Gramaglia, M., et al. Flexible connectivity and QoE/QoS management for 5G networks: the 5G NORMA view, in *2016 IEEE International Conference on Communications Workshops (ICC), Kuala Lumpur, Malaysia, 23-27 May 2016, pp. 373-379*

DOI: https://doi.org/10.1109/ICCW.2016.7503816

Flexible Connectivity and QoE/QoS Management for 5G Networks: the 5G NORMA view

Marco Gramaglia, Ignacio Digon, Vasilis Friderikos, Dirk von Hugo, Christian Mannweiler, Miguel A. Puente, Konstantinos Samdanis, Bessem Sayadi

***Abstract*—The goal envisioned by 5G NORMA is to develop a novel, adaptive and future-proof 5G mobile network architecture. In order to fulfill these requirements, 5G NORMA envisions an extremely flexible architecture to be deployed in a multi-tier distributed data-center. In this paper we focus on the novel mobility management schemes, QoE/QoS, Control and Orchestration mechanisms that are being developed in 5G NORMA. These modules, that follow the software-defined principle, jointly optimize core and access functions. The final result is a modular architecture that adapts to the requirements of very heterogeneous services, while allowing multiple tenants to share network resources among them, providing hence the flexible connectivity needed by future 5G Networks.**

I. Introduction

The world is currently experimenting an extraordinary shift towards mobile networking: the amount of mobile traffic per month is expected to grow to 30 exabytes by 2020, almost ten times more than the values recorded in 2015 [1]. Moreover, the mobile industry already powers 3.6% of the world's GDP at an expected growth rate of 5% [2]. In this context 5G NORMA aims at bringing an unprecedented re-engineering of the mobile networking architecture for the mobile industry to remain competitive.

The 5G NORMA[1] project is working towards this goal: developing a novel mobile network architecture that provides the needed adaptability while optimally using the available resources. The architecture envisioned by 5G NORMA is able to handle traffic demand fluctuation resulting from the heterogeneous and rapidly changing services, possibly modified according to the local context. Figure 1 shows the fundamental entities included in the 5G NORMA architecture : (i) the ***edge cloud***, that is the set composed by the base stations and remote servers deployed at the radio or aggregation sites; (ii) the ***network cloud***, one or more data- centers deployed at central sites; and (iii) the ***controller*** that organizes and executes the functions needed by the network and which is usually co-located in the network cloud.

Marco Gramaglia (corresponding author) is with Universisdad Carlos III de Madrid and IMDEA Networks Institute, mail: mgramagl@it.uc3m.es. Ignacio Digon is with Telefonica I+D, mail ignacio.digonescudero@telefonica.com. Vasilis Friderikos is with King's College London, mail: vasilis.friderikos@kcl.ac.uk. Dirk von Hugo is with Telekom Innovation Laboratories, mail: Dirk.von-Hugo@telekom.de. Christian Mannweiler is with Nokia Networks, mail: christian.mannweiler@nokia.com. Miguel A. Puente is with ATOS, mail: miguelangel.puente@atos.net. Konstantinos Samdanis is with NEC Laboratories Europe, mail: samdanis@neclab.eu. Bessem Sayadi is with Alcatel-Lucent Bell labs, mail: bessem.sayadi@alcatel-lucent.com

[1]https://5gnorma.5g-ppp.eu/

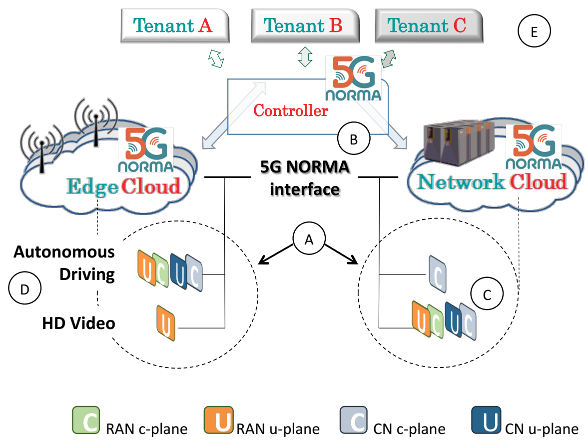

Fig. 1. Main innovations of 5G NORMA concept.

Figure 1 also shows the five 5G NORMA main pillars, 3 innovative *enablers*:
i) the **Adaptive (de)composition and allocation of mobile network functions** between the edge and the network cloud depending on the service requirements and deployment needs (A in Figure 1), ii) the **Software-Defined Mobile network Control and Orchestration** which applies the SDN principles to mobile network specific functions (B), and iii) the **Joint optimization of mobile access and core network functions** localized together in the network cloud or the edge cloud (C);

and 2 innovative *functionalities*:
i) the **Multi-service and context-aware adaptation of network functions** to support a variety of services and corresponding QoE/QoS requirements (D), and ii) the **Mobile network multi-tenancy** to support on-demand allocation of radio and core resources towards virtual operators and vertical market players (E).

