Opportunities and Challenges of Joint Edge and Fog Orchestration

Luca Cominardi[1,*], Osamah Ibrahiem Abdullaziz[2], Kiril Antevski[1], Shahzoob Bilal Chundrigar[2], Robert Gdowski[2], Ping-Heng Kuo[3], Alain Mourad[3], Li-Hsing Yen[4], Aitor Zabala[5]

[1] University Carlos III of Madrid, [2] ITRI, [3] InterDigital Europe LTD, [4] National Chiao Tung University, [5] Telcaria SRL

***Abstract*—Pushing contents, applications, and network functions closer to end users is necessary to cope with the huge data volume and low latency required in future 5G networks. Edge and fog frameworks have emerged recently to address this challenge. Whilst the edge framework was more infrastructure-focused and more mobile operator-oriented, the fog was more pervasive and included any node (stationary or mobile), including terminal devices. This article analyzes the opportunities and challenges to integrate, federate, and jointly orchestrate the edge and fog resources into a unified framework.**

***Index Terms*—Edge, Fog, NFV, MEC, SDN, Orchestration, Management, Virtualization**

I. Introduction

The research and development of fifth-generation mobile network (5G) spans across a large variety of new use cases which go beyond the natural evolution of voice and data delivery in 4G mobile networks [1]. Diverse 5G scenarios, such as multi-access network integration, even across operators and less trusted networks, massive Internet of Things (IoT), localized real-time control, vehicular communication, etc. pose significant challenges to the 4G monolithic and centralized network architecture, both in terms of flexibility and scalability, making new services hard to introduce and scale.

To increase flexibility in service offerings and network management, the European Telecommunications Standards Institute (ETSI) Network Function Virtualization (NFV) Industry Specification Group (ISG) pioneered the idea of bringing virtualization capabilities into mobile operator networks [2]. By decoupling the network functions from the underlying hardware platform, NFV allows operators to dynamically deploy services in response to the needs of the traffic. In addition to NFV, ETSI Multi-access Edge Computing (MEC) ISG brings computing capabilities close to the end users to cope with the ever-increasing amount of data (e.g., generated by IoT) and the low latency required by some use cases (e.g., vehicular communication) [3]. NFV and MEC jointly represent a paradigm shift for mobile operator networks, which evolve from a centralized architecture based on monolithic and hardware-integrated functions to a software-based distributed architecture. Such evolution enables a common hosting environment, namely edge computing, at the network edge characterized by low latency and high bandwidth as well as real-time access to radio network information. Network functions and software applications can be hence deployed close to the end users, thus alleviating congestion at the mobile network core and serving efficiently local purposes, such as data aggregation for IoT, localized real-time control, and single aggregation point for multi-access connectivity.

Recently, fog computing gained considerable traction in the industrial community as demonstrated by the newborn OpenFog consortium [4]. Fog computing distributes computing, storage, control and networking functions closer to the users along a cloud-to-thing continuum which also envisions the collaborative usage of a multitude of end user or near-user edge devices to carry out a substantial amount of those tasks. It is noteworthy that non-stationary and volatile devices are also considered in fog computing, for example when apparatus are hosted on moving devices (e.g., car, train, mobile user) or are battery-powered (e.g., IoT). While edge computing focuses on operator networks and related use cases, fog computing focuses more broadly on enterprise use cases, which may not be necessarily related to mobile networks (e.g., smart cities, remote surveillance, etc.). Nonetheless, edge and fog present a significant synergy: they both focus on bringing networking and computing capabilities closer to the user. Nowadays, edge and fog computing are stand-alone domains that require separate deployments eventually contending for the same physical resources (e.g., spectrum). The lack of integration poses numerous challenges to the effective usage of those resources in addition to the cost-effectiveness of having multiple separate physical deployments.

Integrating resources belonging to distinct administrative domains is a challenge that goes beyond the pure technological dimension and involves trust relationships between parties. At this end, federation provides the means for integrating multiple administrative domains at different granularity into a unified platform where the federated resources can trust each other at a certain degree, whereas the federation trust is the embodiment of a service/business-level agreement or partnership between two organizations [5]. Fig. 1 shows the edge and fog resources (in blue) which may be federated among themselves and interact with centralized core and cloud domains (in grey) for offering a real cloud-to-thing continuum.

