Research Article

QoE-Oriented Mobile Edge Service Management Leveraging SDN and NFV

Shuping Peng,[1] Jose Oscar Fajardo,[2] Pouria Sayyad Khodashenas,[3] Begoña Blanco,[2] Fidel Liberal,[2] Cristina Ruiz,[3] Charles Turyagyenda,[1] Mick Wilson,[1] and Sunil Vadgama[1]

[1]*Fujitsu Laboratories of Europe, Fujitsu UK, Hayes Park Central, Hayes, UK*
[2]*University of the Basque Country (UPV/EHU), Bilbao, Spain*
[3]*I2CAT Foundation, Barcelona, Spain*

Correspondence should be addressed to Jose Oscar Fajardo; joseoscar.fajardo@ehu.es

5G envisages a "hyperconnected society" where trillions of diverse entities could communicate with each other anywhere and at any time, some of which will demand extremely challenging performance requirements such as submillisecond low latency. Mobile Edge Computing (MEC) concept where application computing resources are deployed at the edge of the mobile network in proximity of an end user is a promising solution to improve quality of online experience. To make MEC more flexible and cost-effective Network Functions Virtualisation (NFV) and Software-Defined Networking (SDN) technologies are widely adopted. It leads to significant CAPEX and OPEX reduction with the help of a joint radio-cloud management and orchestration logic. In this paper we discuss and develop a reference architecture for the orchestration and management of the MEC ecosystem. Along with the lifecycle management flows of MEC services, indicating the interactions among the functional modules inside the Orchestrator and with external elements, QoS management with a focus on the channel state information technique is presented.

1. Introduction

5G [1], a future end-to-end ecosystem, creates transformative economic and social effect, leading to a "hyperconnected society" [2] and offering enormous potentials for customers and industry. 5G promises wireless access beyond humans to any entity that may benefit from being connected, which often is referred to as the "Internet of Things (IoT)" [3]. Envisioned by 5G-PPP, 7 trillion wireless devices serving over 7 billion people will be interconnected by very dense deployment of wireless communication links [1]. A wide variety of IoT applications will become prevalent in various verticals such as smart grid, smart city, e-Health and telemedicine, and automotive [3]. However, in order to support the needs of such huge number of connected devices, the future 5G wireless infrastructure should possess not only the highly scalable capacity but also features like low latency, high mobility support, and high reliability and resiliency.

To find a viable solution, in 2014, the European Telecommunications Standards Institute (ETSI) proposed a way to integrate cloud computing capabilities into the edge of the mobile network as Mobile Edge Computing (MEC) and established the MEC ISG [4] with the vision of "offering application developers and content providers cloud computing capabilities and an IT service environment at the edge of the mobile network," which "is characterized by ultralow latency and high bandwidth as well as real-time access to radio network information that can be leveraged by applications." The European 5G Infrastructure Public Private Partnership (5G PPP) research body recognized MEC as one of the key emerging technologies for 5G networks [5], economically justified by the Network Functions Virtualisation (NFV) and Software-Defined Networking (SDN) technologies.

MEC Network Services (NSs) are offered on a virtualised platform in the form of Virtual Network Function (VNF) chains. In a simple form, a VNF is a Virtual Machine (VM)

with a certain specification that runs a piece of software implementing the functionality of a network middle-box, for example, firewall. NFV (being standardized by ETSI NFV ISG) frames a solution able to manage and orchestrate the complete lifecycle (i.e., instantiation, termination, failure detection, and scaling) of NSs as well as individual VNFs [6]. In addition, to enhance the flexibility and dynamism, SDN provides a programmable platform to support connectivity between chained VNFs.

By nature, MEC is a mixed cloud radio environment; thus a holistic management/orchestration system should be able to perform radio resource management together with the proper management of other hardware resources. Of course, radio resource management by itself is a wide area ranging from efficient handover process to handling Quality of Service (QoS) and Quality of Experience (QoE) for the different mobile end users. Generally speaking, the radio management system is able to efficiently share the available radio resources among the different users, taking into account both the channel quality and the traffic demand of each mobile user. The introduction of MEC opens new management opportunities, enabling a quick personalisation of application-level data according to the context of the users. This application-level processing leads to different hardware requirements over time in order to react to radio channel variability. Thus, this holistic management/orchestration system should be able to exploit a control loop between the available radio performance for the set of mobile users and the allocated hardware resources for the different edge services.

The paper is an effort to draw a framework for the MEC management/orchestration system. To do so, it is organised as follows: Section 2 reviews the current state of MEC, SDN, and NFV and their standards activities being progressed in ETSI and Open Networking Foundation (ONF). Section 3 describes the overall scenario covering the key elements in the MEC ecosystem (incl. Orchestrator, SDN, and NFV). The Orchestrator reference architecture and its key components and interfaces are further detailed in this section. Section 4 shows the lifecycle management flows, that is, creation, termination, and modification of the MEC services. Section 5 presents potential use cases to demonstrate the benefits of MEC for different vertical sectors, while Section 6 focuses on the use of radio channel quality reports in MEC and its associated overhead. Section 7 provides a realistic example of the overhead of using radio channel quality reports in edge service instances taking into account different radio access network virtualisation and deployment models. Finally, Section 8 concludes this paper.

2. SDN and NFV for MEC Ecosystem

2.1. MEC Ecosystem. Information Technology (IT) and telecom networks are converging, which brings new capabilities and possibilities. A key transformation has been the emergence of the MEC paradigm, that is, the ability to run IT-based facilities at the edge of the mobile network and applying the concepts of cloud computing to perform specialized network tasks over multipurpose IT hardware resources [4].

The close proximity of MEC system to the end devices considerably improves the QoE (e.g., reduced latency) and offloads traffic from the backhaul and core networks. Additionally, bringing intelligence at the edge of radio access network (RAN) offers the capability to develop new types of context-aware and added-value services [7]. This ecosystem opens up new business opportunities and creates new value chains that supports the mutual coexistence of multiple actors such as Over the Top Players (OTT), service providers, mobile network operators, and different network, and IT equipment vendors.

Depending on the access technology, there are different ways to implement MEC. Outdoor macrocell deployments embed computing and virtualisation capabilities directly into RAN elements, for example, evolved NodeB (eNB). For indoor deployments, such as WiFi, MEC consists of powerful on premise gateways with dedicated intelligence.

ETSI MEC ISG has listed use cases [4] that can benefit from the deployment of MEC system, for example, distributed content and DNS caching, application-aware performance optimization, active device tracking, augmented reality content delivery, and video analytics. In Section 5, we will present two use cases to highlight the benefits of MEC from this perspective.

2.2. NFV. As mentioned above, cloud computing concept is at the heart of MEC. NFV [8] is a concept that provides guideline for NS/VNF implementation on standard commodity devices such as servers, switches, and routers. ETSI has put great efforts in standardizing the different aspects of NFV, for example, NFV Management and Orchestration (MANO) architecture [9]. From a general perspective, the MANO architecture is responsible for NS lifecycle management, that is, to instantiate, terminate, monitor, scale individual VNFs as well as VNF chains (NS). NFV has been significantly pushed forwards by industry, especially telecom network operators, in recent years, aiming for benefits such as the following:

(i) Innovative services and applications, such as targeted services based on geography or customer sets

(ii) Dynamic and flexible service chaining of VNFs

(iii) Reduced equipment costs (CAPEX) and OPEX (e.g., lower power consumption, minimum and maintenance cost)

(iv) Leveraging on standardized functional components and interfaces

(v) Reduced time to market

(vi) A new actor added in the business value chain, that is, VNF provider.

