5G PPP Architecture Working Group

View on 5G Architecture

This document will be updated after a public consultation, pls download the final version after the EuCNC 2016 conference (July 1st) from the following link:

https://5g-ppp.eu/white-papers/

Table of Contents

List of Acronyms and Abbreviations

3GPP	3rd Generation Partnership Project
4G	4th Generation Mobile Network
5G	5th Generation Mobile Network
5G PPP	5G Infrastructure Public Private Partnership
API	Application Program Interface
ARN	Active Remote Node
BBU	Baseband Unit
BSP	Business Service Plane
BSS	Business Support System
CAPEX	Capital Expenditure
CN	Core Network
CO	Central Office
CoMP	Coordinated Multipoint
CP	Control Plane
CPRI	Common Public Radio Interface
C-RAN	Cloud Radio Access Network
DC	Data Center
D-RAN	Distributed Radio Access Network
DC	Data Centre
DCB	Data Center Bridging
DWDM	Dense Wavelength-Division Multiplexing
E2E	End to End
EPC	Evolved Packet Core
EPS	Evolved Packet System
FEC	Forward Error Correction
FTTH	Fiber-to-the-Home
G.fast	Transmission Technology for Telephone Lines up to 1 Gbit/s
HAL	Hardware Abstraction Layer
HARQ	Hybrid Automatic Repeat Request
IEEE	Institute of Electrical and Electronics Engineers
LTE	Long term Evolution
LTE-A	LTE-Advanced (a.k.a. 4G)
MAC	Medium Access Control
MANO	Management and Orchestration
MEC	Mobile Edge Computing
MME	Mobility Management Entity

MEF	Metro Ethernet Forum
MSMP	Multi-Service Management Plane
NGFI	Next Generation Radio Interface
NG-PON2	Next Generation PON Technology based on WDM and TDMA
OBSAI	Open Base Station Architecture Initiative
OFDM	Orthogonal Frequency-Division Multiplexing
OPEX	Operational Expenditure
OLT	Optical Line Terminal
ONF	Open Networking Foundation
ONU	Optical Network Unit
ORI	Open Radio Interface
OSS	Operations Support System
PDCP	Packet Data Convergence Protocol
PNF	Physical network Function
PON	Passive Optical Network
RAN	Radio Access Network
RFB	Reusable Function Block
RLC	Radio Link Control
ROADM	Reconfigurable Optical Add-Drop Multiplexer
RRH	Remote Radio Head
RRC	Radio Resource Control
SDK	Software Development Kit
SDN	Software Defined Networking
SDO	Standards Developing Organization
TSN	Time-Sensitive Network
TDMA	Time-Division Multiplexing
TSON	Time Shared Optical Network
MANO	Management and Orchestration
MIMO	Multiple-input Multiple-output
OSS/BSS	Operations and Business Support Systems
VDSL	Very High Speed Digital Subscriber Line
vCore	Virtual Core
VIM	Virtualized Infrastructure Manager
VNF	Virtual Network Function
VNFM	VNF Manager
WAN	Wide Area Network
WDM	Wavelength-division Multiplexing
WSS	Wavelength-Selective Switch

1 Introduction

The European Union funded 5G Public Private Partnership (5GPPP) is an important initiative where public and private sectors in Europe work together to develop 5G and secure the European leadership. Several projects[1] have received support to work on areas ranging from physical layer to overall architecture, network management and software networks. This is very important because 5G is not only a new radio but also a framework that integrates new with existing technologies to meet the requirements of 5G applications. The 5G Architecture Working Group as part of the 5GPPP Initiative is looking at capturing novel trends and key technological enablers for the realization of the 5G architecture. It also targets at presenting in a harmonized way the architectural concepts developed in various projects and initiatives (not limited to 5GPPP projects only) so as to provide a consolidated view on the technical directions for the architecture design in the 5G era.

The current white paper focuses on the produced results after one year research mainly from 16 projects working on the abovementioned domains. During several months, representatives from these projects have worked together to identify the key findings of their projects and capture the commonalities and also the different approaches and trends. Also they have worked to determine the challenges that remain to be overcome so as to meet the 5G requirements. The goal of 5G Architecture Working Group is to use the results captured in this white paper to assist the participating projects achieve a common reference framework. The work of this working group will continue during the following year so as to capture the latest results to be produced by the projects and further elaborate this reference framework.

The 5G networks will be built around people and things and will natively meet the requirements of three groups of use cases:

- Massive broadband (xMBB) that delivers gigabytes of bandwidth on demand
- Massive machine-type communication (mMTC) that connects billions of sensors and machines
- Critical machine-type communication (uMTC) that allows immediate feedback with high reliability and enables for example remote control over robots and autonomous driving.

The demand for mobile broadband will continue to increase in the next years, largely driven by the need to deliver ultra-high definition video. However, 5G networks will also be the platform enabling growth in many industries, ranging from the IT industry to the automotive, manufacturing industries entertainment, etc.

5G will enable new applications like for example autonomous driving, remote control of robots and tactile applications, but these also bring a lot of challenges to the network. Some of these are related to provide low latency in the order of few milliseconds and high reliability compared to fixed lines. But the biggest challenge for 5G networks will be that the services to cater for a diverse set of services and their requirements. To achieve this, the goal for 5G networks will be to improve the flexibility in the architecture.

The white paper is organized as follows. In section 2 we discuss the key business and technical requirements that drive the evolution of 4G networks into the 5G. In section 3 we provide the key points of the overall 5G architecture where as in section 4 we elaborate on the functional

---

[1] 5G PPP Phase I Projects - https://5g-ppp.eu/5g-ppp-phase-1-projects/

architecture. Different issues related to the physical deployment in the access, metro and core networks of the 5G network are discussed in section 5 while in section 6 we present software network enablers that are expected to play a significant role in the future networks. Section 7 presents potential impacts on standardization and section 8 concludes the white paper.

2 Challenges and requirements and key differentiating characteristics

2.1 5G Key Requirements

5G networks are expected to offer the opportunity to launch, efficiently and cost-effectively, numerous new services thus, creating an ecosystem for technical and business innovation. In addition, the 5G infrastructures will provide tailored network solutions specialized to **support vertical markets** such as automotive, energy, food and agriculture, healthcare, etc. ([2-1]) Moreover, it will be necessary to **accelerate the service delivery** to all the involved stakeholders. It is exactly the need to support a diverse set of vertical industries and simplify their provision that calls for new advanced architectural frameworks for the processing and transport of information. Contrary to the evolution of previous generations of mobile networks, 5G will require not only improved networking solutions but also a sophisticated integration of massive computing and storage infrastructures.

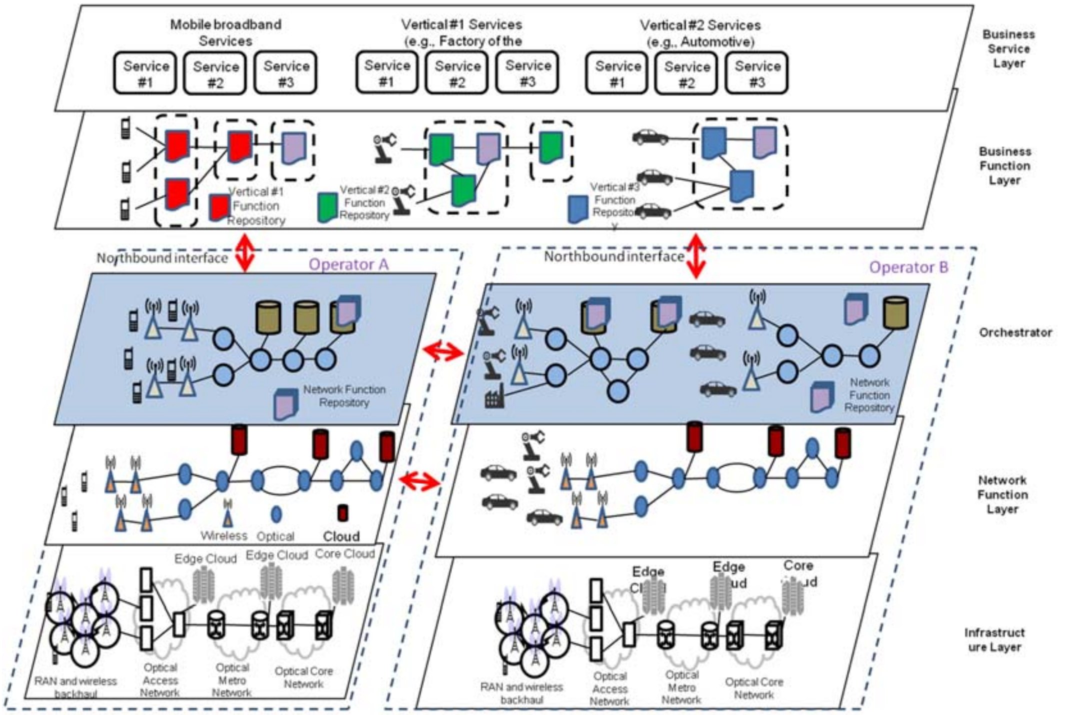


Figure 2-1: The 5G ecosystem

It is anticipated that service providers will require **access to resources of the underlying network and computing infrastructure**. Thus, the infrastructure providers will expose, via northbound interfaces, their telecommunication systems to typical mobile broadband or new vertical service providers. This will allow **multi-tenancy and multi-service support,** as well as access to either mobile or converged fixed-mobile access networks where different networking policies will be enforced. In Figure 2-1, following the definitions of ([2-1]), the top layers (i.e., business service and functions layers) are involved with the specification and implementation of the business processes and the provision of application related functions organized in function repositories. The service providers may offer their services through one or multiple telecommunication operators. A telecommunication operator may also have the role of a service provider as it the case today.

To serve such a diverse ecosystem, the telecommunication operators will have to deploy **orchestrator functions** that will allocate appropriate computing and network resources to the services targeting diverse and dedicated **business driven** logical networks. These logical networks (network function layer in Figure 2-1), so-called **network slices**, will contain specialized networking and computing functions that meet the desired KPIs of the service providers. Note that in cases where a single infrastructure provider is not on its own able to support the requirements of a service provider, 5G networks will support **cross-domain orchestration** of services and resources over multiple administration domains allowing for **flexible sharing schemes**. The implementation of these schemes will also require interworking among operators in the network function layer as well (e.g., setting up SDN rules).

The abovementioned evolutions will also have to operate in a **ubiquitous** and **energy-efficient** way. Moreover, the 5G system will have to be designed in a **future-proof way** so as to enable smoother transitions in future generations.

A faster service instantiation will call for **new trust models** to support new business and service delivery models in an evolved cyber-threat landscape. This new environment also calls for innovative solutions to address the increasing societal concerns regarding **user privacy.**

The abovementioned ecosystem is the anticipated outcome of addressing challenges derived from a large number of new use cases. During the past years, several organizations (e.g., [2-2]), forums (e.g., [2-3]), or research projects worldwide, have been trying to identify the new use cases and the requirements these will impose to the network architecture. Currently, all active projects, funded by the EU under the 5G Public Private Partnership, define and analyze the envisioned 5G use cases related to their research areas. Although a plethora of use cases has been already defined, a first-level grouping into three main categories, based on the key considered services, is widely adopted. The categories are: Extreme mobile broadband (xMBB); massive machine-type communications (mMTC); and ultra-reliable machine-type communications (uMTC). However, an analysis based solely on this grouping is not sufficient, since different use cases may have different characteristics (e.g., mobility and data traffic patterns) and hence different values for requirements (e.g., delay, reliability, user throughput etc). The extreme diversities of services, as well as the vast number of end devices that will have to be supported, yield a unprecedented set of requirements that has to be taken into consideration.

2.2 5G Design Objectives

In 5G networks, spectrum availability is one of the key challenges of supporting the enormous mobile traffic demand. Nowadays, the current spectrum is crowded already. Especially in very dense deployments it will be necessary to go higher in frequency and use larger portions of free spectrum bands. This means that 5G networks will **operate in a wide spectrum range with a diverse range of characteristics**, such as bandwidth and propagations conditions. Thus, appropriate mechanisms are needed that today do not exist in the current 4G systems. Another potential solution could be the adoption of some appropriate spectrum-sharing technique. This implies that the new 5G architecture should allow spectrum to be managed more efficiently, by accurately monitoring spectrum usage and by enabling sharing strategies in mobile networks. Note that in [2-1] it is clearly mentioned that “Concerning bands above 6 GHz the relevant WRC-15 resolution calls for appropriate studies to be conducted and completed in time for WRC-19 to determine the spectrum needs and the appropriate sharing and compatibility conditions with the incumbent services.” Alternatively, to improve spectrum usage efficiency mechanisms are under investigation to facilitate multi-RAT resource allocation. This could exploit spectrum in licensed, unlicensed and/or lightly-licensed bands. It would allow prioritization and allocation of traffic across heterogeneous access technologies in a dynamic way to diversified spectrum resources.

These spectrum characteristics, as well as the diverse use case requirements, will require the ability to concurrently support multiple instances of differently parameterized network functions, or even the introduction of novel network functions. The exact parameterization and the placement of these functions will depend on the deployment of the available hardware, the nature of the communication links and the required topology. Moreover, current research activities suggest that 5G networks should provide the means for highly **efficient transmission and data processing**. Examples of this feature include realizations of network functions inside the radio protocol stack, allowing e.g. fast access of devices for mMTC with extremely low overhead for the control plane signaling. Moreover, low latency solutions are being investigated that involve placing network functions closer at the edge of the access network. The framework of **Mobile Edge Computing (MEC)** will also play an important role for meeting a number of crucial requirements extending the Infrastructure-as-a-Service concept up to the last mile.

To achieve all the above capabilities, most likely new paradigms and enablers such as **SDN and NFV** need to be followed and supported. In such an ultra-flexible environment, it is necessary to consider new solutions, such as the **separation of user and control planes,** and possibly, **re-definition of the boundaries between the network domains** (e.g. radio access network and core network). Although improving the flexibility of the future networks is a definite requirement, the added complexity that it will introduce (e.g., number of new interfaces, novel network management functions, security and trust issues) has to be carefully studied and evaluated. Most likely, different domains (edge, access, transport, core, services) of 5G networks will offer different levels of flexibility. Furthermore, 5G networks will have to offer solutions to support **different air interface variances** in an efficient way. This requires certain innovations, including the configuration of the air interface using different numerologies, waveforms, etc., evolved resource management solutions for heterogeneous environments, mechanisms for integrating the control and user plane with legacy or non-5GPPP systems, etc. Also 5G should provide an efficient **interworking between 5G and an evolution of LTE** as the latter could already meet the requirements for some of the uses cases discussed for 5G like the Narrow Band Internet-of-Things (NB-IoT).

Moreover, 5G networks will have to address the **complexity** of advanced communication modules and different antenna types with different beam-forming capabilities. Examples of this are multi-antenna schemes with large antenna arrays, massive MIMO and clustering of millimeter-wave access points addressing the coverage and mobility needs by using beam-steered antenna patterns. Furthermore, depending on the use case and deployment scenario, it is needed to support different antenna types, e.g. omni-directional antenna patterns, low/high gain beam forming antenna pattern, flexible/fixed beam forming pattern, and analog/digital/hybrid beam forming is required, depending on the use case and deployment scenario

Another new feature that distinguishes 5G networks from legacy systems is the native and efficient support of communication schemes like **multi-connectivity** (e.g. communication of a single user with two or more different network nodes operating in different RATs, which may also employ high or very high frequencies). Multi-Connectivity is a key technology to fulfil 5G requirements related to data-rate, latency, reliability and availability. Also, 5G will support novel schemes like the **network-controlled device-to-device (D2D)** communication, including point-to-point, multi-cast and broadcast communication. Other novel mechanisms include **device duality** schemes, where a device can act both as a “normal” end user device (including sensor types) and as a network node extending the infrastructure part of the system. These schemes will have to be supported over a **wide range of physical deployments**, from distributed base stations to centralized cloud-RAN deployments or distributed edge clouds.

Different types of backhaul, such as converged optical and wireless transport network solutions, will be also supported taking the trade-off between delay and capacity into account. **Self-backhauling**, where devices can act as base stations and self-establish wireless backhaul links to suitable donor base stations, is regarded as another important feature. The 5G architecture will provide inherent means for convergent fixed-mobile networking. Operators will be able to use the same physical network to provide access to fixed and mobile users. Ethernet is expected to be used as a common transport platform, allowing the integration of new and existing transmission technologies. Virtual networks can then be operated in parallel slices on the same physical network. A superordinate network operation and management functionality will be provided based on SDN over the same infrastructure, as only a fraction of the overall telecom traffic is mobile. Fixed-mobile convergence enables the mobile network to reuse the existing fixed network infrastructure for the rollout. It is also important to provide a consistent and continuous service experience for all end users, independent of the underlying access network.

The 5G networks should also support more sophisticated mechanisms for traffic differentiation than those of legacy systems, in order to fulfill diverse and more stringent end-to-end Quality of Service (QoS) requirements. Note though that 5G networks will have to provide for the **separation and prioritization of resources on a common infrastructure** for operational and security purposes. The **support of the slicing framework** will have to take into consideration these QoS requirements**.**

Access networks and network equipment deployed at the edge are specialized devices with ‘hard-wired’ functionalities. Any adaptation to the ever-increasing and heterogeneous market requirements implies a huge investment to change and deploy hardware. One potential solution could be the **virtualization of part of the communications infrastructure** (e.g. core/edge segments and access points/macrocells); but other innovative solutions like the **appropriate use of small cell infrastructures** should also be examined.

The new services are expected to be “manufactured by software”, hosted in a multi-domain “infrastructure factory” where resources and network functions are dynamically and flexibly traded and provisioned. This new situation calls for **End to End Resource, Infrastructure and Service Orchestration** (i.e., multi-domain orchestration of diverse programmable infrastructure domains, possibly belonging to different administrations/operators). Also, control and business parameters need to be exchanged to realize integrated services involving multiple infrastructure owners. This will allow **overcoming the Over-The-Top (OTT) issue**, where application providers send traffic over the top of the Internet, across multiple networks to end users without any delivery guarantee. 5G networks will have to support a significant number of new services through multiple tailor-made environments. This calls for scalable **new advanced autonomic network management** platforms. Furthermore, this involves the collection and processing of large data volumes from the 5G network, and the development of a system for managing network nodes while supporting federated network management. This is crucial for guaranteeing QoS even when the network context changes. Towards this end, investigations are ongoing as to what extent 5G networks and devices (e.g. using over-the-air programmability) can be software-configurable and to what extent software platforms can be hardware-agnostic. Self-organized capabilities enable the network to efficiently predict demand and to provide resources, so that it can heal, protect, configure and optimize accordingly. The platforms will do this by generating the minimum cost on network equipment (CAPEX) and operations cost (OPEX), whilst keeping QoS tailored to user demand with adequate resources. The operational cost includes network resource allocation, service provision and monitoring, performance degradation and energy efficiency. Moreover, the management platforms will offer network resilience mechanisms, such as the identification of network errors, faults or conditions like congestion or performance degradation. Also, they will identify serious security issues such as unauthorized intrusion or compromised network components, and liaise with autonomic network management to formulate and take appropriate action. The overall objective is to create a cognitive and autonomic management system developed through the application of policies that can self-adapt to the changing conditions of the network and to the external environment in which the network operates, via a well-defined set of self-organizing functions. These platforms also need to support **multi-tenancy** environments.

[2-1] 5G empowering vertical industries (February 2016) https://5g-ppp.eu/wp-content/uploads/2016/02/BROCHURE_5PPP_BAT2_PL.pdf
[2-2] 3GPP TR 22.891 V2.0.0, Technical Specification Group Services and System Aspects; Feasibility Study on New Services and Markets Technology Enablers; Stage 1 (Release 14), 2/2016
[2-3] NGMN Alliance, “NGMN 5G White paper”, February 2015

3 Overall Architecture

This section discusses topics related to the overall 5G architecture and its impact on (i) Mobile Networks, (ii) Physical Networking and Computing Facilities, (iii) Service & Infrastructure Management and Orchestration and (iv) Hosting and Deployment Systems.

