Banchs, A., et al. A NOvel radio multiservice adaptive network architecture for 5G networks, in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring), Glasgow, Scotland, 11-14 May 2015 [proceedings], pp. 1-5*

DOI: https://doi.org/10.1109/VTCSpring.2015.7145636

A NOvel Radio Multiservice adaptive network Architecture for 5G networks

Albert Banchs
University Carlos III of Madrid
IMDEA Networks institute, Spain

Markus Breitbach
Deutsche Telekom AG
Germany

Xavier Costa
NEC Network Laboratories
Germany

Uwe Doetsch
Bell Labs, Alcatel-Lucent
Germany

Simone Redana
Nokia Networks
Germany

Cinzia Sartori
Nokia Networks
Germany

Hans Schotten
University of Kaiserlautern
Germany

Abstract**—** This paper proposes a conceptually novel, adaptive and future-proof 5G mobile network architecture. The proposed architecture enables unprecedented levels of network customisability, ensuring stringent performance, security, cost and energy requirements to be met; as well as providing an API-driven architectural openness, fuelling economic growth through over-the-top innovation. Not following the 'one system fits all services' paradigm of current architectures, the architecture allows for adapting the mechanisms executed for a given service to the specific service requirements, resulting in a novel service- and context-dependent adaptation of network functions paradigm. The technical approach is based on the innovative concept of adaptive (de)composition and allocation of mobile network functions, which flexibly decomposes the mobile network functions and places the resulting functions in the most appropriate location. By doing so, access and core functions no longer (necessarily) reside in different locations, which is exploited to jointly optimize their operation when possible. The adaptability of the architecture is further strengthened by the innovative software-defined mobile network control and mobile multi-tenancy concepts.

Keywords—5G architecture, multi-service architecture, adaptive decomoposition and allocation, software-defined mobile network control, mobile multi-tenancy

I. INTRODUCTION

Mobile networks are becoming one of the most important enablers for the ongoing digitization of business and daily life and the transition towards a connected society. Driven by the increasing adoption of smart-phones and tablets, consumer driven high volume services such as video streaming will continue to **push fast growth of the global mobile IP traffic**. In addition, services in high priority areas for our society such as education, health, government, public protection and disaster relief, and public transportation or in business areas such as smart grids, transportation, automotive, vehicular communication and automation, increasingly rely on a mobile infrastructure. Here, as well as human centric traffic, machine type devices will increasingly create mobile traffic. These machine type communications (MTC) (also known as Machine to Machine (M2M)) will show significantly different characteristics from today's dominant human-to-human (H2H) traffic and will thus result in **new and versatile requirements** and challenges for mobile networks from a technical and business point of view.

As predicted by Cisco in June 2014 [1], mobile data traffic will increase globally 11-fold between 2013 and 2018, and grow three times (!) faster than fixed IP traffic. Due to the broad adoption of flat rates, mobile operators will have to support this growth in mobile data volume without respective increases in revenues. **Cost-efficiency** will, thus, remain a **key challenge** for future network developments. Another major challenge to be addressed is **energy-efficiency**, i.e., the need to support growing mobile data volumes without increasing the energy consumption which translates to greener operations as well as cost savings. ch

The growth in mobile traffic will not be homogenous: Busy-hour Internet traffic is expected to grow more rapidly than average Internet traffic [1]. The uptake of MTC services will, in addition, result in locally and over time varying characteristics of the mobile traffic. Thus, a future mobile infrastructure will not only have to support a fast growing overall mobile data volume and a significantly increased number of connected mobile devices [2] at significantly improved energy- and cost-efficiencies, but it also has to provide **the capability to flexibly adapt to dynamically fluctuating traffic demands (over time, location and characteristics) and a broad range of potentially new service requirements** of future service portfolios. In addition, the wider use of machine type communications requires wider geographic coverage, with implications for network architecture.