2.3. SDN. Besides NS lifecycle management, communication and networking between VNFs in a chain are an important task on MEC ecosystem. Intercepting traffic and forwarding it to the correct NS is a challenge that can be solved with the help of SDN [10] technology. SDN abstracts the network infrastructure from applications and enables network programmability by decoupling the network control plane from its data/forwarding plane. The network control intelligence
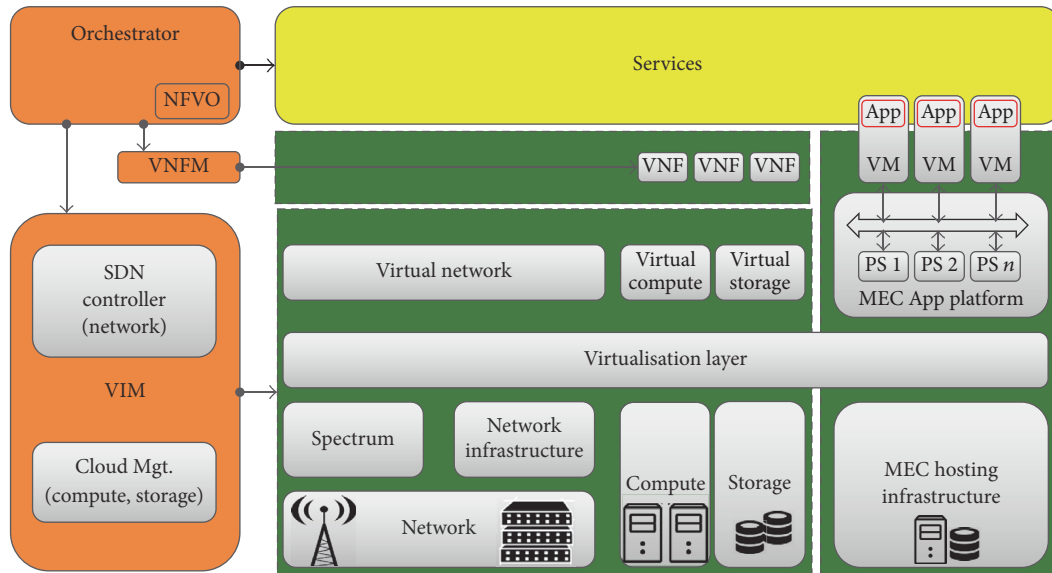
Figure 1: Overall scenario of the mobile cloud environment.

is logically centralised in SDN controller that has a global view of the underlying network (consisting of physical or virtual nodes). By taking advantages of standard southbound protocols such as OpenFlow [11], SDN greatly simplifies the control and management of the heterogeneous network devices [11] and enables vendor-agnostic programmable network environment. In the same way, on the MEC ecosystem, SDN can be employed to control and manage the network that interconnects distributed MEC servers and storage, establishing dynamic and on-demand network connectivity for chaining the VNFs and providing services. SDN will bring the following benefits for the MEC ecosystem:

(i) Rapid innovation through northbound APIs to support Third-party applications

(ii) Centralised control and management of network devices from multiple vendors

(iii) Global view of network status

(iv) Automated establishment and fast (re-)configuration of on-demand programmable network connectivity

(v) Fine granular network control with the ability to apply policies at different layers

(vi) Multitenancy support by network virtualisation

(vii) Improved QoE levels.

Besides that, in an advanced form, a cloud SDN controller can be merged with the RAN data/control plane logic. In this case, a single SDN controller is able to handle all data forwarding responsibilities, for example, from eNB to the network core or over the cloud-enabled RAN (e.g., to support MEC edge services). This can significantly improve networks flexibility and agility, which leads to better QoS/QoE. In this sense, some techniques such as channel state information (CSI) can be adopted to enhance radio resource utilization in edge services.

3. Overall Scenario

3.1. Overall Architecture. Figure 1 shows a mobile cloud environment. Compute and storage resources are generally located in remote centralised data centres, while end users are interconnected to those resources through diverse network infrastructures (e.g., IP, Ethernet, and optical). Alternatively, IT assets can be also hosted at the mobile network edge (e.g., at a LTE eNB) [12].

To address the limited IT resources per virtualised node (LTE eNB), clustering them has been proposed. Over the heterogeneous physical infrastructure/resources, the virtualisation layer (e.g., Hypervisor) enables the creation of virtual resources (incl. network, compute, and storage). Enabled by NFV, VNFs can be instantiated on the commodity servers, switches, and routers, which opens great opportunities to new and emerging innovative services. The MEC application platform is sitting on top of all the available resources (physical or virtual, network or compute/storage) to provide MEC platform services such as Radio Network Information Services (RNIS) and to enable third-parties to develop MEC applications through standard APIs.

In order to efficiently control and manage this heterogeneous cloud environment, an effective control and management system becomes essential. Industry has made great efforts in pushing forwards technologies such as SDN controller [11] and NFV MANO (Management and Orchestration) [8]. To better fit into the mobile cloud environment, the functionalities of SDN and NFV also need to be integrated seamlessly.

In Figure 1, we show a possibility for the mentioned seamless integration. The Virtualised Infrastructure Manager (VIM) is managing both virtual and physical resources, containing Cloud Management System (Cloud Mgt., e.g., OpenStack [13]) and SDN controller (e.g., OpenDaylight [13, 14]) that is in charge of the programmable network interconnections.
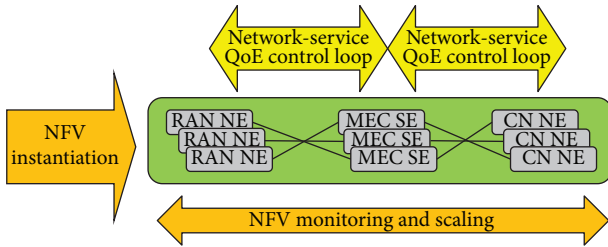
Figure 2: Overall edge service management framework.

On top of that, at the logical level, the VNF Manager (VNFM) is taking over the lifecycle of individual VNFs. NFV Orchestrator (NFVO), adopted from MANO, provides an overall service management view, including lifecycle events of NS (chain of VNFs) plus high level radio management decisions (e.g., some centralised Self-Organising Network (SON) features).

3.2. Edge Service Management. The proposed system integrates the management of network elements (NE) and Service Elements (SE) at different levels, as depicted in Figure 2. In this sense, the concept of network service involves both network and service instances, deployed in runtime over the available general-purpose infrastructure, which allow provisioning the target service with cross-domain adaptive capabilities.

First of all, the NFV management elements are in charge of receiving a new service request and instantiating the required NEs and SEs to cope with the request. This set of NEs and SEs composes a new and isolated network slice. In the example provided in Figure 2, the network service requires a series of VNFs acting as RAN NEs, some VNFs to cover the Core Network (CN) functionalities, and a series of VNFs that implement the required MEC-enabled edge service instances or MEC SEs. The instantiation of this network service slice involves both NFV and SDN procedures, in charge, respectively, of creating the VNFs and of dynamically applying the traffic steering rules to transport the data flows over the most convenient data paths.

Once the basic network service slice is created and activated, all the NE and SE nodes are endowed to interchange some control information in order to optimally provision the target service in multiuser scenarios, over the allocated underlying resources. This feature is especially relevant in mobile networks, where the limited radio resources are shared among multiple mobile users that experience variable radio quality conditions over time. In order to properly control and manage the QoE in the network slice, the network elements need to be adapted to the different service requirements and the service elements need to dynamically adjust the service configuration to the network capabilities over time.

Finally, the system needs to take control over the achieved QoE over the service lifetime. It may happen that the allocated infrastructure resources are not enough to comply with the service QoE requirements, for example, due to an unexpected increase of users. For instance, an increased number of mobile users with low channel quality may require additional hardware resources to implement real-time transcoding of multimedia streams. In those cases, the continuous NFV monitoring capabilities can react by temporarily reallocating extra resources to the network service.

4. Orchestrator Architecture and Operation

4.1. Orchestrator Architecture and Its Key Components and Interfaces. By taking advantages of the control and management capabilities of SDN and NFV, we are able to control and manage the heterogeneous resources and VNFs. However, in order to realize a holistic MEC ecosystem that supports services in a dynamic and optimal manner, we will need a "brain", that is, the Orchestrator, that has a global view of the status of the resources and VNFs and is able to produce intelligence for managing services (i.e., instantiating, terminating, modifying, monitoring, and charging). The efficiency of orchestration is extremely important for the MEC ecosystem. Otherwise, the low latency benefit brought by MEC will be compromised.