Different services may use different allocation of network functions in radio (RAN) and core (CN) both for user (u-) and control (c-) plane according to their Quality of Experience (QoE) / Quality of Service (QoS) needs. This capability is one of the key functionalities used by 5G NORMA to provide the needed flexibility to the future 5G Network architecture. The advantages of this approach are manifold, ranging from network related metrics (i.e., increased capacity, reduced latency, etc.) to business related ones, like cost efficiency or the reduced service creation time.

This paper focuses on the ***controller*** entity of the architecture: an overview is provided in Section II, while Section III focuses on the blocks devoted to the flexible support of heterogeneous services, describing functionalities like the selection of the mobility management scheme, the configuration wireless control functions or network orchestration. Some joint-optimization challenges are discussed in Section IV and, finally, Section V concludes the paper.

II. Controller Architecture

The recent advances in the network *softwarization* solutions (for both *programmability* and *virtualization*) provide the fundamental elements needed to build a flexible and context-aware network architecture. Moreover, some of the concepts devised in recent research projects, namely EU FP7 METIS [3] and EU FP7 iJOIN [4] can be considered as starting point for 5G NORMA innovations.

The scope of iJOIN [2] was focused to small-cell networks where small-cells are connected through heterogeneous backhaul to the core network. One of its key-innovations is the RAN as a service (RANaaS) platform (an extension of the C-RAN solution [5]) composed by physical radio access points and virtual eNodeB controlling radio processing units (iRPU). This split, in essence, allows for a flexible split between remote and central sites, a strategy followed also by 5G NORMA with the *central* and *edge* cloud concepts.

The goal of METIS [3] was to provide a novel mobile network architecture under different viewpoints (Functional, Logical Orchestration and Control Architecture, and Deployment). Like iJOIN also METIS focused on the RAN part, hence not all the function required for a complete 5G Mobile Network were covered: the final outcome is the definition of a 5G Orchestrator of Network Function and the underlying transport layer using Software Defined Networking principles.

Finally, the work undergoing at Open Network Foundation (ONF) has many similarities with the software-defined mobile network control concept proposed by 5G NORMA. Indeed, new ONF extensions aim to use the spirit of *softwarization* to provide flexibility in the implementation of mobile network functions other than routing and forwarding, which is one of the goals of 5G NORMA.

A. NFV/SDN Framework

The network architecture envisioned by 5G NORMA has to support many services with different stringent requirements in terms of latency, throughput and availability, in conjunction with the multi-tenancy and network slicing [6]. This requires the re-design of the current mobile network architecture to move from current *network of entities* architecture to a *network of functions* architecture in order to support the required flexibility and programmability features.

To this end, 5G NORMA leverages on the close interaction of two key enablers for this transformation: Network Function Virtualization (NFV) [7] and Software Defined Networking (SDN) [8]. NFV is a network evolution towards a cloud architecture that uses standard general purpose hardware to deploy network functions. The NFV working group at ETSI defines an architecture framework, called MANO, which describes the Virtual Network Function (VNF) execution environment, their management and orchestration. SDN is deployed to connect the chain of different vNFs needed by a service and to provide a better service-aware traffic steering. Thanks to its capability to connect and control the resources (network capacity, virtual machines, storage capacity), SDN and NFV are hence complementary technologies that allow the deployment of new services and the operation of the network.

B. Design principles

As mentioned above, 5G mobile network architecture will have to support a multitude of services with significantly different requirements. The following dimensions elaborate where 5G systems, like 5G NORMA, need to exhibit a future-proof design that allows sufficient flexibility:

Multi-tenancy capabilities allow several service providers operating on top of a shared infrastructure. The range of tenants is diverse, ranging from mobile network operators (MNOs) to OTT service providers to companies from vertical industries. This also results in varying levels (depths) of service and resource control to be exposed to tenants.

A shared infrastructure leverages the economies of scale to be expected when hosting multiple logical mobile networks. The infrastructure consists of heterogeneous hardware resources (general-purpose as well as dedicated, special-purpose hardware) and necessary software for hosting mobile network functions. The infrastructure as a whole is provided by several infrastructure providers, e.g., MNOs or 3rd party providers.

Efficient network control allow for a sufficient abstraction of controllable resources and functions and expose uniform control APIs on different abstraction and architectural levels. Thus, it allows for, e.g., cross-domain orchestration of network functions and services, flexibility in function decomposition and placement, and customized business service composition.

The fragmentation of administrative domains increases complexity. Vertically, at least business service providers, network service providers and infrastructure providers have to be differentiated. Depending on the type of tenant, some or even all can collapse into one. Horizontally, multiple providers of each type co-exist.

