JITEL 2017
XIII Jornadas de Ingeniería Telemática
27-29 Septiembre 2017, Valencia (España)

Actas de las XIII Jornadas de Ingeniería Telemática (JITEL 2017), Valencia (España), 27-29 de Septiembre de 2017.

ISBN: 978-84-9048-595-8
DOI: http://dx.doi.org/10.4995/JITEL2017.2017.6587

Quality of Service (QoS) oriented management system in 5G cloud enabled RAN

Rubén Solozabal, José Oscar Fajardo, Bego Blanco, Fidel Liberal
University of the Basque Country (UPV/EHU)

{ruben.solozabal, joseoscar.fajardo, begona.blanco, fidel.liberal}@ehu.eus

***Abstract*—This paper analyze different techniques to implement Quality of service (QoS) on multi-tenant 5G networks. Describes the architecture of the next generation mobile network based on cloud-enabled small cell deployments and also proposes an hybrid-cloud solution coexisting with centralized cloud RAN (C-RAN), in order to achieve a gradual implementation of the technology. In this context, the work here presented deals with the challenges of preserving the quality of experience in a multi-tenant cloud enable RAN bearing in mind the Key Performance Indicator (KPI) agreed in the Service Level Agreement (SLA). To achieve this goal, QoS should be managed at different levels of the architecture. Feedback should be given between learning modules in order to analyze the results and infer enhanced decision rules which may conclude in an architecture replacement.**

***Keywords*—Network Function Virtualization, Network Service instantiation, cloud-enabled small cells, QoS/QoE C-RAN, Mobile Edge Comuting**

I. Introduction

The flexible Radio Access Network (RAN) proposed in 5G will leverage Software Defined Network (SDN), Network Function Virtualization (NFV) and Mobile Edge Computing (MEC) principles for a simplified network deployment and management, enhancing CAPital EXpenditures (CAPEX) / Operating expense (OPEX) efficiency. Intelligent 5G centralized RAN systems will concentrate processing resources together in shared data centers not only in order to reduce deployment costs, but also to provide low latency connections between different RAN processing units. Traditional deployments of specialized devices with 'hard-wired' functionalities will be replaced by general-purpose reconfigurable computing assets. Making use of cloud computing, SDN and NFV, the centralized RAN will become a Cloud RAN (C-RAN) [1]. The softwarization of the network functions will enable the automation of network service provisioning and management, thus, the adaptation to growing and heterogeneous market requirements at a lower cost.

To facilitate the adaptation of the current architecture to the proposed, SESAME project[2] analyzes the development of multi-tenant cloud enabled RAN (C-RAN) through the evolution of the architecture of traditional commercial Small Cells (SC) to Cloud Enabled Small Cells (CESC). A CESC is a multi-operator SC that integrates a virtualized execution platform to support the executions of novel applications in the network edge using NFV and SDN technology. The proposed solution extends the Small Cell as a Service (SCaaS) model, which provisions of shared radio access capacity to mobile network operators in localized areas. Efficient management of resources, rapid introduction of newer network functions and services, easy of upgrade and maintenance and CAPEX/OPEX reduction are only few examples of various benefits that the proposed solution provides.

Despite the potential technical benefits, viability of the solution strongly depends on several factors such as the guarantee of the Service Level Agreements (SLAs). A SLA which captures the particular Key Performance Indicators (KPI) of a delivery –scope, quality, and responsibilities– can play a significant role towards business success. In this paper several techniques are proposed at different levels of the next generation mobile architecture in order to improve the overall quality of experience.

This paper is organized in six sections. First, Section II deals with cloud architectures evolution. Then, Section III analyzes the service provisioning models over the proposed architecture and Section IV discusses the definition of network services. Next, Section V proposes hybrid cloud approach for the evolution to 5G C-RAN and Section VI introduces the implementation of a QoS into the service life-cycle management. Finally, Section VII summarizes the main conclusions.

II. Multi-tenant Cloud Enabled RAN

Traditionally, actual installation of physical infrastructure is needed to provide coverage in one Point of Presence (PoP). Such an ownership increases operator's CAPEX and significantly hampers business agility, particularly when considering the high degree of cell densification. In a multi-tenant scenario, an infrastructure provider can grant

EDITORIAL UNIVERSITAT POLITÈCNICA DE VALÈNCIA

access to third parties such as network operators, service providers or Over-The-Top (OTT) players. Sharing the physical infrastructure increases service dynamicity and reduces the overall cost and energy consumption compared to the case where parallel systems are installed in one PoP to support connectivity for different parties.

