Cognitive management of multi-service multi-tenant 5G mobile networks

Christian Mannweiler, Henning Sanneck
Nokia Bell Labs, Munich, Germany
{christian.mannweiler,henning.sanneck}@nokia.com

Abstract

Fifth generation (5G) mobile networks are expected to support diverse use cases beyond traditional voice and mobile broadband services. Efforts to connect anyone and anything have gained significant traction in several industries, among them automotive, logistics, manufacturing, and robotics. Exemplary use cases include eHealth, IoT (Internet of Things) applications, and V2X (vehicle-to-anything). Therefore, 3GPP LTE/EPC will have to evolve towards a flexible mobile network accommodating novel architectural principles, while at the same time maintaining backward compatibility. More specifically, in order to avoid deploying multiple physical networks addressing the respective use case, a single infrastructure needs to host multiple, logically separated dedicated networks, also referred to as network slices. This imposes new challenges on network management. While the objective is to manage each of the logical networks in a mostly isolated manner, the shared use of the underlying infrastructure will make "cross-slice" management functions mandatory. Further, the novel 5G mobile network architecture shall support legacy radio access technologies as well as novel radio interfaces, e.g., based on mm-wave or cm-wave transmissions. It should accommodate emerging processing paradigms such as mobile edge computing (MEC) and Cloud RAN (radio access network), while enabling flexible deployment patterns based on small, micro, and macro cells and allowing programmability to support very different requirements in terms of latency, robustness, reliability, and throughput. Cognitive network management is an important component to make such complex systems operable.

1 Introduction

Future mobile networks will be subject to manifold technical requirements with respect to throughput, latency, reliability, availability, as well as operational requirements such as energy-efficiency and cost-efficiency. These requirements result from an increasing diversity of services carried by the mobile network as well as novel application areas such as industrial communications, vehicular communications, or e-health [1]. In order to provide cost- and energy-efficient solutions, it is necessary to avoid dedicated physical mobile network deployments for each use case. Rather, one infrastructure should have the ability to host multiple logical networks at the same time, thus allowing multi-service and multi-tenancy deployments [2]. Hence, there is the need for a flexible, scalable, and inter-operable mobile network system. Flexibility is required to make sure that one mobile network instance can appropriately be customized to the network environment of a particular use case, e.g. radio access technology, transport network capacity, or access point density. Scalability is required because the actual quantitative technical requirements may differ significantly, e.g. while mobile broadband services have to accommodate a moderate number of subscribers per cell, massive machine-type communication services assume several hundreds of devices per cell. Additionally, inter-operability with existing deployments (backward compatibility) as well as novel technologies (future proofness), are required in order to efficiently use resources [3]. For example, different industries may follow their own standards which still need to be integrated into the same system such as 3GPP LTE and IEEE 802.11.

This paper outlines how the 5G NORMA mobile network architecture incorporates important principles from cognitive network management and software-defined mobile network control (SDM-C). It depicts the tight integration of programmable network management and control and how automation is exploited to reap the advantages of the so-called 'management-control continuum' [4] by enabling a flexible, scalable, and inter-operable mobile network. Efficiently managing network slicing allows for providing customized logical mobile network instances which suit each individual application without modifying the mobile network standard for each use case.

The remainder of this article is structured as follows: Section 2 describes the main enablers of a multi-service mobile network architecture. Section 3 elaborates on multi-tenancy and sharing approaches for 5G mobile networks. The role of an Umbrella and Service Management function for cognitive end-to-end management of network slices is discussed in Section 4. Conclusions and next steps are compiled in Section 5.

2 Key enablers for a novel mobile network architecture

In order to fulfill 5G objectives in terms of network flexibility, scalability, operability, and automation, 5G NORMA has identified three technical enablers, namely software defined mobile network control, adaptive composition and allocation of mobile NFs, and joint optimization of mobile access and core [6].