Today, mobile network operators subsume many of the aforementioned aspects in the notion of network slicing, an evolution of network sharing, which has been a key business model for mobile network operators to reduce deployment and operational costs. In any case, holistic, end-to-end management and orchestration frameworks and dynamic software-defined mobile network control become crucial factors.

On the management plane, this paper therefore proposes a network-wide orchestration, management, and inter-slice coordination architecture extending the ETSI NFV MANO framework. On the control plane, the examples of mobility management and QoE/QoS control illustrate the design principles of software-defined mobile network control.

C. Building Blocks Overview

Figure 2 shows an overview on the proposed architecture blocks as well as the interactions among them. 5G NORMA

[2]http://www.ict-ijoin.eu
[3]http://www.metis2020.com

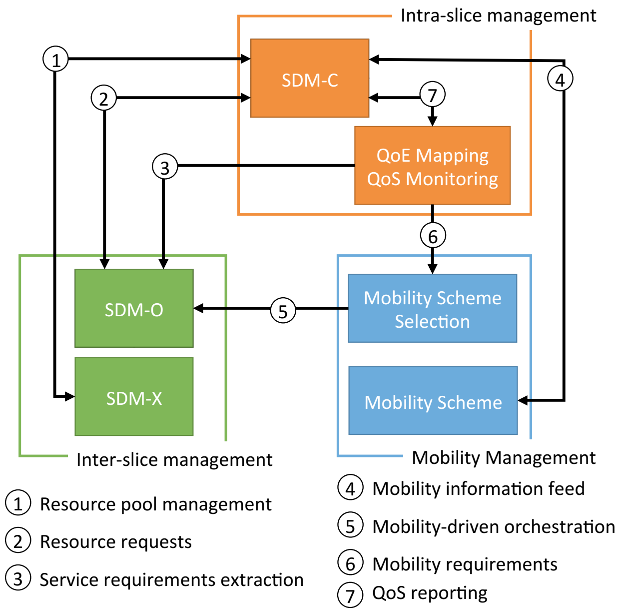

Fig. 2. The 5G NORMA flexible connectivity and QoE/QoS management building blocks.

provides flexible connectivity for 5G Networks using 5 building blocks, that will be detailed throughout this section: the Software-Defined Mobile Network Controller (SDM-C), Orchestrator (SDM-O), Coordinator (SDM-X) and the modules in charge of Mobility Management and QoE/QoS Mapping & Monitoring. Mobility Management also considers potential multi-path issues of e.g., devices connected concurrently via multiple interfaces, and of migration of functions between edge clouds.

Figure 3 shows the life-cycle of a network slice creation and operation using the 5G NORMA proposed architecture. The SDM-O (see Sec. III-D) is the entity that interfaces to the business domain. It handles slices creation requests associated to a well-defined service (e.g., Vehicular, IoT), possibly belonging to several tenants. For every request the orchestrator maps the abstract service requirements in terms of KPI (i.e., parameters as fixed in the service level agreement, SLA), to real requirements that are used to build the actual chain of virtual Network Functions (vNF), starting from a template library. This process is somehow similar to the one being investigated by the IETF Service Function Chaining (SFC) WG[4]. SDM-O orchestrates the needed vNF, optimally locating them in the infrastructure (some optimization use cases are discussed in Sec. IV).

SDM-O has a complete picture of the network: it manages the resources needed by all the slices of all tenants, optimally configuring them in order to reduce the amount of used resources. On the other hand, SDM-C (see Sec. III-C) manages the resources within a network slice. It also instantiates the forwarding path used to realize an SFC (Service Function Path, SFP, in the IETF SFC WG) taking into account the constraints and requirements defined by the SDM-O.

[4]https://datatracker.ietf.org/wg/sfc/

The information used to define those constraints is gathered from the QoE/QoS Mapping module (see Sec. III-B), a module that is also in charge of continuously analyzing the status of a network slice and reporting to SDM-C. Based on those reports, SDM-C may adapt to the new situation either by reconfiguring some of the vNF it manages (i.e., changing the pre-scheduler, asking for a less aggressive MCS), or by reconfiguring some paths using a SDN- alike technique or, at last instance, by asking for more resources to the Orchestrator.

The Mapping & Monitoring module is not the only entity reporting to the SDM-C. Also the Mobility Management module (see Sec. III-A) reports to the SDM-C about mobility management specific issues. The Mobility Management module is also used by the Orchestrator to optimize the migration of mobility related vNF among different edge clouds to fulfill specific QoE/QoS constraints.

Finally, network slices reshaping (i.e., scale-in and scale-out operations) requests are managed by the SDM-O if requests regard computational resources, or by the SDM-X (see Sec. III-E) if shared resources such as spectrum are needed.