5G networks are conceived as extremely flexible and highly programmable E2E connect-and-compute infrastructures that are application- and service-aware, as well as time-, location- and context-aware. They represent:

- an evolution in terms of **capacity**, **performance** and **spectrum access** in radio network segments; and
- an evolution of native **flexibility** and **programmability conversion** in all non-radio 5G network segments: Fronthaul and Backhaul Networks, Access Networks, Aggregation Networks, Core Networks, Mobile Edge Networks, Software Networks, Software-Defined Cloud Networks, Satellite Networks and IoT Networks.

5G Architecture enables new business opportunities meeting the requirements of large variety of use cases as well as enables 5G to be future proof by means of (i) implementing network slicing in cost efficient way, (ii) addressing both end user and operational services, (iii) supporting softwarization natively, (iv) integrating communication and computation and (v) integrating heterogeneous technologies (incl. fixed and wireless technologies)

These qualities give 5G networks a number of advantages. One is a high degree of **flexibility**. They serve highly diverse types of communication – for example, between humans, machines, devices and sensors - with different performance attributes. They also enforce the necessary degree of flexibility, where and when needed, with regard to capability, capacity, security, elasticity and adaptability.

5G networks represent a shift in networking paradigms: a transition from today’s “network of entities” to a **“network of (virtual) functions”**. Indeed, this “network of (virtual) functions”, resulting, in some cases, in the decomposition of current monolithic network entities will constitute the unit of networking for next generation systems. These functions should be able to be composed on an “on-demand”, “on-the-fly” basis. In fact, a research challenge consists in designing solutions which identify a set of elementary functions or blocks to compose network functions, while today they implemented as monolithic

Further advantages emerge in the areas of **management**, **control** of systems and **resources**. 5G networks enable the uniform management and control operations that are becoming part of the dynamic design of software architectures. They can host service executions in one or more slices.

Since no single solution fits all needs as captured in section 2, the following Figure 3-1 presents the various relevant views on emerging 5G designs. The network softwarization and programmability is one of the basis for the 5G architecture design and impacts all identified views, as presented in subsection 3.1. The views are then further elaborated in the following subsections, in particular aspects about the logical and functional view are introduced in subsection 3.2, the physical resources and infrastructure control views are in subsection 3.3 and finally the system management and applications and business services view in subsection 3.4. More detailed analysis of those architectural views are then presented in sections 4–6.

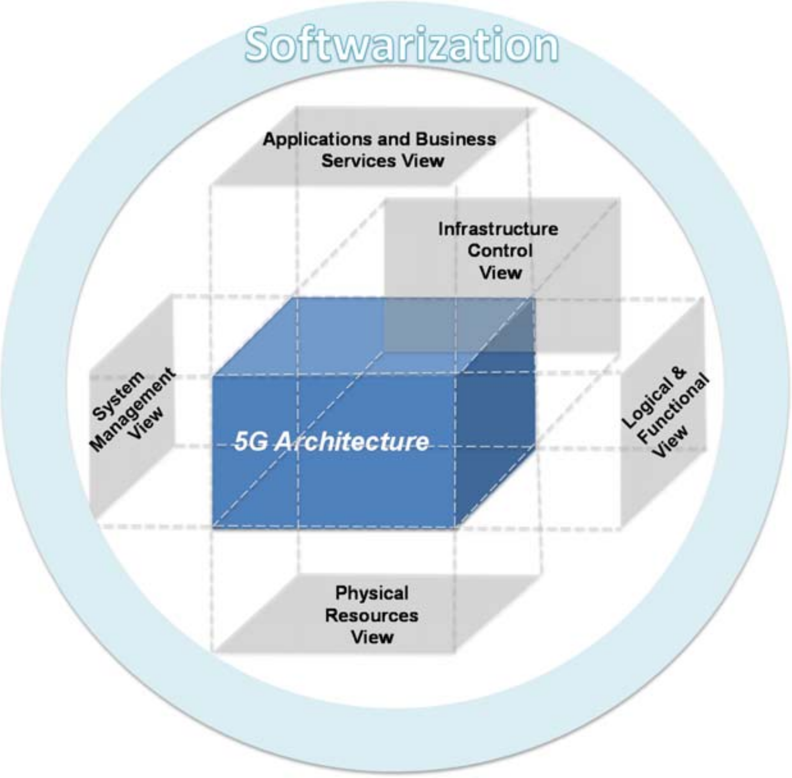


Figure 3-1: 5G Architecture Views

3.1 Network Softwarization and Programmability

The proposed framework is aimed at all 5G Network segments: *Radio Networks, Fronthaul & Backhaul Networks, Aggregation and Core Networks, Network Clouds, Mobile Network (i.e. a combination of network segments where the last link is wireless - a radio network) and enabling technologies like Mobile Edge Networks, Service/Software Networks, Software-Defined Cloud Networks, Satellite Networks, IoT Networks*.

The perspectives of this proposal are described as separate **planes**. Although separately defined, the planes are not completely independent: key items in each are related to items in the other planes. However, the planes are sufficiently independent to simplify reasoning about the complete system requirements. The interworking between planes is manifested by groups of interfaces (i.e. reference points) that would be used for exchange of information and/or controls between separate (sub)systems sharing boundaries. The projected separation of concerns in distinct planes are: *Application and Business Service Plane, Multi-Service Management Plane, Integrated Network Management & Operations Plane, Infrastructure Softwarization Plane, Control Plane and Forwarding/Data Plane*.

The proposed framework for network softwarization and programmability is presented in Figure 3-2, where each plane is exemplified.

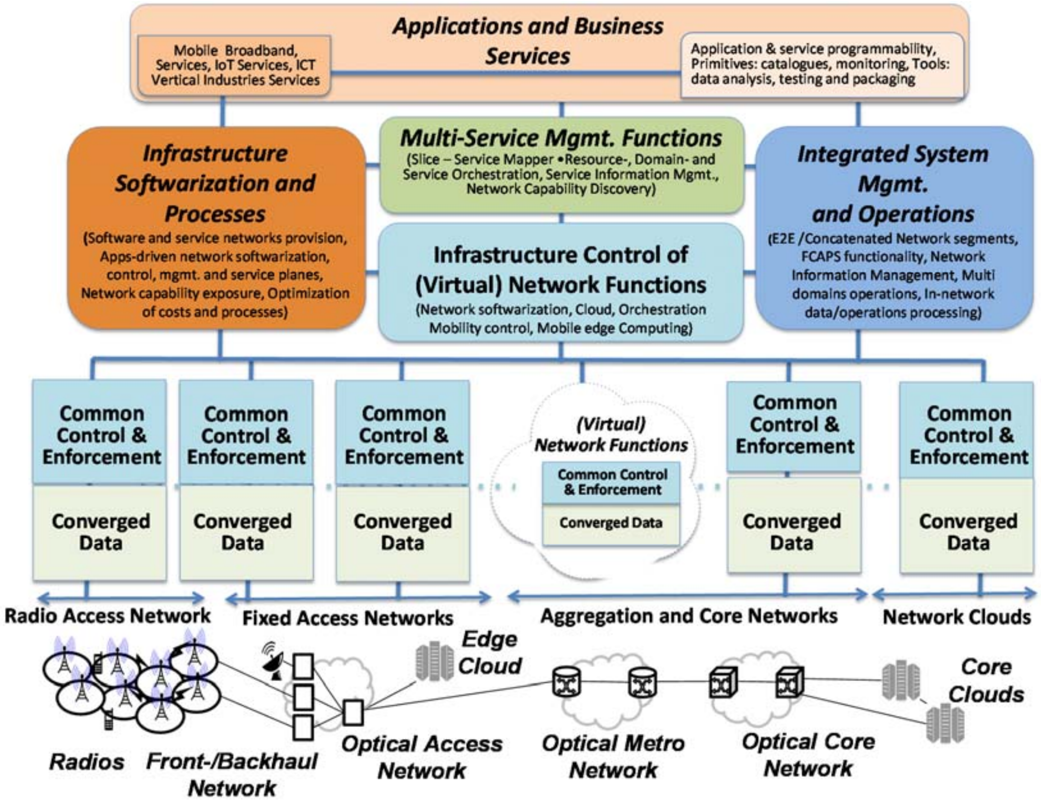


Figure 3-2: Network Softwarization and Programmability Framework

The following are key system-differentiating functions of the network softwarization and programmability framework:

- **5G Converged Data Plane functions** are distributed to the edges of a common core network, resulting in creating a distributed flat network. The control plane functions, which are responsible for mobility management, QoS control, etc., direct the user traffic to be served agnostically to the access networks to which it is attached. They also integrate heterogeneous technologies (including fixed and wireless technologies).
- **5G Infrastructure Softwarization Plane functions**, which are responsible for native network softwarization (network virtualization, functions virtualization, programmability, etc.) in all 5G network segments, enable effective integration of communication and computation.
- **'Network of (virtual) function'** is supported as the unit of networking in 5G Networks.
- **Network architecture can evolve and change** rather than needing to be replaced.

This network softwarization and programmability framework is based on the following separation in distinct planes:

Application and Business Service Plane – Defines and implements the business processes of the services along specific value chains. A service in the 5G context is a piece of software that performs one or more functions, provides one or more APIs to applications or other services of the same or different planes to make usage of those functions, and returns one or more results. Services can be combined with other services, or called in a serialized manner to create a new service. An application in the 5G context is a piece of software that utilizes the underlying services to perform a function. Application operation can be parameterized, for example, by passing certain arguments at call time, but it is meant to be a standalone piece of software; an App does not offer any interfaces to other applications or services.

Multi-Service Management Plane – The functions and interfaces in this plane are used to set up and manage groups of network instances and/or nodes. More specifically, the setup consists of creating/installing/arranging NFs and interfaces according to the available physical and virtual resources. It also comprises the set of functions associated with the network operations, such as fault management, performance management and configuration management. It further includes Slice –Service Mapper functions, Resources, Domain and Service Orchestration functions, Service Information Management functions and Network Capability Discovery functions. It also includes the lifecycle management of individual network functions and mobile network instances as a whole. In current mobile networks, this role is often performed by the Operations Support System (OSS). The idea is to enable the creation, operation, and control of multiple dedicated communication service networks running on top of a 5G E2E infrastructure.

Integrated Network Management & Operations Plane – Enables the creation, operation, and control of dedicated management functions operating on top of a 5G E2E infrastructure; and the collection of resources required for managing the overall operation of individual network devices. It further includes E2E Network segments management, FCAPS functionality, Monitoring operations, Network Information Management, In-network data and operations processing and Multi domains management operations

Infrastructure Softwarization Plane – Enables the provisioning and operation of software and service networks. It facilitates the operation of end-to-end heterogeneous networking and distributed cloud platforms, including physical and logical resources and devices. It includes software for designing, implementing, deploying, managing and maintaining network equipment, network components and/or network services by programming. The software utilize features such as flexibility and rapidity all along the lifecycle of network equipment/components/services, in order to create conditions that enable the re-design of network and services architectures, optimize costs and processes, allow self-management and bring added value to network infrastructures [3 -15]. It further includes provision of software and service networks, application driven network softwarization, S/W Programmability of Software Networks, dynamic deployment of new network and management services (i.e. which could be executed in data, control, management, service plane), network capability exposure, E2E slice provisioning and control in software networks.

***Infrastructure Control Plane* -** The collection of functions responsible for controlling one or more network devices. Control Plane instructs network devices, network elements, and network functions with respect to processing elementary data units (packets, frames, symbols, bits, etc.) of the user/data/forwarding plane. The control of (virtual) network functions include Control of Network Softwarization functions, Control of Orchestration functions, Control of Mobility control functions, Cloud Control functions, Mobile Edge Computing Control functions and adaptors to different enforcement functions. The control of (virtual) network functions is generally 5G-applicable, and they are separated from the control and enforcements functions which are network segment-specific. The control plane interacts primarily with the forwarding plane and, to a lesser extent, with the management plane.

***Forwarding Plane / Data Plane* -** The collection of resources across all network devices responsible for forwarding traffic.

3.2 Impact on Mobile Networks

The evolution of the mobile network architecture is driven by the need to provide communication services for a manifold of applications. In addition to that, **network slicing** is also an important part of the overall 5G architecture that addresses the deployment of multiple logical networks as independent business operations on a common physical infrastructure. One of the goals would be to provide network slices flexibly meeting the wide range of use cases that the 2020 timeframe will demand ([3 -1][3 -2][3 -3][3 -8][3 -16]). To this end, a “5G slice” could be composed of a collection of 5G network functions (NF) and specific radio access technology (RAT) settings that are combined together for a specific use case and/or business model. Furthermore, since the concept of network slicing was initially proposed to be adopted by the 5G core network (CN), NGMN uses the term “end-to-end (E2E) network slicing” to refer to the overall system design concept, including both CN and RAN aspects [3 -7][3 -4][3 -2]. In that context, network slices must fulfill a set of requirements such as the need for sharing and efficiently reusing resources (including radio spectrum, infrastructure, and transport network); differentiation of traffic per slice; visibility of slices; protection mechanisms among slices (a.k.a. slice isolation); and support for slice-specific management ([3 -19]). The support for E2E network slicing appears as one of the key requirements in 3GPP, it is still under discussion how exactly network slicing would impact the RAN design, on both the access network and user equipment (UE) sides, although these concepts are currently under investigation ([3 -21][3 -2]).

In order to address the imposed 5G requirements, a novel 5G mobile network architecture is foreseen which provides the means to support the expected service diversity, flexible deployments, and network slicing.

There is a common understanding that key technology options towards flexibility are the usage of a multi-service and context-aware adaptation of network functions, adaptive (de)composition and allocation of mobile network functions, software-defined mobile network control, and joint optimization of mobile access and core network functions [3 -22]. The 5G mobile network architecture would include both physical and virtual network functions, as well as edge-cloud and central-cloud deployments. Further, it is clear that that the 5G mobile network needs to integrate LTE-A evolution with novel 5G technologies [3 -2] on RAN level (according to the current definition of RAN), whereby RAN level integration would go far beyond existing interworking between access technologies, fulfilling the vision of what NGMN calls a “5G RAT family” ([3 -23]).

It is currently being assumed by 3GPP for the Next Generation Architecture that a logical CN/RAN split will exist (possibly with some change in the exact logical split between RAN and CN), allowing for an independent evolution of both RAN and CN, and for cross-layer optimizations in some deployments when the functions are co-located. As shown in Figure 3-3, this setup could make use of a S1* CN/RAN interface and X2* inter-node RAN interface as studied in [3 -20].

There is also ongoing research ([3 -8]) of concepts aiming to provide a high degree of architecture flexibility, e.g., a flexible assignment and integration of RAN and CN functions. Focus of future research work is to develop all options and compare them in terms of flexibility, complexity and cost involved in meeting the requirements of future uses cases.

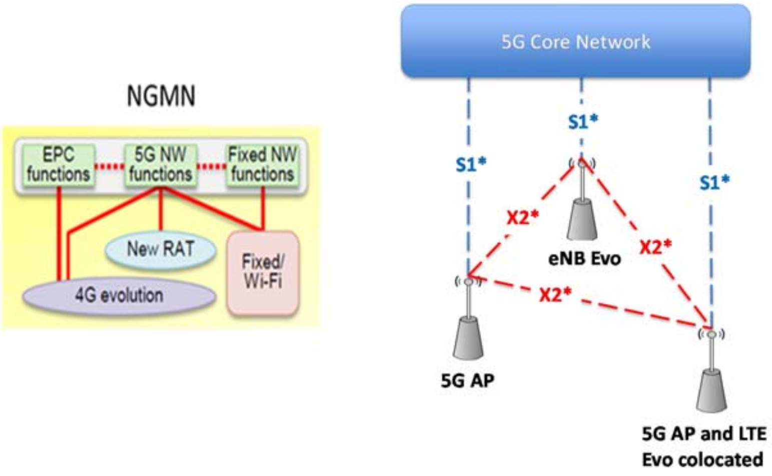


Figure 3-3: Network elements and interfaces based on the logical CN/RAN split, as in [3 -1].

The architecture in Figure 3-4 shows both mobile network functionality and management and orchestration functionality. This builds on ETSI-NFV principles and entities which it extends with the introduction of the E2E Service Management & Orchestration module as well as a programmable controller in order to flexibly configure and control Virtualized and Physical Network Functions (VNFs and PNFs) ([3 -21][3 -15]).

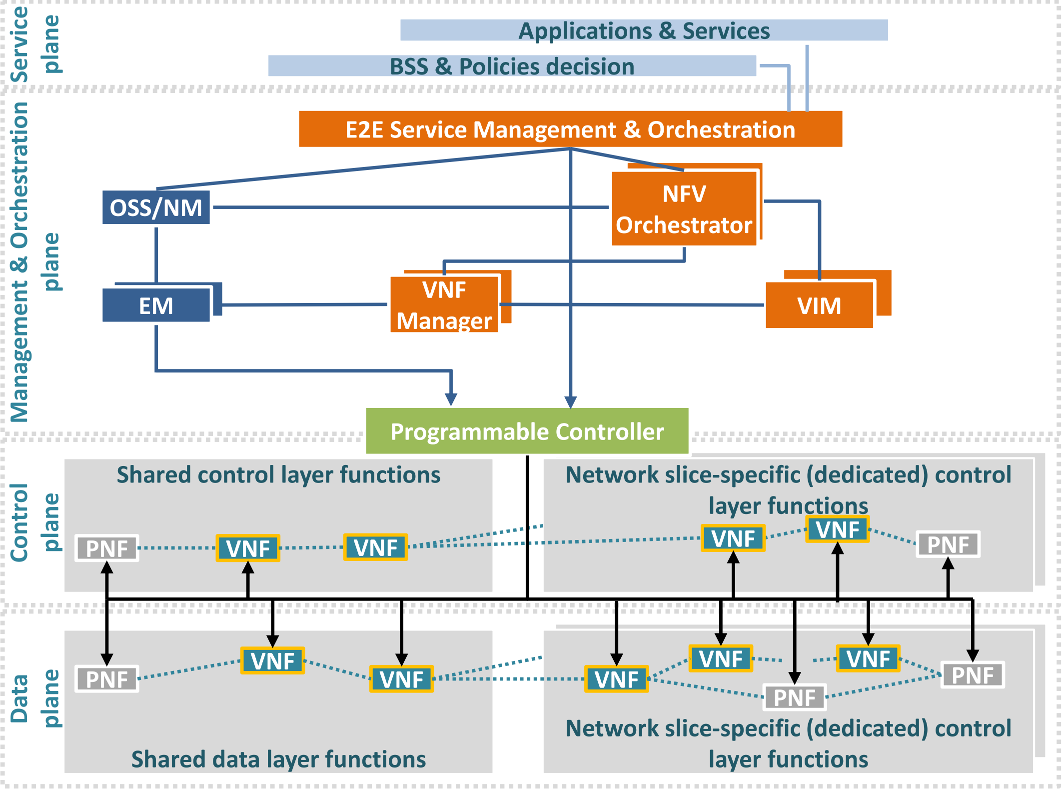


Figure 3-4: Framework for control, management and orchestration of network functions.

The **separation of control and user plane**, as introduced through **software-defined networking** (SDN), will also impact the 5G mobile network, which may similarly split functionality and provide corresponding interfaces.

Beside the radio access and core network, the **transport network will play a key role in 5G** to flexibly and dynamically address the requirements of future mobile networks. In order to support the required flexibility, a **unique packet-based network** is required. Three main types of interface are envisioned: packetized CPRI, next generation fronthaul interface (a new functional split within RAN), and backhaul. In order to address these interfaces, traffic class concepts will be introduced. Furthermore, to efficiently support network slicing by the transport network too, the concepts and systems of SDN and network functions virtualization (NFV) will be supported by the transport network, e.g. by separating the control and data planes through common packet-based data path abstraction. This **unified data and control plane** interconnects distributed 5G radio access and core network functions, hosted on in-network cloud infrastructure. The 5G transport network will consist of **integrated optical and wireless network infrastructure**.

