Contents

Acronyms

AN Access Network. 7

BH Backthaul. 4

CP Control Plane. 7

E2E End to End. 4

FH Fronthaul. 4

IaaS Infrastructure-as-a-Service. 4

MANO Management and Orchestration. 6

MNO Mobile Network Operators. 3

MTA Multi-Tenancy Application. 7

NBI Northbound Interface. 6

NF Network Functions. 7

NGMN Next Generation Mobile Networks. 3

NS Network Services. 4

NSIL Network Service Instance Layer. 3

SBI Southbound Interface. 7

SDN Software Defined Network. 4

SIL Service Instance Layer. 3

UP User Plane. 7

VI Virtual Infrastructures. 4

VNF Virtual Network Functions. 4

1 Introduction

Mobile networks are a key element of today's society, enabling communication, access and information sharing. Moreover, traffic forecasts predict that the demand for capacity will grow exponentially over the next years, mainly due to video services. However, as cellular networks move from being voice-centric to data-centric, operators' revenues are not able to keep pace with the predicted increase in traffic volume. Such pressure on operators' return on investment has pushed research efforts toward designing for 5G novel mobile network solutions able to open the door for new revenue sources. In this context, the network slicing paradigm has emerged as a key 5G disruptive technology addressing this challenge. Network slicing for 5G allows Mobile Network Operators (MNO) to open their physical network infrastructure platform to the concurrent deployment of multiple logical self-contained networks, orchestrated in different ways according to their specific service requirements; such network slices are then (temporarily) owned by tenants. The availability of this vertical market multiplies the monetization opportunities of the network infrastructure as new players may come into play (e.g., automotive industry, e-health) and an higher infrastructure capacity utilization can be achieved by admitting network slice requests and exploiting multiplexing gains. With network slicing for 5G networks, different services (e.g., automotive, mobile broadband, or haptic Internet) can be provided by different network slice instances. Each of these instances consists of a set of virtual network functions that run on the same infrastructure with a tailored orchestration. In this way, very heterogeneous requirements can be provided on the same infrastructure, as different network slice instances can be orchestrated and configured separately according to their specific requirements. Additionally, this is performed in a cost-efficient manner as the different network slice tenants share the same physical infrastructure. A network slice is defined by Next Generation Mobile Networks (NGMN) as "a set of network functions, and resources to run these network functions, forming a complete instantiated logical network to meet certain network characteristics required by the Service Instance(s)." According to NGMN, the concept of network slicing involves three layers, namely, (i) service instance layer, (ii) network slice instance layer, and (iii) resource layer. The Service Instance Layer (SIL) represents the end user and/or business services provided by the operator or the third-party service providers, which are supported by the Network Service Instance Layer (NSIL). The NISL is in turn supported by the resource layer, which may consist of the organic resources such as compute, network, memory, storage, etc., or it may be more comprehensive as being a network infrastructure, or it may be more complex as network functions. Figure 9.1 depicts this concept where the resources at the resource layer are dimensioned to create several sub network instances, and network slice instances are formed that may use none, one, or multiple sub network instances. Figure 1 depicts this concept where the resources at the resource layer are dimensioned to create several sub network instances, and network slice instances are formed that may use none, one, or multiple sub network instances. The end goal of network slicing in 5G mobile networks is

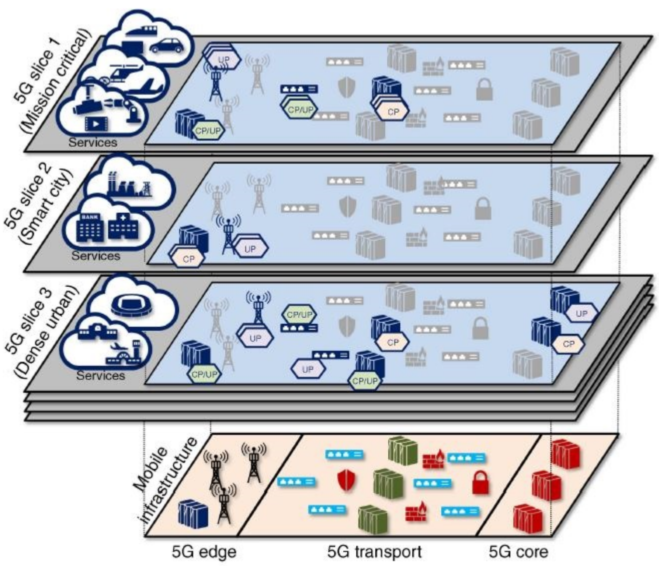

Figure 1: Example of network slicing in 5G.