A reference architecture of the Orchestrator is shown in Figure 3, with its key components and interfaces. We try to reuse the already defined standard NFV reference points/ interfaces [8] (i.e., Or-Vi, Or-VNFM, NFVI, and Ve-VNFM-vnf) and repositories (i.e., NS catalogue, VNF Catalogue, NFV Instances, and NFVI Resources). For the MEC ecosystem, the messages being exchanged over the interfaces can be customized to better suit MEC with the potentials to further improve the system efficiency. For instance, the NFV service chaining on the radio processing could use a relatively fixed descriptor (PHY→MAC→RLC→PDCP→DPI). The Orchestrator will also need to be able to interact with the legacy OSS/BSS and the orchestrators of other systems (e.g., backhaul, core, and remote data centres) that are in similar or equivalent positions.

Inside the Orchestrator, we include nine key components that are more service-oriented, as shown in Figure 3. Note that the NFVO is not separately specified; rather an overall system Orchestrator is presented here. Some components can be merged due to implementation considerations.

The service requests will be first handled in a Request Handler or a Portal. The SLAs and Policies will be checked via SLA Manager (SLA Mgr.) and Policy Manager (Policy Mgr.), respectively. Taking the information exposed from VIM and VNFM about the available resources and VNFs, the service chaining and Service Mapping modules will be able to design/create NFV service chains and map the created services to the underlying available resources. Here the mapping is still on the service level, which only specifies the required resource type and amount. The actual resource mapping will be performed by VIM and VNFM. Service Lifecycle Manager (Service Lifecycle Mgr.) is responsible for the lifecycle management of the services such as service instantiation, modification, operation, and termination. During the service modification, the Service Scaling module will be called to perform the optimal scaling against dynamic demands. The services will be monitored during their lifetimes, in terms of, for example, performance and security. Accordingly, services will be charged against SLAs.
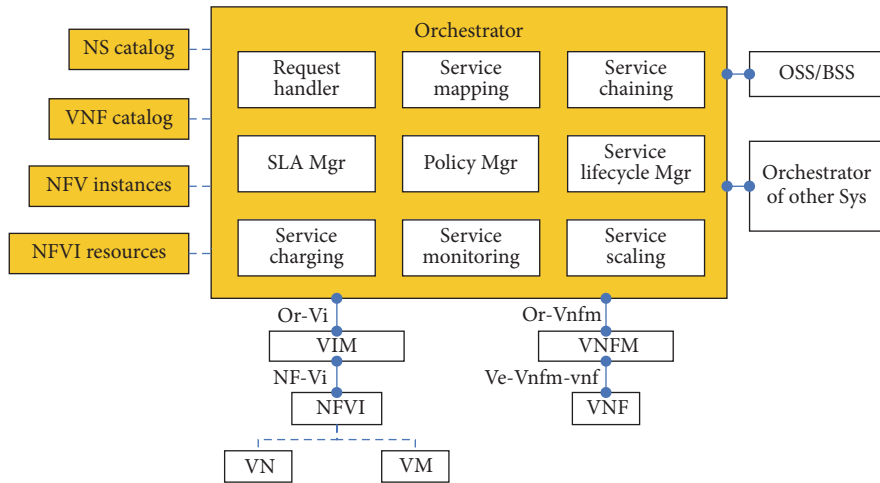
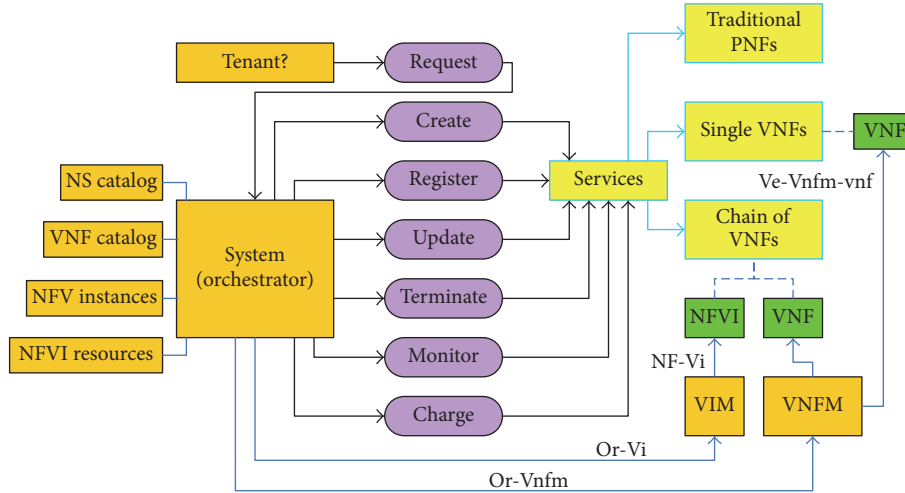FIGURE 3: Orchestrator architecture (Incl. key components and interfaces).

FIGURE 4: Lifecycle management of both services and resources.

4.2. Lifecycle Management of Services and Resources. In this section, to further justify the specified functional modules inside the Orchestrator and their internal interactions as well as externally with other elements in the MEC ecosystem, we created a number of lifecycle management flows for both services and resources. In Figure 4, the key lifecycle management actions and the involved elements are captured, that is, a *tenant* "Requests" the *System (Orchestrator)* to "Create," "Register," "Update," "Terminate," "Monitor," and "Charge" *Services* (incl. those that are composed with traditional Physical Network Functions (PNFs), a single VNF, or a chain of VNFs). The tenant here covers a wide range (e.g., end users, service providers, or virtual network operators), which is dependent upon certain situations.

The particular lifecycle management flows are further detailed in the following subsections (i.e., (A) Service Instantiation, (B) Service Modification, (C) Service Operation, and (D) Service Termination).

4.3. Service Instantiation. We divide the service instantiation into three stages, that is, Service Request, Service Provisioning, and Service Activation, as shown in Figure 5.

4.3.1. Service Request. The tenants send service requests to the system (Orchestrator in particular). The services are described by using a predefined/standard template with metadata (e.g., SLAs) specified.

4.3.2. Service Provisioning. When the Orchestrator receives the requests via the Request Handler, it will first perform the security check (e.g., authorization) as well as SLAs and Policies check by interacting with the SLA Mgr. and Policy Mgr., respectively. If the service request passes the checks, it will be further processed. Otherwise, the request will be blocked at this stage.

To start the process, the tenant, by going through the NS catalogue, will be exposed with all the available services that
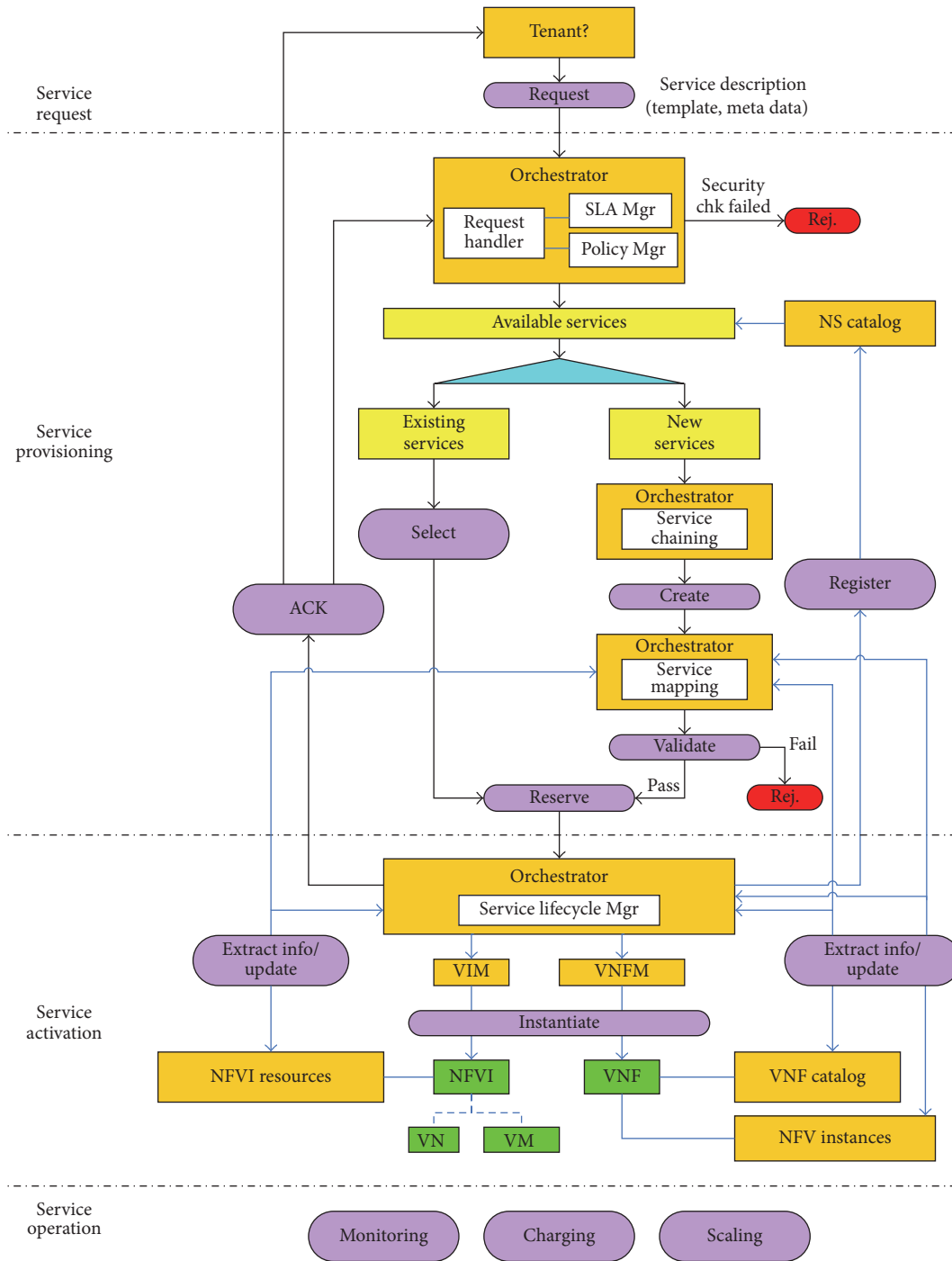
FIGURE 5: Workflow of service instantiation.