As a result of the broadened service portfolio to be supported, the **set of key performance indicators (KPIs) to be considered in the 5G design process needs to be extended to capture the new requirements**. The evolution of mobile networks from 2G to 3G to 4G and LTE-Advanced was driven by requirements resulting from a continuously broadened service portfolio. Most of these services were consumer driven, typically human-centric services as voice, web browsing, and recently in particular video streaming, which now accounts for about 50% of the overall mobile data volume. **Key performance indicators** for network design were peak data rates, average and cell-edge user throughput and overall cell throughput. The need to consider these KPIs will continue for the 5G network design. However, in addition new KPIs need to be defined to measure **massive uptake of machine-type traffic supporting new vertical user groups** in industry, public administration, and business. Here, network availability, coverage (both deep indoors and for sparse rural areas), robustness and reliability, support for dramatically increased numbers of devices, long stand-by times, almost zero-complexity of devices, extremely low latency and efficient support for cloud services must also be considered. This is in combination with increased requirements on energy- as well as cost-efficiency.

A major difference to today's rather homogenous human-driven traffic demand is that these new service classes typically show demand pattern dynamically varying over time and location. Future network architectures will have to efficiently adapt to these changing demand patterns since network optimisation for some of the above listed new design criteria will be difficult to achieve and will demand balancing multiple, perhaps contradictory, requirements. As such, the adaptation will have to be fast, local, and context- and service-aware. Thus, the projected heterogeneity of the new and legacy use cases, services, business models, infrastructure usage approaches and radio access concepts along with the variability of the emerging traffic demands requires a level of **adaptability that is not supported by today's architectures**. It is thus the key objective of this paper to contribute to a successful all-entailing 5G design process by proposing the key building blocks of an API-driven customisable 5G mobile network architecture meeting these requirements. To this end, the key challenges that the design process are:

- To provide an architecture that enables an efficient integration and usage of the advanced radio concepts developed for 5G and beyond as well as a smooth integration of legacy RATs;
- To allow the support of a broad versatility of services expected for the future; and
- To enable an economically successful introduction of 5G by (i) supporting a cost- and energy-efficient 5G operation, (ii) an early, step-by-step introduction of 5G solutions without requiring to wait for nation-wide deployments, as well as, and (iii) a scalable architecture that cost-effectively supports a range of deployment models
- To provide performance improvements with respect to capacity, scalability, latency, cost-efficiency, service creation, security and energy efficiency.

II. Architectural Approach

The main goal of this paper is propose a **multi-service mobile network architecture** that adapts the use of the mobile network (RAN and CN) resources to the service requirements, the variations of the traffic demands over time and location, and the network topology (including the available front/back-haul capacity). The proposed architecture is called 5G NOvel Radio Multiservice adaptive network Architecture (5G NORMA).

The key idea behind 5G NORMA for achieving the above objective is to **decompose the mobile network functions** (including access and core functions) and **adaptively allocate** them to the access network or central cloud, depending on (i) the specific service and its requirements, e.g., bandwidth and latency; and (ii) the transport network capabilities (e.g., available front/back-haul capacity). This adaptive allocation of functions brings several advantages:

- When service requirements and backhaul capacity allows **centralizing** the functionality in the network cloud, better scalability and pooling gains can be obtain from moving all the functionality to the cloud.
- When services have special requirements that require moving part of the functionality to the access, or backhaul constraints do not allow full centralization, gains can be obtained by using a **fully or partially distributed configuration**. Achievable benefits can for example be lower latencies, enabling autonomous operation of edge clouds, or offloading the backhaul.

To implement the above adaptive (de)composition and allocation of network functions, the architecture applies concepts from software-defined networking (SDN) and network virtualization (NFV). With the resulting paradigm, which we refer to as **software-defined mobile network control**, the functionality executed in the network cloud is designed following a software-defined approach: it relies on a well-defined, 'programmable' interfaces with the access node that allow to flexibly adapt the network behaviour by modifying the centralized functionality only [3].

As a result of the flexible allocation of functions in the access and network cloud, the separation between Radio Access Network (RAN) and core network (CN) is blurred. Indeed, with the proposed architecture, RAN and CN functions can be executed in the same location, either the access or the network cloud. Thus, 5G NORMA opens the opportunity to **jointly optimize access and core functions**, taking advantage of the fact that these functions are no longer separated in different locations but may be executed next to each other.