Beyond the pure centralization of an eNB functions, one of the emerging technologies to cope with more personalized and user-centric service provisioning is the novel MEC. This may be exploited to deploy proximity-enabled services with close-to-zero latency characteristics. Regardless of the adopted architecture for C-RAN, MEC-driven service instances must be deployed over the cloud resources available at the RAN side.

In this centralized solution, the upper RAN functions are located in powerful data centres that are ideally connected to the RRHs through high-speed and low-latency fronthauls. Yet, high fronthaul delays may degrade the performance of certain novel edge services that require close-to-zero latencies as prescribed by 5G objectives. Alternatively, nowadays CESCs architecture may become better suited for deploying mobile edge services. In that case, some processing and storage resources are placed close to the RRH, and thus, the fronthaul delay is significantly reduced. Deploying huge data centres implies a series of requirements in terms of space, energy, etc. Hence, this second option envisages the deployment of a series of HW resources with limited capacity and requirements and in a distributed configuration.

Figure 1 shows an architecture to consolidate multi-tenancy in the mobile communications infrastructures based on a substantial evolution of the SC towards cloud-enabled devices, as proposed in the SESAME project.

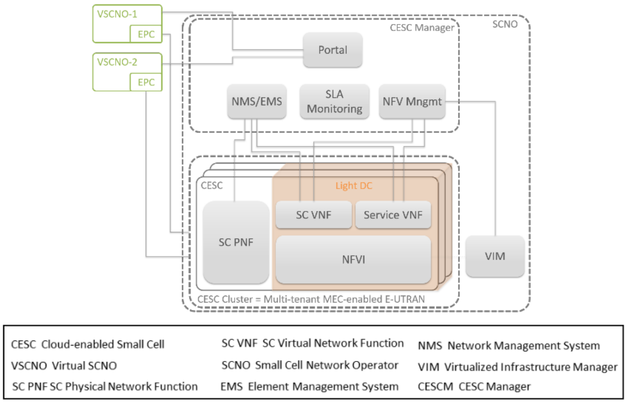
VSCNO-1
EPC
VSCNO-2
EPC
CESC Manager
SCNO
Portal
NMS/EMS
SLA Monitoring
NFV Mngmt
CESC
Light DC
SC VNF
Service VNF
SC PNF
NFVI
VIM
CESC Cluster = Multi-tenant MEC-enabled E-UTRAN
CESC Cloud-enabled Small Cell
VSCNO Virtual SCNO
SC PNF SC Physical Network Function
SC VNF SC Virtual Network Function
SCNO Small Cell Network Operator
EMS Element Management System
NMS Network Management System
VIM Virtualized Infrastructure Manager
CESCM CESC Manager

Fig. 1. Multi-Tenant cloud enabled RAN using SCs

The key element of this architecture is the CESC, owned by a Small Cell Network Operator (SCNO), which consists of a micro server integrated with the small cell to support both radio connectivity and edge services. It foresees the split of the small cell into Physical Network Functions (PNF) and Virtual Network Functions (VNF)[3], enabling a multi-tenant environment in support of a Multi-Operator Core Network (MOCN)[4]. This will allow Virtual Small Cell Network Operators (VSCNO) not only to support connectivity but also to provide added value mobile edge services in a PoP.

Resources on a single micro server (i.e. RAM, CPU, storage, HWA) might not be enough to support the mobile edge computing services of all tenants. CESC clustering enables the creation of a micro scale virtualised execution infrastructure in the form of a distributed data centre, denominated Light Data Centre (Light DC), enhancing the virtualisation capabilities and processing power at the edge. The hardware architecture of the Light DC envisages that each micro server will be able to communicate with all others via a dedicated LAN/WLAN guaranteeing the latency and bandwidth requirements needed for sharing resources. It provides also the backhaul connections to the operators Evolved Packet Core (EPC).

In the context of multi-tenant cloud enabled RAN, a Network Service (NS) is understood as a chain of PNFs and VNFs that jointly supports data transmission between a User Equipment (UE) of an operator and the operator's EPC, with the possibility to involve one or several service VNFs in the data path. It clearly highlights that, beyond the conventional orchestration and management of the cloud resources in a virtualised environment, the proposed solution entails a series of specific challenges such as the dynamic composition of the Light DC resources based on the status of CESC cluster(s), coordination of specific type of resources (radio-related resources, service-related HWA, etc.) and isolation of dedicated network slices to each tenant.