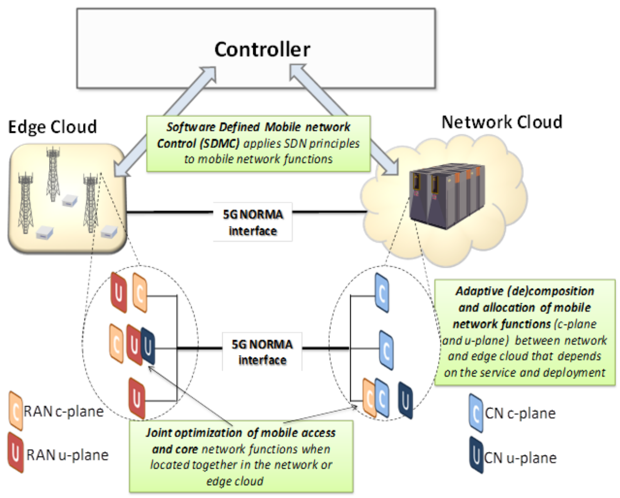

Figure 1: 5G NORMA concept and enablers

2.1 Software-defined mobile network control

5G NORMA architecture incorporates the Software-defined Mobile Network Control (SDM-C) concept which generalizes the software-defined networks (SDN) paradigm [5] towards mobile network functionality.

SDM-C resembles SDN in that mobile network functionality is split into (i) those functions that are being 'controlled'; and (ii) those functions that 'control' the overall network and are executed at the controller, and (iii) control application that incorporate the 'control intelligence'. However, the scope of 'controlled' functions is not limited to data plane functions but also includes control plane functions.

To realize SDM-C, where mobile network functionality is controlled by possibly external software located in the controller, it is essential to specify an interface between the controller and the mobile NFs that (i) is standardized and supported by all deployed equipment, and (ii) provides sufficient flexibility to obtain the desired behavior of 'controlled' functions. In this way, SDM-C effectively provides network programmability capabilities, possibly allowing third parties, e.g., virtual operators, vertical market players, or OTT providers, to modify network functions on-demand.

Indeed, the control of wireless networks comprises many functions, possibly including channel selection, scheduling, modulation and coding scheme (MCS) selection, and power control. With a software-defined approach, a large subset of these functions could be performed by a programmable central control, which provides very important benefits for the operation of the mobile network. The key advantages resulting from incorporating the proposed approach include the following:

- **Flexibility**: A current problem for mobile network operators is the high amount of capital and operational expenditures of their networks independent of the actual traffic load and service usage. By means of SDM-C, operators are able to fit the network to their needs by re-programming the controller and thus reducing costs, while also enhancing reliability through higher levels of automation.
- **Unified umbrella management**: Adopting a logically centralized control and management unifies heterogeneous network technologies and provides an efficient umbrella management and control of heterogeneously deployed networks, reflecting evolving traffic demands, enhanced mobility management and considering dynamic radio characteristics.
- **Programmability**: Allowing third parties to acquire network resources on-demand satisfying their individual SLAs. In addition, programmability can enhance the user perceived QoE by customizing the network function accordingly.
- **Enabling new services**: By modifying the behaviour of applications that run on top of the SDM-C controller's northbound interface, many new services that were not included in the initial architecture design can be enabled by modifying the network behavior and adapting its capabilities for the introduction of new services within few hours instead of weeks **Error! Reference source not found.**.
- **Performance**: By adapting the functions such as scheduling or channel selection to the specific needs of the applications or the scenario, significant performance gains can be achieved. For instance, the controller has a global view of the network, which allows for optimizing the resource allocation and scheduling across multiple base stations (BSs).

2.2 Adaptive (de)composition and allocation of mobile network functions

Service flexibility is supported by decomposing the mobile NFs, including access and core functions, which are usually associated to a network element and adaptively allocate them to the edge cloud or central cloud, depending on

- the specific service and its requirements, e.g., bandwidth and latency;
- the transport network capabilities, e.g., available network capacity and latency.

The adaptive composition and allocation of NFs enables several advantages. If service requirements and backhaul capacity are sufficient to allow for centralizing the functionality in the central cloud, better scalability and pooling gains can be obtained from moving major parts of the functionality to the cloud. In contrast, if services impose specific constraints that require moving part of the functions to the access, or backhaul constraints do not allow for fully centralizing mobile network functionality, then gains can be obtained by using a fully or partially distributed configuration. Achievable benefits include lower latencies, autonomous operation of edge clouds, or offloading the backhaul and the central cloud. Finally, there is no need to define a single general purpose function per task, e.g., forward error correction, link adaptation, or scheduling, that is suitable for all physical deployments and/or services, but instead multiple different functions can exist, each one optimized to its specific deployment scenario and supported services. Hence, this optimization may allow for providing multiple lightweight and possibly stripped-down versions of a function compared to a single complex multi-purpose function.