In addition adapting the network functions to the different services, the above technologies also serve as an enabler for **mobile network multi-tenancy**, i.e., to flexibly share the network infrastructure among different operators. Indeed, 5G NORMA treats the network elements as a pool of resources that can not only be adaptively assigned to different services but also to different operators.

It follows form the above explanation that the 5G NORMA architecture relies on five key innovations, three of which can be considered as **<u>innovative enabling technologies</u>**: (i) adaptive (de)composition and allocation of mobile network functions, (ii) software-defined mobile network control (SDMC); and (iii) joint optimization of mobile access and core network functions; while the other two can be considered as **<u>innovative functionalities</u>**: (iv) multi-service and context-aware adaptation of network functions; and (v) mobile network multi-tenancy.

In the following we describe the key innovations in more details. Section III focuses on the three innovative enablers of 5G NORMA and provides some technological insights into their concept and design, while Section IV highlights the benefits resulting from the innovative functionalities of 5G

NORMA that are implemented by means of the enablers. Finally, Section V presents the 5G NORMA ***functional architecture***, which describes the specific mobile network functions of the 5G NORMA architecture in order to provide the desired innovative functionality building on the enabling technologies.

III. Innovative Enablers

Figure 1 illustrates the fundamental blocks of the 5G NORMA architecture: (i) the ***edge cloud***, which is the integration of base stations and distributed servers at the radio or at the aggregation sites, (ii) the ***network cloud***, which consists of servers at central sites, and (iii) the (centralized) ***controller***, which is responsible for controlling the functions executed in the edge and network clouds and is usually co-located in the network cloud. Along with these blocks, the figure also illustrates the "3" enabling technologies of 5G NORMA (i) **adaptive (de)composition and allocation of mobile network functions** between the edge and the network cloud depending on the service requirements and deployment needs; (ii) **Software-Defined Mobile network Control (SDMC)**, which applies the SDN principles to mobile network specific functions; and (iii) **joint optimization of mobile access and core network functions** localized together in the network cloud or the edge cloud. As we can see from the figure, different services (three examples are illustrated in the figure) use a different allocation of RAN and CN functions (both user- and control-plane): while for some services most of the functionality may be located in the edge cloud, for others it may be located in the network cloud. In the following, the key concepts behind the three innovative enablers of 5G NORMA are explained in more detail.

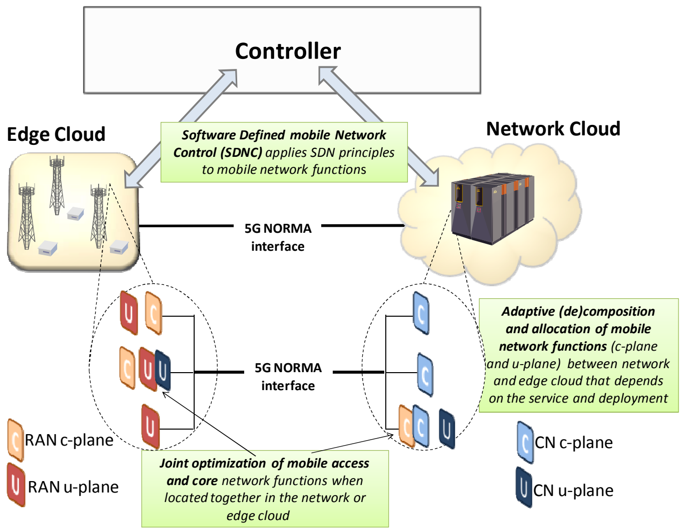

Fig. 1. NORMA concept based on the "3" innovative enablers.

A. *Adaptive (de)composition and allocation of mobile network functions*

Centralized RAN architectures [4] have recently attracted considerable attention as a technology that leverages cloud techniques to centralise the computational resources of the mobile network. However, current centralized RAN solutions have been designed with the assumption that the backhaul is composed of fibre links and hence mobile network functions location is static, which is unlikely to hold in future very dense deployments relying on wireless, bandwidth-limited, backhaul links.