All management tasks, e.g. resource allocation and service lifecycle management over the distributed infrastructure, are carried out by a centralized unit called CESC Manager(CESCM). A single instance of CESCM is able to operate over several CESC clusters, each constituting a Light DC, through the use of a dedicated Virtual Infrastructure Manager (VIM) per cluster. CESCM is the main management component in the architecture, covering the orchestration, management and configuration of NSs. The CESCM has a high-level knowledge of the virtual and physical resources available on the C-RAN environment, including the radio access functionalities. CESCM is composed of the following modules:

- The NFVO is the entity in charge of NS lifecycle management (creation, termination, monitoring, scaling etc.) via coordination between ETSI MANO elements, such as VIM, EMS and VNFM. NSs are defined in the form of NS descriptors (NSD), which contain VNF descriptors (VNFD) – defining required IT resources needs to be dedicated for a VNF as well as its specific functionality – and connectivity between VNFs.
- The VNFM is the entity in charge of the lifecycle management of the VNFs, from deployment to termination, keeping track of their status to adjust their configuration if needed.
- The EMS is the entity in charge of the key functionalities as fault, configuration, accounting, performance and security (FCAPS). It manages the traffic between

the different network elements, coordinating configuration of multiple devices. The EMS associated to radio functions also includes autonomous self-x functionalities to reconfigure the mobile network.

- The SLA component enhances service reliability providing monitoring mechanisms to evaluate the performance of NSs in the radio and cloud environments. It communicates with the NFVO, notifying faults in the system for it to perform the appropriate actions that assure the QoS guarantees of each service in a multi-tenant environment.

In order to communicate CESCM and CESC cluster, the VIM as described in the ETSI[5], is the responsible of managing the virtualized infrastructure, that includes the catalog of the allocated resources, forwarding graphs and chaining rules among VNFs and repository of resources, to provide optimized features. VIM is the software entity that monitors and manages the Network Functions Virtualization Infrastructure (NFVI) (i.e., Light DC) and performs the lifecycle management of the virtual units that will host the VNFs. This centralized administration of virtual resources across multiple localized infrastructure, so that instances can be administrated in a coordinated way, provides the flexibility and scalability needed to optimize and maximize the use of such resources. The VIM is enhanced with SDN component for the networking aspect. The controller takes into account the physically distributed NFVI and the stringent requirements in RAN performance metrics. Moreover, it uses SDN for propagating the VNF chaining requests to the NFVI in order to properly manage the networking resources within the Light DC.

III. SERVICE PROVISIONING MODELS

From the business perspective, three major role players are identified. Function provider (FP) is the VNF developer which sells/develops VNFs. Service Provider (SP), is the one who composes NS –i.e. chain of VNFs, PNFs– with the available VNFs and offers them to the customer. Customer is the one who purchases NSs. In multi-tenant cloud enabled RAN, there are two main possible ways to form a joint radio-cloud model, as illustrated in Fig. 2.

- Mobile Edge Computing as a Service (MECaaS): This model has been inspired mainly from the MNO-MVNO business relationship. Briefly, in this model, MVNO relies completely on the infrastructure and other services provided by the MNO. VSCNO asks for high level KPIs on the SLA. Here, VSCNO only has an overall vision of the system and SCNO has to provide enough support, i.e. both in terms of hardware and number/composition VNF chains (i.e. NS), to meet the agreed KPIs. Performance reports are provided to VSCNO on time intervals (even real time). In simple words, with this model, VSCNO does not chain VNFs to form a mobile edge service, and a high level KPI view is enough for it to request a service without going to details.
- CESC Infrastructure as a Service (CESCIaaS): In this model, VSCNO on SLA asks for connectivity in a certain coverage area according to the elements and for aggregated cloud resources on the Light DC, e.g. a certain amount of GB of storage, of RAM, etc. This model corresponds with the famous Infrastructure as a Service (IaaS) paradigm, which is one of the three fundamental service models of cloud computing[6]. With this model in place VSCNO can compose VNF chains on demand. As a consequence, any VNF instantiation (depending on the used hardware resources) consumes a portion of available VSCNO's aggregated resources. Therefore, the deployment of VNF chains is conditioned to the amount of requested resources by the VSCNO.