to be able to realize End to End (E2E) network slices starting from the mobile edge, continuing through the mobile transport (Fronthaul (FH) /Backthaul (BH)), and up until the core network (CN). The allocation of a slice involves the selection of the required functions, their constrained placement, the composition of the underlying infrastructure, and the allocation of the resources to fulfill the services' requirements, for example, bandwidth, latency, processing, resiliency. We consider two main network slicing services that enable different degrees of explicit control and are characterized by different levels of automation of the mobile network slices management: 1) The provision of Virtual Infrastructures (VI) under the control and operation of different tenants-in line with an Infrastructure-as-a-Service (IaaS) model[1], that is, creation of a network slice instance. 2) The provision of tenant's owned Network Services (NS), that is, creation of a service instance. In the former service, the deployment of a mobile network deals with the allocation and deallocation of VIs.The logical entities within a VI encompassing a set of compute and storage resources are interconnected by a virtual, logical network (i.e., virtual nodes are interconnected by virtual links over the substrate network). The Vis can be operated by the tenant via different Software Defined Network (SDN) control models. In the latter, NS are instantiated directly over a shared infrastructure, and as a set of interrelated Virtual Network Functions (VNF) connected through one or more VNF forwarding graphs. Multi-tenancy is an characteristic that can be applied to both kinds of services, guaranteeing separation, isolation, and independence

[1]form of cloud computing that provides virtualized computing resources over the internet

between different slices coupled with the efficient sharing of the underlying resources for both VI and NS concepts. In this context, a tenant is a logical entity owning and operating either one or more VIs or one or more network services. A tenant can be associated with an administrative entity (e.g., mobile virtual network operators) or user of a given service (e.g., over-the-top service providers).

2 Architecture for Network Slicing

The necessary architecture involves the aspects of resource virtualization, virtual infrastructure, and network service management. The design proposed in Fig.

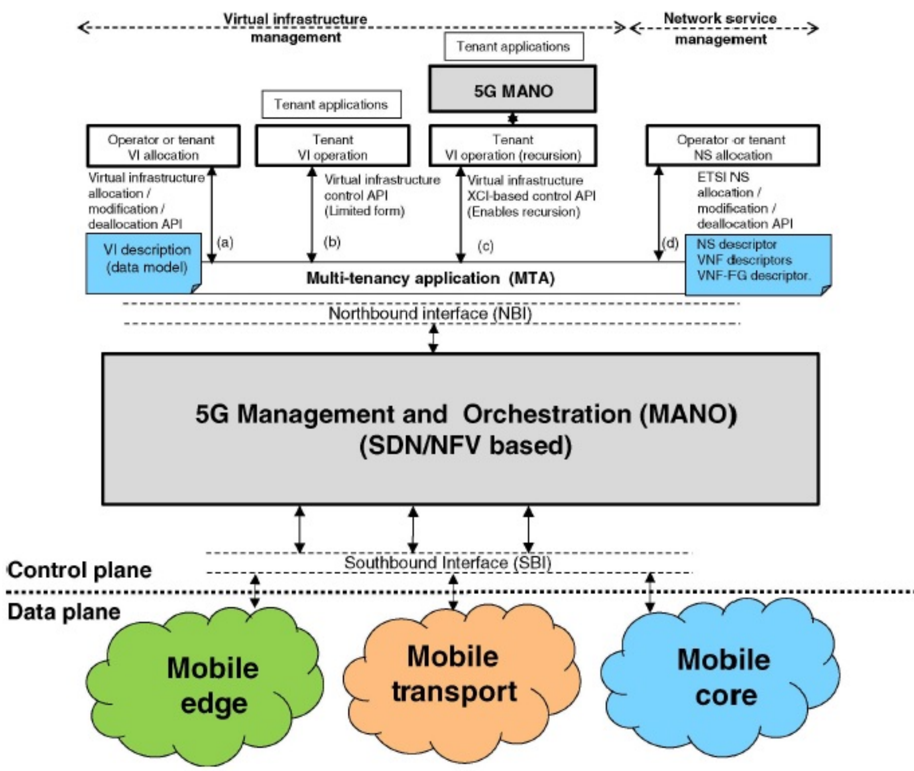

Figure 2: Architecture for network slicing.

2 follows the SDN principles of (i) data and control plane fully decoupled, (ii) control logically centralized, and (iii) applications having an abstracted view of resources and states. The data plane is the resource layer which includes mobile edge, mobile transport, and core. The infrastructure is composed of links, forwarding nodes (e.g., switches and routers), cloud nodes (e.g., data centers), and so on, comprising a set of network, computing, and storage resources. The control plane is divided into two layers: an application layer at the top and the 5G Management and Orchestration (MANO) platform below. The design of the MANO is based on the ETSI management and network orchestration framework with integrated SDN-based control. The MANO provides an abstracted view of available resources and states and control, and management functions to an ecosystem of applications, via a Northbound Interface (NBI). On the other

hand, the MANO is connected to the data plane elements via a Southbound Interface (SBI) to execute control and management functions (e.g., OpenFlow, SNMP, OVSDB) on the actual hardware components. With respect to the Multi-Tenancy Application (MTA), it implements the multi-tenancy support by coordinating and managing tenants access to a shared infrastructure, performing resource isolation between instances assigned to different tenants, and delivering multi-tenancy-related services, such as the allocation and operation of Vis, by means of dedicated APIs in cooperation with the data plane, enforcing this logical separation. As shown in Fig. 2 such APIs depend on the actual service: for the control of a VI or NS lifetime, instantiation, modification, and deletion.