5G NORMA develops a virtualization of mobile network functions which allows for a **flexible decomposition of mobile network functions between the radio access and the central cloud infrastructure**, supporting fully distributed, partially distributed, and fully centralized deployments. The decision of the degree of centralization (i.e. allocation of Network Functions) takes into account computational requirements, existing or required backhaul deployments, service and application requirements, and costs. This decision is driven by the optimization of metrics such as latency, QoE, resource utilization, energy efficiency, etc., and it is further dependent on the spatial and temporal characteristics as well as traffic fluctuations of the RAN. For instance, for latency-critical services we may allocate functions in the radio access to minimize delay, while we may allocate them in the central cloud for other services for efficiency reasons. This is especially important given the widening range of applications for 5G.

The decomposition and allocation of mobile network functions can be (broadly) based on the service requirements and backhaul deployment. For instance, we may choose different configuration settings as a function of the backhaul characteristics, i.e., the transport network connection to the antenna site, e.g., a fully distributed configuration, a fully centralized one and an intermediate one. The network rate and network latency are key in the selection of the configuration to be used, since (i) the more centralized the configuration is, the more information needs to be sent to the network cloud, and hence the larger the rate required, and (ii) the interaction between the lower layers of the protocol stack have very strict timing requirements, and hence splitting these lower layers between the edge and the cloud is only possible when fronthaul/backhaul latency is sufficiently low. In addition to the backhaul deployment, the decomposition and allocation of functions also depends on the service requirements (see Section IV).

B. *Software-Defined Mobile network Control (SDMC)*

Current Software-Defined Networking (SDN) implementations focus mostly on wired networks to separate routing control (which decides how to handle the traffic) from forwarding (which routes traffic according to decisions by the control plane).

5G NORMA applies these same principles to wireless functions beyond routing, where benefits of this technology are even more significant than for wired networks, as the control functionality of wireless networks include many **wireless related functions than just routing control** [5]. This includes time critical functions (such as scheduling control, MCS selection and HARQ processing) and other less time critical (like Radio Resource Control, power control and handover decision and execution). With the SDMC concept, all these functions can be implemented more easily by a programmable central control, which provides very important benefits for the flexible operation of the wireless edge network.

C. *Joint optimization of mobile access and core network functions*

Today, the physical separation between mobile access and core network functions limits the interaction between RAN and Core Network functions requiring the specification of complicated interfaces between those functions, which in turn delays innovation and technology uptake.

The adaptive (de)composition and allocation of mobile network functions described above may pool – in the edge or in the central cloud processor – functions that were located in the RAN in traditional architectures with other functions that were located in the CN. This not only blurs the separation between access and core network, but also provides new opportunities to jointly **optimize the operations of mobile access and core network functions** which previously were dealt with separately because of their different locations.

IV. Innovative Functionality

The objective of the three innovative enablers described in the previous section is to enable the two innovative functionalities of 5G NORMA, namely (i) *multi-service- and context-aware adaptation of network functions*, and (ii) *mobile network multi-tenancy*.

The two innovative functionalities of 5G NORMA are illustrated in Figure 2. The figure shows three examples of services that would probably be served with different configurations in terms of the decomposition and allocation of functions:

- *Vehicular communications*, which requires very low latency and hence would be better served by moving all the functionality into the edge so that delays are minimized; since this kind of services typically consist of separate packets that do not require session continuity, even CN c-plane functions can be (partially) moved to the edge in this case;
- *Tactile communications*, which also requires very low latencies and (like in the previous case) would better be served by moving all the functionality to the edge; however, in this case session continuity is required and hence CN c-plane functions need to stay in a central location;
- *Internet access*, which does not have any stringent latency requirements and hence can be served by moving all the functionality to the cloud, which provides efficiency gains.

Of course, the specific design of the optimal allocation for each of these services is something that needs to be carefully designed for each specific service, and the purpose of the above examples is no other than illustrating the variety of possible configurations that may be used.

The figure also shows the *multi-tenancy* feature of 5G NORMA. As it can be seen in the figure, multi-tenancy is supported by placing the different operator modules ('tenant') on top of the controller, which is the responsible module for adaptive allocating the mobile network resources to the different 'tenants' or operators.