2.3 Joint orchestration of mobile access and core

A joint orchestration aims to overcome the drawback of static distribution of mobile access and core NFs into specified infrastructure entities or network elements. Thus, all data forwarding and data processing functions across mobile access and core are considered jointly for fulfilling service and architecture requirements.

5G networks need to support new use cases, network deployment scenarios, and business models in addition to those of current mobile broadband access and IP connectivity. Therefore, the 5G network architecture must allow for: (i) flexible structuring or restructuring of the network with support of various RATs and inter-RAT interworking between 5G and legacy technologies; and (ii) flexible capability to segregate mobility management from service control, i.e. packet forwarding and processing, across different mobility control areas or administrative domains. It is practically required to distinguish functions that are executed closer to the access point on bare metal or edge cloud, and those functions which are executed more centrally in the central cloud. So far, there exists a logical split between radio access and core network which basically enables an independent evolution of both, it allows for integrating different RATs, and it enables multi-vendor interoperability. These characteristics are maintained by the 5G NORMA architecture.

A key technology to enable the joint optimization across different domains is software-defined mobile network control and orchestration and the smart functional decomposition of NFs. This alleviates a problems that is experienced in current standards where static function splits are too restrictive for supporting novel services or services with very divergent requirements efficiently. The joint optimization across different network domains herein focuses on the following areas:

- Providing on-demand adaptive NFs dedicated and optimized for specific services;
- Providing optimized Quality of Service/Experience (QoS/QoE) support with flexible aggregated service flow, i.e., enhanced bearer service model, in-service-flow QoS differentiation and multi-connectivity;
- Providing enhanced support for network slicing and multi-tenancy; and
- Providing enhanced support for mobility, load-balancing, and resource management.

3 Multi-tenancy in mobile networks

In contrast to IT environments that largely rely on homogenous hardware, e.g., x86 processor architectures for computing resources or quasi-standard database technologies for persistent storage, mobile networks exhibit a far greater diversity in terms of utilized hardware and software technologies. As a consequence, abstraction by means of virtualization is less common, particularly in the radio access network (RAN). Since virtualization serves as the main enabler for multi-tenancy and network slicing by instantiating additional VNFs for the given network instance, sharing of non-virtualized resources and functions become critical issues to be solved for truly end-to-end network slicing.

3.1 Resource sharing

Mobile networks rely on diverse hardware infrastructure resources. They include both general purpose and specialized hardware that comprise memory, compute, storage, networking, and other fundamental capabilities. Examples include hardware based on x86 architecture, non-x86 instruction set architectures (e.g., reduced instruction set computer (RISC) and ARM), programmable, purpose-built hardware exhibiting a tight coupling between hardware and software systems (e.g., DSPs, FPGAs), and non-programmable, purpose-built hardware for a dedicated processing function, exhibiting very limited (or no) configurability (e.g., radio frequency (RF) components and associated spectrum resources).

For resources that are eligible to virtualization, sharing can implemented fairly easily. Virtualized resources can be partitioned into (almost) arbitrary portions and assigned to a resource consuming entity, usually a virtualized network function. In case a given VNF is required in two separate logical network instances, it can be instantiated multiple times, possibly with even different configuration. For physical network functions (e.g., a monolithic eNodeB), which are characterized by a tight coupling of software and hardware systems (cf. Figure 2), this approach cannot be followed. Joint, multi-tenant usage of PNFs therefore requires alternative approaches, such as sharing policies and procedures controlled by the element manager (EM) and SDM-C and executed by, e.g., the radio scheduler. From a management and control perspective, it therefore becomes necessary to differentiate between shared and dedicated network functions.

3.2 Shared and dedicated network functions

Network slices, which are defined as logical end-to-end mobile networks [2], operate on top of a (partially) shared infrastructure, as outlined in Sec. 3.1. Further, they are composed of shared and dedicated as well as physical and virtualized network functions. For the control of shared functions, 5G NORMA has introduced a modified SDM controller, referred to as SDM-X ("SDM Coordinator"). As shown in Figure 2, SDM-C and SDM-X

- control the (virtualized or physical) network function,
- do not control and orchestrate underlying network function virtualization infrastructure (NFVI) resources.

In case of VNFs, this implies that SDM-C/X only control software running inside a VM/container (also referred to as the mobile function or application logic, cf. Figure 3). The virtualization container of the VNF (VM, Docker container, etc.) is controlled by the network function virtualization (NFV) management and orchestration (MANO) entities, i.e. Virtual Infrastructure Manager (VIM) and VNF Manager (VNFM) [7]. This ensures a consequent split between management/control of mobile network functions on the one hand and NFVI resources on the other. For PNFs, which exhibit a tight coupling of hardware and software, SDM-C/X control the integrated hardware/software system of the network functions.