The paper is organized as follows: Section II presents related work in the state-of-art. Section III exposes the opportunities enabled by a joint orchestration of edge and fog

* Corresponding author. Email: luca.cominardi@uc3m.es
This work has been partially funded by the H2020 collaborative Europe/Taiwan research project 5G-CORAL (grant num. 761586).

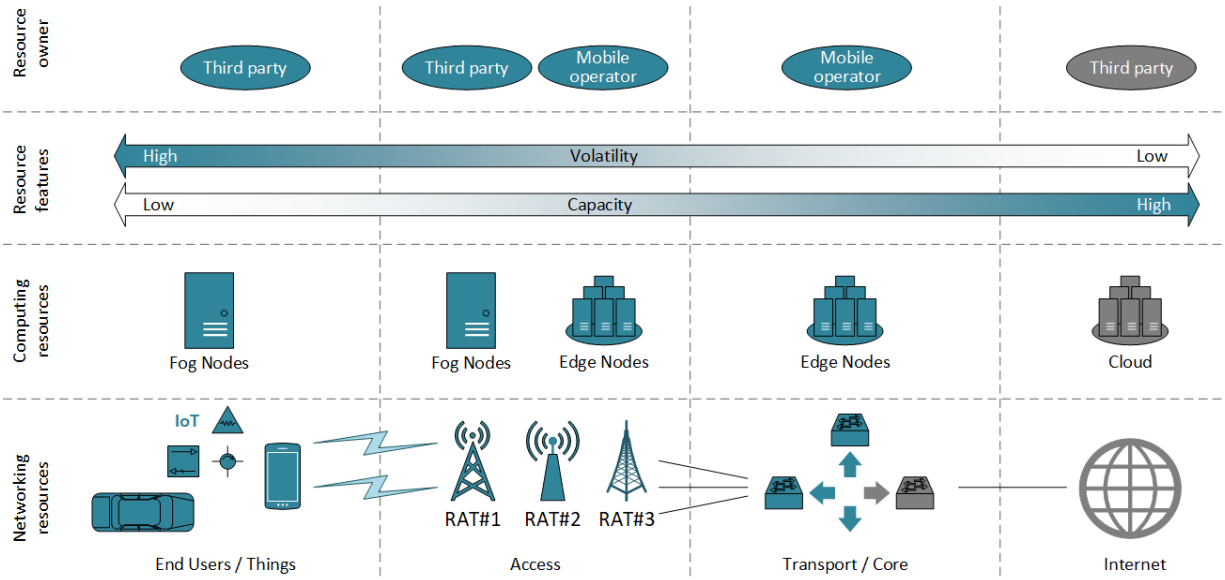

Figure 1: Edge and Fog resources and characteristics.

resources. Section IV reports the challenges of achieving such integration, and finally, Section V draws the conclusions.

II. Related Work

This section provides an analysis of state of the art on topics related to the opportunities and challenges of joint edge and fog orchestration. These topics range from cloud level federation, edge-and-fog orchestration jointly with an analysis of the ETSI NFV and MEC frameworks.

The broad necessity of cloud service providers to fulfill their sparse customers, their needs and concerns in line with customer's data ownership and scarcity of standards defining inter-clouds service interfaces, has led to the adoption of decentralized cloud federations. A cloud federation can integrate a pool of diverse services from multiple service providers that self-govern each other by using well-defined interfaces and agreements between them [6].

Cloud federation is a key enabling technology for cooperative service deployment. In a dynamic fashion, it allows heterogeneous and independently administrated clouds to interact and share resources with each other. Federated clouds offer an integrated cloud service by federating infrastructures provided by different cloud service providers. The ability of cloud federation to share cloud resources among participating service providers improves resource utilization and enhances elasticity and reliability of cloud service. Federated clouds also enable new business opportunities.

Virtualization technologies and its orchestration, including the use of virtual machines and containers, play a major role in the provisioning of elastic mobile services in federated clouds. In this case, the federation mechanisms should include functionalities such as deployment, runtime management and monitoring, termination, authentication, access control and live migration of services in remote clouds [7].