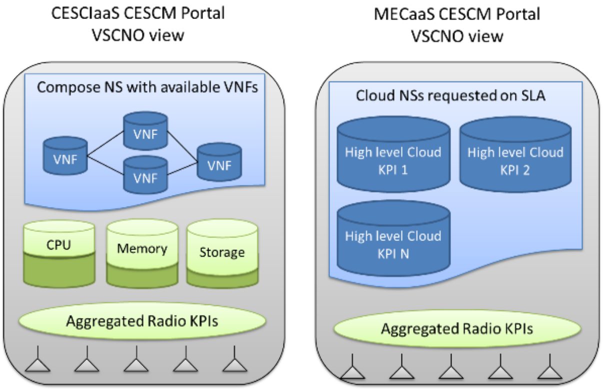

Fig. 2. VSCNO's CSCM portal view on multi-tenant cloud enabled RAN

Figure 2 shows the dashboard view of VSCNO based on the two mentioned joint cloud-radio service provisioning models.

IV. NS DEFINITION

Fig. 3 shows a conceptual view of the ecosystem, which hosts three different VSCNOs over the shared NFVI, i.e. Light DC. The network service is described as a collection of VNFs (including radio related and service related instances) required to deploy a complete 5G mobile service for the end users of the VSCNO. To form the multi-tenant scenario depicted in Fig. 3, NFVO needs to process the functional chaining of VNF requested by each VSCNO and to trigger its instantiation to the VIM that manages the NFVI. Therefore, a NS can be characterized through a series of radio-level and service-level KPIs that are captured in the SLA between the SCNO owning the NFVI and the interested VSCNO. A NS is defined through its associated Network Service Descriptor (NSD).

The main building blocks of a NSD are the VSCNO Network Connectivity Topology (NCT) and the VNF Forwarding Graph (VNF-FG) descriptors. The NCT determines the complete list of VNFs and their Connection Points (CPs), as well as the possible interconnections through a series of Virtual Links (VLs). In this sense, the VSCNO NCT can be seen as the virtual network slice assigned to that VSCNO in the CESC Cluster. VIM will map the logical request to the actual hardware by

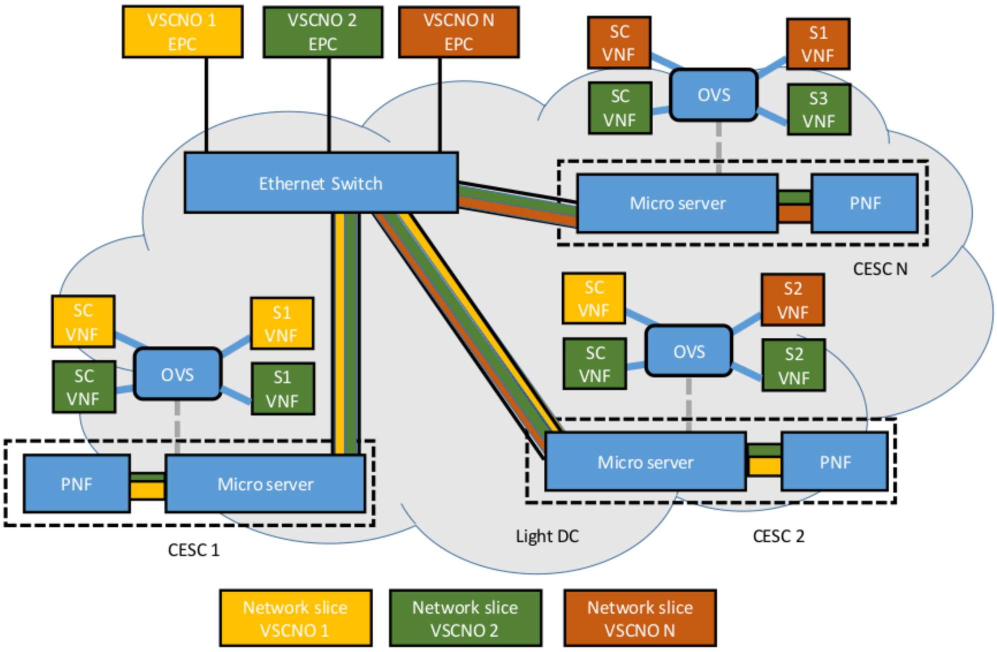

Fig. 3. Edge network services

instantiation of VNFs, which may run in different micro servers. The distributed nature of the edge cloud introduces a novel challenge on the hardware resource allocation. At the end of the placement process, the created NS for the VSCNO can be observed as a separate underlying virtual LAN instantiated and ready to use. However, the actual data flows between VNFs need to be enforced (VNFs can be seen as the origin and destination of data flows). The SDN controller (integrated into the VIM) implements the forwarding rules necessary to move packets according to the VNF-FG.