SDM-C directly interfaces with VNFs dedicated to a particular network slice. In order to access shared functions, e.g., for including them in a particular network slice; an SDM-C needs to use the according interfaces of the SDM-X, which therefore also coordinates between multiple SDM-C instances and the associated network slices. Table 1 summarizes the control scope of SDM-X and SDM-C, depending on whether a network function is physical or virtualized a whether it is shared or dedicated. Typically, PNFs are rather shared among network slices while VNF instances are usually dedicated to a particular network slice.

Table 1: Management and control scope of SDM-C and SDM-X

	Shared functions	***Dedicated functions***
PNF (network element)	**SDM-X** Control of integrated software/hardware system	**SDM-C** Control of integrated software/hardware system
VNF	**SDM-X** Control of mobile function software only	**SDM-C** Control of mobile function software only

As depicted in Figure 2 and similar to the SDN principle, both SDM-X and SDM-C expose northbound interfaces (NBI) to their respective control applications. Control applications can be roughly grouped into three categories:

- **Network management and orchestration applications**: Depending on the scope of the application, element management, OSS (operation support system), or service orchestration entities can act as control applications. The OAM type 1 interface (Itf-S) is replaced an SDM controller and the according southbound interface (SBI, Itf-S*, see Figure 3), e.g., to enforce FCAPS management decisions at the respective network function/network element.
- **Control plane applications**: By splitting the c-plane functionality of a mobile network into a controller and several control application, the complexity and heterogeneity of underlying resources and controlled functions can be shielded from the control logic, which is part of the application. The NBI constitutes a single interface towards the control application for programming the network behavior. If the control logic shall be changed, in many cases it is sufficient to only re-configure or exchange the application, not the entire c-plane function. These functions have to come with according 'plug-ins' for the controller so that additional protocols and functionality can be supported on the controller's SBI.
- **3rd party applications**: Tenants operating their own logical mobile network (virtual mobile network operators (MVNO), businesses from vertical sectors, over-the-top (OTT) service providers) can get the possibility to control selected parts of their network slice themselves, depending on their expert level and the MNO's willingness to authorize 3rd parties for network control.. This can range from very limited control with only few configuration choices to a nearly full control of network functionality.

Besides realizing the network programmability paradigm, SDM-C and SDM-X are an efficient means to realize the management-control continuum.

4 Umbrella & service management

A 5G Umbrella & Service Management function serves two major objectives: (i) it shall hold an end-to-end management and orchestration view of a network slice and (ii) for 3rd parties, it should serve as the "entry point" into the telecommunications service provider's administrative domain in order to request the commissioning and operation of a network slice.

As such, it must employ highly automated processes for the configuration, performance, fault, and lifecycle management of end-to-end network slices. Further, it works as an umbrella system for evolved Operation, Administration, and Maintenance (OAM) as well as NFV orchestration systems. Figure 5 depicts an according functional design of a Network Management and Orchestration System. It includes two major pillars (OAM functions and NFV functions) and the end-to-end service management on top. OAM functions, such as element manager (EM), manage the mobile function (MF) logic, which is the monolithic network element (NE) in case of physical network functions and the network function (NF) software in case of virtualized NFs. The NFV functions orchestrate NFV infrastructure and platform resources and their mapping to VNFs.

The individual functions, end-to-end network slice management, evolved OSS, and service orchestration are depicted in Figure 3 and described in more detail in the following subsections.

4.1 End-to-end network slice management

The primary task of end-to-end network slice management is the composition of VNFs and PNFs for forming a logically self-contained mobile network instance. Operated by the MNO, it exposes interfaces for network slice creation, operation, and termination request from internal and external stakeholders (e.g., vertical sectors, OTT providers), thereby accommodating different expert levels.