III. Architecture Blocks

A. Mobility management

The challenging requirements in terms of mobility support in a future 5G system are less related to the increased data volume than the expected amount of different terminal types (beside handhelds and smartphones many new machine-type devices are foreseen) and a broad range of services and mobility demands which range from zero or nomadic mobility only to high speed (e.g., vehicular) terminals. While the connection characteristics change during movement across multiple cells and points of attachment, the session has to continue surviving, for example, an endpoint address change. Also seamlessness with respect to the user or application experience may differ, demanding for minimal disruption without loss of information or being able to cope with “break before make” where a higher layer protocol cares for data completeness.

An efficient solution for serving all varying requests cannot be a single universally applicable mobility function, but has to follow a modular approach to adapt the network configuration according to the respective service demands (i.e., the so-called mobility on-demand [6]). The challenge is to identify the actual demand as precisely as possible and select the best fitting solution depending on the overall scenario. Several criteria can be taken into account in the selection process, including the characteristics of the terminal and its environment (e.g., smartphone or sensor device in a pedestrian movement or attached to a car) as well as the network conditions (e.g., load of neighboring cells and radio technology or specific parameters such as bandwidth and latency) but also performance requirements of the application (e.g., in terms of connection reliability and session continuity). Potential solutions may be based on the Distributed Mobility Management (DMM) [9] approach as discussed at IETF (Internet Engineering Task Force) where a split of control and user plane functions and basic modular logical entities for anchoring, location and forwarding management are proposed.

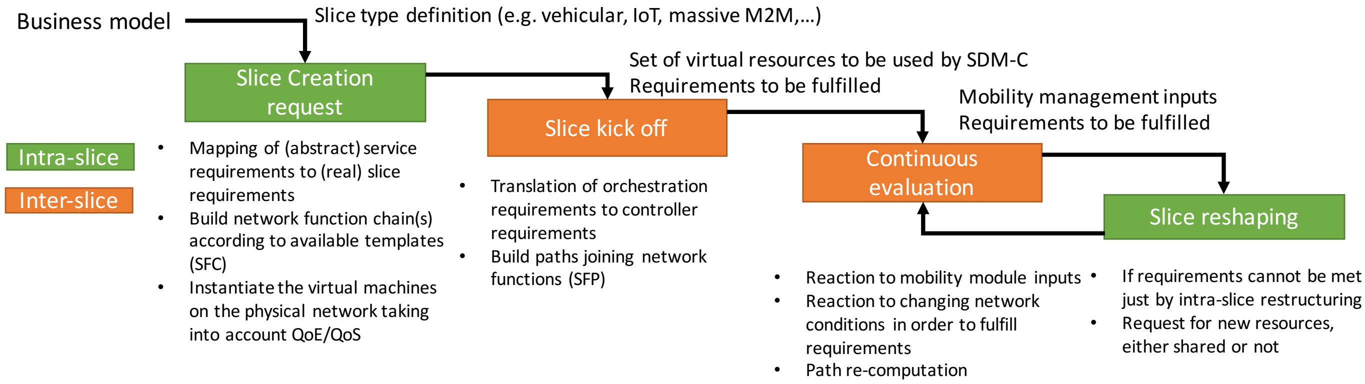

Fig. 3. The lifecycle of a network slice in 5G NORMA.

As part of the mobility scheme selection, special consideration has to be given to concurrent utilization of multiple links to enhance throughput and reliability. A special version of such multi-connectivity is already available as enhanced carrier aggregation in LTE-A where different cells and frequency bands can serve a UE simultaneously and known as offload or local break-out when a cellular device with multiple interfaces connects (e.g., to home WLAN to access the Internet). A commonly managed standardized solution without need for customer interaction and covering heterogeneous access domains has still to be developed. To provide extremely low delay connectivity (e.g., as demanded by tactile Internet applications) the content and processing resources have to be provided near the edge of the access link. Such distributed functionality in the Mobile Edge Cloud (MEC) [10] has to be migrated efficiently in case of session endpoint movement but also due to load balancing reasons in case of resources shared between multiple slices. To keep latency and jitter in required limits demands for new mechanisms when taking the movement decision.

The major demands to the new mobility management protocol are, beside providing a high flexibility, also to fulfill high availability and reliability (e.g., by avoiding a single point of failure), achieve high resource efficiency (e.g., low signaling overhead, avoiding too much tunneling, minimum redundancy in case of error retransmissions), and to not increase the operational burden (e.g., in terms of deployment and network management effort). An example approach for allowing better service awareness could be to consider new mobility options e.g. in DMM as is e.g. proposed in [11] in terms of QoS specific parameters. Additional challenges are discussed in Section IV.