V. EVOLVING TO 5G C-RAN - HYBRID CLOUD APPROACH

Hybrid approaches have been proposed to deal with the transitions from 4G specific hardware based architectures to software based 5G platforms. Both centralized and distributed clouds will coexist during this evolution to a completely softwarized central C-RAN. Multiple clouds can work together under orders form the same orchestrator to develop a hybrid cloud. In this model, VNFs can be spread between centralized and CESC distributed clouds. Depending on the functional split, central cloud can take part in a different layer of the softwarized upper layer protocols. Depending on the case, it can process just the VNFs form the upper protocol layers or the whole softwarized stack as initially proposed[7]. A Hybrid NFV manager is proposed to orchestrate both clouds in an unified manner.

As lower layers on the protocol stack are virtialised, centralized clusters will take on major relevance as the RAN solution.

VI. QOS OVER MULTI-TENANT C-RAN

Ensuring the QoS per tenant base, assuring the SLA, is another important aspect in the service lifecycle management. QoS assurance demands establishing a feedback loop consist in three main steps:

- Monitoring: a phase in which performance metrics are collected from the radio/cloud/software elements (e.g. SC physical network function, VMs, etc.) and handed over to the next step (decision-making). Depending on the NS deployment and nature the metrics to be collected may vary each time.
- Decision-making: a phase in which performance metrics collected in the previous step are processed. Depending on the situation and available resources, a decision will be taken to ensure the level of QoS (with the help of a dedicated algorithm). Besides available resources, in a multi-tenant scenario, the decision-making process needs to take into account the status of other tenants.
- Reaction: upon making a decision, the management/orchestration system needs to coordinate the interaction with the other lower level modules such as Element Management System (EMS), VNFM and VIM to react appropriately.

Implementing this QoS loop means adding the radio dimension to the standard cloud orchestration system defined by the European Telecommunications Standards Institute (ETSI), and/or to shift the traditional radio network management mentality towards a cloud-oriented mind-set. For instance, leveraging on the radio traffic profile over a certain period (e.g. a day, a week, etc.), it would be possible determining when and where cloud services need to be scaled up/down in a Point of Presence (PoP).

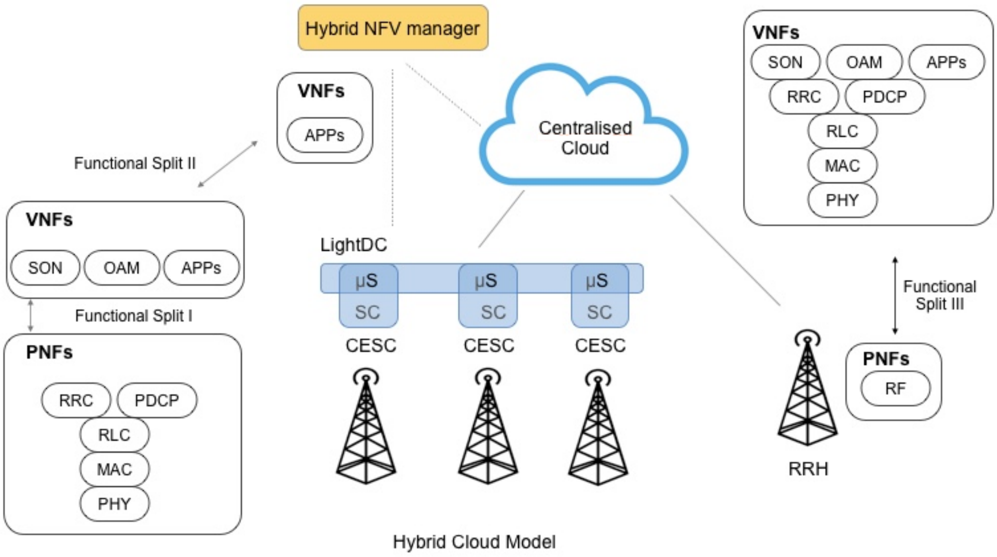

Fig. 4. Hybrid C-RAN

Decision-maker could be placed in several parts along the control plane (VIM, VFNs and NFVO) and the NMS in the management plane[8]. When a change on the architecture is process, a feedback should be given to the learning module in order to analyze the results and infer enhanced decision rules.