3.3 Impact of Physical Networking and Computing Facilities

In order to meet the enormous growth of mobile data traffic, traditional wireless macro-cell networks need to be transformed into architectures comprising large numbers of small cells complemented with macro cells for ubiquitous coverage. Traditional Radio Access Networks, where Base Band Units (BBUs) and radio units are co-located, suffer several limitations. Aiming to address these limitations and take advantage of pooling and coordination gains, **Cloud Radio Access Networks (C-RANs), with the option of flexible processing splits**, have been proposed. The remote processing requirements for operational network purposes (C-RAN) together with the need to support a wide variety of compute and storage end user services, introduce the need of high bandwidth transport connectivity, with stringent delay and synchronization requirements between the radio units and the remote compute and storage resources. In addition, elastic resource allocation in the transport network ([3 -6]) becomes critical to the realization of statistical multiplexing gains.

3.3.1 Integration of Network, Compute and Storage Capabilities

The overall 5G vision involves a converged heterogeneous network environment, integrating a wide variety of network technologies for radio access with wireless and wired transport solutions interconnecting a huge number of vastly different end-devices and users, including compute and storage resources. These resources are called to support a combination of end-user and operational services such as C-RAN and the associated split options, and can be hosted either by micro-Data Centers (DCs) - referred to as Mobile Edge Computing (MEC) - or at remote regional and central large-scale DCs. MEC provides IT and cloud-computing capabilities ([3 -11][3 -15]) within the Radio Access Network (RAN) in close proximity to mobile subscribers, while regional and central DCs support the same type of services but can be accessed through the RAN and the transport network. In this environment, a common infrastructure can be used to handle both end-user and operational network services, maximizing the associated sharing gains, improving efficiency in resource utilization, and providing measurable benefits in terms of cost, scalability, sustainability and management simplification. This infrastructure deployment is illustrated in the “Physical Infrastructure” layer in Figure 3-5. The introduction of flexible processing splits ranges between the “traditional distributed RAN” case, where “all processing is performed locally at the access point (AP)”, to the “fully-centralized C-RAN” case where “all processing is allocated to a central unit (CU)”. The optimal “split” can be decided dynamically, based on a number of factors such as the transport network characteristics, the network topology and scale, and the type and volume of services that need to be supported. The required flexibility can be provided through programmable digital hardware, able to support flexible reconfiguration of hardware-accelerated (HWA) and software baseband functions, which can be partitioned at different levels in order to serve most different KPIs. Note that there are preferred options depending on the targeted KPI. To address the challenge of managing and operating such complex heterogeneous infrastructure efficiently, a high degree of flexibility, agility and adaptivity is required in the functions that a network can perform. Therefore concepts such as network softwarization provide a promising way forward. In view of this, the 5G vision involves the adoption and integration of specific technical approaches supporting this paradigm, such as SDN and NFV. In SDN, the control plane is decoupled from the data plane and is managed by a logically centralized controller that has a holistic view of the network. Besides placing HWA functions physically in the network, NFV enables the execution of software-based network functions on commodity hardware (general-purpose servers) by leveraging software virtualization techniques. Through joint SDN and NFV developments, supporting a set of management and control plane functionalities such as those illustrated in Figure 3-5, significant benefits can be achieved. These benefits are associated with flexible, dynamic and efficient use of the infrastructure resources; simplification of the infrastructure and its management; increased scalability and sustainability; and provisioning of orchestrated end-to-end services. Using these technology solutions, operational and business models such as multi-tenancy can be supported through network slicing and virtualization.

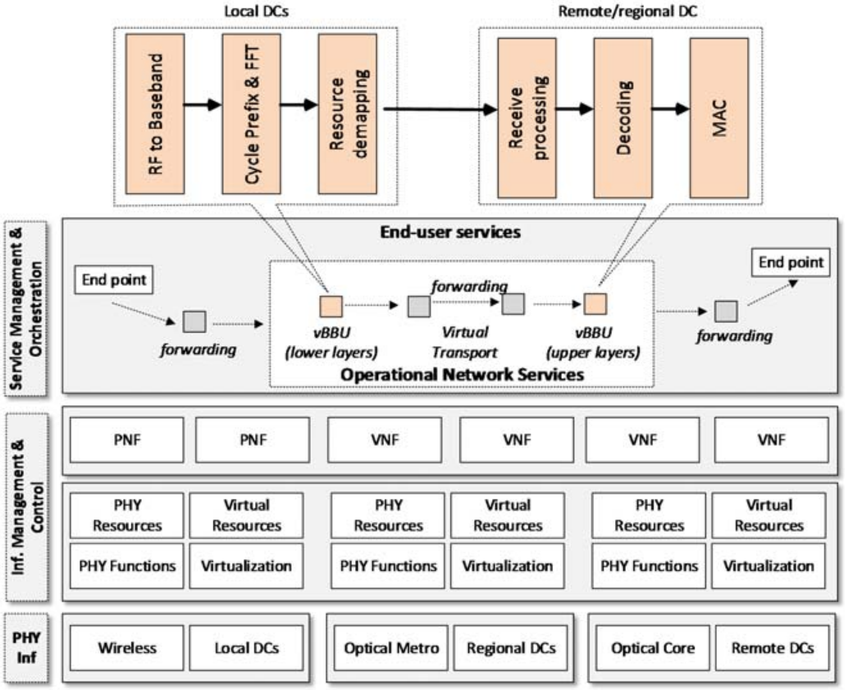


Figure 3-5: 5G infrastructure supporting integrated networking and computing facilities

3.3.2 Networking and Processing Needs

In the context of the integrated 5G infrastructure described above, flexible access to, and unified management of, both network and compute resources play a key role. The 5G infrastructure will have a set of service requirements that need to be satisfied. In response to this, different types of VNFs can be deployed and chained together, each having specific processing and bandwidth requirements as shown in Figure 3-5. To support the required services, sufficient network bandwidth and compute/storage capacity must be allocated; while relevant infrastructure slices can be created, corresponding to specific physical resources, for the interconnection and deployment of VNFs. The physical resources that should be reserved to ensure the delivery of the required services depend on characteristics and requirements of the users and services, as well as the availability of network and compute/storage resources and the relevant cost.

3.3.3 Deployment Options

In the environment described above, flexibility, elasticity and dynamic configurability across data, control and management planes allow optimal access to and utilization of the available infrastructure resources. It is clear that the optimal infrastructure design and configuration will differ depending on the service requirements and characteristics, including both end-user and operational services that need to be seamlessly supported. Identifying the optimal infrastructure design and operational configuration involves satisfying a set of predefined KPIs that span energy consumption/efficiency levels, end-to-end latency in service delivery, minimized capital and operational expenditures, and overall mobility, scalability and sustainability requirements. To ensure that optimal performance is achieved under different and dynamically varying service volumes and requirements, a flexible architecture is proposed, integrating a wealth of next-generation radio access and wireless, as well as wired transport network technologies together with compute and storage resources. This will allow dynamic and flexible choice of best fit options to optimally meet very diverse requirements at different points in time. Hardware configurability, adaptivity and elasticity features will play a key role in addressing these requirements - supported by the deployment of novel technologies and concepts such as network softwarization that rely on decoupling of data, control and management planes and leveraging of technologies such as network slicing.

3.4 Impact on Service & Infrastructure Management and Orchestration

In the following, subsection 3.4.1 introduces the multi-service control and management for single domain while the multi-domain is introduced in subsection 3.4.2. Security aspects are treated in subsection 3.4.3.

3.4.1 Multi Service Control and Management

In Section 3.1, a uniform service framework is constructed by multiple logically separated planes for 5G services provisioning. This subsection covers the Multi-service Management plane, the Integrated Management and Operation plane and the Application and Business Services Plane as shown in Figure 3-2, and describes service and resource orchestration cross 5G network segments. One of the main design objectives is to increase the flexibility and programmability of 5G networks with a novel Service Development Kit, a novel, modular Service Platform and Service Orchestrator, and a novel Integrated Infrastructure Management. It will bridge the gap between telecom business needs and operational management systems. The expected key functionality and systems ([3 -9][3 -13][3 -14]) are represented by the Service Development Kit, the Management System ([3 -10]) and the Service Platform, including: a customizable Service Orchestrator, a Resource Orchestrator, a Service Information Base, and various Enablers as represented in Figure 3-6. The figure also shows the heterogeneity of the physical resources underlying the 5G infrastructures and related 5G network segments: radio networks, access networks, aggregation networks, core networks, software networks, data centre networks and mobile edge computing clouds. The **Multi-Service Management plane** is responsible for the creation, operation, and control of multiple dedicated communication network services running on top of a common infrastructure. Functionality for this plane includes: infrastructure abstraction; infrastructure capability discovery; catalogues and repositories; a large number of service and resource orchestration functions such as plugins; information management functionality; and enablers for automatic re-configuration of running services (i.e. part of the integrated management plane). It interworks with **Application and Business Services plane** that maintains 5G application-related functions, organized in Repositories, and DevOps tools necessary for the creation and deployment of services. Functionality for this plane includes DevOps functionality: Catalogues, Monitoring data analysis tools, testing tools, Packaging tools, Editors and primitives for Application & Service programmability. Figure 3-6 depicts the way in which 5G manages various underlying systems.

In conclusion, one of the main design objectives in 5G Networking and 5G Multi-Service Control & Management is efficient integration of service programmability, domain orchestration functionality and DevOps functionality. This will maximize the predictability, efficiency, security, and maintainability of operational processes.

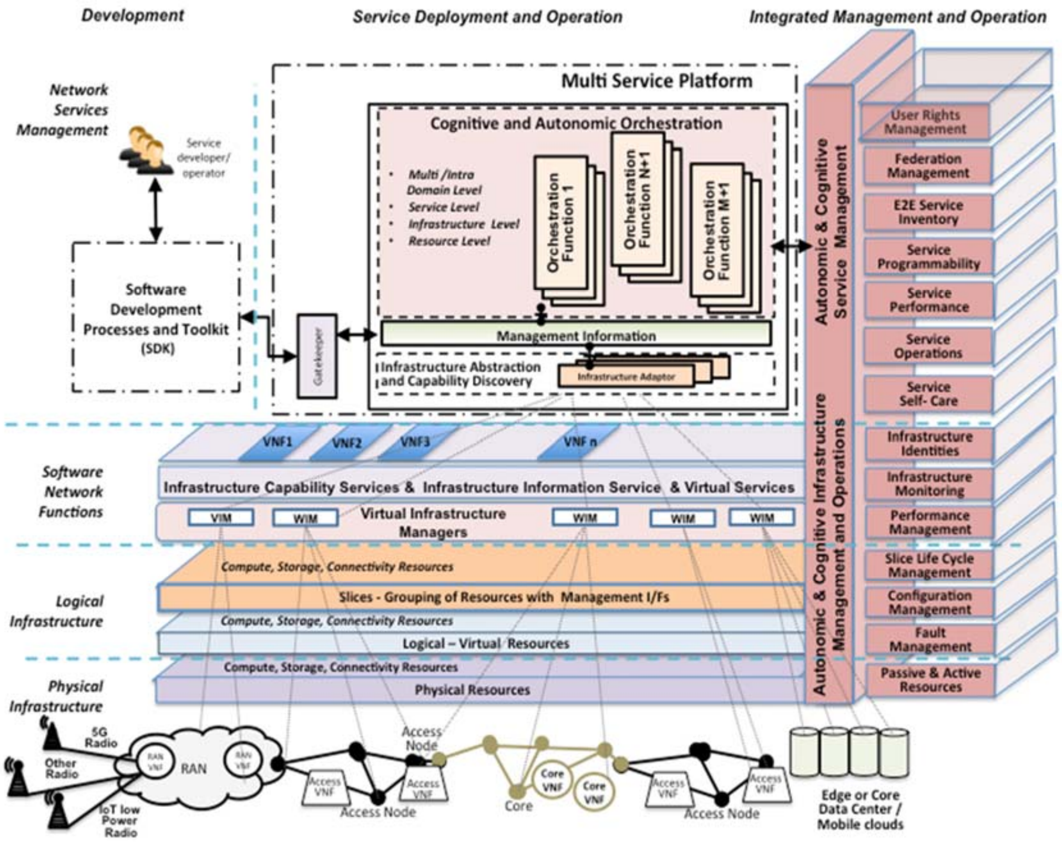


Figure 3-6: 5G Service & Infrastructure Management and Orchestration Architecture

3.4.2 Multi-Domain Architecture

Multi-domain orchestration refers to the automated management of services and resources in multi-technology environments (multiple domains involving different cloud and networking technologies) and multi-operator environments (multiple administrative domains) which includes operation across legal operational boundaries. The scope of the end-to-end multi-domain management and orchestration plane ([3 -12]) involves diverse concepts summarized in Figure 3-7. It represents the reference architectural framework for organizing the components and interworking interfaces involved in end-to-end management and orchestration in multi-domain environments. At the lower plane there are resource domains, exposing resource abstraction on interface I5. Domain orchestrators perform resource orchestration and/or service orchestration exploiting the abstractions exposed on I5 by resource domains.

A Multi-domain Orchestrator (MdO) coordinates resource and/or service orchestration at multi-domain level, where multi-domain may refer to multi-technology (orchestrating resources and/or services using multiple domain orchestrators) or multi-operator (orchestrating resources and/or services using domain orchestrators belonging to multiple administrative domains). The Resource MdO belonging to an infrastructure operator, for instance operator A, interacts with domain orchestrators, via interface I3 APIs, to orchestrate resources within the same administrative domains. The MdO interacts with other MdOs via interface I2-R APIs (business-to-business or “B2B”) to request and orchestrate resources across administrative domains. Resources are exposed at the service orchestration level on interface Sl-Or to Service MdOs. Interface I2-S (B2B) is used by Service MdOs to orchestrate services across administrative domains. Finally the Service MdOs expose, on interface I1, service specification APIs (Customer-to-Business or “C2B”) that allow business customers to specify their requirements for a service. The framework also considers MdO service providers, such as Operator D in

Figure 3-7, which do not own resource domains but operate a multi-domain orchestrator level to trade resources and services.

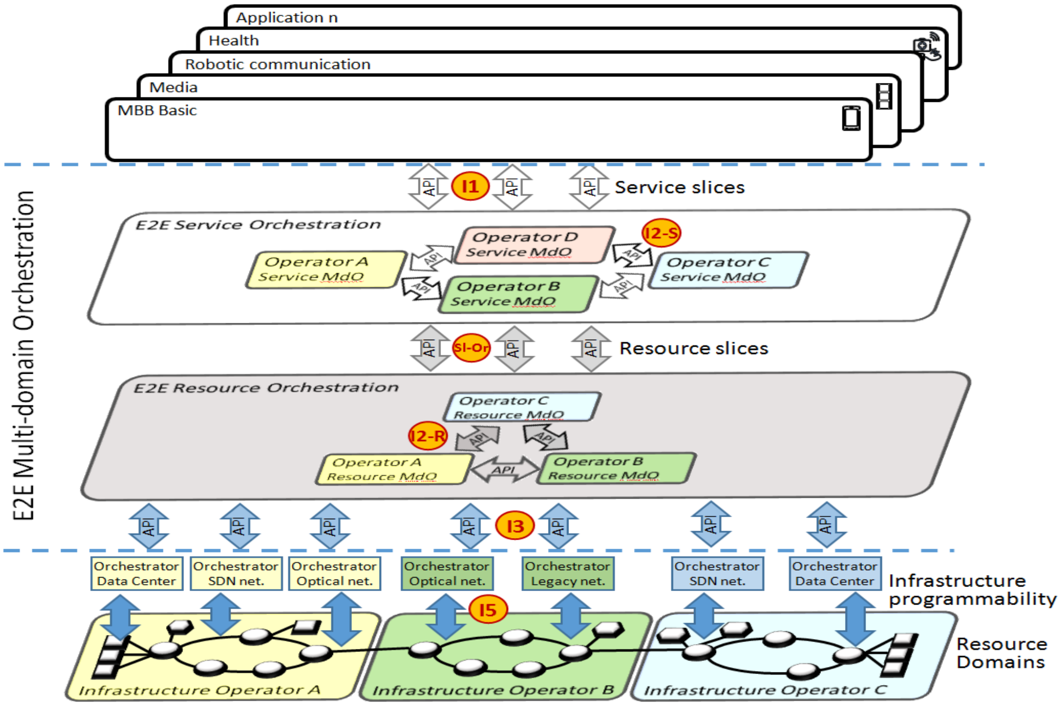


Figure 3-7 - E2E Multi-Domain Management and Orchestration of different infrastructure domains belonging to different operators

3.4.3 Network Security Considerations

The security design of current mobile systems was geared towards the build-up of a successful ecosystem, offering trustworthy communication services to users in all corners of the world. The evolution from 2G to 3G required the addition of new security features in order maintain trustworthiness in the presence of emerging threats, such as false radio base stations and encryption in communications, among others. While the security design of 4G is much more sophisticated, the 4G security architecture is more a consequence of the need to maintain security in a flatter network architecture, where user data is more exposed at the network edges. Mobile system security has so far arguably been more of an added support function than a driver.

When going from 4G to 5G, mobile users will be provided with higher bit-rates, and in consequence the surface of exposure will be significantly higher when it is combined with the estimated exponential growth of the number of devices using 5G networks. 5G security is now being analyzed ([3 -5]) and defined as a set of qualitative aspects such as:

- New business and trust models, including multi-domain and multi-service models where the re-use of the 5G technology outside the telecom sector (e.g. industrial control, public safety, utilities, etc.) needs to be considered; and where the presence of new types of actors, among a myriad of other factors, will even redefine what an “operator”, “user” and “device” are.
- New service delivery models based on virtualization, network slicing and other “aaS” technologies.
- Users need to be considered significantly differently, with respect to what we have learned from previous network generations; trustworthiness among users cannot today be taken for granted.

- Concerns about user privacy are higher than ever, e.g. following allegations of mass surveillance.
- The threat and risk landscape faced by a system supporting critical infrastructure services is very different from that of a system offering public voice/Internet services: the incentive for cyber-attack is much higher and the consequences of successful attacks are much more far-reaching.
- Lighter and faster encryption systems and algorithms should be considered in order to face the significant latency requirements imposed.

The existing 3GPPP security architectures are still very much tied to the traditional operator-subscriber trust model, and safeguard a single offered service: basic IP/telecom connectivity. They cannot simply be re-applied to new networks. There is therefore a need to design a security architecture for a new business and trust model, and one which is flexible enough to allow extensibility rather than “patching”.

The updated security considerations should not only entail new trust models, where mobile infrastructures are shared by multiple virtual mobile telecommunication providers, but also take into account novel technological approaches such as multi-tenancy, network slicing, network virtualization and other novel technologies related to 5G architectures, which could anticipate new types of threat specific to 5G networks. This is why it is critical to control the exposure surface to such attacks, and to provide proactive mechanisms to protect against them. Autonomic self-protection capabilities in the 5G network that might defend users against infrastructure attacks (such as a distributed denial-of-service attack), as well as providing self-healing capabilities to the 5G Network, are a key aspect of the network intelligence expected in the novel 5G technologies [3 -15].

3.5 Impact on Hosting and Deployment

As already stated, NFV enables the execution of software-based network functions on commodity hardware (general-purpose servers) by leveraging software virtualization techniques. In this manner, flexible, dynamic and efficient use of the infrastructure resources can be achieved, facilitating the provisioning of orchestrated end-to-end services. The implementation of virtual functions should not be agnostic to the hosting platform capabilities: features of the latter should be taken into consideration during the development process and before the actual deployment in an operational environment. Moreover, hosting functions on virtual machines can be considered as a process directly related to resource assignment and management ([3 -17]). Therefore, a number of management and configuration challenges arise in this case. For example, in OpenStack, service chain links between functions require significant manual configuration. When functions are deployed across both legacy infrastructures and SDN, the management scope is extended to cover legacy technology as well.