are currently being supported by the system. Two situations can be faced by the Orchestrator.

(i) If the requested service already exists, it will be selected from the NS catalogue, and the required amount of resources will be reserved.

(ii) If the requested service is new to the system, it will be created by calling the service chaining module in

the Orchestrator. A VNF forwarding graph (FG) will be designed if multiple VNFs are needed to form a service chain. Then the VNF FG will be forwarded to the Service Mapping module to get validated through mapping the service against the available resources and VNFs, which are exposed in the repositories (e.g., NFVI Resources, VNF Catalogue, and NFV Instances defined by NFV). If the validation is successful, the
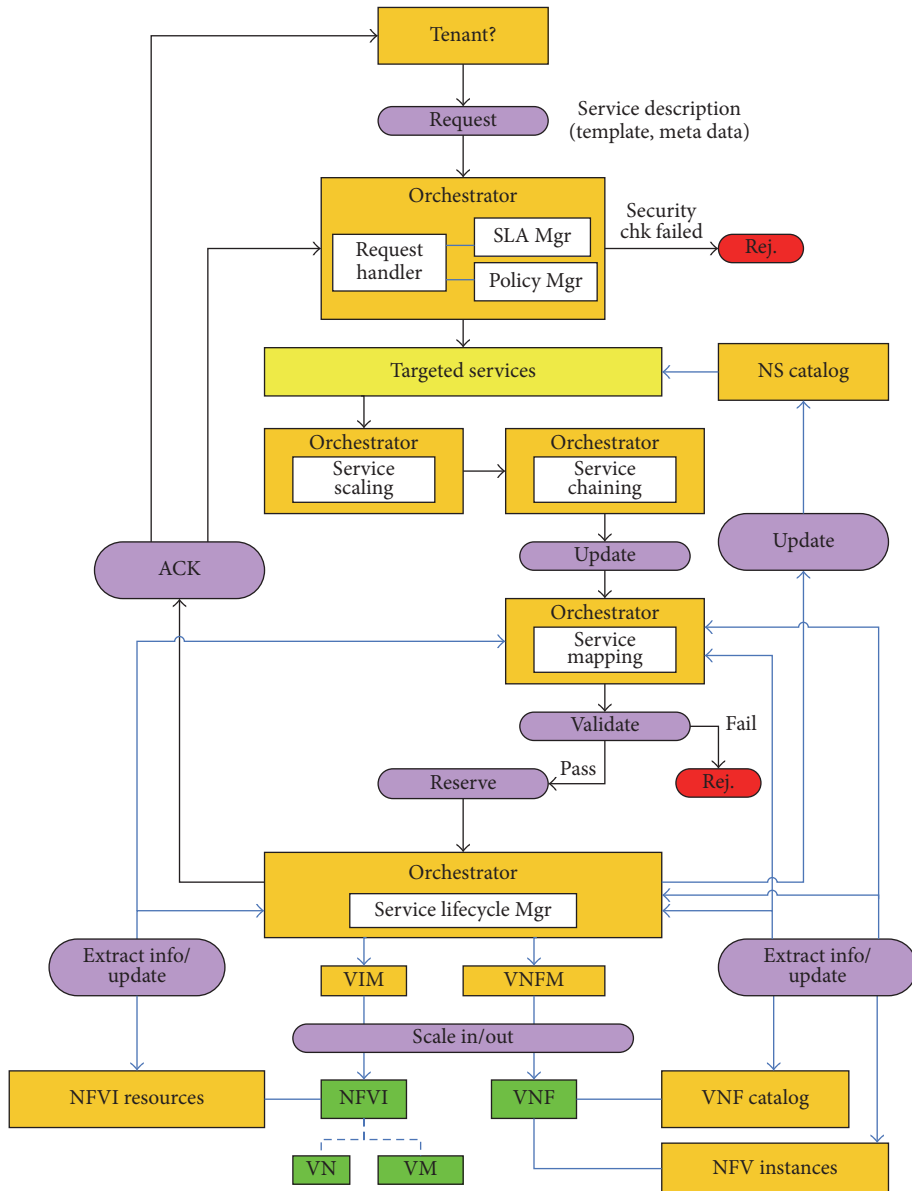
FIGURE 6: Workflow of service modification (scaling in/out).

relevant elements will be notified and the required amount of resources will be reserved. Otherwise, the service request will be rejected.

4.3.3. Service Activation. At this stage, in order to activate the requested service, the Orchestrator will interact with external elements such as VIM and VNFM to instantiate virtual resources and VNFs through its Service Lifecycle Mgr. module. Accordingly, the relevant resource/service repositories will be updated.

Once the service is activated and up running, it enters to the Service Operation stage that mainly involves Service Monitoring and Charging. In the MEC ecosystem, such a dynamic environment, the services will need to be modified/updated in a high frequency to fulfil the tenants'

constantly changing requirements and utilize the limited resources at the edge of the mobile network in an optimal manner.

4.4. Service Modification. The service modification could be requested by either the tenant or the system itself. In Figure 6, we show that the request is raised by the tenant, for which the security check against the modification needs to be performed again. However, it can be easily applied to the other situation, for which the process will start with finding the targeted services from the NS catalogue.