Starting at the lower level on the proposed architecture, SDN controller should provide QoS. Networking between VNFs in a chain is an important task in the cloud architecture. Intercepting traffic and forwarding to the correct NS is a challenge resolved by the SDN controller. By taking advantages of standard protocols such as OpenFlow, SDN controller can prioritize traffic according to QoS levels.

At a higher lever, mapping mechanism deals with the allocation of the softwarized components of a NS into the resources of the CESC cluster. In other words, where instantiate the VNF chain that composes a NS. Placement algorithm checks the available virtual resources in the NFVI catalog and the instantiation requirements. Bearing that in mind allocates the resources in the optimal location (see Fig. 5).

Placement algorithm resides inside the VIM. The objective is to place the VNFs chain in order to minimize the end-to-end delay. This placement algorithm could dynamically be recomputed in order to find actively, and not only when the NS is instantiated, the optimal placement in order to achieve the best QoE. Reordering VNFs while executing is possible thanks to live-migrations techniques inside the cloud enable SCs. To achieve a fast migration some requirements have to taken when designing the cloud architecture[9].

The algorithm execution time is critical. Some algorithms deploy new instances over the cloud resources available at the RAN side. In other words, placement of service VNFs that conform the aforementioned edge services are allocated in the remaining infrastructure. In order to not recompute the whole optimal location of all VNFs. There is a commitment between the handover introduced in the system to live-migrate some services and the benefits of QoE obtained.

In the CESCM, VNFM is in charge of the lifecycle management of the VNFs. VMFM is able to apply policies for NS-level rescaling and reconfiguration to achieve high resource utilization. General purpose clouds introduce automatic rescaling algorithms. Generally a minimum and maximum number of intestacies is given to the VMFM in order to select the optimum QoE. In some cases, VNF must be rescaled so it generates a new template indicating the VIM the new architecture. Letting him to compute the best placement for the needed instances.

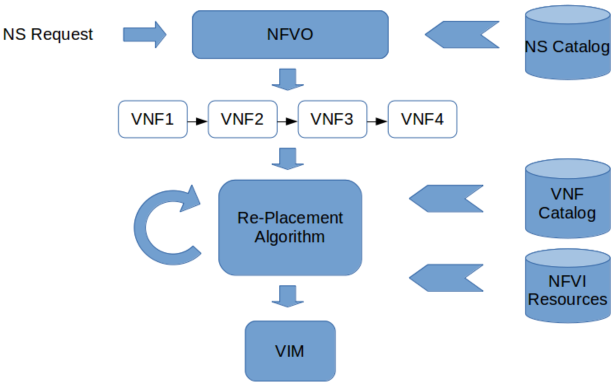

Fig. 5. Replacement algorithm

NMS-EMS are responsible for the management of the network slices that server the NSs. In the proposed solution[10] (Fig. 6) the Metric Aggregator (MA), is responsible for combining and filtering the collected monitored parameters and associates them with the running services over the platform. MA continuously processes the collected monitoring values for the QoS or SLA evaluation.

The Decision Support System(DSS), as shown in Fig. 6, main responsibility is to detect the level of severity on the QoS evaluation process done in MA and decide whether a reactive or a proactive action is needed. Basically, such a decision will be made based on the high level SLA agreements made with VNOs.

According with the decision some downstream reconfiguration must be done: i- NFVO should reconfigure the flow of data in a NS (i.e. changing the SDN rules), ii- VIM could migrate the NS within the PoP, from one PoP to another, iii- and VNFM scale up/down the whole NS (i.e. instantiation of a parallel service or terminating a running one).

NMS manages the interaction between NSs instantiated by different VSCNOs. And it must be done in a prioritize manner. Different pricing schemes could be applied to prioritize clients according to the QoS offered.