Furthermore, performance issues need to be examined. The available resources offered for hosting virtual functions should be carefully assigned. Sharing of resources could be complicated, and a factor leading to performance degradation. In addition, the hosting environment should support and satisfy requirements regarding data, resource allocation, dependencies, availability, and other attributes. As such, the underlying hardware characteristics (e.g. processor architecture, clock rate, memory bandwidth, speed) have a profound impact on the performance of the virtual functions. The hosting environment also affects the interconnection of the functions (e.g. functions on the same physical servers or LANs). In this case, connectivity issues are raised. Moreover, overall efficiency should be pursued. For example, poor reuse of software components (deployed and configured once for a specific server) should be avoided. The environment should be flexible, to allow the quick creation/release/execution of virtual functions, as the deployment and configuration of software is a complex and costly operation.

Cost and flexibility of deployment will definitely be important factors in 5G networks, requiring a shift toward software-based implementations and virtualization technologies ([3 -8][3 -9][3 -12]). For instance, third parties could deploy functionality dynamically, while the platform (e.g., a base station or a server connected to it) could run these concurrently, at the same time providing isolation via virtualization technologies. Furthermore, deploying virtual network functions through a virtual environment improves portability, while each function remains unaware of the underlying operating system. However, a major deployment issue is the efficient placement of the virtualized functions ([3 -18]). The objective is to find a suitable resource allocation for these functions within the hosting environments. Moreover, they can be deployed in different ways and each way has its own advantages and drawbacks. Functions executed on bare-metal are characterized by predictable performance, but resource isolation or security can be at risk. In addition, in this case, the software can be considered as OS-dependent. On the other hand, to achieve easy deployment, the software itself should decrease the dependency on OS kernel code, middleware, and hardware. Hybrid deployments can also be exploited, where part of the service is provided by physical hardware.

Finally, containers have been widely used for providing isolated hosting environments (user-space instances). They are extremely efficient (fast create/start/stop/delete operations) and exhibit high performance (small delay, high throughput, low memory footprint). In addition, they can be reusable and shareable. However, despite their advantages, a few critical issues can be identified, such as i) allocation of (CPU) resources, ii) communication, and iii) live migrations.

3.6 References

[3 -1] METIS II Project - https://metis-ii.5g-ppp.eu
[3 -2] COHERENT Project – www.ict-coherent.eu
[3 -3] mmMAGIC project - https://5g-mmmagic.eu
[3 -4] 5G-Xhaul Project - https://5g-ppp.eu/xhaul/
[3 -5] 5G-ENSURE Project - https://5g-ppp.eu/5g-ensure/
[3 -6] Flex 5Gware Project - www.flex5gware.eu/
[3 -7] 5G-Crosshaul Project - https://5g-ppp.eu/xhaul/
[3 -8] 5G-NORMA Project - https://5g-ppp.eu/5g-norma/
[3 -9] 5G SONATA Project - www.sonata-nfv.eu/
[3 -10] 5G COGNET Project -https://5g-ppp.eu/cognet/
[3 -11] 5G SUPERFLUIDITY Project - https://5g-ppp.eu/superfluidity/
[3 -12] 5GEx Project- https://5g-ppp.eu/5gex/
[3 -13] SONATA Deliverable D2.2 “Architecture Design” – December 2015
[3 -14] SONATA Deliverable D2.1 “Use Cases” – December 2015
[3 -15] SELFNET Project - https://5g-ppp.eu/selfnet/
[3 -16] 5G CARISMA project - http://www.charisma5g.eu
[3 -17] Yousaf, Faqir Zarrar, and Tarik Taleb. "Fine-grained resource-aware virtual network func-tion management for 5G carrier cloud." IEEE Network 30.2 (2016): 110-115.
[3 -18] Moens, Hendrik, and Filip De Turck. "VNF-P: A model for efficient placement of virtualized network functions." Network and Service Management (CNSM), 2014 10th International Conference on. IEEE, 2014.
[3 -19] I. D. Silva, G. G. Mildh, A. Kaloxylos, et el, “Impact of Network Slicing on 5G Radio Access Networks”, EuCNC 2016, Athens
[3 -20] EU PROJECT METIS-II, 5G RAN Architecture and Functional Design White Paper, https://metis-ii.5g-ppp.eu/wp-content/uploads/5G-PPP-METIS-II-5G-RAN-Architecture-White-Paper.pdf
[3 -21] EU PROJECT 5G NORMA, Functional Network Architecture and Security Requirements, 31/12/2015, https://5gnorma.5g-ppp.eu/wp-content/uploads/2016/01/5G_NORMA_D3.1.pdf

[3 -22] P. Rost, A. Banchs, I. Berberana, M. Breitbach, M. Doll, H. Droste, C. Mannweiler, M.A. Puente, K. Samdanis and B. Sayadi, “Mobile Network Architecture Evolution towards 5G,” in IEEE Communications Magazine, vol. 54, no. 6, June 2016.

[3 -23] NGMN Alliance, “5G White Paper”, March 2015 https://www.ngmn.org/5g-white-paper.html

4 Logical and Functional architecture

This chapter lists the current trends identified by the 5G PPP projects exploring the design of the logical and functional architecture for 5G networks. It starts by explaining the key differences between the notion of “network functions” in 5G and in legacy systems in section 4.1, followed by some examples of network functions that may be tailored to different services in section 4.2. Building upon this common understanding on network functions in 5G, section 4.3 elaborates on key design paradigms related to the logical architecture, while section 5.4 discusses issues related to potential logical entities and interfaces. Considerations on the protocol stack architecture and multi-connectivity are presented in section 4.5. Section 4.6 concludes this section by referring to logical functions related to orchestration and network management.

4.1 General Considerations on (Virtual) Network Functions in 5G

There is wide consensus that network functions will have a very different nature in 5G than in previous cellular communications generations, and should follow different design paradigms

The notion of “network functions” in 5G will not only relate to connectivity, but also to computation and storage in all 5G network segments ([4-1], [4-2]). More precisely, network functions will provide typical connectivity-related services, such as filtering and forwarding, packet inspection, stream handling for signal processing purposes etc. Moreover, in 5G networks they will also provide complex functions, like web servers or data base functionality inside or at the edge of the network, spanning both stateless and stateful functions. One specific class of network functions in 5G will be the “**Virtual Network Functions (VNFs)”**. They are represented by one or more virtual machines running different software and processes on top of industry-standard high-volume compute platforms, switches and storage, or cloud computing infrastructure. These are capable of implementing network functions traditionally implemented via custom hardware appliances and middleboxes (e.g. router, NAT, firewall, load balancer, etc.). The VNFs will play an important role especially in the design of CN functions.

Network functions in 5G will be designed to cater for very diverse service requirements, and will be mapped to physical architecture in a service-specific way ([4-3]). The support of very diverse services may for instance be enabled by having certain network functions dedicated to specific services, and/or by designing network functions that are parametrizable to suit different services. In this context, it may be beneficial to define sets of basic/elementary “Reusable Function Blocks” (RFBs) as the building blocks used to compose high-level functions. An RFB is here seen as the generalization of the concept of VNF. Some RFBs could be designed to support a wide range of services (through appropriate parameterization), while others are dedicated to particular services.

In general, it is understood that network functions will also be mapped to the physical architecture depending on the use case, service-specific requirements, and the physical properties of the existing deployments. Further, these will be individually instantiated for each logical network running on the same infrastructure. A coexistence of different use cases and services will hence imply the necessity of using different VNF allocations within the same network.

Network functions in 5G will be more strongly decoupled from physical architecture than in legacy systems ([4-4][4-15]). Traditionally, mobile network functions are implicitly grouped into network entities via specification of their interconnections, where each entity is responsible for a pre-defined set of functions. Accordingly, the degrees of freedom for assigning network functionality to physical network entities are very limited. For instance, 3GPP Evolved Packet Core (EPC) elements like gateways may be collocated with base stations in 3GPP Evolved Packet System (EPS), but moving only parts of the functionality of a gateway or Mobility Management Entity (MME) within a physical base station would require additional modification of 3GPP interfaces. Furthermore, traditional RANs, where Base Band Units (BBUs) and radio units are co-located, suffer several limitations including: i) increased CAPEX and OPEX due to often underutilized dedicated resources; ii) limited scalability and flexibility; iii) lack of modularity and limited density; iv) increased management costs; and v) inefficient energy management due to lack of resource sharing. To address these limitations, Cloud Radio Access Networks (C-RANs) have recently been proposed. In C-RAN, distributed access points, referred to as remote radio heads (RRHs), are connected to a BBU pool, the Central Unit (CU), through high bandwidth transport links known as fronthaul (FH). However, as currently such deployments use non-virtualized baseband processing at the central location, this is rather about relocating functionality, which does not exploit all the characteristics of cloud computing and is unable to realize e.g. pooling gains. Despite these gains it is worth to remember that simplicity was a keyword in the 4G design when a flat architecture has been proposed, considering the flexibility the centralized architecture in 3G had. Therefore, a balance between flexibility and complexity needs to be taken into account.

In 5G systems, network functions are to be designed to allow a maximally flexible instantiation or even dynamic (re)allocation of functions (i.e. logical entities) to physical entities, enabled by the following guidelines:

- avoidance of strict timing relations between network functions and protocol stack layers, and design of network functions which in current systems operate synchronously with the radio (e.g. RLC functionality) to operate asynchronously to the radio or with otherwise relaxed timing constraints.
- design of network functions such that these are either able to adapt to the physical architecture in which they are used (i.e. maximally exploiting centralization and pooling gains when possible, while showing graceful performance degradation when mapped to a decentralized physical architecture with non-ideal structural properties and physical interfaces involved), or are replaceable by alternate network functions specifically optimized for these non-ideal environments.
- support of the on-demand composition of network functions and network capabilities.
- support of the design, implementation, deployment, management and maintenance of network functions by software programming, exploiting characteristics of software such as flexibility and rapidity of design, development and deployment throughout the lifecycle of network functions etc.

Ultimately, this can be seen as replacing a "**network of entities**", as in legacy systems, by a "**network of (virtual) functions**."

Clearly, some network functions have such strong timing relations with the radio, or depend so strongly on e.g. hardware acceleration, that it is challenging to virtualize them. In fact, despite the extensive effort carried out by companies and research groups specialized in software acceleration on commodity computing platforms, the gap between HW-based and SW-based implementation is for some functions still significant and possibly not going to decrease in the future ([4-5]). For this reason, **there is the general accepted notion of physical and virtual network functions**.

Traditionally, the decoupling of logical functionality from its physical realization has always required dedicated security mechanisms. For example, access control mechanisms and encryption are required to allow sensitive data to be stored or communicated on physically exposed/shared media such as radio links or shared disks. The fact that 5G networks and functions will, to an even higher degree, be provided as logical/virtualized concepts, emphasizes both the required scope and the criticality of security. This criticality will further increase due to the needed support for mission-critical services and the need for slice isolation. As such, most of the necessary security functions in current networks are flexible in the sense that they can "move along" with the motion of the functionality that they protect. But this does not mean that security is agnostic of the physical realization of the logical architecture. On the contrary, the 4G standardization decision to place the PDCP/user plane in the eNodeB led to extensive work in defining additional and rather sophisticated security measures to make this physical realization acceptable. Thus, while it is desirable to define a security architecture that is flexible and extensible, allowing e.g. functional end-points to be re-allocated due to mobility or traffic optimization, a complete independence of the physical architecture cannot be obtained. In fact, security itself can never be fully virtualized, as software can never protect itself completely. Certain aspects of the logical security architecture will be dependent on some form of hardware root-of-trust, e.g. for key-management, software verification, secure boot, etc. ([4-6]).

4.2 Service-Tailored Radio Access and Core Network Functions

In this section we elaborate in more detail on specific examples of network functions that could be tailored to specific service needs in 5G. In general, there is the common understanding that specific services will likely reuse the same functionalities as other services for a large portion of the protocol stack, differing only for a smaller number of functionalities. For instance, it may be possible to use flexible air interface numerology. Also, depending on the detailed service needs it may be possible to further select among different coding strategies, MIMO modes and framing structures optimized for throughput, delay, or reliability. However, upper layer packetization function may still be the same for a wide range of services, which would allow for reusing the same software implementation ([4-7]).

As mentioned in previous sections, 5G will encompass multiple segments and layers, where the aforementioned notion of reusable network functions can be applied. For example, in the transport segment comprising backhaul and fronthaul networks, reusable functions can be in the form of a virtualization substrate offering each tenant or slice a particular transport network abstraction, slice specific QoS mechanisms, supported for example by different transport tunnels, or slice specific SLA monitoring mechanisms. However, the notion of service-tailored functions is expected to be most relevant for the Radio Access and Core Network segments. In this context, Table 1 lists network functions that could differ for different services or the environments in which they are used.

Table 1. Potential service-specific flavors of network functions.

1. Type of network function		2. Possible service-specific flavor
Core network	Value added services	Parental Control (e.g., user context for children and requested service dependent optional part of a service chain), DPI, Video optimizers, firewalls, service chaining in GiLAN.
	Authentication, Authorization, Accounting (AAA) and security	Service-specific access control and accounting/charging policy functionality and placement. Service specific security (e.g. a slice with no encryption and/or with added data integrity).
	Traffic control	QoE, QoS, mapping, monitoring, flow processing/policing and enforcement done in a service-specific way. Tighter mobile network – Transport interaction.

	Mobility management	Mobility management function design and selection may be service-specific, to allow for a higher degree of customization, e.g., network-slice-specific or radio-access-specific mobility management
General connectivity	Connectivity model	bearer-based (for high throughput services) or connection-less (for IoT).
	Multi-Connectivity	Multi-connectivity at different network layers (micro/macro), spectrum (sub-6 GHz/mm-wave), user plane (MAC/RLC/PDCP), technologies (WiFi/LTE) depending on service, deployment and RAT
Spectrum Access		Service-dependent operation in licensed, unlicensed, license-assisted spectrum, or time-frequency multiplexed in common spectrum
RRC related	Mobility	No (metering), local (enterprises), in groups (trains, buses), very high velocity (cars/trains/aircraft), on demand (tracking sensors) or classical mobility (pedestrian broadband)
	Cell discovery	sub-6 GHz MIMO (broadcast), massive MIMO mm-wave (sub-6 GHz assisted), small cells in ultra-dense networks (via macro coverage layer) cell discovery, or IoT (no cell discovery)
PDCP		Potential service-specific omitting of header compression and ciphering
RLC		Unacknowledged mode only (e.g. sensor) or acknowledged mode only (e.g. mission-critical services)
MAC / PHY	Carrier Aggregation	Carrier aggregation may not be needed in each scenario as it also impacts battery consumption; it could further include very distinct spectrum.
	Multi-Cell Cooperation	Service, load, deployment and channel-dependent tight cooperation (symbol-synchronized operation, RNTIs/scrambling/CSI-RS/scheduling/precoding coordination up to joint Tx/Rx CoMP) or loose cooperation (ICIC)
	H-ARQ	Optimized for spectral efficiency (massive broadband) coverage (sensor, IoT), reliability (mission critical services) or latency (tactile Internet)
	RACH	Prioritized RACH schemes to achieve service prioritization especially in mMTC scenarios
	Coding	Block codes for short (sensor) transmissions, turbo-codes for high throughput.
Software Defined Mobile Network Control Applications		Service (or network slice)-specific control applications use the northbound interface of a controller to allow for configuration, optimization, etc. of network functions (i.e., extending the SDN paradigm to any network function); applicable to network functions from both c- and u-plane, e.g., programmable session and mobility

	management
Intrusion Prevention and Intrusion Detection	Service-specific access control to allow or block access to deployed services and service-specific monitoring to detect suspicious traffic. Configuration of the Intrusion Prevention and Intrusion Detection VNFs will depend on the type of services that require protection and on the security policies set.

4.3 Key Logical Architecture Design Paradigms

Having discussed both general and specific considerations regarding network functions in 5G, we will now elaborate on key design paradigms related to the overall logical and functional architecture that have been identified by a wide range of 5G PPP projects.

There is a common understanding that in the context of 5G, the traditional logical network architecture, where network functions are grouped into logical entities, which are defined irrespective of service needs and are typically closely related to physical entities, will be replaced by a more flexible architecture. This will allow a service- or slice-specific grouping of network functions to logical entities, and the mapping of logical to physical architecture which is in full accordance with the envisioned **ETSI NFV architecture framework** ([4-8]). A key aspect in this context will be **infrastructure programmability**, i.e. the capability to tailor control functions and data plane functions according to the network reality and service needs and on a per-slice basis. Infrastructure programmability is seen as the enabler for end-to-end orchestration of resources and services.

There is also consensus that the 5G logical architecture should foresee a **split of control and user planes**, enabling individual scalability of both planes and logical centralized control ([4-7], [4-9]). This will also be a necessary approach to providing a unified control framework for 5G networks. In the RAN, and especially in the context of ultra-dense small cell networks, there are various considerations to let, e.g. macro-cells handle the control plane and small cells provide the user plane, allowing for a dynamic activation and deactivation of small cells, more efficient mobility management, increased mobility robustness, and increased control plane capacity. This approach appears particularly relevant in the context of mmWave small cells. The exact extent of control signaling handled by the macro cells (for instance, whether this is limited to higher protocol stack layers such as PDCP and RRC, or also spans lower layers), is still under investigation. In general, one has to consider that some radio control functions, e.g. scheduling, are tightly coupled with the user plane, and may thus, have to be physically co-located.

Building on these general design paradigms, we will now venture into specific considerations of logical entities, interface and protocol stack architecture in 5G.

4.4 Considerations regarding Logical Entities and Interfaces

Currently, 3GPP is assuming that a single standardized RAN component, like an eNB, would allow a high degree of deployment flexibility where very flexible implementations and deployments would be allowed. In other words, that would specify simply an inter-node interface to support mobility and multi-connectivity. However, in order to enable a high degree of implementations and deployments 5G-PPP is currently investigating the following innovations regarding logical entities and interfaces characterizing the 5G mobile network:

- **Within the RAN**, the currently deployed D-RAN or C-RAN implementations are not considered optimal to address the radio technologies considered for 5G. Hence, new functional splits between Remote Units (RUs) and Centralized Unit (CUs) are currently being investigated by several 5G PPP projects.

- **Regarding the E2E logical architecture**s, several alternatives are currently being discussed comprising a logical split between RAN and CN functions, following traditional approaches (e.g., like the today's LTE CN-AN split) or following new schemes where mobile network elements are decomposed into functional blocks that can be instantiated at different aggregation levels. Both approaches offer different advantages and their applicability is still an open issue for evaluation.

Irrespective of service- and slice-specific tailoring and chaining of network functions, and their flexible mapping to physical architecture, interoperability between vendors will require that certain logical interfaces between network functions are standardized, or that at least de facto standards are established. The challenge of avoiding many additional interfaces due to an increased number of mapping options is being investigated and may be addressed by a flexible container protocol on user and control plane.

Within the RAN, different considerations on possible function splits and logical interfaces are being pursued:

- **State of the art: function split within PHY layer (e.g. based on CPRI or OBSAI interface)**. A classical solution in the context of C-RAN is to draw an interface within the PHY, such that A/D conversion and down/up-conversion are performed in remote radio units (RRU), while all other processing functions are centralized. The corresponding physical fronthaul interfaces are standardized through the Common Public Radio Interface (CPRI), the Open Base Architecture Initiative (OBSAI) and the Open Radio Interface (ORI), with CPRI currently being the most frequently used standard. The main disadvantage of this solution is that the bandwidth requirements on the fronthaul interface scale with the number of antennas (or rather baseband signal processing chains in the context of analog or hybrid beamforming) and the system bandwidth, rendering the approach challenging in 5G, especially in the context of massive MIMO and mmWave communications.
- **Investigated alternative 1**: Function split between PHY and MAC (see Figure **4-1**). If fiber fronthaul is available and a high degree of centralization is desired, yet the aforementioned scaling issues are to be avoided, it is possible to have an interface where MAC functionality, HARQ and FEC are centralized, but modulation, precoding and other PHY functions are performed at the remote radio units. This will allow an entire cloud RAN to be considered as a giant base station having many distributed antennas, but without the high bandwidth requirements on the fronthaul interface that the legacy solutions pose. This alternative is under investigation in [4-4], where further ideas include a flat architecture using only eNBs and advanced gateways (aGW).
- **Investigated alternative 2: Function split between synchronous and asynchronous functions (e.g. RLC and PDCP) ([4-7], [4-9], [4-10])**. If fiber fronthaul is not available (see Figure **4-1**), hence low latency connectivity between centralized processing and RRU cannot be guaranteed, an option is to perform the function split between the functions that are closely tied to the radio (e.g. PHY, MAC and RLC related functions, which are typically synchronous with the radio), and functions that are less tightly tied to the radio (e.g. PDCP and RRC functions, which are in LTE asynchronous with the radio). The latter group of functions could then be virtualized and centralized over many cells, and run e.g. on mobile edge clouds; while the former group would be distributed and run on the actual physical access nodes. An example could be a dense deployment of mmWave small cells, where fast traffic re-routing between the cells would be desired, and hence the RRC and PDCP functionality could be performed by centralized logical user plane and control plane entities. As mentioned earlier, in section 4.1, a key aim in 5G is in general to design network functions such that strict timing constraints between functions related to different protocol stack layers are alleviated or avoided, in order to allow for multiple possible function splits and hence multiple options for logical interfaces between centralized RAN clouds and access nodes.

