEE Transactions on Computers*, 65(5):1440–1452. DOI: 10.1109/TC.2015.2435781.

Yang, S., Li, F., Shen, M., Chen, X., Fu, X., and Wang, Y. (2019). Cloudlet Placement and Task Allocation in Mobile Edge Computing. *IEEE Internet of Things Journal*, 6(3):5853–5863. DOI: 10.1109/JIOT.2019.2907605.

Yao, H., Qiu, C., Fang, C., Chen, X., and Yu, F. R. (2016). A Novel Framework of Data-Driven Networking. *IEEE Access*, 4:9066–9072. DOI: 10.1109/ACCESS.2016.2624781.

Yazdinejad, A., Parizi, R. M., Dehghantanha, A., and Choo, K. R. (2019). Blockchain-enabled authentication handover with efficient privacy protection in sdn-based 5g networks. *IEEE Transactions on Network Science and Engineering*, pages 1–1. DOI: 10.1109/TNSE.2019.2937481.

Ye, H., Li, G. Y., and Juang, B.-H. F. (2019). Deep reinforcement learning based resource allocation for v2v communications. *IEEE Transactions on Vehicular Technology*, 68(4):3163–3173. DOI: 10.1109/TVT.2019.2897134.

You, D., Doan, T. V., Torre, R., Mehrabi, M., Kropp, A., Nguyen, V., Salah, H., Nguyen, G. T., and Fitzek, F. H. P. (2019). Fog Computing as an Enabler for Immersive Media: Service Scenarios and Research Opportunities. *IEEE Access*, 7:65797—-65810. DOI: 10.1109/access.2019.2917291.

You, T., Umair, M., and Hong, Y. (2017). Mobile crowd sensing based on CICN. In *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1272–1275. DOI: 10.1109/ICTC.2017.8190919.

Yu, Q., Ren, J., Fu, Y., Li, Y., and Zhang, W. (2019). Cybertwin: An origin of next generation network architecture. *IEEE Wireless Communications*, 26(6):111–117. DOI: 10.1109/MWC.001.1900184.

Yuan, Z., Jin, J., Sun, L., Chin, K. W., and Muntean, G. M. (2018). Ultra-Reliable IoT Communications with UAVs: A Swarm Use Case. *IEEE Communications Magazine*, 56(12):90–96. DOI: 10.1109/MCOM.2018.1800161.

Zaidi, Z., Member, S., Friderikos, V., Yousaf, Z., Fletcher, S., Dohler, M., and Aghvami, H. (2018). Will SDN Be Part of 5G ? *IEEE Communications Surveys & Tutorials*, 20(4):3220–3258. DOI: 10.1109/COMST.2018.2836315.

Zeng, Y., Zhang, R., and Lim, T. J. (2016). Wireless communications with unmanned aerial vehicles: Opportunities and challenges. *IEEE Communications Magazine*, 54(5):36–42. DOI: 10.1109/MCOM.2016.7470933.

Zhang, C. and Zheng, Z. (2019). Task migration for mobile edge computing using deep reinforcement learning. *Future Generation Computer Systems*, 96:111–118. DOI: 10.1016/j.future.2019.01.059.

Zhang, F., Liu, G., Fu, X., and Yahyapour, R. (2018a). A Survey on Virtual Machine Migration: Challenges, Techniques, and Open Issues. *IEEE Communications Surveys and Tutorials*, 20(2):1206–1243. DOI: 10.1109/COMST.2018.2794881.

Zhang, F., Nagaraja, K., Zhang, Y., and Raychaudhuri, D. (2012). Content Delivery in the MobilityFirst Future Internet Architecture. *2012 35th IEEE Sarnoff Symposium*, pages 1–5. DOI: 10.1109/SARNOF.2012.6222763.

Zhang, H., Quan, W., Chao, H., and Qiao, C. (2016). Smart identifier network: A collaborative architecture for the future internet. *IEEE Network*, 30(3):46–51. DOI: 10.1109/MNET.2016.7474343.

Zhang, L., Afanasyev, A., Burke, J., Jacobson, V., kc Claffy, Crowley, P., Papadopoulos, C., Wang, L., and Zhang, B. (2014). Named data networking. *ACM SIGCOMM Computer Communication Review*, 44(3):66–73. DOI: 10.1145/2656877.2656887.

Zhang, Q.-Y., Wang, X.-W., Huang, M., Li, K.-Q., and Das,

S. K. (2018b). Software Defined Networking Meets Information Centric Networking: A Survey. *IEEE Access*, 6:39547–39563. DOI: 10.1109/ACCESS.2018.2855135.

Zhang, X., Wang, H., and Zhao, H. (2018c). An SDN framework for UAV backbone network towards knowledge centric networking. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 456–461. DOI: 10.1109/INFCOMW.2018.8406959.

Zhang, Y. and Ansari, N. (2014). On protocol-independent data redundancy elimination. *IEEE Communications Surveys Tutorials*, 16(1):455–472. DOI: 10.1109/SURV.2013.052213.00186.

Zhang, Y., Deng, R., Bertino, E., and Zheng, D. (2019). Robust and universal seamless handover authentication in 5g hetnets. *IEEE Transactions on Dependable and Secure Computing*, pages 1–1. DOI: 10.1109/TDSC.2019.2927664.

Zhao, C., Dong, M., Ota, K., Li, J., and Wu, J. (2019a). Edge-MapReduce-Based Intelligent Information-Centric IoV: Cognitive Route Planning. *IEEE Access*, 7:50549–50560. DOI: 10.1109/ACCESS.2019.2911343.

Zhao, Z., Cumino, P., Souza, A., Rosário, D., Braun, T., Cerqueira, E., Gerla, M., do Rosário, D., Braun, T., Cerqueira, E., and Gerla, M. (2019b). Software-defined unmanned aerial vehicles networking for video dissemination services. *Ad Hoc Networks*, 83:68–77. DOI: 10.1016/j.adhoc.2018.08.023.

Zheng, K., Zheng, Q., Chatzimisios, P., Xiang, W., and Zhou, Y. (2015). Heterogeneous Vehicular Networking: A Survey on Architecture, Challenges, and Solutions. *IEEE Communications Surveys and Tutorials*, 17(4):2377–2396. DOI: 10.1109/COMST.2015.2440103.

Zhu, H., Xu, J., Liu, S., and Jin, Y. (2021). Federated learning on non-iid data: A survey. *Neurocomputing*, 465:371–390. DOI: 10.1016/j.neucom.2021.07.098.

Zink, M., Sitaraman, R., and Nahrstedt, K. (2019). Scalable 360 video stream delivery: Challenges, solutions, and opportunities. *Proceedings of the IEEE*, 107(4):639–650. DOI: 10.1109/JPROC.2019.2894817.

Zipf, G. K. (1950). Human behavior and the principle of least effort. cambridge, (mass.): Addison-wesley, 1949, pp. 573. *Journal of Clinical Psychology*, 6(3):306–306. DOI: 10.1002/1097-4679(195007)6:3<306::AID-JCLP2270060331>3.0.CO;2-7.