Once the service that needs to be modified is identified, the process will first go through the Service Scaling module that specifies the modification parameters (e.g., bandwidth, compute/storage capacity, and function update), which may
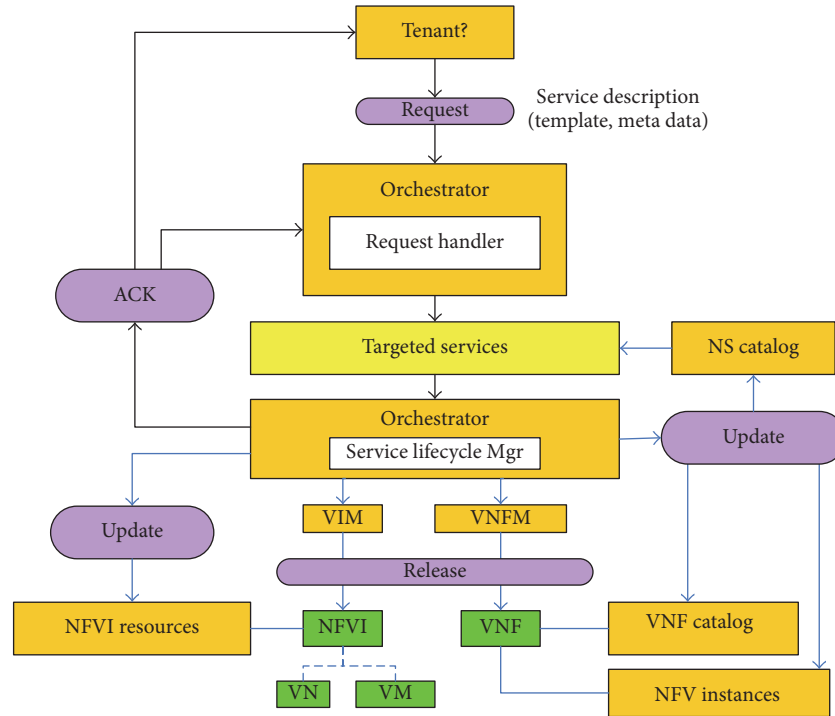
FIGURE 7: Workflow of service termination.

lead to different types of modification such as updating VNF chain by adding/dropping VNFs and updating network connectivity. The service chaining module is called to update the original VNF chain, and then the Service Mapping module to perform the updated mapping against the currently available resources and VNFs in order to validate the modification. If the validation is passed, the Orchestrator will start interacting with VIM and VNFM to scale in/out the virtual resources and VNFs and update the relevant repositories, as shown in Figure 6.

4.5. Service Operation. When the service is finally up and running, the tenant will be charged against the signed SLAs (context-aware KPIs), which can be taken care of by the Service Charging module. The Service Monitoring module will be used to monitor the service in terms of, for example, its performance during its lifetime.

4.6. Service Termination. As shown in Figure 7, once the tenant sends a request to terminate a service, the Orchestrator will search for the targeted service within the NS catalogue in order to obtain the information about the involved elements, resources, and VNFs. Then the Service Lifecycle Manager in the Orchestrator will interact with the VIM and VNFM to release the corresponding resources and VNFs and update the relevant repositories.

In the next section, we describe a use case focusing on IoT to show how the MEC ecosystem could be deployed in the real practice and what benefits it can bring to IoT and explore the key challenges, especially on the part of orchestration.

5. Potential Use Cases

MEC arises as a significant step towards provisioning of a more personalised mobile service. The potential benefits of using MEC instances both in the uplink and downlink directions are significant.

5.1. Uplink Direction MEC Use Cases. Figure 8 illustrates a potential use of MEC to deploy a data preprocessing instance in the uplink direction of the RAN. Massive IoT, either active or devices such as sensors and machines, will compose a significant integral part of 5G infrastructure. Besides the internal Machine-to-Machine (M2M) communications, the devices will collect huge amount of data (e.g., from sensors) regarding themselves and/or their surrounding environments (e.g., temperature and humidity). The data traditionally is sent to a remote data centre (cloud) for further analysis and process, which may lead to IoT devices reacting to an event. Considering the huge number of connected devices and the increasingly complicated tasks posed to IoT devices nowadays, centralised remote analysis of IoT inputs may not result in appropriate actions to different events.

An improved QoS in this type of service would involve many features like latency, security, and so forth. Naturally, by placing intelligence (e.g., data processing services) at the edge of network and in the proximity of IoT devices, fast responses to the real-time events can be more effectively supported. MEC IT resources can provide edge data processing and decision-making. This will offload traffic from the central processing and the backhaul/core network. The configuration of IoT devices can also be done at the edge
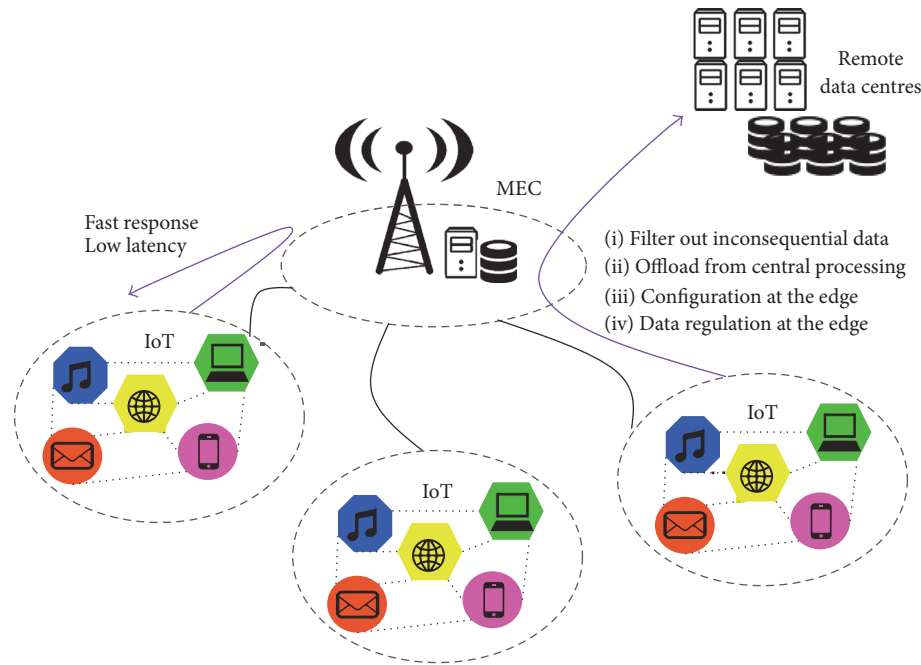
FIGURE 8: Orchestration enabling effective interactions between MEC and IoT.

by taking advantages of the MEC. It will potentially make the IoT devices simpler, smaller, and cheaper. By adopting intelligent orchestration systems, IoT domains will become less vulnerable and more secure in a cost-effective manner.

5.2. Downlink Direction MEC Use Cases. Besides the mentioned improvements in the uplink, MEC instances can also play a significant role in the dynamic management of data transmissions in the downlink, leading, for example, to an enhanced adaptation of multimedia flows based on the contextual information of the different devices or users. The converged radio-cloud Orchestrator may use contextual information to improve QoS/QoE via a coordinated service-radio resource management. One among many possibilities for MEC services is the use of the channel state information (CSI) to provide a good estimation of the radio channel quality that each user is experiencing. This process is performed within the RNIS (Radio Network Information Service) and is integrated as one of the modules of the MEC Application Platform (see Figure 1). This service is implemented as a VNF and is available to the rest of the MEC services of the RAN.

Taking this CSI into account, a MEC service instance would be able to adapt the transmissions and the application-level configuration to the specific data rate achievable by each mobile user. But the use of CSI in RAN is nontrivial, since it depends on the deployment alternatives of the eNB.

To better explain this specific use case, next section provides an overall review of the channel quality reporting in 4G Long Term Evolution (LTE) networks, to later focus on the implications of sharing this type of information in different MEC service deployment schemes.

6. Use of CSI in MEC: Performance Analysis

Previous sections have dealt with the most relevant concepts related to mobile edge service instantiation and showed the potential benefits for QoS/QoE management with a proper use of user contextual information during runtime operation.

This section studies the performance impact of a MEC RNIS service that gathers and analyses CSI data to be used in the service-level VNFs. As an example, the RNIS information could be used by a virtual transcoding MEC service that adapts the transmissions and higher-level configuration of multimedia flows to the channel status of the different mobile users, taking into account the overall radio resource allocation. In order to fulfil this objective, the IoT-enabled radio base stations need to share some channel quality information with the MEC platform. This issue is nontrivial since different eNB virtualisation/deployment options put some conditions on the way to access CSI data with different performance implications.

First, we describe how UE-generated CSI data is used in currently deployed 4G mobile broadband networks, in order to aid the radio scheduling process. Thus, the first subsection described the most relevant CSI parameters to be considered in different 4G technology options.

Second, we analyse how these CSI data can be accessed from the MEC platform in order to enable the common RNIS service that feeds the different service VNFs. Taking into account the current virtualisation/centralisation trends, we analyse different eNB virtualisation alternatives and the impact on the availability of CSI at the MEC platform.

Next, we provide more detailed information concerning the use of CSI data in mobile edge services with special focus on the trade-off between CSI data accuracy and the

associated network overhead in the mobile fronthaul, taking into account different LTE and LTE-A deployment options.