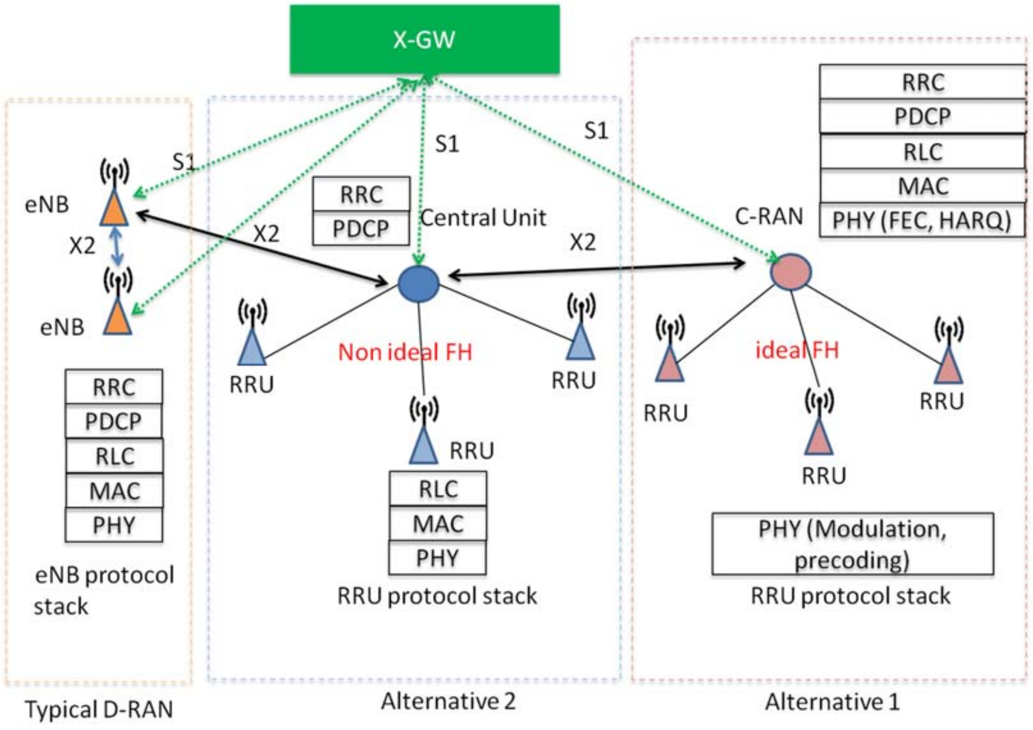


Figure 4-1: Possible function splits in the RAN.

A key factor in determining the degree of centralization of RAN related functions is the type of backhaul or fronthaul available. An initial analysis of the transport requirements imposed by future 5G sub6 and mmWave RAN technologies can be found in [4-11].

Regarding the **logical interface between core network and radio access network**, a typical baseline assumption applied in multiple 5G PPP projects is to consider an evolution of existing interfaces, such as the S1 interface in the EPS ([4-12]), as already elaborated in subsections 3.2.

Projects (e.g., [4-4]) are investigating the benefits of different implementations compared to what is currently deployed in the field, for 4G networks. In one example, X2 and S1 traffic are routed through the fixed network, which is often shared with other operators and services so that security is an issue. If backhaul encryption is used, it starts above the eNB and ends behind the advanced gateway (referred in Figure 4-1 with a generic term xGW). Any traffic (including X2) is routed via the evolved packet core, behind the advanced gateway, which can be hundreds of kilometers away. In 5G deployments, X2 could be routed through the nearest common aggregation node in the fixed network, while bypassing the access gateway. X2 traffic could further be encrypted separately from S1 at the distributed eNB ([4-4]).

Another approach ([4-3][4-16]) under investigation is the following. In conjunction with the decomposition of mobile network elements (both c-/u-plane vertical split and horizontal decomposition), the evolution towards ‘cloudified’ networks is envisioned that it will dramatically change the way mobile network functionality is deployed and geographically distributed, also over time. In the context of this work, roughly, three different locations of cloud resources are distinguished: (1) edge cloud co-located at the antenna site, (2) edge cloud at an aggregation site, and (3) the central cloud. Virtualized network functions from both RAN and CN running on generic network functions virtualization infrastructure resources can be instantiated at any of these locations according to the requirements of the telecommunication network service. In theory, this allows for a considerably higher number of deployment options for any given subset of network functions, therefore providing degrees of freedom to achieve improved service customization, load balancing, and multiplexing gains.

This approach strives for smaller functional blocks consisting of groups of atomic functions. Accordingly, a generic interface connecting functional blocks needs to support different combinations/chains of functions and/or functional blocks. For example, such an interface may define a basic (mandatory) set of information elements (IEs) and primitives and additional sets of information elements and primitives depending on the reference point, i.e. the two atomic functions that the specific interface instance inter-connects. On an additional level, an interface could even be negotiated between two functions. Standardized information element sets and primitives are needed where function blocks of different vendors are chained and need to interoperate, while chaining of a single vendor's functions may also be proprietary. Nevertheless, the standardized basic set of IEs, primitives, and the base protocol that is run between functional blocks to convey the information, could be reused following the above mentioned framework. Although this approach promises several advantages, its implications in terms e.g., of standardization and added complexity, as well as the quantification of the expected benefits are still under investigation.

4.5 Considerations on RAN Protocol Stack Architecture and Multi-Connectivity Aspects

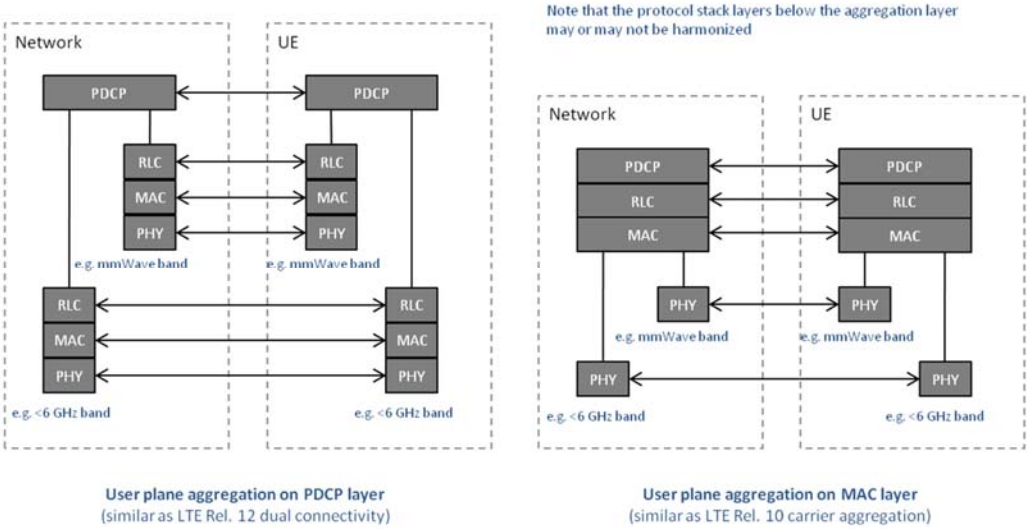
Most research efforts on 5G protocol stack architecture take orientation in the LTE architecture, but consider changes, as related for instance to the role of different protocol stack layers. For example, there are considerations to perform dynamic traffic steering, i.e. the assignment of services to suitable radio access technologies, on MAC level ([4-13]), and hence on a much faster timescale than in LTE, where traffic steering is typically conducted in the form of handover between technologies.

There is a common understanding that, despite the need to have network functions tailored towards different services and their related requirements, or to different frequency bands and cell types, there should be a large extent of harmonization, or commonality, between the relevant protocol stack layers. Although, radio access technologies related to communication below and above 6 GHz, may likely utilize different physical layer numerology and different signal processing approaches, higher protocol stack layers and related network functions should ideally be very similar, for the sake of a lean standards specification, reduced infrastructure and device complexity. This similarity could be obtained by having a large set of common network functions (or "reusable function blocks", as introduced in Section 4.1) that can be parameterized to support different services, bands and cell types. Clearly, one has to be careful not to sacrifice the performance of the 5G system for individual services, bands or cell types; hence, a key research question still to be tackled is to find the right trade-off between harmonization and specialization of network functions.

In this context, it is rather straightforward that for novel radio technology introduced in 5G, one could likely strive for and obtain a large extent of similarity already by design, while between evolved legacy technology (such as LTE-A) and novel 5G radio technology, a large extent of similarity of network functions (in particular on lower-protocol stack layers) may be difficult to achieve - and possibly not even desirable, as this may pose too strong backward-compatibility constraints on novel 5G radio technology.

A common protocol stack layer specification is also a pre-requisite for supporting user plane aggregation or control plane integration, in the form of multi-connectivity approaches across multiple cells or multiple frequency bands. In this respect, the following options are being considered:

The **user plane aggregation among multiple novel 5G radio technologies** could take place on MAC, RLC or PDCP level (or their 5G equivalents), given that the radio technologies are harmonized on and above the protocol layer of aggregation. MAC layer aggregation has the potential to enable tighter integration features like cross-carrier scheduling, but may be challenging in the context of e.g. PHY layers with very different frame structure. Also, user plane aggregation on MAC or RLC layer would typically only be possible in co-located deployments and/or deployments with good backhaul quality. PDCP-level aggregation can enable several features similar to MAC-level aggregation (not necessarily with the same gains) except cross-carrier scheduling; with the benefits of being likely more suitable for distributed deployments with non-ideal backhaul and not requiring the harmonization of the lower layers of the radio technologies. Figure **4-2** shows UP aggregation on PDCP and MAC layers as two potential aggregation options for novel radio technologies. Note that the figure applies both to different 5G radio technologies or multiple instances of the same (e.g. multi-cell connectivity). **User plane aggregation among evolved LTE-A and novel 5G radio** is currently foreseen to be most viable on PDCP layer, though also aggregation on MAC layer is being investigated [4-17].



Note that the protocol stack layers below the aggregation layer may or may not be harmonized

User plane aggregation on PDCP layer
(similar as LTE Rel. 12 dual connectivity)

User plane aggregation on MAC layer
(similar as LTE Rel. 10 carrier aggregation)

Figure 4-2 Options for user plane aggregation among novel 5G radio technologies.

Regarding PDCP-level user plane aggregation, one of the related enhancements that are proposed in the context of mmWave communications is to combine the ([4-14]) dual connectivity options 1a (bearer level split at Serving gateway, S-GW) and 3c (packet level split as PDCP protocol data units in the Master eNode B, MeNB) to allow both options simultaneously. In addition to this, a variation of option 3c may be explored where the traffic is split in the Secondary eNode B (SeNB) instead of the MeNB to increase flexibility to the network, as illustrated in Figure 4-3([4-9]).

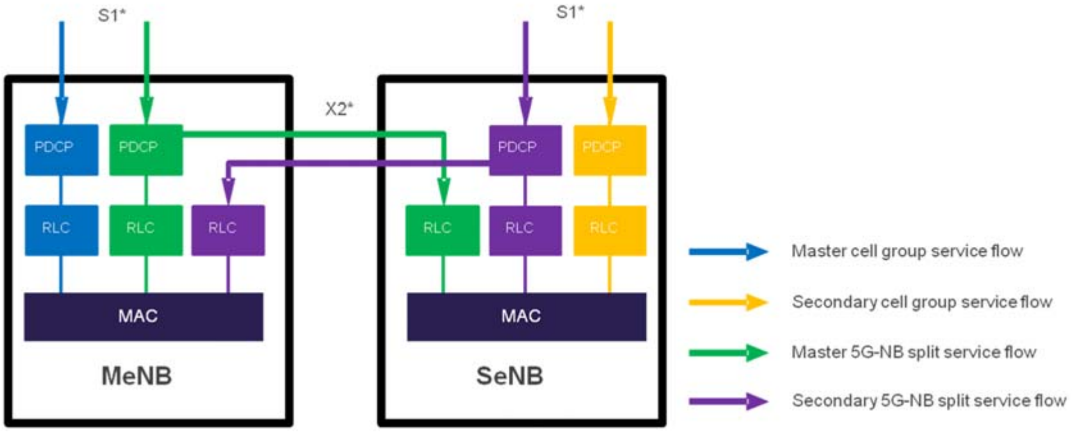


Figure 4-3 Considered radio flow split for enhanced dual connectivity.

Regarding **control plane integration among novel 5G radio instances**, it is currently foreseen that there will be one single RRC protocol instance supporting the user plane aggregation previously described. This would be equivalent to the LTE Release 12 solution, where the single RRC entity at the network side resides at the MeNB, while the SeNB communicates with the MeNB over the X2 interface to support the configurations of lower-layer parameters. Finally, for **control plane integration among evolved LTE-A and novel 5G radio**, the following options are considered:

- A single RRC instance exists at the UE and at the RAN either for the new 5G AI or for the evolved LTE-A;
- Two RRC instances exist at the UE and the RAN, one for the new 5G AI and another for the evolved LTE-A.

In general, various options for handling RRC signaling are investigated in different projects, including RRC diversity or fast control plane switching between cells.

4.6 Orchestration and Network Management Functions

For 5G networks there will be increasingly diverse network resources. Resources will be a heterogeneous mix of technologies and capabilities, possibly located in multiple administrative domains. The programmability of the infrastructure is the enabler for end-to-end orchestration of resources and services. The "end-to-end" requirement implies the capability of realizing multi-domain orchestration of heterogeneous programmable infrastructure domains, possibly belonging to different administrations/operators ([4-1]).

A language for describing the interrelation of network functions, and ultimately for allowing the orchestrator to communicate with the data and control planes in the network, is key for portability and interoperability. While the idea is simple, its actual specification may be anything but trivial ([4-5]).

The role of the network management functions/orchestrator is to build complex network functions and services from less complex/primitive network functions. In this process, the orchestrator has to consider service specific requirements, e.g. latency, physical locations of specialized hardware, etc. This is done through the entire lifecycle of a function/service, i.e. deployment, operation, monitoring and termination. In addition, it analyzes the network situations in real time, diagnoses and predicts existing or emerging network issues, and determines and coordinates reactive or proactive actions to resolve issues.

In general, computational and storage requirements can be expressed as a function of current load (e.g., number of connections, aggregate data rate, etc.), making the approach easily applicable to different kinds of contexts. On the other hand, it is futile to think that all such diverse networks can be treated identically, since there is no one-size-fits-all solution for all such types of networks. Rather, expert domain knowledge, for example from a radio access expert or from a core network expert, will have to be brought into achieve high-quality solutions. This expert knowledge is not only needed in the actual network functions themselves (example: a RAN signal processing function behaves very differently from a core network optical routing function); it is also needed in the way services are orchestrated.

Orchestration activities are processes that will be carried out in future networks in order to achieve increased flexibility and better utilization. For different research projects and at different levels below follows an indicative list of activities that are of a particular interest ([4-2], [4-3]):

- Service orchestration - The prime goal is to adapt a service, typically composed of various types of functions, to the current system and load situations. This task encompasses indicatively: scaling up or out individual functions; restructuring a service graph; modifying the placement of functions inside an actual network; reroute traffic to or between different instances of such functions, or to adjudicate resources between competing services; auxiliary functionality, including lifecycle management of individual functions. Also, umbrella functions for e2e service/network slice management and cross-tenant orchestration (Inter-slice Orchestration, Umbrella and Service Management) are needed. For this, enhanced MANO functions will be needed.

- Service- and network function-specific orchestration algorithms - Since personalized treatment of services and even functions is assumed for 5G systems, it is necessary for an orchestration function to offer the appropriate means to achieve this (e.g., let a service provide its own placement algorithm to be executed in a secure fashion by the orchestration platform).
- Resource orchestration - Orchestration of resources (i.e. connectivity, compute, and storage resources) across multiple network functions in the same domain or across multi-domains. This controls automated lifecycles of logical resources, including the lifecycle of slices in one or multiple resource domains hosting a network function or a topology of network functions.
- Implementation of Network & Service management - The ability to reroute traffic between services and functions is an important part of new 5G management functionality. The same is also true for are autonomic capabilities like monitoring, optimization, configuration, fault resolution, and SLA operations. An orchestration platform hence has to interwork with the new 5G Management and Operation functions, and also provide convenient interfaces and interoperability with an existing, legacy network and service management system.
- Uniform management enablers – In general, an orchestration platform could generalize over network management and service management, implementing them itself or providing access to and interfaces with such subsystems.

4.7 References

[4-1] Carlos J. Bernardos, Olivier Dugeon, Alex Galis, Donal Morris, Csaba Simon and Robert Szabo, “5G Exchange (5GEx) – Multi-domain Orchestration for Software Defined Infrastructures“, *Special Session 9, Introducing THE 5G-INFRASTRUCTURE-PPP – Launching the European 5G Initiative – Part I, EUCnC 2015*, Paris, France, 1 July 2015

[4-2] EU PROJECT SONATA, http://www.sonata-nfv.eu/about#skip-link

[4-3] EU PROJECT 5G NORMA, Functional Network Architecture and Security Requirements, 31/12/2015, https://5gnorma.5g-ppp.eu/wp-content/uploads/2016/01/5G_NORMA_D3.1.pdf

[4-4] EU Project CHARISMA, http://www.charisma5g.eu/wp-content/uploads/2015/08/CHARISMA-D6.1-v1.0.pdf

[4-5] EU PROJECT Superfluidity, http://superfluidity.eu/

[4-6] EU PROJECT 5G-ENSURE, http://www.5gensure.eu/

[4-7] EU PROJECT METIS-II, 5G RAN Architecture and Functional Design White Paper, https://metis-ii.5g-ppp.eu/wp-content/uploads/5G-PPP-METIS-II-5G-RAN-Architecture-White-Paper.pdf

[4-8] ETSI, Network Functions Virtualisation (NFV); Management and Orchestration, ETSI GS NFV-MAN 001 V1.1.1 (2014-12)

[4-9] EU PROJECT mmMAGIC, Deliverable D3.1 “Initial concepts on 5G architecture and integration” – March 2016

[4-10] Alexandros Kostopoulos, George Agapiou, Fang-Chun Kuo, et al.,, “Scenarios for 5G Networks: the COHERENT Approach,” 23rd International Conference on Telecommunications, Thessalonki Greece, May 2016

[4-11] 5G-XHaul Project, Deliverable D2.1”Requirements Specification and KPIs Document”, submitted on March 1st, 2016.

[4-12] 3GPP TS 36.300, Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2, (Release 13), 2016

[4-13] EU PROJECT METIS-II, Deliverable 5.1, Draft synchronous control functions and resource abstraction considerations

[4-14] 3GPP TR 36.842 Study on Small Cell Enhancements for E-UTRA and E-UTRAN – Higher plane aspects, 2014

[4-15] EU PROJECT Selfnet, https://selfnet-5g.eu/

[4-16] P. Rost, A. Banchs, I. Berberana, M. Breitbach, M. Doll, H. Droste, C. Mannweiler, M.A. Puente, K. Samdanis and B. Sayadi, “Mobile Network Architecture Evolution towards 5G,” in IEEE Communications Magazine, vol. 54, no. 6, June 2016.