Many existing works in the literature develop frameworks and architectures to enable provisioning and management of services in federated clouds. Depending on the cooperation model of participants, cloud federation can be classified into several types. The first one is a horizontal federation, where participants cooperate on a peer-to-peer basis. This type of federation well applies to the case of federated mobile edge systems. The second type is a vertical federation, where participants are entities in a hierarchy, like hybrid cloud [11][12] which combines the services provided by a private cloud and a third-party public cloud. This type of federated edge-and-fog architecture refers to the federation between edge and fog systems, between central cloud and edge system, or between central cloud and fog system. Finally, the third type of federation comprises both horizontal and vertical federations.

Most existing federated clouds fall into the category of the vertical federation. For instance, Follow-Me Cloud (FMC) [8] proposes an architecture for federated cloud and distributed mobile network environment which allows the services delivery through an optimal service anchor and the possibility of following mobile users as they roam through federated cloud environments. FMC utilizes Markov-Decision-Process to make cost-effective and performance optimized migration decisions. Furthermore, challenges which cloud providers may face when participating in a federated cloud environment include the heterogeneity of cloud management systems and models describing the services. To resolve this issue, [9] proposes a coordinated application deployment system (CADS) to enable the description of the desired service deployment in form of a topology model. In this way, CADS provides interoperability in the deployment of services in federated clouds.

The NFV ISG defines a Management and Orchestration (MANO) framework [2] for deploying network services on an NFV environment. Nowadays, NFV MANO scope is limited to a single mobile operator network. To overcome such limitation, an NFV Work Item has been recently approved with the aim of enabling the management and orchestration across multiple operators [10]. Although logical inter-connection between

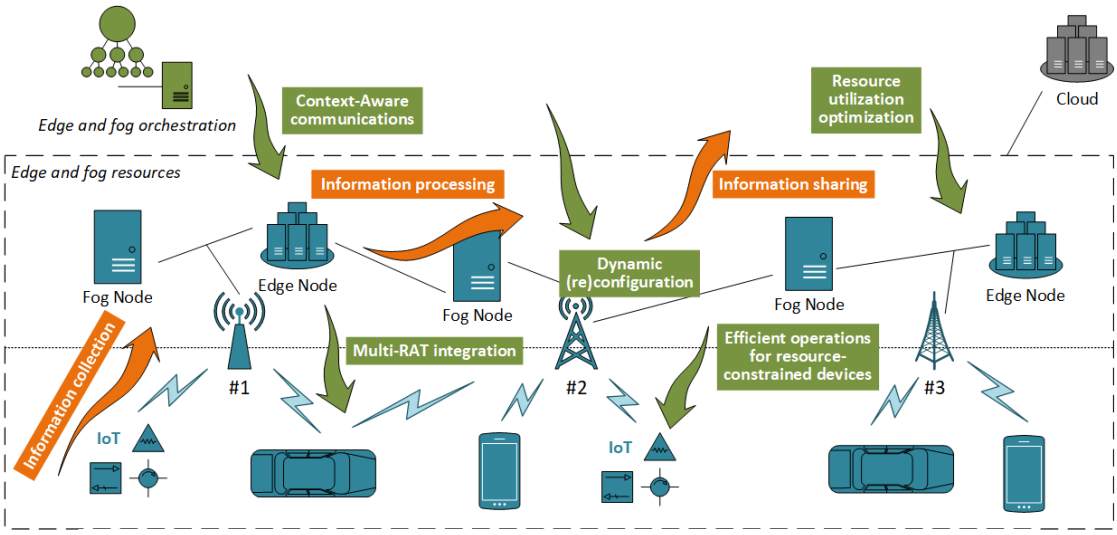

Figure 2: Edge and Fog joint orchestration opportunities.

different mobile operators is being defined, integration with third-party domains (e.g., fog or cloud) is still not considered. Like NFV, MEC framework [3] only considers a single network operator domain and does not consider integration with third-party domains like fog. Finally, although ETSI MEC and NFV enable mobility of applications and services, it is only within the boundaries of the stationary edge resources of the mobile operator and volatile resources are not considered.

III. Opportunities

By bringing fog computational resources into the vision of networking in 5G and beyond, several opportunities can be anticipated to enhance the system efficiency and performance. This section walks through a few potential benefits of joint edge and fog orchestration illustrated in Fig. 2.