6.1. Channel State Information in Mobile Networks. In modern mobile broadband networks, the radio resource management (RRM) and other functions are placed at the eNB with direct access to feedbacks provided by UEs in the form of channel state information (CSI). In LTE networks, the eNB is able to perform adaptive modulation and coding (AMC) in order to dynamically adapt the radio transmissions to the actual channel quality experienced by UEs [15].

However, using perfect AMC with ideal CSI knowledge is unrealistic in real-world deployments, since this would entail a series of performance bottlenecks especially in the UL reporting channels. CSI reporting is a complex subject with multiple alternative configuration options, which have evolved along the different 3GPP releases in accordance with the evolution of transmission schemes supported in LTE. Each transmission scheme requires a specific set of CSI data to perform AMC and scheduling of radio resources [15].

In brief, CSI reports generated by UEs can be periodic (sent to the eNB with a configured CSI reporting periodicity) or aperiodic (upon request by the eNB).

In addition to reporting periodicity, the amount and type of CSI information determines the trade-off between reporting granularity and overhead. Different types of CSI parameters are required in different transmission schemes, and CSI information may be provided as wideband (WB) or subband (SB) information. These parameters directly impact the expected Block Error Rate (BLER).

The most relevant CSI parameters for selecting the target data rate in the next Transmission Time Interval (TTI) are as follows:

(i) Channel Quality Indicator (CQI) is a measure of the coding efficiency expected to be supported by an UE, based on actual channel quality measurements in the current TTI.

(ii) Rank Indication (RI) is used in Multiple Input Multiple Output (MIMO) schemes to determine the number of simultaneous transmissions that can be received at the same time in the current conditions.

The combination of CQI and RI values determines the target bitrate at each moment. Taking into account the most typical transmission schemes used in currently deployed LTE networks, different level of granularity is required for the CSI processes:

(i) Single Input Single Output (SISO) lead to only one data transmission each time. Therefore, only one CQI value is required per UE and per considered bandwidth part.

(ii) In MIMO with Transmission Mode 3, the eNB is able to transmit two signals at the same time for each UE, containing either the same data (for redundancy and enhanced resiliency) or different data (for increased data rate). In this case, the UE needs to report the RI value (RI = 1 for redundancy and RI = 2 for increased
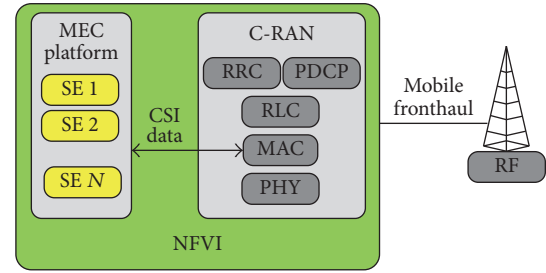

FIGURE 9: Fully centralised mobile network architecture.

data rate) and a single CQI value per bandwidth part. Thus, both RI = 1 and RI = 2 lead to the same amount of data bits sent in each transmission.

(iii) In MIMO with Transmission Mode 4, the transmission scheme for RI = 1 takes the same input CSI parameters, that is, RI = 1 and a single CQI per bandwidth part. However, the transmission scheme associated with RI = 2 requires two different CQI values per transmission layer, since each layer will be adapted to transport the maximum achievable number of bits at each moment.

(iv) LTE Advanced (LTE-A) systems introduce the possibility of aggregating the radio transmission capabilities of different LTE carriers of the same operator through Carrier Aggregation (CA) techniques. In this case, the UE must report the CSI data of the different carriers to the primary cell, with the consequent increase in the uplink signalling overhead. To cope with this problem, some reference configurations [16] propose to space the periodic reporting samples (from 5 ms, for one carrier, to 10 ms with two carriers and 30 ms with three carriers).

Although the CSI reporting mechanisms are designed to feed the LTE scheduler in its RRM decision-making process, the possible use of this type of low-level information at higher layers of the protocol stack has been widely studied. For example, a multimedia server can adapt the served quality levels of the different media streams according to the estimated achievable data rate, or the operation of transport protocols can be fine-tuned to the current radio conditions of the different users.

6.2. Access to CSI Data from MEC Instances. To gain a proper level of personalisation in the service delivery, MEC service instances, for example, a virtual transcoding unit, need to access the CSI feedback of the different mobile users to some extent. Generally speaking, ETSI MEC principles propose to access this information directly from the eNB nodes rather than from the UEs. Depending on the deployed C-RAN architecture, this information exchange may impose different type of requirements in the interfaces between the eNB and the MEC platform.

Figure 9 illustrates the case of a fully centralised Cloud-RAN (C-RAN). The eNB functionalities are split into a Base-Band Unit (BBU), which runs in the NFVI, and the Remote
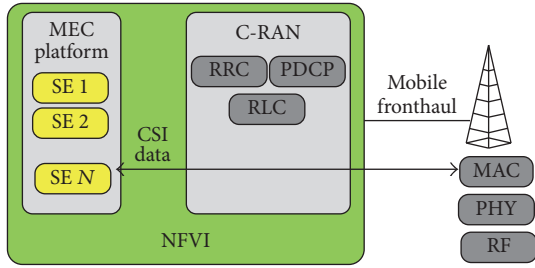
FIGURE 10: Partially distributed mobile network architecture.

Radio Header (RRH), which is deployed in remote locations for performing the radio transmissions. Considering that the MAC layer resides in the BBU and the MEC platform is deployed over the same shared NFVI, the interface for accessing the required CSI data can be a simple API.

Figure 10 depicts the case where the functional split of the eNB is made over the MAC layer. In this case, the low-level CSI data remains at the RRH and therefore accessing to CSI data from the MEC platform consumes part of the capacity of the mobile fronthaul.

In the latter situation, the trade-off between CSI accuracy and network overhead is a key metric of the overall system performance. A higher degree of granularity on CSI data may lead to optimized traffic delivery from the MEC service instances, at the cost of increased overhead in the mobile fronthaul. On the contrary, a lower degree of CSI data would reduce the overhead while the service-level performance may still be accurate enough. Therefore, the optimal operational point eventually depends on the specific network and services.

6.3. Use of CSI for MEC: Required CSI Data Volume. This section analyses the required CSI data volume to feed the RNIS module in both centralised and distributed C-RAN.

According to Figures 9 and 10, the propagation of CSI data from the eNB to the MEC platform introduces some level of overhead in the system. In the first case, this feedback entails copying data within the NFVI, while the second case imposes some data transfer through the network links.

Depending on the LTE/LTE-A transmission scheme, the UEs are configured to send different amount of CSI data to the eNB. In this paper, we only consider the CSI parameters that determine the achievable bitrate for each user, namely, the RI and the CQI values. Other CSI information related to the coding of transmitted signals is thus not considered. Additionally, it must be noted that the following numbers are obtained for an LTE bandwidth of 20 MHz.

(i) SISO transmissions:

    (a) WB: a unique CQI value for the whole bandwidth, resulting in 4 bits per CSI report;

    (b) SB: a unique CQI value of 4 bits per bandwidth part and the subband identification of 4 bits for each of the 4 subbands, leading to a total of 32 bits per CSI report.
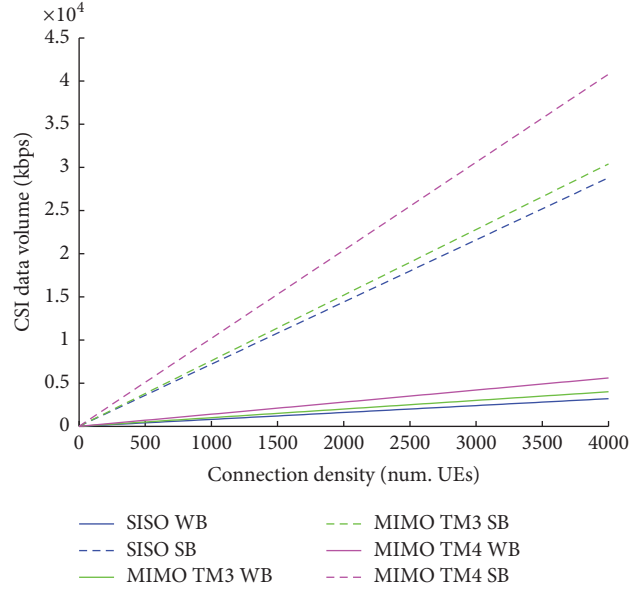

FIGURE 11: CSI data volume at CRR of 5 ms with different LTE transmission schemes and increasing number of mobile users.