[4-17] A. Ravanshid, P. Rost, D. S. Michalopoulos, V. V. Phan, H. Bakker, D. Aziz, S. Tayade, H. D. Schotten, S. Wong, and O. Holland, Multi-Connectivity Functional Architectures in 5G, International Conference on Communications (ICC), workshop on 5G Architecture, 2016

5 Physical architecture

This section considers the physical architecture for 5G mobile radio networks. This architecture comprises both the radio access network itself and its interconnection to the core network functions where these functions are deployed at distributed or centralized nodes in the fixed network. To deploy these functions, the fixed network encompasses several aggregation nodes that offer computing and storage capabilities. These capabilities can be used flexibly for the efficient operation of the mobile network. As described in section 4, depending on the use case, efficient operation can be achieved by centralized or distributed network functions together with virtualization of the transport network resources.

A major challenge in the 5G mobile radio access network is the efficient integration of an additional layer of small cells into the existing macro-cell network. Besides using the classical distributed RAN also for small cells, cloud-RANs (C-RAN) are considered as an innovative approach in which small cells are deployed as remote radio heads (RRHs) connected to a centralized macro-cell via a fronthaul interface. Moreover, as the fixed network comprises a heterogeneous set of technologies that is integrated by using Ethernet as a common transport platform, Ethernet is also an interesting possibility for the fronthaul transport. This might be an attractive choice as current fronthaul interfaces – e.g., CPRI, ORI, or OBSAI – may encounter capacity bottlenecks when confronted with 5G scenarios. Hence, a new functional split that shifts more processing into the RRHs and, hence, also needs more transport capacity, appears to be an interesting option. In consequence, such C-RANs need to be integrated efficiently into the classical distributed RAN (D-RAN) architecture.

This section is organized as follows: In section 5.1, the requirements of the new 5G RAN architecture are outlined. The C-RAN architecture is analyzed in section 5.2, and it is discussed why a new fronthaul interface could be needed in the logical architecture. In section 5.3, the physical substrate is highlighted, onto which 5G will most likely be implemented including aggregation in the access-, metro- and core domains. Finally, in section 5.4 it is investigated here the new network functions in 5G can be placed physically.

5.1 Radio access network

To address 5G challenges ([5-1]), a combination and integration of new radio technologies with existing technologies is anticipated, as also explained in Section 4. Firstly, **new types of frequency bands like micro- and millimeter waves are expected to be used. These will make small cells even smaller and denser than in current setups**. Also, the adoption of **massive MIMO** systems will necessitate **more efficient interference management schemes**, e.g., by coordinated multi-point (CoMP) techniques ([5-2]). This interference coordination has to happen across systems (e.g., across macro and small cells). In fact, interference between heterogeneous macro and small cells is only an exemplary aspect that has to be coordinated tightly. Another example is the co-deployment of LTE-A and novel 5G radio.

While these goals could be reached in conventional D-RANs, C-RAN concepts have been also proposed. C-RAN promises significant CAPEX and OPEX advantages for operators by centralizing hardware and by significantly reducing energy consumption ([5-3]). Towards 5G, these tradeoffs between these schemes are analyzed, and intermediate solutions are also currently under investigations..

The desired RAN flexibility imposes requirements on what technologies should support. On the other hand however, the options for realization are expected to set limits on what RAN technologies can offer in real deployments. The ability to resolve these dependencies in a real system, by using **different deployment scenarios over a single infrastructure**, is a core advance of 5G over previous architectures.

To elucidate these dependencies, let us consider four scenarios where the concrete situation of a RAN (e.g., load, channels) leads to different choices of processing locations. A first important scenario is to place core network functions at the aggregation points in the decentralized network. Another scenario instead applies centralized joint processing at the macro cells for the subordinate small cells. A third scenario considers multiple superimposed clusters of centralized processing, each serving multiple adjacent radio sites. The fourth scenario includes sites connected via wireless backhaul links to other sites that have a fixed backhaul connection, imposing additional considerations and constraints on what to process at which node. The general intention is that **the 5G RAN design must be able to support and leverage all these scenarios**. Moreover, a joint support of both backhaul and fronthaul over the same network infrastructure is expected to be needed ([5-4]).

5.2 Physical architecture to support cloud-RAN: A new fronthaul interface

As said in the previous subsection, the 5G RAN architecture must able to support adaption to the service and infrastructure requirements. For example, neither D-RAN nor C-RAN are always optimal. Whilst D-RAN is best for supporting low latency, C-RAN is more suitable for high spectral efficiency, high energy efficiency and reduced cost.

More specifically, implementing fronthaul-based C-RAN faces formidable challenges in a converged fixed-mobile network architecture as envisioned for 5G. It would have to transport waveform samples as ordinary payload ([5-5]), summing up to required data rates exceeding 1 Tbit/s per macro-cell site. Hence, there is a **need to reduce the optical bandwidth utilization** of the fronthaul.

To this end, **one main idea is to split the baseband processing** while maintaining the performance of the radio network ([5-6][5-7]). The range of split options ranges from the D-RAN, where all processing is performed locally at the remote radio head (RRH), to the fully-centralized C-RAN, **where some or all signal processing is shifted from a RRH to a base band unit (BBU) in a central unit**. The optimal allocation of processing functions, executed locally or remotely – i.e. the optimal "split" – can be decided based on a number of radio link operation factors, such as the interference situation of the mobile users (which may or may not require centralized processing); or by fixed network factors, such as transport network characteristics, network topology and scale, as well as type and volume of services that need to be supported ([5-8]).

To bridge between these conceptual extremes (D-RAN and C-RAN), a flexible architecture – also referred to as *dynamic Cloud-RAN* in this section– can be implemented using the concepts of virtualization and software-defined networking. Ideally, waveform samples carried in the fronthaul network need no longer be processed by a physical BBU – instead they could be processed by a virtual network function (VNF) that implements BBU functions. Of course, this will not be achievable for all functions using software alone. Some functions need dedicated hardware support, such as encryption, HARQ/user queues and FEC. The rest of the functions can run on commodity computer hardware and be placed at a suitable network node so that fronthaul requirements (e.g. latency and jitter) can be fulfilled. Therefore, the BBU that is responsible for a particular mobile user is no longer tied to one fixed physical location but instead it is split into physical and virtual network functions. The virtual ones can be moved dynamically in accordance with the network and service requirements. Therefore, data centers become part of the dynamic Cloud-RAN ([5-14]).

Synchronization is another challenge, as it is currently conveyed via CPRI to the RRH. By using Synchronous Ethernet (SynchE) and providing a common time reference by means of the IEEE 1588 precision time protocol, it is assumed that similar synchronization performance can be achieved in the RAN as for CPRI, and at similar cost ([5-8]). The requirements of the new fronthaul are considered also by the Time-sensitive Networks (TSN) task group in the IEEE.

Although, an intermediate option between D-RAN and C-RAN seems very promising. some questions have still to be answered. Due to the time variance of the mobile radio channel, joint control of radio links should always be performed as close as possible to the mobile user in order to react on changing channel and interference situations. But for mobile users, the best point for coordination point is likely to move in the network.

Note the IEEE 1904.3 standardization group is working on fronthaul over Ethernet. The Next Generation Radio Interface (NGRI) White Paper ([5-9]) is followed by the new IEEE 1914.3 standardization group aiming at defining a new fronthaul interface, also including the flexible split options. **It is natural that Ethernet transport technologies (which inherently offer features for virtualization, switching and sharing of network infrastructures) are provided by the IEEE**, and that the principal definition of the functional split can be independent of the waveform. But transport blocks and specific control information of the radio link can be very different for each wireless technology. **The new fronthaul interface may be specified by the standardization forums responsible for the radio link, i.e. 3GPP**, as part of making its wireless technology ready for C-RAN deployment.

5.3 Fixed network

Although an end user experiences mobile communication as a wireless technology, only the last hop is actually wireless so that data is mostly transported via a wired infrastructure. The is due to a basic constraint: As radio spectrum is limited, the sooner mobile operators can offload their traffic into a wired network, the better the performance and spectrum utilization will be. As a consequence, most operators are currently deploying FTTH and 4G in parallel, as support of the 4G network by an FTTH backhaul. **It is intuitive that 5G targets seamless integration of fixed and mobile users via converged fixed-wireless technologies** (see Figure 5-1). The fixed network comprises core, metro and access domains. While the core and metro domains are commonly realized with optical transport, the access domain uses a heterogeneous set of transport technologies. In the context of 5G, adopting optical transport also in the access domain is a key enabler as it offers high capacity. But optical transport as such lacks flexibility in general. **The necessary flexibility for 5G deployments can be realized by a combination of passive and active electro-optical technologies introduced into the access and metro domains**: Passive optical network (PON), active remote node (ARN), and dynamic frame-based optical switching nodes – they are investigated in more detail below.

5.3.1 Heterogeneous access domain

In 5G, the fixed access network interfaces the radio link with the core network. Radio heads and base stations can be connected via heterogeneous transport technologies. These include fixed links like dedicated fiber, VDSL/G.fast, coax and plastic optical fibers, together with wireless alternatives such as micro- and mm-wave and optical wireless links, and also radio over fiber as a combination of radio and optical technologies. **The heterogeneous set of technologies is aggregated in the access domain for further transport over a single-mode optical fiber**.

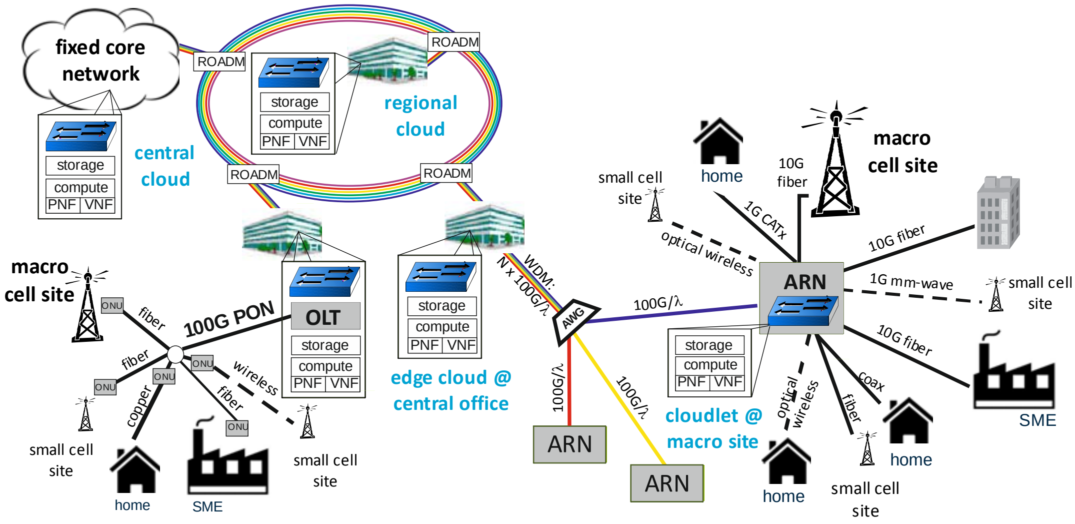


Figure 5-1 - Physical architecture of a converged fixed-mobile network for 5G.

Two main technologies are widely used here for this transport: **Passive optical networks (PONs) or Active Remote Nodes (ARNs)**. A PON consists of an optical line terminal (OLT), an optical link to a passive power splitter deployed at a remote site and individual fibers to optical network units (ONUs). **PONs share the optical bandwidth by a flexible assignment of transmission opportunities among the ONUs, thus enabling statistical multiplexing.** Note that the number of ONUs in a PON is physically limited due to the path loss at the splitting point. Current PON technology needs further evolution towards 5G. NG-PON2 introduced wavelength-division multiplexing (WDM) to multiply capacity to N*x*10 Gbit/s. In order to reach multiple 100 Gbit/s for a macro-cell with many small cells ([5-1]), orthogonal frequency division multiplexing (OFDM) promises higher capacity per wavelength by a more efficient use of the optical bandwidth, together with reduced-bandwidth processing at ONUs. 20 GHz bandwidth, frequency-selective link adaptation and polarization multiplex are current ideas for reaching more than 100 Gbit/s per wavelength in a future PON.

Instead of using a PON, an ARN can be deployed if powering is available; this is plausible as base stations (BS) require powering anyway. **An intuitive idea is to co-locate the ARN at the macro cell sites and to serve carrier-grade Ethernet and mobile radio services jointly at the ARN**, **in order to maximize statistical multiplexing already at the network edge and to consolidate traffic that is transferred toward higher aggregation nodes.** In the uplink towards the (CO), a WDM-PON is employed so that legacy fiber infrastructures can be utilized. The main challenge is to provide the multiple 100 Gbit/s data rates required for 5G at low-enough cost over the 10-20 km WDM uplink. This is already identified in recent research leveraging recent developments for interconnects inside data centers and extending their longer reach.

Both the PON and ARN architectures end in the so-called central office (CO). **Note that operators are currently aiming to rebuild their COs into remote data centers, where further compute and storage capabilities will be available soon.** In ARN-based deployments, network resources can be deployed even closer to the end user. The ARN is a carrier-grade Ethernet switch that allows naturally the addition of storage and compute capabilities at the network edge, nearest the radio links, and any other distributed network function that providers may want to develop. **In addition, hence, each macro-cell site can be developed into a small data center, ensuring ultra-low latency**.

5.3.2 Flexible metro domain

The metro network domain provides connectivity between the access branches and the regional/core data centers (DCs) as well as to the global internet (Figure 5-1). **Current metro solutions deploy dense wavelength-division multiplexing (DWDM) in metropolitan areas.** Up to 100 wavelength channels can be transported in parallel over the same fiber to accommodate the large aggregated traffic from the access branches. Besides advantages for longer distances typical in the metro domain (using coherent optical transport technologies), DWDM leads to formidable aggregate capacity in the Tbit/s range over a single optical fiber. This high aggregate capacity makes DWDM especially suitable for supporting broadband services in densely populated scenarios addressing 5G requirements.

However, the huge capacity of DWDM metro networks comes at the price of low flexibility. Optical switching can be performed e.g. by using wavelength-selective switches (WSS) and by reconfigurable optical add/drop multiplexers (ROADMs). But these optical devices are relatively slow; it is hardly possible to offer optical packet switching in this way. **Given the very diverse requirements of operational and end-user services in the context of 5G, there is a need for new approaches, deploying more dynamic and flexible solutions that offer higher granularity at the sub-wavelength level and more elasticity in the metro domain** ([5-12]). In view of these requirements, dynamic frame-based WDM metro solutions are currently being investigated, including the concept of the time shared optical network (TSON), to provide variable sub-wavelength switching granularity and the ability to dynamically allocate optical bandwidth elastically ([5-13]).

Between the CO uplink and the metro ring, ROADMs select one or more wavelengths for the particular branch in the access domain – i.e. the metro ring further aggregates the traffic**. The optical transceiver industry and the academic community are actively searching cost-effective solutions offering 100 Gbit/s per wavelength that work over distances typical in metro networks, e.g. several tens of kilometers.** However, the required combination of capacity, flexibility, low cost and distance is not feasible with today's technologies. **Further advances in optical devices** – e.g., tunable lasers, ROADMs, and fast optical switches – **as well as ultra-high speed digital signal processors with low cost and reduced power consumption will be necessary** to support the required functionality and acceptable OPEX and CAPEX.

5.3.3 Integrating across access, metro, and core technologies

It is common to use Ethernet transport over all media in the fixed network as it offers low-cost interfaces and simplifies network integration. Ethernet is the primary network protocol also in data centers for server-to-server communications. The Data Center Bridging (DCB) Task Group in IEEE 802.1 has developed a set of standards with the objective of creating a converged data-center network infrastructure using Ethernet as the unified fabric. Due to its widespread use, Ethernet is a cost-effective technology.

Hence, **Ethernet is seen as the common technology for integrating heterogeneous technologies in the access, metro and core domains**. Note also that the Metro Ethernet Forum (MEF) defined a number of additional profiles which enable the infrastructure operator to define different virtual network topologies (such as ring, bus and tree) in parallel to each other in the same physical infrastructure. **Using Ethernet as the common transport platform implies the use of Ethernet in the fronthaul as well.**

Towards the metro and core network domains, WDM is additionally used for network integration. DWDM also allows different virtual network topologies (such as ring, bus and tree) on the same optical fiber infrastructure. **While WDM is less flexible than Ethernet due to the time needed to reconfigure ROADMs and WSSs, this might be overcome by (1) using combined electro-optical solutions offering more flexibility such as the TSON described above and (2) placing an Ethernet switch or router also into the metro domain.**

5.3.4 Support for multiple services

In the previous sections, the term "network slicing" was used to refer to the ability to create various logical architectures on top of a single physical architecture. **A similar concept to network slicing was previously discussed in the fixed network under the term "open access", allowing the service providers to use a common physical infrastructure provided by a network provider, making the deployment of parallel physical networks unnecessary.** Network slicing and open access have the potential to enhance the commercial success of 5G wireless networks by means of spectrum and infrastructure sharing.

To this end, different SPs need only to deploy a single converged fixed-wireless network and then offer their services, optimizing their CAPEX and OPEX by sharing the converged network's resources. **One way to support such multi-vendor scenarios is by centralizing control and management via an SDN approach.** This is further eased by operation via Metro Ethernet technologies as the common transport platform together with a Hardware Abstraction Layer (HAL). In this way, vendors can integrate seamlessly with the joint control and management plane.

5.4 Mapping of network functions to physical resources

This subsection describes the set of resources available in the physical network that can be used and supervised by management and orchestration to operate the mobile network for different use cases and tenants. Therefore, the different categories of infrastructure resources are discussed and the types of hardware on which they can be implemented.

5.4.1 Deployment opportunities for computing resources

The physical deployment opportunities offered by the fixed network to 5G suggest five different computing resource categories: central, regional, edge cloud and cloudlet nodes and physical network functions (PNF).

A central cloud node lives in a centrally located data center. Such a center hosts a large collection of processing, storage, networking, and other fundamental computing resources. On such a node, tenants are allowed to deploy and run arbitrary software, e.g., operating systems and applications. Typically, only a few central cloud nodes are found in a nationwide operator network.

A regional cloud node is available in densely populated metropolitan, urban, and sub-urban areas, e.g. **attached to a metro ring**. Besides hosting network functions, these nodes host software deployed and run on behalf of a consumer, again including operating systems or applications. The number of regional cloud nodes is at least one order of magnitude higher than the number of central cloud nodes.

An edge cloud node is implemented inside an access branch of the fixed network, serving e.g. a city quarter, and thus even closer to the end user. In typical deployments, it would be situated **inside the CO.**

A cloudlet (nano cloud) is a mobility-enhanced small-scale cloud data center that is located at the edge of the network **collocated with the macro cell sites**. In the case of active deployment based on the ARN, both the cloudlet and the macro-cell sites would be co-located with the ARN. The main purpose of the cloudlet is supporting resource-intensive and interactive mobile applications by providing powerful computing resources to mobile devices and IoT devices with lower latency.

Physical network functions describe functions that are realized directly in hardware, without using virtualization techniques. That could be due to legacy implementation; it could also be due to the fact that such functions are not amenable to virtualization on standard commodity hardware. This second case describes more strictly the notion of a purely physical network function – one that *cannot* be virtualized even if one wanted or only at unacceptable

CAPEX/OPEX. This is often due to a tight coupling between hardware and software, the need to use specific acceleration engines (e.g., DSP or FPGA), or very tight latency or real-time requirements as often found in signal processing.

Central, regional, edge clouds and cloudlets provide physical resources to execute VNFs for management and orchestration. **From the core downwards to the fixed access domain, central, regional, edge clouds and cloudlets grow in heterogeneity of hardware and hypervisor**; the geographical deployment and topological structure become more complex as well. Hardware includes both commodity and specialized hardware components like memory, compute, storage, networking (in both kinds of hardware), or accelerators (mostly in specialized hardware).