A. Context-Aware Communications and Computations

Context-aware communications and computations open a new degree of freedom in optimizing the network performance based on context information extracted from the underlying infrastructure of computing, storage and networking resources. This raises the opportunity of developing new algorithms to optimize the network performance based on the learning and intelligence derived from the context information of the edge-and-fog system. For example, where the edge-and-fog system is likely to have multiple co-existence radio access technologies (RATs), one can envision efficient multi-RAT management and coordination algorithms by leveraging on the radio information extracted from each RAT. Artificial intelligence and machine learning based optimization are examples of tools that can also be deployed in here.

B. Resource Utilization Enhancement

Instead of solely relying on the computing substrates in the edge data centers, edge-and-fog orchestration allows the distribution of the various computing and networking tasks across both edge and fog resources, including any type of devices that possess networking and computing capabilities. This creates a larger pool of resources distributed near the end users enabling higher multiplexing gains, greater utilization efficiency of the resources, and a larger pool of cooperating resources for executing certain functions or tasks tailored to the needs of the applications and end users. Such paradigm may create new business models wherein terminal devices can also participate in the pool of edge and fog resources in return of incentives (e.g. service subscription reduction), which helps infrastructure providers decrease the deployment and maintenance cost of their edge data centers.

C. Efficient Operation for Resource-Constrained Devices

In 5G, various categories of devices are envisioned. These range from vehicles and drones, to smartphones, tablets, and laptop computers, to IoT devices such as sensors or actuators. Clearly, some of these devices will have limited computational capability and battery (the so-called "resource-constrained" devices) due to their low-cost nature. With an edge-and-fog orchestration, these resource-constrained devices can now rely on the edge and fog resources to execute some of their computationally and power demanding tasks. This presents an opportunity for low-cost devices to remain intelligent and run advanced applications despite their limited capabilities.

D. Flexible and Scalable Functionalities

Edge-and-fog's NFV inspired architecture and technologies aspire to become a viable proposition to enable flexibility and scalability of the envisioned 5G system. The foreseen joint orchestration would allow traffic engineering between nodes, thus setting roots to flexible behavior and scalability of the system. Indeed, dynamic allocation of the computing and networking resources can be used for example to prioritize edge and fog resources in an area of higher demand which may lead to more optimized resource utilization. A concentrated traffic or computing request can be directed to a limited number of edge and fog resources while others will shift to idle mode or be switched off, thus improving the overall energy conservation of the system. Furthermore, software migration

and placement capability of the orchestration process allows for a seamless transfer of intelligence between geographically disparate nodes. This tackles the variable application delay constraints. Software components can be placed in the user vicinity fulfilling its latency requirements. Together with the ability to create interrelation (chain) of functions and applications and then map it into an underlying substrate of computing and networking resources, such ecosystem will be able to handle any 3rd party driven dependency between functions and applications whilst preserving the scalability of the solution at the same time.

IV. Challenges

Joint orchestration of edge and fog computing is still new, and it does not have yet a corresponding framework defined. Such framework will need to satisfy requirements of real-time communication utilizing edge nodes, federation among multiple stakeholders, and dynamic resource discovery of volatile and non-volatile resources. Deployment strategies are also needed such as where to place the workload, connection policies, and when to use edge or fog nodes accounting for their heterogeneity. To define this framework, we identify the following seven research challenges to be addressed.

A. Federation mechanisms

Federation is a process where different entities negotiate terms and conditions with a goal to form an alliance of trust and start sharing resources between each other. The result, the federation, should be beneficiary for all included entities. The key elements are the trust between entities and maintenance of negotiated conditions for long-term federation. Enabling trust between different entities is a challenge that can be solved using centralized or decentralized solution. The centralized solution is through single dedicated entity (server, repository) managed by the trusted organization. It demands high level of maintenance and strict security policies. Additional resources may be needed to ensure scalability of the system. The decentralized solution is through a peer-to-peer network of trusted entities that maintain highly distributed repository. Although it considers a complex setup operation and high-security risks, recent advances in trust-enabling technologies (e.g., Blockchain, Bitcoin, Ethereum) prove the contrary [13][14]. The distributed repository can be deployed fast using current infrastructure as well as secure setup of the peer-to-peer network of entities. The distributed repository is in line with the edge-and-fog architecture where the risks can be addressed through the scalability, storage and speed of the peer-to-peer solution (i.e., Blockchain) or through a combination of smart contracts, REST APIs and web applications (i.e., Ethereum).