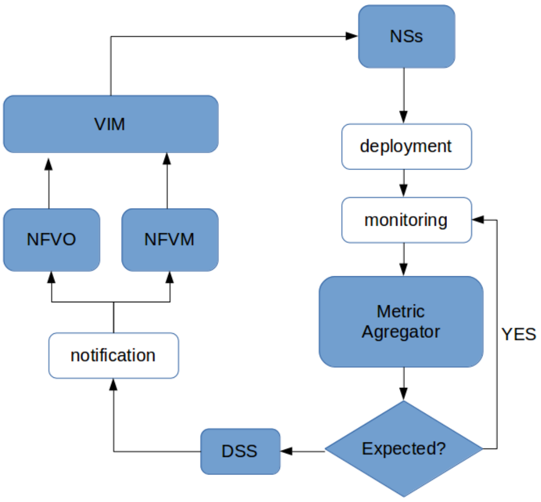

Fig. 6. QoS feedback loop

VII. CONCLUSIONS

This paper has analyzed the evolution of mobile networks to cloud architectures. Proposes an hybrid-cloud coexistence until centralized cloud solutions were extensively applied. It describes the proposed provisioning models and analyzes the components in the proposed cloud RAN architecture in which QoS mechanisms should be applied in order to achieve the KPI agreed with the VSCNOs.

As a result, a vertical management system interaction is needed to reciprocally manage QoS of the overall system in a coordinated manner. QoS mechanism implemented in each level should has its own actuation point in order to improve the QoS locally and therefore improve the global user expedience. In addition an upstream information flow of the decisions taken locally is needed to allow higher levels to achieve global optimization. And also a downstream flow is needed to fix the degrees of freedom at each level the optimization could be made.

VIII. ACKNOWLEDGEMENTS

The research leading to these results has been supported by the EU funded H2020 5G-PPP projects SESAME (Grant Agreement n 671596) and ESSENCE project (Grant Agreement no 761592) and by the Spanish Ministerio de Economia y Competitividad (MINECO) under grant TEC2016-80090-C2-2-R (5RANVIR).

REFERENCIAS

[1] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, “Cloud radio access network (c-ran): a primer,” *IEEE Network*, vol. 29, no. 1, pp. 35–41, Jan 2015.

[2] B. Blanco, J. O. Fajardo, I. Giannoulakis, E. Kafetzakis, S. Peng, J. Pérez-Romero, I. Trajkovska, P. S. Khodashenas, L. Goratti, M. Paolino, E. Sfakianakis, F. Liberal, and G. Xilouris, “Technology pillars in the architecture of future 5g mobile networks: Nfv, mec and sdn,” *Computer Standards & Interfaces*, vol. 54, pp. 216 – 228, 2017, sI: Standardization SDN&NFV. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0920548916302446

[3] 3GPP, “Network Sharing: Architecture and functional description,” 3rd Generation Partnership Project (3GPP), TS 23.251, 12 2013.

[4] “Virtualization in Small Cell Networks,” *Small Cell Forum*, 2015.

[5] ETSI, “Network Functions Virtualization(NFV): Management and Orchestration,” *Small Cell Forum*, 2014.

[6] S.Sharma, “Evolution of as-a-Service Era in Cloud. A review on as-a-Service framework.” *Center for Survey Statistics and Methodology, Iowa State University*, USA 2015.

[7] J. O. Fajardo, F. Liberal, I. Giannoulakis, E. Kafetzakis, V. Pii, I. Trajkovska, T. M. Bohnert, L. Goratti, R. Riggio, J. G. Lloreda, P. S. Khodashenas, M. Paolino, P. Bliznakov, J. Perez-Romero, C. Meani, I. Chochliouros, and M. Belesioti, “Introducing mobile edge computing capabilities through distributed 5g cloud enabled small cells,” *Mobile Networks and Applications*, vol. 21, no. 4, pp. 564–574, Aug 2016. [Online]. Available: https://doi.org/10.1007/s11036-016-0752-2

[8] B. Blanco, J. O. Fajardo, and F. Liberal, *Design of Cognitive Cycles in 5G Networks*. Cham: Springer International Publishing, 2016, pp. 697–708. [Online]. Available: https://doi.org/10.1007/978-3-319-44944-9_62

[9] “OpenStack Live-migration.” [Online]. Available: https://docs.openstack.org/admin-guide/compute-configuring-migrations.html

[10] P. Sayyad, B. Blanco, I. Taobada, M.-A. Kourtis, G. Xilouris, I. Giannoulakis, E. Jimeno, I. Trajkovska, J. O. Fajardo, E. Kafetzakis, J. Garcia, F. Liberal, A. Whitehead, and M. Wilson, “Service management and orchestrarion over multi-tenant cloud-enabled ran,” March 2017.