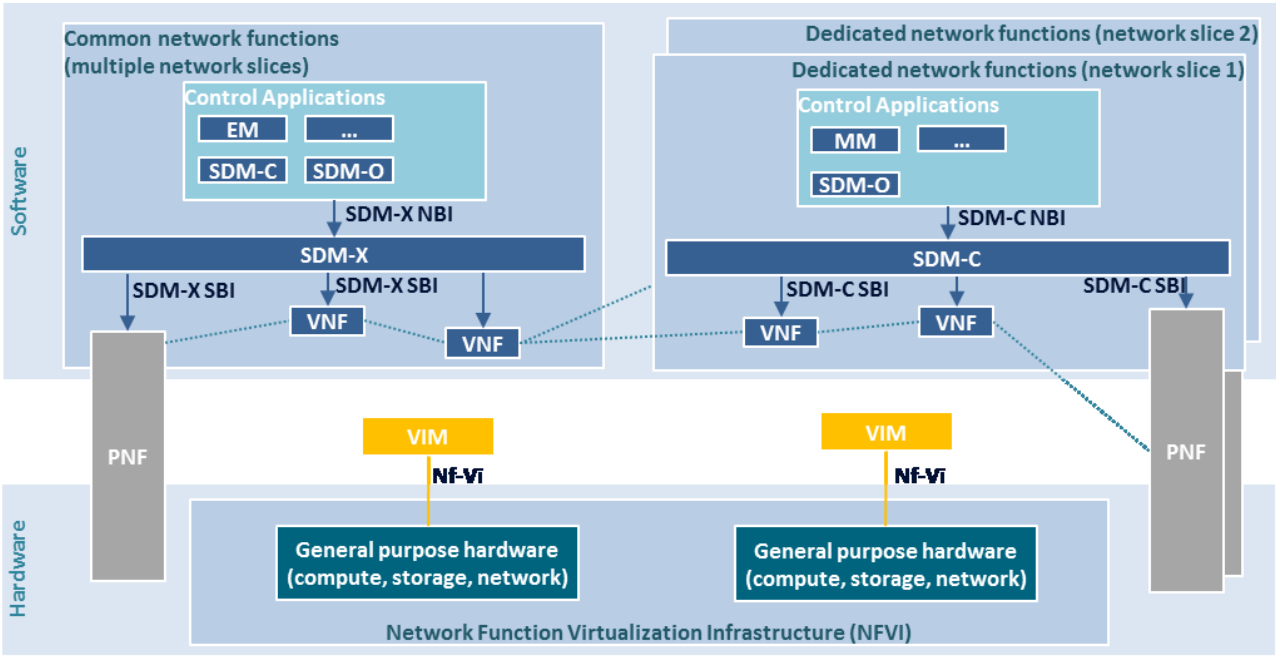

Figure 2: Management and Control of VNFs and PNFs

For example, a creation request can consist of direct network slice selection or, alternatively, of information on service requirements such as service level agreements and key quality indicators that describe the required characteristics of the requested telecommunications service. By (partially automated) means of business and policy decisions, network slice management maps a request to a set of network functions that, as a whole, form the network slice blueprint. The output is an annotated network slice instance that carries information

- on dedicated and shared NFs from both the control plane and the user plane,
- resource reservation and prioritization policies in case of shared functions,
- on the set of SDM controllers and coordinators as well as authorized control applications,
- for differentiating virtual and physical NFs included in the network slice, and
- further NF configuration information.

The annotation is also used to delegate ('dispatch') the management of the individual functions to the appropriate management and orchestration pillar, i.e. network management for PNFs as well as the application logic of VNFs and orchestration functions for the lifecycle management of VNFs. The latter consists of managing the virtualization container (e.g., virtual machines) that serves as execution platform ('VNF-PLAT') of the VNF software, see Figure 3.

4.2 Network management

In order to efficiently operate multiple logical mobile network instances independent from each other on top of the same infrastructure, current network management systems have to evolve towards a higher degree of automation and cognition in the execution of FCAPS management for mobile NFs. First, this includes the capability of differentiating between dedicated functions and resources of the network that can be managed independently, i.e., without impacting other network instances ('network slice-specific management'), and shared functions and resources, where the mutual impact has to be coordinated and prioritization policies and rules have to described and enforced ('cross-network slice management'). Second, increasing requirements on the 5G network in terms of adaptability and reactiveness will blur the demarcation line between management and control tasks, leading to a tighter integration of the formerly separated planes and a more frequent execution of self-configuration and self-optimization decisions. In 5G NORMA, these two developments are addressed by the software-defined mobile network coordination (SDM-X) entity, which executes a programmable control over the network function via the SBI and can be instrumented by both management and control plane SDM applications.