B. QoE/QoS Mapping and Monitoring

Quality of Experience (QoE) is defined as "the degree of delight or annoyance of a person whose experiencing involves an application, service, or system" [12]. This contrasts with Quality of Service (QoS), which concerns objective and technical metrics at network (delay, jitter, packet loss, etc.) and application level (frame rate, resolution, etc.). Lately, QoE is becoming the ultimate item to be delivered to end-users.

The driving paradigm is to keep users satisfied abstracting from the objective QoS factors that cause that satisfaction. This allows having satisfied users while allocating the minimal amount of resources for that purpose, thus reducing costs, avoiding churn and increasing energy efficiency. This is particularly interesting in wireless networks, where radio resources are scarce. 5G networks will have to cope with unprecedented *densification* levels, causing the access network to account for the major energy consumption share [13]. In this scenario a moderate reduction in the data rates can lead to large energy (and therefore costs) savings [14], which fosters the utilization of QoE-based management approaches in future mobile network architectures.

The identification of service-relevant QoE metrics and modeling of how these are affected by the different QoS metrics is a key aspect. QoE metrics decompose individual user experience in different dimensions, which are perceivable and nameable on a service and context basis, and whose combination forms the subjective quality perceived by the user.

The QoE/QoS mapping process can be modeled in two steps: first, the QoS metrics modulate the set of QoE metrics; second, the combination of these QoE metrics forms the end user QoE. In current works, utility functions have been widely used to map one or several QoS metrics to QoE [15]–[17] and different methods to map QoE metrics to an overall QoE score are described in [18].

Figure 4 depicts a high level view of the QoE/QoS mapping, which architecturally is strongly related to the SDM-C block. It also shows different monitoring mechanisms that can be used for QoS monitoring. In order for the mapping module to perform the QoE/QoS mapping, it needs to know the current QoS status of the network and service. It may also include the terminal, which is the component that receives the QoS that is actually delivered to the end-user. Monitoring poses many challenges for the forthcoming 5G networks with the multiplication of the number of devices and the need of providing management processes (such as monitoring) with minimal signaling overhead. SDN principles also involve central controllers which may become easily a bottleneck if the number of elements they control is large.

To that extent, 5G networks shall provide monitoring mechanisms and protocols for an efficient monitoring, in which

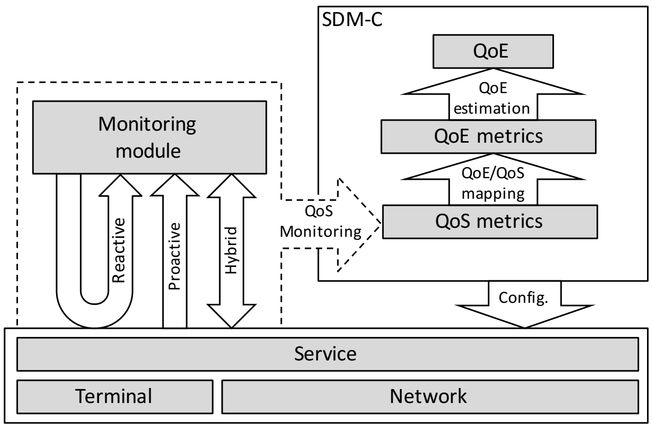

Fig. 4. QoE/QoS mapping and monitoring.

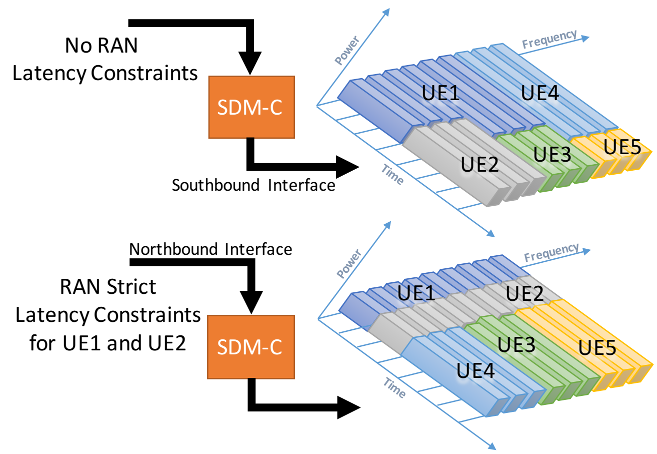

Fig. 5. An example of SDM-C operation.

the trade-off between the monitoring accuracy and signaling overhead is optimized.The usual architecture for monitoring systems is mainly based on two module types: a central monitor and distributed monitoring agents. The monitoring strategies can be classified into three types: (i) reactive, in which the monitor polls the agents; (ii) proactive, in which the agents proactively send data to the monitor; and (iii) hybrid, a combination of the two. In this sense, hybrid monitoring is the one that is widely open for innovative monitoring algorithms and protocols and it is the one on which 5G NORMA monitoring work will focus.