Identifying the required mix of VNFs and PNFs in the 5G network architecture and implementing them exemplarily in a representative scenario are critical research topics on the path towards successful 5G deployment. In this way, the specialized hardware and software requirements become well understood and sufficient hardware can be made available at the central, regional, edge cloud and cloudlet sites during the deployment.

5.4.2 Support for dynamic Cloud-RANs

5G requires a very flexible architecture for implementing the different scenarios and specifically the *dynamic Cloud-RAN* described in subsection **Error! Reference source not found.** ([5-15]). **In the following, it is described how the whole 5G mobile network could be organized as a *dynamic Cloud-RAN* with distributed network functions.**

Investigations indicate that small cells can be efficiently controlled by the superordinate macro cells, with best results ARN-based deployments. In this case, the mobility management and handover functions could be implemented as VNFs (eventually including PNFs) **in the cloudlet collocated with the macro cell site.**

If the user moves to the cell edge covered by the macro-cell, **coordination with adjacent macro cells** is needed. These are **best controlled by using computing resources attached to the CO** where a whole mobile network segment containing all base stations connected to this CO can be managed. Accordingly, the user is then handed over from the cloudlet to the edge cloud.

Mobile network segments are formed by COs, to which macro cell sites are connected. If users move between such segments, handed over from the edge cloud in the CO, they **are controlled by the regional cloud attached to the metro ring**. Same will happen for the handover from the regional to the central cloud.

While users move between the cells, their network functions should be moved to the appropriate central, regional, edge cloud or cloudlet in the physical network. PNFs have to be implemented at each cloud physically, while the mobility management and handover control functions could be implemented as a distributed network function on the virtualized storage and compute capabilities in the appropriate cloud.

In case of handover, the required **VNFs and PNFs are already pre-configured** at the destination cloud, **so that they are already instantiated and immediately available**. Handovers can then be implemented by changing the address of user packets after the corresponding management functions have been taken over by the destination cloud.

For low-latency, additional means may be required. Note that during handover in 4G, lost packets are transferred via the X2 interface. **5G may need decentralized security** ([5-5]) **and fast direct X2 paths** established through the fixed access network so that the core is bypassed. In the dynamic Cloud-RAN, the X2 path is the link **between the source and destination clouds**. Furthermore, in 5G the transport of lost packets to the destination cloud may use look-ahead techniques.

5.5 References

[5-1] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, M. Fallgren, „Scenarios for the 5G Mobile and Wireless Communications: The Vision of the METIS Project, IEEE Comun. Magazine, vol. 52, no. 5, pp.26-35, May, 2014.

[5-2] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfröjd, T. Svensson, The role of small cells, coordinated multipoint, and massive MIMO in 5G, Commun. Magazine, IEEE, vol.52, no.5, pp.44-51, May 2014.

[5-3] Chi-Lin I, Jinri Huang, Ran Duan, Chunfeng Cui, Jesse (Xiaogen) Jiang, Lei Li; Recent Progress on C-RAN Centralization and Cloudification, IEEE Access, Vol. 2, 2014, pp. 1030-1039.

[5-4] ICT-317669 METIS, Deliverable 6.4, Final report on architecture, January 2015.

[5-5] V. Jungnickel, K. Habel, M. Parker, S. Walker, C. Bock, J. Ferrer Riera, V. Marques, and D. Levi, Software-defined Open Architecture for Front-and Backhaul in 5G Mobile Networks, in Proc. 16th IEEE International Conference on Transparent Optical Networks (ICTON), 2014.

[5-6] U. Dötsch, M. Doll, H-P. Mayer, F. Schaich, J. Segel, P. Sehier, Quantitative Analysis of Split Base Station Processing and Determination of Advantageous Architectures for LTE, *Bell Labs Tech. J*, vol. 18, no.1, pp. 105-128, May 2013

[5-7] D. Wübben, P. Rost, J. Bartelt, M. Lalam, V. Savin, M. Gorgoglione, A. Dekorsy, G. Fettweis, Benefits and Impact of Cloud Computing on 5G Signal Processing: Flexible centralization through cloud-RAN, *IEEE Signal Proc. Mag*., vol.31, no.6, pp.35-44, Nov. 2014

[5-8] N. J. Gomes, V. Jungnickel, P. Chanclou, J.-P. Elbers, P. Turnbull, A Flexible, Ethernet Fronthaul for 5th Generation Mobile and Beyond, Optical Fiber Conference 2016 (invited paper).

[5-9] China Mobile, Alcatel-Lucent, Nokia Networks, ZTE, Broadcom, Intel, White Paper of Next Generation Fronthaul Interface, Version 1.0, Juni 4, 2015.

[5-10] D. Verchere, Cloud Computing over Telecom Network, Optical Fiber Communication Conference (OFC), 2011, USA, paper OMW1.

[5-11] European Commission, Digital agenda for Europe: 5G Infrastructure Public Private Partnership (PPP): The next generation of communication networks will be 'Made in EU', February 2014.

[5-12] A. Tzanakaki, M. Anastasopoulos, D. Simeonidou, I. Berberana, D. Syrivelis, T. Korakis, P. Flegkas, D.C. Mur, I. Demirkol, J Gutierrez, E. Grass, Q. Wei, E. Pateromichelakis, A. Fehske, M. Grieger, M. Eiselt, J. Bartelt, G. Lyberopoulos, E. Theodoropoulou, 5G Infrastructures Supporting End-User and Operational Services: The 5G-XHaul Architectural Perspective, ICC 2016, W01. Workshop on 5G Architecture (5GArch).

[5-13] B. R. Rofoee, G. Zervas, Y. Yan, M. Anastasopoulos, A. Tzanakaki, S. Peng, R. Nejabati, and D. Simeonidou, Hardware Virtualized Flexible Network for Wireless-DataCenter (invited), IEEE/OSA J. Opt. Commun. Netw., March 2015, vol. 3, pp A526 - A536.

[5-14] A. Oliva, X. Costa-Pérez, A. Azcorra, A. Di Giglio, F. Cavaliere, D. Tiegelbekkers, J. Lessmann, T. Haustein, A. Mourad, P. Iovanna, Xhaul: toward an integrated fronthaul/backhaul architecture in 5G networks, IEEE Wireless Commun. 22(5): 32-40 (2015).

[5-15] CHARISMA Deliverable D1.1: CHARISMA intelligent, distributed low-latency security C-RAN/RRH architecture.

6 Software Network Technologies

Software network technologies are introduced as fundamental enablers to realize the requirements of programmability, flexibility (e.g., re-configurability, reusability and infrastructure sharing), adaptability (e.g., self-configuration, self-protection, self-healing and self-optimization) and capabilities (e.g., mobile edge computing, network slicing, autonomic network management) expected to be inherent in 5G networks.

The benefits of softwarization in 5G include operational expenditure (OPEX) and capital expenditure (CAPEX) reduction, speedy service creation and deployment, efficient service life-cycle management, energy consumption reduction, and improved quality of experience for users, among others. Indeed, softwarization is recognized as one of the key features of 5G networks as it drives the paradigm shift of mobile network design and implementation. This section provides more details on software network technologies, as such as well as their impact on the network architecture. Expectations are that a number of enablers would be required such as multi-tenancy management, multi-domain orchestration, end-to-end network slicing, uniform virtualization and abstraction facilities.

6.1 Softwarization in 5G

Softwarization of networks includes the implementation of network functions in software, the virtualization of these functions, and the programmability by establishing the appropriate interfaces. As mentioned in section 3, a new softwarization and programmability framework in the 5G system architecture is expected in order to manage 5G networks. This softwarization requirement is also identified by ITU-T Study Group13, which has recently launched the IMT-2020 Focus Group on network aspects to analyze how emerging 5G technologies will interact in future networks [6-2].

Network softwarization is an approach to use software programming to design, implement, deploy, manage and maintain network equipment/components/services. It takes advantage of programmability, flexibility and re-usability of software for rapid re-design of network and service architectures. The goal of network softwarization is to optimize processes in networks, reduce their costs, and bring added value to network infrastructures.

Leveraging virtualization technologies, softwarization is one of the key enablers for unifying the 5G end-to-end service platform, and for realizing network slicing as a service. Softwarization evolves networks into the management and orchestration of complex software systems, encompassing and harmonizing what hitherto was thought of as inseparable domains: network- and resource-oriented functions and application-oriented functions. This enables developers and operators to better match needs and capabilities, building application-aware networks and network-aware applications. This joint expressive power will be one of the main drivers of innovations enabled by 5G.

While softwarization plays this key role for 5G network management and service provisioning, it is important to note the variety of needs for softwarization in different segments of 5G networks. 5G network segments include radio access networks, core networks, transport networks, network clouds, mobile edge networks and Internet. Certainly, each segment has its own technical characteristics, and thus different requirements of softwarization. The software network technologies applied in 5G network segments are illustrated in Figure 6-1. In the following subsections, the views of 5G PPP on softwarization in radio access networks, mobile edge networks, core networks and transport networks are examined.

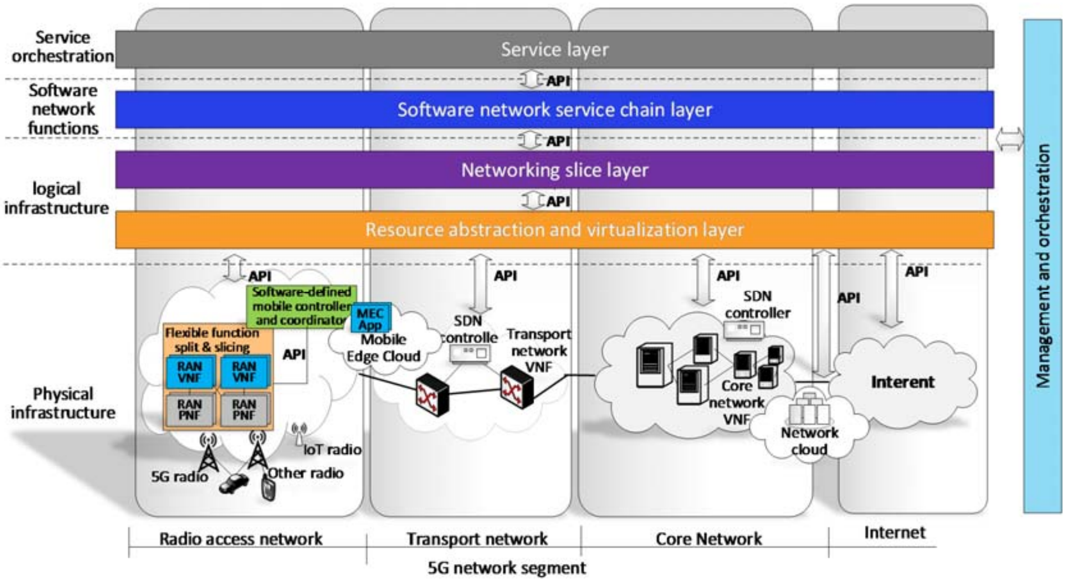


Figure 6-1: Software network technologies in 5G overall architecture.

6.1.1 Softwarization in radio access networks

The fundamental system requirement for 5G RAN is unprecedented agility in spatial, temporal and frequency dimensions. We identify four design aspects for 5G RAN that will greatly benefit from the introduction of software network technologies, namely i) flexibility in spectrum management; ii) fine-grained network programmability; iii) dynamic provisioning of network slices; and iv) heterogeneous and dense deployments. These four aspects are hardly orthogonal, and thus, design tradeoffs emerge.

5G RAN should support a wide range of physical deployments, and be able to maximally leverage centralization, while also supporting distributed base stations and being able to operate over non-ideal backhauls. A key enabler for this is the implementation of some radio functions as VNFs, allowing these to be flexibly shifted toward or away from the radio edge, depending on the physical architecture and specific application requirements.

From the 5GPPP viewpoint, the following considerations are important regarding the softwarization of radio network functions. Preliminary analyses in [6-1] concluded that functions that are asynchronous to the radio interface – in LTE these are packet data convergence protocol (PDCP) and radio resource control (RRC) functions related to measurement control and reporting, handover preparation and execution, dual connectivity, random access, RRC state transition etc. – are most suitable to be implemented as VNFs and possibly centralized, as they typically require low data rates on their interfaces, and scale with the number of users and not the overall traffic. Further, these functions can typically cope with relatively larger latency (e.g. tens of milliseconds in LTE).

Synchronous functions in LTE are PHY, medium access control (MAC) and radio link control (RLC) functions such as scheduling, link adaptation, power control, interference coordination etc. These typically require high data rates on their interfaces, scale with traffic, signal bandwidth and the number of antennas, and benefit from hardware acceleration. The potential for centralization is here most pronounced in deployments with low-latency and high-bandwidth backhaul, due to timing and real-time processing requirements. Partially centralized solutions may be an option for lower bandwidth and/or higher latency backhauls

Considering the control plane, the design of the control plane in 5G radio access networks will benefit from SDN principles. To implement software-defined network control in radio networks, wireless functionality will be split between the functions that are being ‘controlled’ and remain relatively stable, and those functions that 'control' the overall network and are executed at the controller. The logically centralized software-defined mobile network control regulates both data plane and control functions, which can be placed arbitrarily in the edge cloud or the central cloud.

6.1.2 Softwarization in mobile edge networks

In line with the vision of the Mobile Edge Computing (MEC) paradigm, e.g., from ETSI MEC Industry Specification Group (ISG), it is widely recognized that mobile edge networks will extend softwarization from the conventional data center to the edge of 5G networks. It will also enable services to be deployed on demand to the most effective locations within the access network according to the requirements of applications e.g., in terms of real-time service delivery for fast and efficient deployment/re-deployment of mobile edge networking and computing, it is essential to develop automatic softwarization mechanisms to establish the required services from scratch (even bare metal) in a timely fashion.

6.1.3 Softwarization in core networks

The majority of the CN and service plane functions are expected to be deployed as VNFs in the 5G timeframe, thus running in virtual machines over standard servers, potentially on cloud computing infrastructures (i.e. data centers). The design of these CN functions will to some extent explore SDN principles, such as data plane/control plane split, fulfilling the envisioned SDN/NFV native architecture. These VNFs can be flexibly deployed in different sites in the operator's network, depending on the requirements with regards to latency, available transport, processing and storage capacity, etc. Different services or network slices can utilize different CN and service plane VNFs, which can be deployed at different network sites.

6.1.4 Softwarization in transport networks

The softwarized, programmable transport networks can act as a platform for applications, user services and network services, to adapt the operation of the transport network to the needs of the RAN. This may be carried out through an appropriate interface, a configurable policy framework, resource discovery and optimization, NFV and an SDN control framework that is separate from the forwarding plane. Transport connections will be dynamic and provisioned on-demand. In addition, a softwarized transport network will allow for tightly coupled interactions with the RAN, whereby transport and RAN could jointly coordinate aspects such as mobility and load balancing, or manage sleep periods of RAN and transport equipment.

6.2 Programmability considerations in 5G

Programmable networks are networks that allow the functionality of certain network elements to be programmable dynamically. Network programmability is a concept that involves network softwarization and virtualization of infrastructure and network functions. It demands systematic splitting and abstraction of network functions, and the corresponding design of APIs and SDKs to face the emerging needs for service flexibility, network efficiency, network security and reliability. It empowers the fast, flexible and dynamic deployment of new network and management services, executed as groups of virtual machines in the data plane, control plane, management plane and service plane in all segments of the network. Through the access to network services via open API and SDK, it will also facilitate the creation of ecosystems which benefit different planes therein.

The programmability of life-cycle management of softwarized services should cover service creation or composition, service provisioning, service monitoring, etc. The benefits of enabling such programmability would be cost-efficient and the uniform way of operating, maintaining, and upgrading physical or virtual network infrastructure and network management applications.

The introduction of NFV is an important step towards generalizing network programmability. While many conceptual problems of NFV have been solved, it is still difficult to program for NFV/SDN architecture. A software development environment tailored for NFV development is required. Well-defined interfaces between the SDK and the orchestration platform are needed in order to support rapid deployment and monitoring of new services and of new versions.

To make programmable networks practical there are still several challenges to be solved. Networking is a field where general-purpose commoditized domain-specific platforms are still lagging behind. In the networking field, bare metal switches are just starting to emerge, and a clear and neat interface between hardware functions and their software configuration and control is still missing. The reason behind this is not only technological, but also revolves around the identification of what should be the network and traffic processing “primitives”, or blocks, supported by a platform, and how such blocks should be combined to produce meaningful and realistically complex network functions.

To solve this, there is a need for appropriate functional decomposition of a higher-level network function into smaller and reusable network and processing blocks, and the identification of how to formally describe and run-time enforce how such blocks combine together to form a desired higher level function. As a solution, architectural components of a programmable network node may include three main elements: a set of basic/elementary (domain-specific) reusable function blocks (building blocks used to compose high-level functions and services, as discussed in section 4); a function block composition language, which provides the platform-agnostic formal description of how such reusable function blocks should be invoked; and a function block composition execution environment in charge of executing such a composition.

Moreover, one key consideration in programmability is how to handle stateful processing in network functions. The current SDN approaches like OpenFlow face limitations to install and enforce stateless packet forwarding rules delivered by the remote controller. It has severe limitations when called to deploy more complex (e.g. stateful) flow-processing and filtering functions. It is necessary to promote a novel network processing architecture able to support a programming abstraction which retains the platform independent features, but at the same time able to execute stateful flow-processing tasks, formalized in a vendor-agnostic manner. This adds to the extensions of SDN discussed (from a different angle) in Section 6.1.1.

Finally, programmability needs to be built into the technology layer itself. For example, in the transport segment, wireless technologies, such as mm-wave links, could be made programmable by building APIs that allow the control plane to steer beam directions, thus effectively controlling the network topology. Optical transport technologies should also incorporate programmability functions, which allow the control plane to allocate bandwidth to connections in a very granular and flexible manner, adjusting both time and frequency resources; as well as to selectively turn off optical transceivers to save energy.

6.3 Resource and service orchestration

Resource and service orchestration are two fundamental aspects in 5G networks. Resource orchestration refers to the integration and coordination of physical or virtual network, compute an storage components available in networks to realize different services, to support network slicing, and/or to achieve certain performance goals. Service orchestration is built on top of resource orchestration to provision services cross multiple network segments and/or different network domains. Software networks technologies are expected to be the cornerstone for flexible resource and service orchestration in the 5G domain. In the following the view of 5GPPP projects on resource and service orchestration are presented. For resource orchestration the focus is on the virtual network function orchestration.

6.3.1 Flexible network function orchestration

Resource orchestration in 5G networks will leverage NFV/SDN technologies. NFV/SDN principles have been elaborated by open source activities, European projects and Standards Developing Organization (SDO) initiatives, aiming toward the realization of Carrier Grade Software Defined Infrastructures (CG-SDI) capable of hosting 5G network functions and next generation services [6-4].

Most NFV orchestration platforms are based on the reference functional architecture Management and Orchestration (MANO) proposed by ETSI NFV Industry Specification Group [NFV14]. The purpose of MANO is to facilitate the management of the resources in the network so as to meet expected demand patterns, maintaining desired QoS & security levels, while making efficient use of resources and minimizing costs. This could be achieved through interactions with the various layers of the MANO stack including the VIM, VNFM and Orchestrator.