Federation mechanisms are determined for the dynamic integration of multiple administrative domains into a unified platform using different granularity in either centralized or decentralized fashion. Different stakeholders expose different capabilities depending on their physical constraints (computing, storage, bandwidth) using different policies. It determines the stakeholders' degree of trust and conditions in which they are willing to join in the federation. The centralized solution holds the trust in a single entity. The terms and conditions are negotiated at a single point, exposing security threat (of a single point of failure) which by default demands maintenance and redundancy (similar to the DNS architecture). The decentralized solution distributes the trust burden to all entities. In this case, overlay peer-to-peer networks can be established between different stakeholders. A stakeholder can maintain several federation networks based on the degree of trust it exposes to each federation (e.g., gold federation, silver federation, etc.) as proposed in [5]. Using one or both approaches, the challenge is to rapidly and efficiently enable the edge and fog systems to dynamically scale up into unique virtualization environment using the heterogeneous and exposed resources thus satisfying user and network demand in highly secure and trusted manner.

B. Dynamic discovery of resources

As aforementioned, the edge-and-fog computing system can be constructed by federating resources via joint orchestration. However, in contrast to edge or cloud computing where the tasks are basically performed by dedicated and static data centers or servers, fog computing could be carried out by mobile and battery-constrained devices that are volatile (i.e. may become available or unavailable spontaneously). Thus, it is key for the orchestration and management system to localize and monitor the available computing and networking resources, the pool of which may consist of both volatile and non-volatile substrates. In particular, the system should be able to identify the resources that have become available for federation, and such identification process is dubbed as "discovery" in this context. As the resources to be discovered may be physical devices that belong to different owners and/or administrative domains, there are two foreseeable challenges.

First, how can devices discover or be discovered by the orchestration and management system? Devices and associated resources may belong to different owners, have stationary or mobile nature, may have different availability or simply communicate by different protocols. The monitoring entity needs means to reach to those heterogeneous devices but also to estimate their stability and trust level in order to support stability of the overall system. Resource discovery in multi-RAT environments may require the use of distinct mechanisms depending on the connectivity availability. For instance, Link Layer Discovery Protocol (LLDP) works well in Ethernet networks while is not applicable to mobile networks (e.g., LTE). Similarly, IP-based mechanisms may not work in environments where Layer 2 security mechanisms are in place (e.g., 802.11 wireless networks).

Second, how can resources be discovered across different (often overlapping) administrative domains? Different management systems of separate administrative domains need to have means to discover each other's resources in order to provide services which require enhanced pool of computing resources as well as to achieve better overall utilization of overall pool of resources (e.g. reduce the energy usage during the idle periods in the network). To establish mutual resource usage, two control planes of two administrative domains must discover each other first. This process is a precondition for further federation process between different systems.

C. Multi-Tenancy

Multi-tenancy refers to the support of co-existing applications requested by different tenants within the same infrastructure. All tenants perceive their resources as dedicated without mutual interference. Multi-tenancy enhances resource utilization and enable business opportunities, which may find its application in edge or fog system alone. It is of greater need in federated edge and fog systems for several reasons. First, there may be multiple over-the-top service (e.g., video, voice, social applications, etc.) providers who operate solely on top of the federated edge and fog systems. Second, there may be multiple industry vertical market players (electricity utility, automotive, e-health, etc.) who exploit federated edge-and-fog system to enhance system reliability and robustness.

However, enabling multi-tenancy demands mechanisms to ensure security, isolation, and privacy among tenants. For network infrastructure, this is known as network slicing. For edge-and-fog infrastructure, we need similar schemes to provide an isolated amount of cloud capacity customized to best suit specific application needs. More specifically, we need the following functions in federated fog and edge systems: *i)* a function that performs admission control based on tenant's need (SLAs) and current status of the infrastructure; *ii)* a function that securely exposes selected service capabilities and management policies with a standard resource descriptor to the tenants for SLA negotiation and matching; and *iii)* a function that provides performance monitoring information to the tenants.