(ii) MIMO transmissions with TM3:

    (a) WB: a unique CQI value for the whole bandwidth, resulting in 4 bits per CSI report. 4 bits of CQI value, with 1 additional bit for RI parameter, resulting in 5 bits per CSI report;

    (b) SB: a unique CQI value of 4 bits per bandwidth part and the subband identification of 4 bits for each of the 4 subbands, with 1 additional bit for RI parameter, leading to a total of 33 bits per CSI report.

(iii) MIMO transmissions with TM4:

    (a) WB: two different CQI values for the whole bandwidth, a reference CQI value of 4 bits, and a differential CQI value of 3 bits. Together with the required RI value, this leads to 8 bits per CSI report;

    (b) SB: considering the two CQI values per subband, the total amount of CSI data is increased to 45 bits per CSI report.

Taking into account the fact that each periodic CSI report is sent by each UE either at the configured CQI Reporting Rate (CRR), usually every 5 ms, Figure 11 illustrates the increase of CSI data volume for the considered transmission modes as the number of mobile users in the cell increases in a fully centralised mobile network architecture. The represented data is the result of adding CSI data volume for the different users with the respective reporting periodicity.

As it can be observed, the increased connectivity density (considered here up to 4000 mobile UEs) leads to considerable volume of aggregated CSI data (up to 45 Mbps for MIMO TM4).
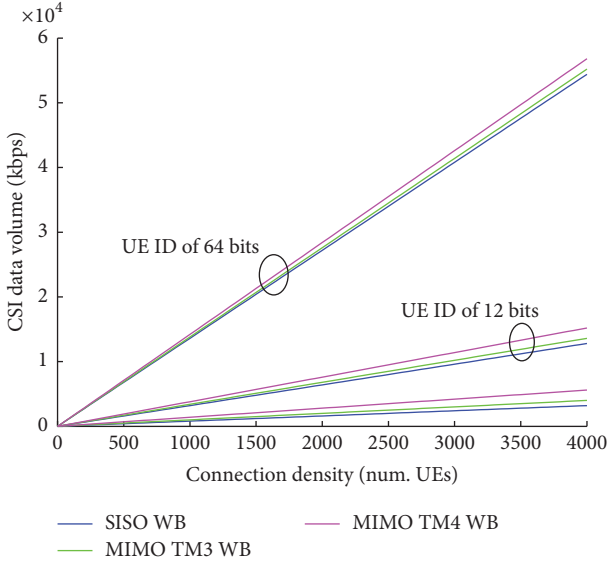
FIGURE 12: CSI data volume at CRR of 5 ms with LTE MIMO TM3, additional user identification bits, and increasing number of mobile users.

FIGURE 13: CSI traces used: evolution of CQI over time.

FIGURE 14: Average throughput miss-estimations for "class-1" and MEC CRR based on instantaneous CQI values.

In distributed deployment scenarios, the CSI feedback needs also to include some identification for the different UEs. In the considered example scenario, a complete differentiation for 4000 devices would require 12 extra bits per CSI report. In a more general case, the use of the international mobile subscriber identity (IMSI) would require 64 bits of additional data for each CSI report.

As a result, Figure 12 illustrates the evolution of the required CSI data volume from the eNB to the MEC platform for up to 4000 mobile devices.

In this situation, the required CSI reporting data would reach 50 Mbps just for the wideband reporting and up to 90 Mbps considering subband reporting.

Generally speaking, the MEC service instances (e.g., the transcoding unit) will use user context information for a high level handling of traffic flows at transport or application levels. For that aim, a rough estimation of the overall instantaneous achievable throughput for each device would be enough, and therefore the averaged CQI information as provided by the wideband CQI could provide appropriate estimations.

Similarly, the granularity of CSI feedback reports from the eNB to the MEC platform is subject to analysis and highly depends on the dynamics of the radio channel quality.

7. Use of CSI for MEC: Performance Results

In order to evaluate the reliability of the macroscopic CQI samples for service-level management, we propose a simple exercise where a RNIS service instance is able to capture CQI values from the eNB at determined sampling rates. Therefore, the eNB is able to perform the channel-aware adaptive modulation and coding based on the fine grain CRR of 5 ms, while the service instance is able to adapt the transmissions to the user based on the coarse grain CRR in the range of 250 ms to 4 s.
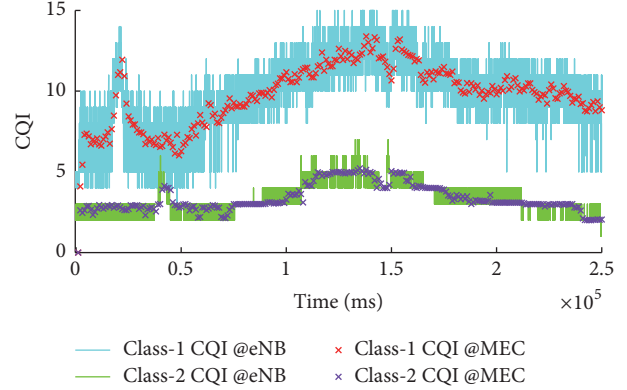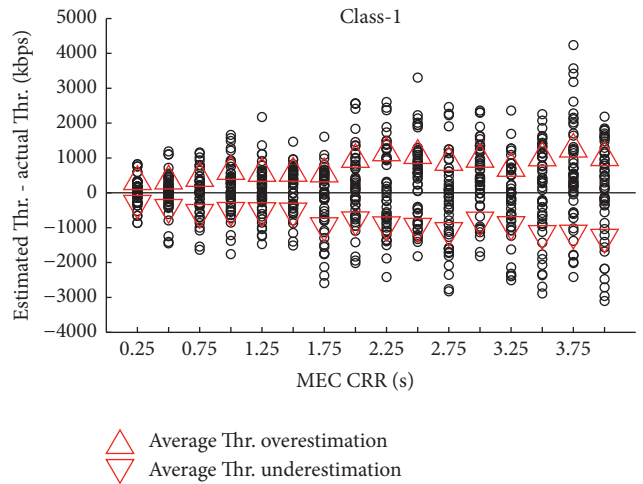
In this exercise, we consider an LTE cell of 20 MHz, with MIMO TM3 transmission scheme of up to 150 Mbps in downlink. In order to inspect the feasibility of reducing the CSI reporting rate at the MEC RNIS instances (MEC CRR), we use a real-world CSI sample trace obtained from live networks [17]. Figure 13 illustrates two different CQI samples. The "class-1" user is associated with a mobile user that wanders around the cell and experiences a high variability of CQI values. Meanwhile, the "class-2" user is associated with a lower variability of the radio quality conditions. For the two conditions, Figure 13 also illustrates the result of sampling the CSI data at a periodicity of one second.

Figures 14 and 15 show the MATLAB simulation outcomes of applying the delayed CQI estimations at MEC CRR level compared to the actual CQI values in the "class-1" radio channel.

In this example, CQI values are generated by double sampling process. First, the UE feedbacks the median CQI values with CRR = 5 ms, and the eNB generates a CQI value at MEC CRR = 1 s by simply copying the instantaneous available CQI value. The effects of inaccurate CQI estimation are just gauged in terms of expected achievable throughput and actually
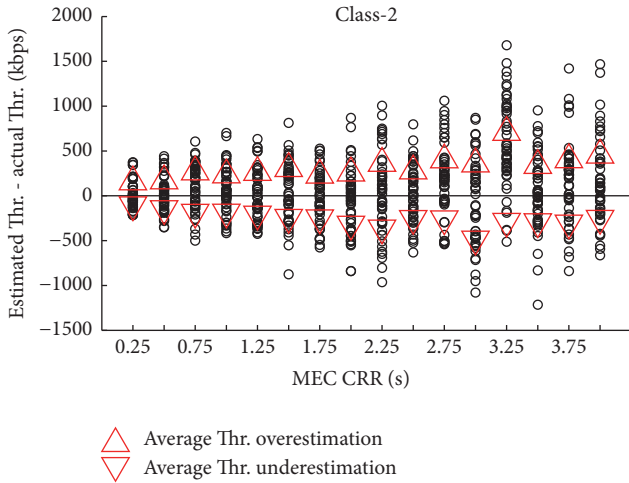
FIGURE 15: Average throughput miss-estimations for "class-2" and MEC CRR based on instantaneous CQI values.