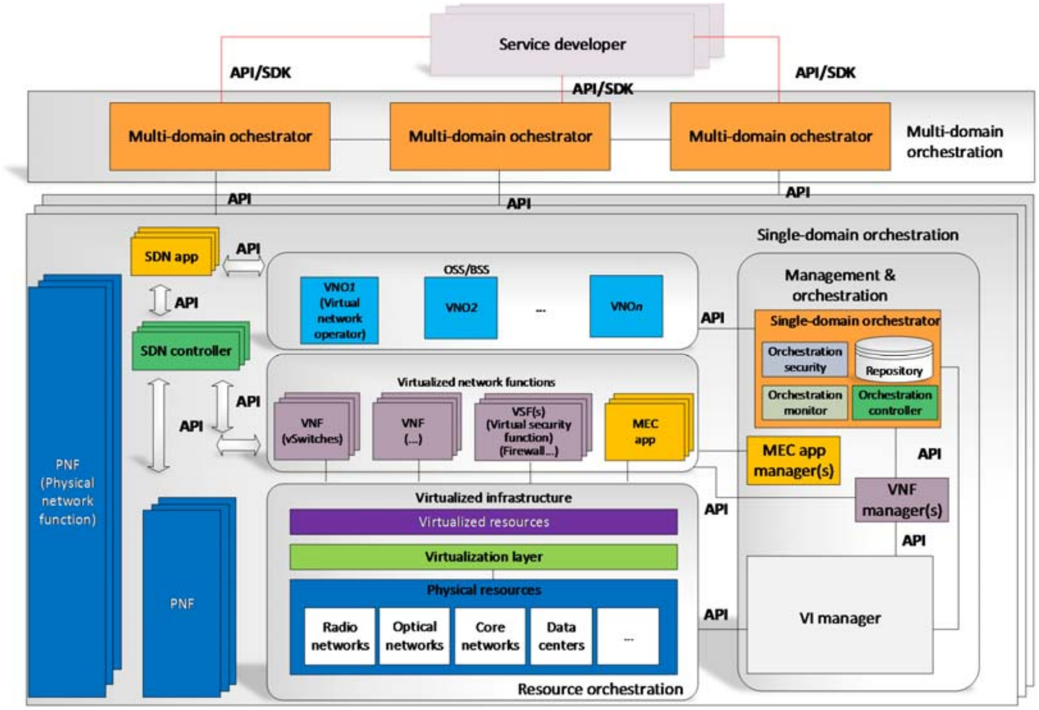


Figure 6-2: NFV/SDN orchestration for services

However, the current MANO architecture will need significant improvement in order to fulfill 5G requirements. This has been one area of focus in 5G PPP. A seamless and flexible combination of NFV and SDN technologies would be a natural target for the 5G orchestration platform. One combined NFV/SDN orchestration platform developed by several 5GPPP projects is illustrated in Figure 6-2, where software network technologies are the key to make the resource and service orchestration. For more information of the orchestration architecture, please refer to Section 3.4. The new NFV and SDN combined orchestration plane will allow the following benefits: the combination of VNFs into service graphs; the lifecycle management of VNFs; the coordination of allocating VNFs to virtualized resources; the homogenized control and management of the hardware resources; and the slicing of resources for supporting multi-tenancy.

The orchestration plane should be highly configurable and modular, supporting various orchestration patterns. It is configurable by the operator of a network through the way that its components interact. This platform is also configurable in that core functional parts – e.g. placement and scaling functions, contextualization, and lifecycle management – can be customized for each running network service individually. It will be able to deal with both stateless and stateful services. It will be able to integrate various monitoring frameworks and use it to trigger its own operations. The platform itself may be able to use a slicing system to run and manage different services in separate slices. It may also need to cope with MEC orchestration needs, addressing MEC applications life-cycle management.

Moreover, the orchestration plan operates under the policy set by infrastructure providers and communicates with the OSS/BSS of the network service provider to report status and possibly to receive requirements.

To support diverse services, multiple network and service functions can be orchestrated by chaining the functions on demand in a specific order, to create control plane and data plane service graphs. The path can be established via the physical connections as well as the logical overlay among the network functions (VNFs) by the SDN controller.

To facilitate this orchestration, network functions are classified based on their different levels of compatibility with either NFV or SDN implementations. For back-compatibility of legacy network functions, such functions are classified into: (i) the physical network functions and elements which cannot be virtualized; and (ii) the virtual network functions and elements that are virtualized. For physical and virtual network functions, different control methods are applied through SDN controllers or orchestration platforms.

Virtual network functions and applications could be distributed at the edges of 5G mobile networks to benefit from the Mobile Edge Computing model. To ensure scalability, these functions could be managed by a pool of SDN controllers or orchestration platforms. The architectural organization and physical locations of these SDN controllers/orchestrators should be optimized to allow high cost efficiency and high performance.

The main challenge in 5G resource orchestration is the implementation of a centralized management and orchestration plane that is able to coordinate IT cloud requirements (compute, storage) and networking requirements (SDN-based WAN). The 5G management and orchestration plane is expected to close the loop between the performance requirements of virtualized service functions and the allocation of resources exposed by a virtualized distributed heterogeneous substrate.

Automated management is necessary to handle the complexity in the 5G resource orchestration platform. Automated management in multi-administration environments involves defining novel business, coordination and information models, trading mechanisms and pricing schemes.

Autonomic management of SDN and NFV components can be implemented in an environment that uses monitoring data collected from the various components and functions in the network, and applies machine learning (ML) techniques and algorithms to develop a model which in turn informs the network management decisions. ML techniques could address the challenge brought by accurate service demand prediction and provisioning in virtualised environments. This should allow the network to resize and provision itself to serve predicted demand according to parameters such as location, time and specific service demand from specific users or user groups. To realize these ML techniques, approaches from Big Data handling are likely to become necessary.

6.3.2 Service orchestration

Softwarization for network management indicates introducing programmability to the complete set of network management functionalities including network configuration, monitoring, maintenance, fault/failure management, security, accounting, among others. Moreover, automatization of these softwarized management functionalities will significantly reduce OPEX and CAPEX whilst offer high flexibilities, increased predictability of management quality, improved productivity and robustness. On top of this, intelligence-based, autonomic operation will need to be in place to drive the automatic deployment of the network services.

A the most promising feature empowered by network softwarization is orchestration that easily, rapidly and flexibly implements incumbent and emerging services. In softwarized 5G networks, end-to-end services will be realized through cross-domain orchestration of services and resources over multiple administrations, as well as multi-domain orchestration over single administrations. The orchestration will allow instantiating end-to-end networks and services into heterogeneous technology domains and multi-vendor environments. Composite services need to take into account the seamless combination of networking with computing and storage.

Service deployment, activation and further management will be viewed as the efficient mapping of service functions onto a virtualized substrate, possibly belonging to multiple infrastructure operators. End-to-end service and resource orchestration will leverage northbound APIs exposed by existing networking and cloud management solutions based on SDN and NFV principles.

In 5G networks, network slicing will be an essential means for service provisioning, in order to accommodate different vertical sectors under a unified network infrastructure view. Network slicing may provide horizontally planned network architectures in telecommunications; and the business model that decouples physical access to the network resources from the actual provisioning and delivery of services on top of these network resources. In such architectures, the owner of the physical infrastructure does not supply services for the network, but separated retail service providers act as virtual network operators. The integral part of the management and operation of such networks is an intrinsic layer of security embedded in the control and management platform.

The softwarized 5G transport network will support network slicing through the concept of transport slices. A transport slice will allow an orchestrator to flexibly connect physically distributed compute resources that host the VNFs belonging to a given tenant or service. Within a transport slice subsets of VNFs could be connected at arbitrary levels, such as layer 2 or layer 3; and topologies, such as point to point, multicast trees or broadcast domains. Transport slices should also support seamless VNF mobility within the slice.

6.4 Security considerations in software networks

Software networks come with both challenges and opportunities for security. Softwarized networks bring the flexibility to implement network functions. This level of flexibility provides new capabilities to protect against new threats coming from the exponential increase of the number of users in the network. On the other hand, softwarization may impact the security of networks and thus new protection capabilities need to be put in place, including advanced management capabilities such as migration, upgrading and reverting changes in softwarized networks.

One of the challenges is the development of mechanisms that protect a network from malfunctioning software, e.g. by detecting and preventing exploits. The standardized open interfaces that support the programmability of the network obviously call for strong mechanisms for authentication and authorization at the control plane of a network. Moreover, the significant requirement of standardization for ensuring inter-operability between all the stakeholders involved is to leverage novel authentication and authorization methods in multi-domain environments.

A "trust stack" anchored in hardware platforms will also be essential for verifying the authenticity of the software, and serving as a root-of-measurement of the trustworthiness of the infrastructure components during orchestration. A secure, efficient, and scalable solution is needed to manage these trust anchors, e.g. when new hardware is deployed. Mechanisms need to be in place that detect and protect against denial-of-service and traffic analysis attacks, etc..

A full and secure utilization of the network-slicing concept built on top of such an infrastructure requires high assurance virtualization components, such as hypervisors. Integrating security into NFV as virtualized security will also bring advantages, and allow security to be tailored dynamically for specific application requirements as part of the normal operation of a network that supports network slicing.

If all these entities and their interplay are properly designed, the benefits will be enormous. Strong isolation between network slices will be guaranteed, allowing mission-critical services to run side by side; cyber-attacks will be confined to slices without impacting other network parts. Furthermore, specialized security services can be provided in those slices that need them. For example, micro-segmentation can be used to protect particularly critical resources. In addition, novel software networking based security paradigms would significantly strengthen security management. For instance, distributed security NFVs can be deployed to resolve distributed denial of service (DDOS) attacks, and when coupled further intelligence, substantially improve self-protection in 5G can become a reality.

6.5 References

[6-1] ICT-317669 METIS, Deliverable 6.4, “Final report on architecture”, January 2015.

[6-2] Study Group 13, ITU-T, http://www.itu.int/en/ITU-T/studygroups/2013-2016/13/Pages/default.aspx

[6-3] Networks Functions Virtualization (NFV); Management and Orchestration, ETSI GS NFV MAN 001 V1.1.1 (2014-12)

[6-4] Marco Gramaglia, Ignacio Digon, Vasilis Friderikos, Dirk V Hugo, Christian Mannweiler, Miguel A Puente, Konstantinos Samdanis, Bessem Sayadi, "Flexible Connectivity and QoE/QoS Management for 5G Networks: the 5G NORMA view", 3rd International Workshop on 5G Architecture, Kuala Lumpur, Malaysia, May 2016.

7 Impact on standardization

In the following sections we have identified the standard organizations impacted by the 5G architecture work. A standardization roadmap for the architectural innovations presented in this white paper will be published by the 5GPPP Pre-standard Working Group later this year.

7.1 Impacted standards organizations

A large number of standards-related organizations is involved in the standardization of "5G". This includes classical SDOs, but increasingly also forums, industry groups, open source initiatives and other types of industry cooperation. Some of these newer types of cooperation have had a major impact on the way standards are being developed, and therefore are likely to be very relevant for 5G standardization. We use the generic term "organization" here for any "body' that creates outputs that are likely to be accepted by the industry as "standard".

Note: Open source initiatives such as OPNFV, OpenDaylight, OpenStack, etc. have gradually become an alternative to the standardization of protocols in formal specifications. However, most of these initiatives base themselves on architecture work in other organizations, e.g ETSI ISG NFV in the case of OPNFV. Therefore it is likely that they will be impacted in an indirect manner by ideas developed in the architecture space, rather than directly.

7.2 Architecture and security groups in standards organizations

A non-exhaustive list of the organizations that are active in the field of 5G architecture and security includes 3GPP, ETSI, IETF, ONF, BBF and oneM2M. In each of these larger organizations there are subgroups that have a specific role on architecture and security.

7.2.1 3GPP

The 3rd Generation Partnership Project (3GPP) unites telecommunications standard development organizations (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC) and provides their members with a stable environment to produce the Reports and Specifications that define 3GPP technologies.

In 3GPP, architecture work related to 5G is split into RAN architecture and system architecture.

3GPP SA2 is responsible for RAN-CN interface, RAN-CN functional split and CN system architecture for the next generation mobile networks that shall support access technologies including the new RAT(s), the evolved LTE and non-3GPP. Many topics under study at SA2 are centered around how 3GPP CN system architecturally shall evolve to work with network slicing, and continue to accommodate features such as eDRX, CIoT, V2X etc. by developing 5G-native session management, mobility management, UE states transition, etc.

3GPP RAN2 and to some extent RAN3 are responsible for the RAN architecture, including possible slicing of RAN.

3GPP SA3 is responsible for security aspects.

3GPP SA5 is responsible for network management and is the main group in which network function virtualization for 3GPP is being discussed.

7.2.2 ITU-T

ITU-T IMT 2020 covers all 5G no-radio segments as far as overall 5G architecture, network softwarization, integrated network management, fixed mobile convergence is concerned

7.2.3 ETSI

ETSI is a regional SDO, which has also introduced Industry Specification Groups (ISGs) as a tool for the industry to create common output in a more lightweight manner than through the traditional Technical Groups and ETSI Projects. ISGs include e.g. ETSI ISG Network Function Virtualization (NFV), ETSI ISG Mobile Edge Computing (MEC), ETSI ISG Microwave Transmission (mWT) and ETSI ISG Next Generation Protocols (NGP). Recently, ETSI has established its first Open Source Group (OSG), ETSI OSG OpenSource MANO (OSM), which is an open source initiative whose objective is to create a production-quality NFV orchestrator based on open-source code that should become a reference implementation of the MANO stack. Among the Technical Groups in ETSI is SES, which is a relevant group for satellite communication.

ETSI ISG NFV has become of major relevance to the industry with respect to virtualization aspects. Within ETSI ISG NFV, the following groups are the most relevant for architecture and security issues:

IFA (Interfaces and Architecture)

EVE (Evolution and Ecosystem)

REL (Reliability Availability and Assurance)

SEC (Security)

7.2.4 IETF

The Internet Engineering Task Force (IETF) and its research organization Internet Research Task Force (IRTF) have among other things been developing technology that 3GPP refers to. The following groups are likely to be relevant for 5G architecture and security issues.

IETF DMM (5G)

IETF SFC (service function chaining)

IETF I2NSF (security)

IRTF NFVRG (virtualisation)

7.2.5 ONF

The Open Networking Forum (ONF) has developed an SDN architecture that may be relevant for SDN and NFV concepts. The ONF Services Area is responsible for both SDN architecture and security.

7.2.6 BBF

Broadband Forum (BBF) is an influential group for fixed networking aspects. The Working Groups that may be relevant are:

BBF Architecture and Migration

BBF Wireline Wireless Convergence (5G)

BBF User Services (IoT)

BBF SDN/NFV

7.2.7 oneM2M

In the field of machine-to-machine communication, or the Internet of Things, there is no single dominant standards organization. Nevertheless, oneM2M is expected to be one of the organizations impacted by 5G architecture and security in this area. The most relevant groups in oneM2M in this context are:

oneM2M ARC (IoT architecture)

oneM2M SEC (IoT security)

oneM2M MAS (IoT Management Abstraction and Semantics)

7.2.8 Other

Additional standards groups that may be impacted include:

TMF ZOOM (SDN/NFV)

8 Conclusions

The current white paper captures the latest findings related to 5G research activities of a number of projects[2] funded by EU. Initially the white paper discusses a number of requirements driven either from a business perspective (e.g., minimize the service provision time, support multi-tenancy) as well as from technical challenges (e.g. use different spectrum bands, operate in ultra-dense environments, support ultra-low delay services at the highest reliability etc). Based on these requirements, we have discussed in detail what is the implication on the overall 5G network including the functional architecture and the physical deployments. Also we elaborate on a number of technical enablers that are expected to help 5G networks meet their goals.

More specifically, the white paper identifies the important role that some general design principles are expected to play in 5G networks like the support of slicing, the CP/UP split and the flexible placement of functions to support different use cases. A crucial enabler to achieve these functions is the use of softwarization and programmability that are expected to cater for the appropriate level of flexibility in 5G networks. Also, the use of SDN and NFV will play an important role, although it is clear that their applicability in the different domains (access, backhaul and core) is expected to have to different weight and different implementations. In such a flexible environment topics like network management and security need to be revisited and new mechanisms need to be introduced.

Also 5G networks will support a heterogeneous environment not only in terms of radio access technologies, but also different numerologies in the 5G air interface variances. Topics like the inter-working with LTE-A or the support of different functional split in C-RAN or D-RAN environments, the convergence of fixed and mobile networks call also for improvements in the existing architectures. Finally, 5G services will require improved orchestration schemes among different administration domains, as well as the full exploitation of cloud facilities at different aggregation points.

The above list of findings is the outcome of the first year of research activities by the 5G-PPP Initiative Projects. The life-time of these projects is typically for two years or more, thus this white paper will be updated during the next year capturing the commonalities among these projects, the different trends researchers are investigating and the advantages and disadvantages of their solutions.

---

[2] 5G PPP Phase I Projects- https://5g-ppp.eu/5g-ppp-phase-1-projects/

List of Contributors

Name	**Company / Institute / University**	**Country**
Editorial Team		
Overall Editors		
Redana Simone	Nokia Bell Labs 5GPPP Architecture WG Chair	Germany
Kaloxylos Alexandros	Huawei Technologies Dusseldorf GmBH 5GPPP Architecture WG Vice-Chair	Germany
Section 2 Editor		
Kaloxylos Alexandros	Huawei Technologies Dusseldorf GmBH	Germany
Section 3 Editors		
Galis Alex	University College London	UK
Rost Peter	Nokia Bell Labs	Germany
Section 4 Editors		
Marsch Patrick	Nokia Bell Labs	Poland
Queseth Olav	Ericsson	Sweden
Section 5 Editor		
Jungnickel Volker	Fraunhofer Heinrich Hertz Institute	Germany
Section 6 Editors		
Chen Tao	VTT	Finland
Chun Fang	EICT	Germany
Section 7 Editor		
Hans van der Veen	NEC	Germany
Contributors and Reviewers		
Alcaraz-Calero Jose M.	University of the West of Scotland	U.K.
Assem Haytham	IBM	Ireland
Bartzoudis Nikolaos	CTTC	Spain
Bianchi Giuseppe	University of Rome	Italy
Buda Teodora Sandra	IBM	Ireland
Bulakci Ömer	Huawei Technologies Dusseldorf GmBH	Germany
Camps Daniel	i2CAT	Spain
Chen Tao	VTT	Finland
Chun Fang	EICT	Germany
Costa Xavier	NEC	Germany
Da Silva Icaro Leonardo	Ericsson	Sweden
De la Oliva Antonio	UC3M	Spain
Demestichas Panagiotis	WINGS ICT Solutions	Greece

Doll Mark	Nokia Bell Labs	Germany
Erez Biton	Nokia	Germany
Escaola Eduard	I2cat	Spain
Estevez David Gutierrez	Samsung	U.K.
Fontes Francisco	Altice Labs, S.A	Portugal
Foteinos Vassilis	Wings ICT Solutions	Greece
Galis Alex	University College London	UK
Garcia Diego Lopez	TID	Spain
Guerzoni Riccardo	Huawei Technologies Dusseldorf GmBH	Germany
Gutierrez Teran Jesus	IHP	Germany
Jungnickel Volker	Fraunhofer Heinrich Hertz Institute	Germany
Kaloxylos Alexandros	Huawei Technologies Dusseldorf GmBH	Germany
Karl Holger	University of Paderborn	Germany
Kuo Fang Chun	EICT	Germany
Mannweiler Christian	Nokia Bell Labs	Germany
Marsch Patrick	Nokia Bell Labs	Poland
Mullins Robert	TSSG	Ireland
Naslund Mats	Ericsson	Sweeden
Nikaein Navid	EUROCOM	France
Parker Mike	University of Essex, Colchester	U.K.
Payaro Miquel	CTTC	Spain
Queseth Olav	Ericsson	Sweden
Redana Simone	Nokia Bell Labs	Germany
Rost Peter	Nokia Bell Labs	Germany
Saavedra Andreas Garcia	NEC	Germany
Safjan Krystian	Nokia Bell Labs	Poland
Serrano Pablo	UC3M	Spain
Spapis Panagiotis	Huawei Technologies Dusseldorf GmBH	Germany
Teran Jesus Gutierrez	IHP Microelectronics Frankfurt/Oder	Germany
Tercero Miurel	Ericsson	Sweeden
Tzanakaki Anna	University of Bristol	U.K.
Vaishnavi Ishan	Huawei Technologies Dusseldorf GmBH	Germany
Vlacheas Panagiotis	WINGS ICT Solutions	Greece
Wang Qi	University of the West of Scotland	U.K.
Xu Lei	IBM	Ireland