D. Multi-virtualization technology coexistence

Virtualization refers to the different approaches for creating a virtual version of networking or computing hardware. There are multiple virtualization techniques whose difference primarily resides in the location of the virtualization layer and the way resources are used. *Full-virtualization* provides a complete abstraction of the physical hardware. This allows software to run on distinct types of hardware without requiring any modification. This is the case of virtual machines. *Hybrid-virtualization* provides an incomplete abstraction of the hardware. This imposes targeted modification to the software to run on different systems. This is the case of virtual machines using specific I/O hardware acceleration extensions. *Para-virtualization* allows software to be executed in isolated domains but does not provide any hardware abstraction (e.g., software is explicitly written for a given operating system). This is the case of containers.

A mix of those virtualization techniques could be present at the same time in the edge and fog domains, especially if the overall ecosystem is the result of federation among multiple organizations. This poses a considerable challenge to the possibility of deploying any application on any node. Indeed, the orchestration system can instantiate applications only if the virtualization substrate is compatible with the application packaging, thus reducing the possibilities of resource optimization. Therefore, developers should package their applications for any possible target system. Such requirement could be relaxed by the usage of automated tools which take care of packaging the same application for multiple target systems. Tools like Vagrant [2] could be hence extended to support multiple virtualization substrates.

E. Functions and applications placement

Servers used to host applications/functions have a finite amount of compute capacity, notably in the case of fog, where resources are on the move and with limited compute capacity. In principle, services can be composed by placing functions and applications into the appropriate edge and fog points-of-presence (PoPs). This requires provisioning of the computing resources. There will be cases where the functions and applications will be hosted in different domains (in the case of federation) and therefore it is necessary to decide which node to be used as edge or fog PoP. Typically, each request will have SLA requirements of latency, throughput and availability targets. The fundamental challenge is to deploy the functions and applications on integrated edge and fog resources while meeting the necessary SLAs. Several issues complicate the optimization of functions and applications placement. First, volatility of the resource in terms of availability, for how long the resource will be available to be part of a service. Second, the workload varies dynamically and with the limited and finite compute capacity, it also becomes difficult to scale the resources up and down depending on the load. Third, there are performance issues that appear by co-locating the virtualization functions and applications onto the same server or node. Fourth is the complexity of optimal placement in a federated environment. Also in dynamic environments, resources may get fragmented which makes it more difficult to place optimally functions and applications. Integer Linear Programming (ILP) is often used for finding the optimal placement of functions and applications in static environments. However, those algorithms require a long process time before reaching to the solutions. To this end, heuristic algorithms are more suitable in dynamic environments where the solution, even if suboptimal, needs to be provided in a short timeframe. Machine learning could be employed to enhance the accuracy of the placement based on historical data.

F. Dynamic service placement and migration

Service placement and migration is the process of transitioning an individual or organizational data across multiple cloud providers. With the advent of edge-and-fog computing, edge and fog nodes become places that users use to have seamless connection to the services with low communication latency. However, edge-and-fog computing brings in yet another challenge of dynamic service placement and migration. As a user moves to different geographical areas, should its service be migrated from one edge or fog node to another? The main challenge introduced here is to maintain relatively low service downtime and overall migration time without impacting the quality of service (QoS). It is challenging to find the optimal decision also because of the uncertainty of the user's mobility along with the transmission cost. In addition, the placement of the selected services needs

[2] https://www.vagrantup.com/

to consider potential mobility patterns, to provide the desired performance to the associated user always.

G. Dynamic Resource Management

Dynamic resource management is the ability to manage dynamically the resources (compute, network, storage) by means of automation and self-allocation mechanisms. In a Multi-RAT environment, one could always think about routing the traffic dynamically from one RAT to another depending on the user's/network demand. In addition, probabilistic assumption on the mean workload needs to be derived at different time resolutions to provide the optimal compute/network resources to the users. One important challenge here is how to manage the fog and edge resources dynamically. This is especially challenging due the heterogeneity and volatility of the edge and fog resources.

H. Security

Any entity involved in the edge-and-fog computing can be possibly malicious, so security issues of the orchestration may mainly come from three aspects involving different entities' interactions: integrating heterogeneous platforms, sharing resources among devices, and hosting third-party applications. They require the authentication between different entities, dynamic resource authorization, and the protection against malicious applications, respectively. In addition, those solutions designed to interwork with the cellular network require to be compliant with the 3GPP standard's security requirements. It can be challenging to fulfill the requirements while keeping the edge-and-fog computing transparent to the 3GPP network architecture [15]. To prevent security threats of the edge and fog computing platforms from propagating towards the existing cellular network, the orchestration shall also provide a firewall-like security middleware between them. Though the software/hardware entities involved in the edge and fog computing solutions can be diverse, the orchestration shall introduce a set of general security requirements and mechanisms to establish a baseline security level.