FIGURE 16: Average throughput miss-estimations for "class-2" and MEC CRR based on instantaneous CQI values.

achievable throughput. Since greedy traffic sinks are used for single-user analysis, this difference of throughputs is computed for the whole train of TTIs and the average difference value is obtained.

The average differential throughput over the whole CQI trace is plotted as a single point. For each coarse grain CRR value, a total of 50 points are represented associated with different initial TTIs within the trace. Additionally, the average of all results is illustrated for the positive values and for the negative values.

As can be observed, the overall trend is that increasing the MEC CRR leads to higher average miss-estimations of the available throughout, and especially to higher variance in the results. Positive average values mean that the achievable throughput estimated by the service instance is higher than the actual achievable throughput, and therefore the eNB would need to cope with traffic demands that cannot be allocated to the mobile user. Negative average values mean that the estimated throughput values at the service instance are lower than the achievable throughput due to the channel quality, and therefore some radio capability would be unused in average.

A more detailed analysis of the average miss-estimations in terms of throughput signifies the difference between using instantaneous or median CQI values for CQI sampling at service level. Median macroscopic CRR values provide better estimations than individual macroscopic CRR values, enhancing the performance in almost 50% in many macroscopic CRR values.

Figure 16 illustrates the average miss-estimations in terms of throughput for both instantaneous and median CG-CQI values and for the two types of channels considered. In this case, median CG-CQI values are generated by computing the median value of the CQI values available at the eNB over 1 s period, which at the same time are median values of the UE-generated instantaneous CQI values over 5 ms periods.

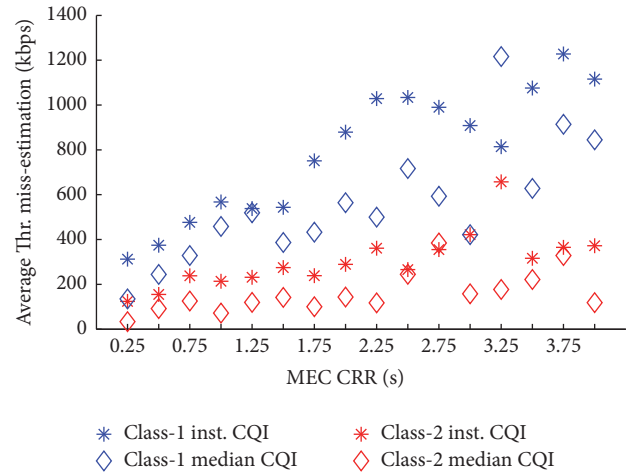Throughput miss-estimation is defined as the average of the absolute values of the average throughput resulting from each individual test, regardless they are overestimations or underestimations. It must be noted that throughput estimations are carried out for LTE users with MIMO transmission scheme and direct CQI to MCS mapping.

The results show that median MEC CRR values provide better estimations than individual MEC CRR values, enhancing the performance in almost 50% in many CG-CRR values. Comparing the two radio channels, the general trend is that throughput estimations are worse as the MEC CRR value increases. Yet, since the "class-2" channel is less variable and with lower CQI values, lower miss-estimations are found in the achievable throughput. Nevertheless, the same trend is observed concerning the benefits of using median MEC CQI values instead of instantaneous values.

8. Conclusions

The emerging 5G technology and MEC concept, leveraging on SDN and NFV technologies, are promising solutions able to fulfil the QoS requirements of diverse IoT verticals (e.g., low latency and huge transmission capacity). Effective orchestration is an essential element to form a holistic MEC ecosystem, which needs to be able to coordinate all the heterogeneous elements, resources, and services in such a complex and sensitive multivendor environment. We presented an overall MEC ecosystem scenario integrating the functional components and interfaces of SDN and NFV. A reference architecture of Orchestrator is given showing key functional modules and internal and external interfaces. A few lifecycle management flows of services and resources are depicted to further validate the proposed functional components in the reference architecture of the Orchestrator. Finally, a use case with a focus on improving QoS highlighted how IoT could benefit from the MEC system. In particular, the use case presented the potential benefits of the application of MEC both in the uplink and in the downlink and highlighted the implications of the access to CSI data from MEC RNIS instances, in order to enable other QoE-oriented

MEC services. After a performance analysis, the results show that median MEC CRR values provide better estimations than individual values and that the throughput estimations gets worse as the MEC CRR value rises.
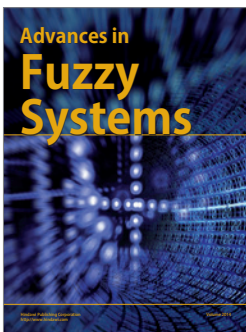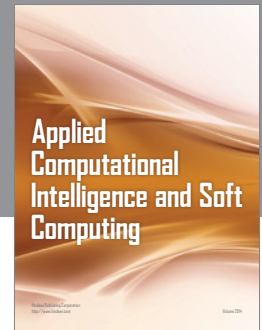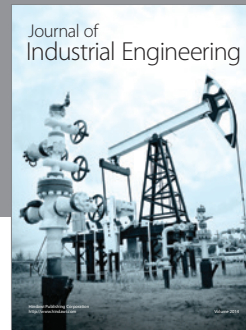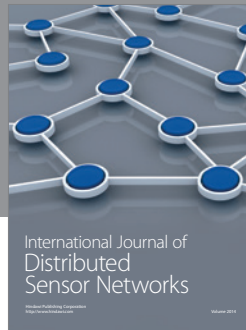
Competing Interests
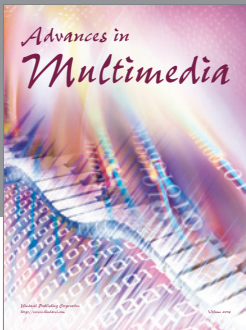
The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

References

[1] 5G Infrastructure Public Private Partnership (5G-PPP), https://5g-ppp.eu/.

[2] Fujitsu, http://science-innovation.co.uk/wp-content/uploads/2015/02/Fujitsu_UK_Ireland_corporate_brochure.pdf.

[3] 4G Americas, *5G Technology Evolution Recommendations*, 4G Americas, 2015.

[4] ETSI MEC ISG, "Mobile Edge Computing-Introductory Technical White Paper," no. 1, September 2014.

[5] 5G Infrastructure Public Private Partnership, *5G Vision: The Next Generation of Communication Networks and Services*, 5G Infrastructure Public Private Partnership, 2015.

[6] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing—a key technology towards 5G," ETSI White Paper 11, 2015.

[7] D. Soldani and A. Manzalini, "Horizon 2020 and beyond: on the 5G operating system for a true digital society," *IEEE Vehicular Technology Magazine*, vol. 10, no. 1, pp. 32–42, 2015.

[8] http://www.etsi.org/technologies-clusters/technologies/nfv.

[9] ETSI NFV ISG, "Network Functions Virtualisation (NFV); Management and Orchestration," ETSI GS NFV-MAN 001 V1.1.1, December 2014.

[10] https://www.opennetworking.org/sdn-resources/sdn-definition.

[11] Open Networking Foundation (ONF), "Software Defined Networking: The New Norm for Networks," April 2012.

[12] J. O. Fajardo, F. Liberal, I. Giannoulakis et al., "Introducing mobile edge computing capabilities through distributed 5G cloud enabled small cells," *Mobile Networks and Applications*, vol. 21, no. 4, pp. 564–574, 2016.

[13] https://www.openstack.org/.

[14] https://www.opendaylight.org/.

[15] 3GPP specification 36.213: Evolved Universal Terrestrial Radio Access (E-UTRA), Physical layer procedures, v 12.11.0, September 2016.

[16] 3GPP specification 36.101: Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception, v 12.13.0, October 2016.

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