C. SDM-Controller

5G NORMA architecture incorporates the Software-Defined Mobile Network Controller (SDM-C) to enable a flexible network management and operation. Following a similar spirit to SDN 5G NORMA envisions the SDM-C, in which mobile network functionalities are split into two categories: those functions that are being "controlled" and remain relatively stable and those functions that "control" the overall network and are executed at the controller. In particular, the SDM-C is specifically devised to control mobile network functionality, and it is not limited to data plane functions but also control plane functions. That is, SDM-C takes care of managing the function within a network slice, including wireless controllers and traffic steering.

The control of wireless functions is the most innovative element in SDM-C. Functions like channel selection, scheduling, MCS selection, and power control may be controlled by the SDM-C according to well-defined control policies. The SDN approach is the inspiration behind the SDM-C, so both a *northbound* and a *southbound* interface are specified. The *northbound* interface is used to control the network operation from a QoE/QoS perspective. For instance, maximum latencies between elements of a SFC can be specified using this interface. The SDM-C elaborates the requirements received through the *northbound* interface and controls the network resources (i.e., transport network, wireless functions, other vNF belonging to its slice) through the *southbound* interface. This interface should be standardized and supported by all deployed equipment and be flexible enough to have leeway for the desired behavior of the network by changing the "control" functions only while leaving the "controlled" ones unchanged.

An example of the SDM-C operation is depicted in Figure 5. If no special requirements are imposed for the latency of users attached to a specific base station, then the SDM-C could take scheduling decisions without considering such constraint. However, if strict latency constraints are required for certain UEs (e.g., UE1 and UE2 are vehicles that need low delays for safety communications), then the SDM-C reconfigures the scheduler vNF to obtain the required latency. In the example, UE1 and UE2 get the whole bandwidth and are scheduled with the highest priority in time.

As the SDM-C exposes through its *northbound* interface a common way of managing all the (heterogeneous) underlying network functions, several advantages can be obtained; ranging from cost reductions (i.e., networks operators do not need bind themselves to specific wireless equipment that does not adapt to changing demands) to reduced service creation time (i.e., the common *northbound* interface simplifies the development, debugging and deployment of new services).

D. SDM-Orchestrator

The Software-Defined Mobile Network Orchestrator (SDM-O) is a network management entity of the 5G NORMA architecture that builds on the top of the current ETSI NFV MANO to enable the support of multi-service and multi-tenancy by the means of network slicing and resource orchestration. The SDM-O combines service and network resource orchestration.

The SDM-O analyzes the requirements of incoming service requests from different tenants, creating on the first stage a VNF forwarding graph including e.g. DPI, video accelerator, etc., that fulfills the Service Level Agreement (SLA) of the incoming request and allocating or importing from a tenant an accompanied policy. Such VNF graph can then feed the network wide orchestrator, which can allocate a network slice to accommodate the requirements of a particular service.

The notion of network slice in 5G NORMA is aligned with the NGMN view [6], which refers to a network slice as a collection of network resources, functions and radio technologies, considering also the selection of a control/user plane split. Hence, the role of the SDM-O is to allocate the network slice capacity and the network functions according to the request as typically laid out in SLAs and provided policies. To accomplish this goal, the SDM-O is equipped with network resource allocation algorithms that forecast the amount of capacity required per network slice. In 5G NORMA different types of algorithms are studied, considering network capacity per slice in isolation or as a pooled resource.

5G NORMA also develops an enhanced concept of network functions beyond LTE. Such network functions can be composed by a customized selection of "atomic" functions (i.e. elementary sub-functions of the legacy LTE ones), combining both radio and core networks. The responsibility of the SDM-O is to select and compose such 5G NORMA network functions and allocate them at particular geographical network locations, considering also their combination into a network service function chain. In addition, the SDM-O, allocates to network slices network functions, which can be shared among particular slices.

Typically, the SDM-O is composed by a two level hierarchy of orchestrators, an inter-slice and intra-slice one. The inter-slice is responsible for the network slice allocation, while the intra-slice one takes care of the life-cycle operations (e.g. relocation of network functions, (re)composition of network functions, etc.), of each individual slice reflecting evolving service demands. The SDM-O can then feed the SDM-C with a service policy provision and such a network function chain, which can provide the arranged linkage of network functions instantiating the network slice. In turn, the SDM-C interacts with SDM-O to indicate the need for scale-in and scale-out of network capacity for a particular network slice.

E. SDM-Coordinator

5G NORMA introduces the Software-Defined Mobile Network Coordinator (SDM-X) for controlling network functions or resources, which are shared among selected network slices. The SDM-X sits on the common control layer which gathers all common network functions and resources among network slices. Such shared functions can either be the virtual or physical network functions that the network slices rely on. No matter the nature of these functions (virtual or physical) they can be core network functions such as Home Subscriber Server (HSS), more specific service related core elements such as Machine-Type Communications (MTC) specific Mobility Management Entity (MME) or wireless resources.