V. Conclusions

The edge and fog are key pillars of future networks where intelligence and innovations will be increasingly applied. There is however not yet a common unified platform that integrates and federates these two pillars together. Whilst the edge is more infrastructure-oriented and hence easier to integrate, the fog tends to be more volatile with resources appearing and disappearing on the go, and belonging to different owners. The opportunities for such unified framework are clearly acknowledged, but there remains to be several challenges that need to be addressed first before such a common framework could emerge. These include: *1)* the dynamic discovery of volatile and non-volatile resources; *2)* the federation of these resources when they belong to different domains and owners; *3)* the support of multi-tenancy in particular for the volatile fog resources; *4)* the customization and interworking of different virtualization technologies suitable to each type of resources (edge and fog); *5)* the dynamic placement of functions and applications across the continuum of fog and edge; *6)* the automation and dynamic allocation and management of the resources; and finally *7)* the security, trust and privacy considerations.

This paper presented these challenges that are being addressed in the framework of the collaborative research project 5G-CORAL [16].

References

[1] 3GPP, "Feasibility Study on New Services and Markets Technology Enablers," 3rd Generation Partnership Project, TR 22.891, Sep 2016.

[2] ETSI, "Network Functions Virtualisation (NFV); Management and Orchestration; Architectural Options," European Telecommunications Standards Institute, GS NFV-IFA 009, July 2016.

[3] ETSI, "Mobile Edge Computing (MEC); Framework and Reference Architecture," European Telecommunications Standards Institute, GS MEC 003, Mar 2016.

[4] OpenFog Consortium, "OpenFog Reference Architecture for Fog Computing," Architecture Working Group, Feb 2017.

[5] C. Simon et al., "5G exchange for inter-domain resource sharing," 2016 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN), Rome, 2016, pp. 1-6.

[6] T. Kurze et al., "Cloud federation," Proceedings of the 2nd International Conference on Cloud Computing, GRIDs, and Virtualization, CLOUD COMPUTING 2011, 2011

[7] R. Moreno-Vozmediano et al., "IaaS cloud architecture: From virtualized datacenters to federated cloud infrastructures," IEEE Computer, vol. 45, no. 12, pp. 65–72, Dec. 2012

[8] Taleb T. et al., "Follow-me cloud: When cloud services follow mobile users," IEEE Transactions on Cloud Computing. 2016.

[9] Panarello A. et al., "Automating the Deployment of Multi-Cloud Applications in Federated Cloud Environments," 10th VALUETOOLS 2017.

[10] ETSI, "Network Functions Virtualisation (NFV); Management and Orchestration; Report on architecture options to support multiple administrative domains," European Telecommunications Standards Institute, DGR NFV-IFA 028, Dec. 2017.

[11] Yuqian Lu et al., "Development of a hybrid manufacturing cloud," Journal of Manufacturing Systems, Oct. 2014.

[12] Jin Li et al., "A hybrid cloud approach for secure authorized deduplication," IEEE Trans. on Parallel and Distributed Systems, 26(5):1206-1216, May 2015.

[13] T. Swanson, "Consensus-as-a-service: A brief report on the emergence of permissioned, distributed ledger systems," Technical Report, Apr. 2015. [Online]. http://www.ofnumbers.com/wp-content/uploads/2015/04/. Permissioned-distributed-ledgers.pdf. [Accessed: 31 Oct 2017].

[14] E. Münsing et al., "Blockchains for decentralized optimization of energy resources in microgrid networks," 2017 IEEE Conference on Control Technology and Applications, 2017.

[15] ETSI, "Mobile-Edge Computing - Introductory Technical White Paper," European Telecommunications Standards Institute White Paper, Sep. 2014.

[16] 5G-CORAL, "A 5G Convergent Virtualised Radio Access Network Living at the Edge," H2020-ICT-2016-2, [Online]. http://5g-coral.eu/. [Accessed: 31 Oct 2017].