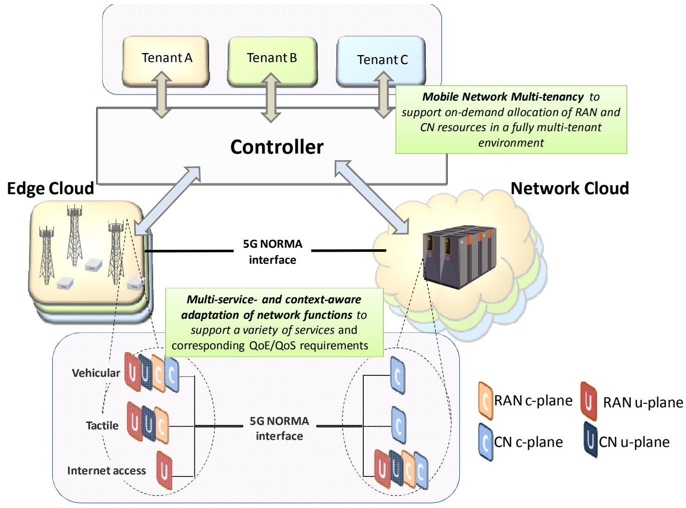

Fig. 2. Multi-service and multi-tenancy architecture.

A. *Multi-service- and context-aware adaptation of network functions*

Although current network architectures and concepts already support a variety of service classes and respective QoE/QoS, network functions typically cannot be reallocated nor adapted to service classes. Current architectures are based on a pre-defined set of bearers, support only limited variations of the topology, and, thus, prevent the networks from efficiently supporting new and dynamically changing service classes.

In contrast to this, 5G NORMA considers fully adaptive, programmable network functions that might – depending on their usage context – be reprogrammed and even relocated. The latter option **allows for a dynamic relocation of network functions between the edge of the network and a centralized cloud infrastructure, thereby enabling low latency communication** and global optimisation.

B. *Mobile network multi-tenancy*

Infrastructure sharing is a key business model for mobile operators (tenants) to reduce deployment and operational costs, driven by the capacity requirements forecasted for future mobile networks as well as the decreasing operators' benefits margin. While passive and active sharing solutions are partially used and standardized today, these sharing concepts are based on fixed contractual agreements with mobile virtual network operators (MVNOs) on a coarse granularity basis (monthly/yearly).

5G NORMA introduces the unprecedented no-human intervention (signalling-based) dynamic sharing concept referred to as **Mobile Network Multi-tenancy** building on the 3GPP SA1/SA5 infrastructure sharing enhancements efforts towards on-demand capacity brokering. Our concept supports classical (e.g. mobile operators) and non-classical tenants, e.g. YouTube, Netflix, utility companies (energy), etc. In our vision, **5G NORMA Multi-tenancy does not only consider RAN resources but a range of other scarce network resources, including backhaul and core network capacity**.

V. FUNCTIONAL ARCHITECTURE

In order to provide the desired features in terms of ***multi-service adaptation*** of network functions and support for ***multi-tenancy***, the 5G NORMA architecture comprises functions such as *Radio Resource Management* (RRM), *scheduling*, *QoE control* and *mobility management*, among others. The design of these functions is performed leveraging on the three key enabling technologies of 5G NORMA, namely (i) *adaptive allocation of mobile network functions*, (ii) *SDMC* and (iii) *joint optimization of access and core*. In the following, we describe the main functions of 5G NORMA, which corresponds to the 5G NORMA functional architecture.

The network architecture includes the following modules: Radio Connection Management, Radio Resource Management (RRM) and scheduling. These can be classified as radio access functions and mobility/QoS management. The key functionality of each of the radio access modules, and the underlying design criteria, are as follows:

- The ***Radio Connection Management*** module is responsible for selecting the available and suitable communication links for a given user or session, including device-to-device, multicast or multi-connectivity links, among others. This module is specifically designed to provide *multi-service adaptation* and one the key criteria behind its design is the coupling of access and core functions to *jointly optimize access and core* operation.
- The ***Radio Resource Management*** (RRM) module includes admission and congestion control functions to determine the resources allocation. This module is key to the *multi-tenancy* support functionality, and its design includes the *adaptive allocation* to the RAN or the cloud depending on the service requirements.
- The ***scheduling*** module is closely linked to RRM and is responsible for scheduling the users' sessions, including link adaptation, HARQ processing, interference coordination (e.g., ICIC) and full joint transmission and reception. The design of this module follows *a software-defined mobile network control* (SDMC) approach, its functionality being split between the RAN and the centralized server.