Considering the case of the shared spectrum, 5G NORMA introduces a higher degree of flexibility in resource scheduling, creating in this way a more robust and customized network slice taking into account the service needs. The role of SDM-X is to enable a rapid, short term scheduling decisions. For both core network and radio cases, the SDM-O decides the shared functions among network slices and provides to the SDM-X the relevant service policy. The SDM-X analyses this information together with the received network MANO requirements in order to decide whether it is necessary or not to modify a network slice capabilities to fulfill the agreed SLAs for the given tenant.

IV. Cross-optimization challenges

We outline below some of the key cross-issues between mobility, multi-path routing and vNF chaining and routing. The aim is to illustrate some generic cases and propel the idea that a joint design is required in order to optimize the performance of a system architecture like 5G NORMA.

A. VNF Location and Chaining Problem

The vNF location and chaining problem requires to find the optimal number and placement of vNFs by taking into account the number of service requests and the ordering of the visiting vNFs in order to minimize the overall network operational costs and the utilization of the network. Depending on the nature of vNFs some more detailed constraints will need to be taken into account such as for example the anti-affinity constraint which requires some vNF that provide a correlated access the physical underlying resources to be implemented in different physical nodes because isolation of the operation of the different VMs might be compromised. This problem is also called the service chaining problem which, as mentioned above, relates to the process of routing a network flow (service) over a number of NFs in a pre-defined order. The service chaining problem can be easily shown to fall into a special category of facility location problems which are in general NP-hard and therefore intractable for pseudo-real time solutions of large network instances [19], [20]. OpenNF proposes an implementation of the control plane for VNFs as well as the network data plane by extending SDN functionalities [21]. Despite the significant attention that this problem has received over the last few years, there has been very little work on its application in wireless mobile networks. This is especially true with respect to the issues of mobility, QoS and multi-path routing that could provide the means for service differentiation and increased levels of network utilization. The following sections provide details on open-ended problems as pertain to the efficient deployment of vNF routing and chaining in mobile networks.

B. VNF Routing & Chaining with Mobility Support

Most of previous proposed solutions on the issue of vNF chaining and routing do not take into account user mobility. When user mobility is taken into account we can potentially have the case where the path between the node of the last-in-order vNF to be visited and the service access router is changed due to a handover to a different access router.

Moving to a new access router means that the above last routing path segment will be changed and therefore the chaining and location of vNF might not entail optimal operation for the network. Hence mobility issues need to be taken into account and potentially a joint optimization scheme should be implemented that takes into account the effect of service migration/handover to a different access router. The above joint design could be implemented in per-network flow basis or for aggregate network flows using statistical information by exploring historical data on aggregate number of handovers in the specific geographical location.

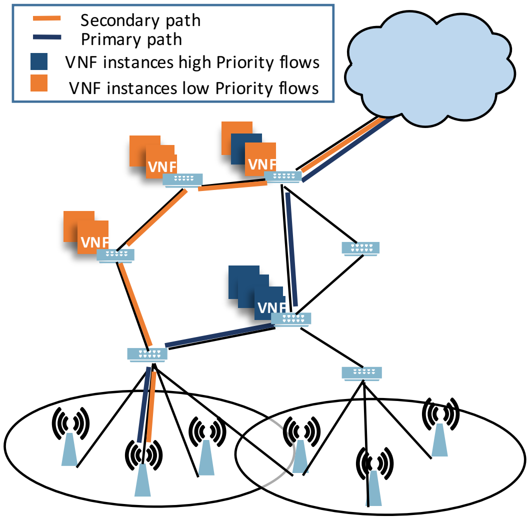

Fig. 6. An illustration of vNF chaining and routing using multi-path routing for service differentiation and better utilization of network resources. As shown in the figure high priority service flows are allowed to use the shortest path (blue lines) and perform vNF chaining whereas low-priority service flows can utilize a secondary shortest path (orange line) to perform vNF chaining.

C. VNF Routing & Chaining with Multi-Path Support

The envisioned decoupling between control and data forwarding plane via SDN allows for the incorporation of novel and flexible routing schemes compared to the current approaches which are mainly based on a single shortest path between two communicating network entities in the network. To this end, multi-path routing is a feature of high promise which has yet to be explored in emerging architectures and an above mentioned programmable forwarding and control plane will propel such solutions from concepts to real-world implementations.

Multi path routing can be utilized for inter-vNF routing to provide efficient utilization of available resources and to provide policies for per-flow treatment based on different service flow priorities. An example of that scenario is depicted in Figure 6, which shows the case of two flows with different priorities and how multi-path routing can be jointly executed with vNF routing and chaining so that network resources are better utilized. In a more general framework VNF chaining and routing can be composed in order to fulfill QoE/QoS constraints and/or requirements.