4.3 Service and resource orchestration

In the NFV orchestration domain, dedicated entities perform service orchestration and resource orchestration. While the former instantiates, scales, and terminates the virtualized network service as a whole as well as the individual functions, the latter maps the resource allocation requests to the available NFVI resources. The resource orchestrator therefore hides the infrastructure's performance characteristics in terms of capacity, throughput, latency, processing capabilities, etc. as well as their geographical and logical distribution across centralized data centers, aggregation sites, and local edge clouds from the service orchestrator.

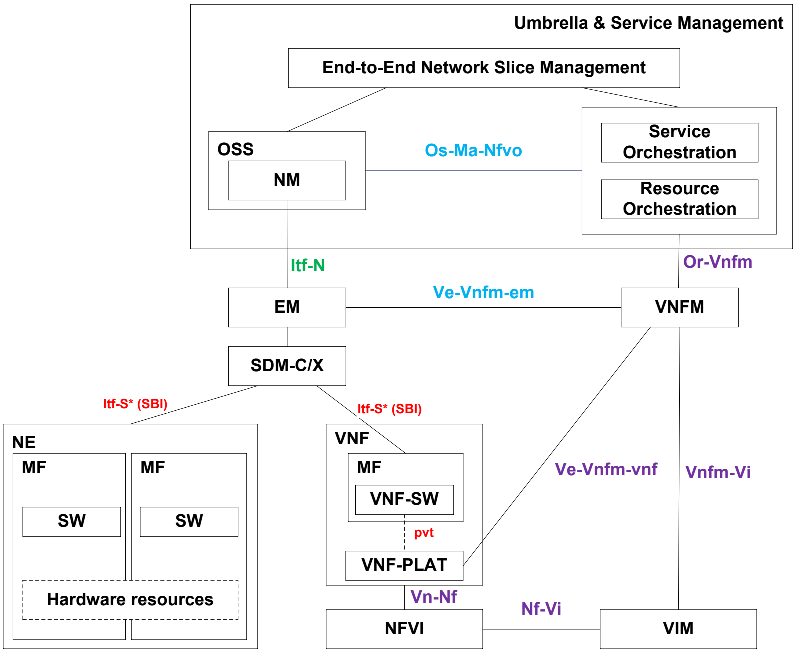

Figure 3: Network Management and Orchestration System

A further task is inter-slice resource orchestration. Local edge cloud deployments, unlike central data centers, are not expected to provide an abundance of NFVI resources. Resource orchestration therefore also controls the virtual resource sharing among slices, i.e., it prioritizes the scaling of network slices. It executes policy-driven decisions to solve conflicting requirements between slices for sharing virtual resources and links, e.g. rules based on different slices' priority policies. In the (rather rare) case that network slices share VNFs (e.g., subscriber databases), it coordinates the access to such functions among network slices.

5 Conclusion

This paper has outlined the architecture of a future-proof cognitive umbrella and service system for managing and orchestrating logical mobile network instances ('network slices') that are deployed on a common infrastructure and share parts of the mobile network functionality. It has depicted the major technical enablers for a highly automated and programmable system, among them software-defined mobile network control and coordination. The specifics of multi-tenancy in mobile networks have been described an according architectural solutions have been provided. In particular, the architecture realizes the management-control continuum for next-generation network management. Finally, the authors showed how customized telecommunication services and limited service configuration possibilities can be provided to external parties, such as, MVNOs, OTT service providers, or business from vertical sectors.
Further work of the 5G NORMA projects includes the further specification of functions, processes, interfaces, and protocols to realize multi-tenant multi-service mobile networks.

Acknowledgment

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

References

[1] NGMN Alliance, "NGMN 5G White Paper", Feb. 17, 2015; https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf
[2] NGMN Alliance, NGMN Network Slicing 'Description of Network Slicing Concept https://www.ngmn.org/uploads/media/160113_Network_Slicing_v1_0.pdf
[3] EU H2020 5G NORMA, "D2.1: Use cases, scenarios, and requirements," September 2015
[4] TM Forum, "IG1118 OSS / BSS Futures: Preparing the Future Mode of Operation R14.5.1 Exploratory Report", 2014
[5] Software-Defined Networking: The new Form for Networks, ONF White Paper, April 13, 2012 https://www.opennetworking.org/images/stories/downloads/sdn-resources/white-papers/wp-sdn-newnorm.pdf
[6] EU H2020 5G NORMA, "D3.1: Functional Network Architecture and Security Requirements" December 2015
[7] ETSI GS NFV-MAN 001, "Network Functions Virtualisation (NFV); Management and Orchestration", V1.1.1 (2014-12), December 2014