The key functionality of the mobility and QoE/QoS management modules and their underlying design criteria are described next:

- The ***Mobility Management*** module is responsible for managing mobility at a per-flow or per-device level. A key feature of its design if the *joint optimization of core and access*, which is implemented through its interaction with the QoE/QoS controller module (described below).
- The ***QoE/QoS Controller*** module is responsible for gathering information on the services requirements and implementing the necessary functionality to ensure that these requirements are met. This function is central to both *multi-service adaptation* and *multi-tenancy*, and its design follows the *SDMC* concept.
- The ***Orchestrator*** module is responsible for the network-wide orchestration to achieve the desired level of flexibility, scalability and robustness. This functionality is closely related to the *multi-service adaptation* and SDMC enabling technologies.

Figure 3 shows the key radio access and mobility/QoS functions of the 5G NORMA architecture described above as well as their inter-relation. It is worthwhile noting that, in addition to the five key innovation criteria that guide the design of these modules, these functions also include some key innovative aspects in their design.

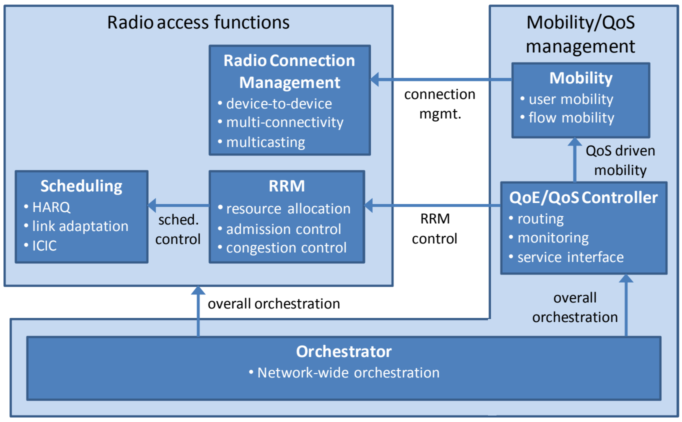

Fig. 3. 5G NORMA Functional architecture.

VI. SUMMARY AND CONCLUSIONS

Upcoming cellular networks will have to satisfy requirements that today's architecture cannot provide in terms of adaptability to a large variety of services and cost-effectiveness. To meet such requirements, a novel architecture needs to be designed – rather than a simple evolution of today's network, this new architecture requires of disruptive concepts that provide the desired level of flexibility and contribute to improve efficiency and (as a consequence) cost.

In this paper we have presented the basic design guidelines of what we believe will be the key building blocks of the future 5G architecture. The key concepts rely on the notions of flexible decomposition and allocation of functions in the access and cloud and software-defined mobile control, which allow for jointly optimizing access and core functionality. Building on these enablers, future networks will be able to provide multi-service adaptation as well as multi-tenancy support.

REFERENCES

[1] Cisco Visual Networking Index: Forecast and Methodology, June 2014.

[2] Nokia white paper, "5G use cases and requirements", 2014.

[3] X. Costa-Pérez, J. Swetina, T. Guo, R. Mahindra and S. Rangarajan. Radio Access Network (RAN) Virtualization for Future Carrier Networks. IEEE Communications Magazine, July 2013.

[4] Mobile Research Institute, "C-RAN The Road Towards Green RAN". Oct. 2011.

[5] C.J. Bernardos, A. de la Oliva, P. Serrano, A. Banchs, H. Jin, L.M. Contreras, J.C. Zuñiga, "An Architecture for Software Defined Wireless Networking," IEEE Wireless Communications Magazine, vol. 21, no. 3, April 2014.