V. Conclusion

In this paper we present the 5G NORMA view on how to provide flexible connectivity and QoE/QoS Management within its envisioned architecture. We describe the main directions that future 5G Networks should be following, and the way 5G NORMA builds toward them. We introduce the fundamental blocks of the 5G NORMA architecture, detailing the role played by each of them, and discuss on additional challenges that future 5G Networks should tackle in order to support the numerous envisioned services with heterogeneous requirements. The future work includes the definition of the granularity of each interface and the internals of the presented entities, within the framework of the proposed architecture.

Acknowledgment

This work has been performed in the framework of the H2020-ICT-2014-2 project 5G NORMA. The authors would like to acknowledge the contributions of their colleagues. This information reflects the consortium's view, but the consortium is not liable for any use that may be made of any of the information contained therein.

References

[1] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 20152020*, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf, Retrieved in March 2016.
[2] *The Mobile Economy 2014*, http://www.gsmamobileeconomy.com/GSMA_ME_Report_2014_R2_WEB.pdf Retrieved in December 2015.
[3] J. F. Monserrat, G. Mange, V. Braun, H. Tullberg, G. Zimmermann and O. Bulakci, *METIS Research Advances towards the 5G Mobile and Wireless System Definition*, EURASIP Journal on Wireless Communications and Networking, 2015,
[4] D. Sabella, P. Rost, A. Banchs, V. Savin, M. Consonni, M. D. Girolamo, M. Lalam, A. Maeder and I. Berberana, *Benefits and challenges of cloud technologies for 5G architecture*, 1st International Workshop on 5G Architecture (5Garch 2015), 2015.
[5] R. Wang, H. Hu, and X. Yang, *Potentials and Challenges of C-RAN Supporting Multi-RATs Toward 5G Mobile Networks*, IEEE Access, 2014.
[6] NGMN Alliance, *NGMN 5G white paper*, https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf, Retrieved in December 2015.
[7] ETSI ISG for NFV, *ETSI GS NFV 002: Network Functions Virtualisation (NFV); Architectural Framework*, January 2015.
[8] ETSI ISG for NFV, *OpenFlow-Enabled SDN and Network Functions virtualization*, January 2015.
[9] D. Liu, J. C. Zuniga, P. Seite, H. Chan and C. J. Bernardos, *Distributed Mobility Management: Current Practices and Gap Analysis*, RFC 7429, January 2015.
[10] W. Shiqiang, R. Urgaonkar, M. Zafer, T. He, K. Chan, and K. K. Leung, *Dynamic Service Migration in Mobile Edge-Clouds*, In Proc. of IFIP Networking. 2015.
[11] M. Liebsch, S. Matsushima, S. Gundavelli and D. Moses, *Protocol for Forwarding Policy Configuration (FPC) in DMM*, IETF, July 2015
[12] P. Le Callet, S. Möller and A. Perkis, *Qualinet white paper on definitions of quality of experience*, European network on QoE in multimedia systems and services, version 1.2, March 2013.
[13] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume and A. Fehske, *How much energy is needed to run a wireless network?*, IEEE Wireless Communications, 18, no. 5 (2011): 40-49.
[14] J. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. Soong and J. Zhang, *What will 5G be?*, IEEE Journal on Selected Areas in Communications, 32, no. 6 (2014): 1065-1082.
[15] P. Reichl, B. Tuffin and R. Schatz, *Logarithmic laws in service quality perception: where microeconomics meets psychophysics and quality of experience*, Telecommunications Systems Journal, Springer, vol. 52, no. 2, Jun. 2011, pp. 587-600.
[16] P. Reichl, S. Egger, R. Schatz and A. D'Alconzo, *The logarithmic nature of QoE and the role of the Weber-Fechner law in QoE assessment*. In Proc. of IEEE International Conference on Communications (ICC), 2010.
[17] M. Fiedler, T. Hossfeld and P. Tran-Gia, *A generic quantitative relationship between quality of experience and quality of service* IEEE Network, 24, no. 2 (2010): 36-41.
[18] A. Raake, *Speech quality of VoIP: assessment and prediction*. John Wiley & Sons, 2007.
[19] N. K. Chowdhury and R. Boutaba, *A survey of network virtualization*. Computer Networks, 54(5):862 876, 2010.
[20] R. Mijumbi, J. Serrat, J. Gorricho, N. Bouten, F. De Turck and R. Boutaba, *Network function virtualization: State-of-the-art and research challenges*, IEEE Communications Surveys Tutorials, 2015.
[21] A. Gember-Jacobson, R. Viswanathan, C. Prakash, R. Grandl, J. Khalid, S. Das and A. Akella, *OpenNF: enabling innovation in network function control*, ACM SIGCOMM, 2014.