2.1 Enablers and Design Principles

Future 5G networks will be built on the previous novel concepts that were not envisioned by the previous generation network architectures. The revolution provided by the introduction of software-defined networking and network function virtualization (NFV)[2] opens the door to a large list of possible applications recalling that the latter focuses primarily on optimization of the network services, instead the former to separate the control and forwarding plane for a centralized view of the network. The fundamental parts involved in the network slicing realization for the future 5G networks are now discussed.

2.1.1 Modularization

The evolution of mobile communication systems towards 5G was intentionally aiming at achieving architecture flexibility, heterogeneous accesses and vertical business integration, leveraging on NFV and SDN. To enable the design of logical architectures tailored to performance and functional requirements of different use cases, the principle of architecture modularization and network function decomposition was proposed at the earliest 5G research stages.Network Functions (NF)s are functional blocks that provide specific network capabilities to support and realize the particular service(s) each use case demands. Generally implemented as software instances running on infrastructure resources, NFs can be physical (a combination of vendor-specific hardware and software, defin- ing a traditional purpose-built physical appliance) and/or virtualized (network function software is decoupled from the hardware it runs on) [8]. In particular, conventionally monolithic network functions are proposed to be split into basic modules defined with proper granularity, both for the Control Plane (CP) and User Plane (UP), thus allowing the definition of different logical architectures via the interconnection of different subsets of CP and UP NFs. In the process of decomposing the NFs into basic modules, the distinction between NFs relating to the Access Network (AN) and core network emerged. To minimize the dependency of the 5G core on the access (and vice versa), and achieve the

[2] NFV and VNF are often used interchangeably

definition of a convergent network[3], providing connectivity via a multitude of accesses not only including cellular radio, a different AN/CN functional split and an interface model are necessary. Besides flexibility, the architecture modularization provides the essentials to support network slicing, as a network slice can be defined as an independent logical network shaped by the interconnection of a subset of NFs, composing both CP and UP, and which can be independently instantiated and operated over physical or virtual infrastructure.

2.1.2 Virtualization: SDN-VNF

Through technologies like software-defined networking (SDN) and network functions virtualization (NFV), network softwarization can provide the programmability, flexibility, and modularity that is required to create multiple logical (virtual) networks, each tailored for a given use case, on top of a common network. These logical networks are referred to as network slices. The concept of separated virtual networks deployed over a single network is indeed not new (e.g., virtual private networks, VPNs); however, there are specificities that make network slices a novel concept. We define network slices as end-to-end (E2E) logical networks running on a common underlying (physical or virtual) network, mutually isolated, with independent control and management, which can be created on demand. Such self-contained networks must be flexible enough to simultaneously accommodate diverse business-driven use cases from multiple players on a common network infrastructure

Future 5G networks will bring the concept of network programmability beyond what is now possible with SDN. While SDN splits routing and forwarding capabilities of a switch and reassigns the former to an SDN controller, this split between logic and agent should be performed for any NF, including the ones related to the CP. That is, the SDN principles are extended to all control and data layers as well as management functions usually deployed in mobile networks. The following three categories can be identified: (i) networking control functions (e.g., mobility and session management, and potentially QoS/QoE control); (ii) connectivity control functions (mainly packet forwarding or SDN-based packet forwarding); and (iii) wireless control functions (e.g., radio link adaptation and scheduling). There are many advantages of implementing selected wireless control functions in a way such that they are not bound anymore to specialized hardware, such as an LTE enhanced Node-B (eNB), but rather become independent pieces of software. Network functionalities are then performed by a (virtualized) programmable and logically centralized controller that abstracts and, thus, homogenizes different network technologies. Such a controller will make network slices programmable by controlling the topology and functionality of the service chains as well as resource control inside the network slices. Further, this approach implies to have a unique control point for the network: By operating a small number of such controllers, network operators reduce the complexity of the network management and control. Dense wireless

[3]Network convergence is the efficient coexistence of telephone, video and data communication within a single network.

networks, as envisioned in 5G, will benefit from this extension of the SDN approach. The control of mobility support schemes and adaptation to dynamic radio characteristics, such as the ones used in a multiple radio access technology Radio Access Technology (RAT) scenario, are performed by the controller that can employ especially tailored algorithms per network slice they are deployed in. Moreover, if needed, VNFs)can be deployed closely to the users (e.g., in a network slice supporting vehicular URLLC) reducing their experienced latency.

2.1.3 Orchestration

3 Network slicing

In order to create tenant- or service-specific networks, NGMN has proposed the concept of net- work slicing [1], as detailed in Chapter 8. While legacy systems host multiple telecommunication services, such as mobile broadband, voice, SMS, etc., on the same mobile network architecture, for instance composed of Long Term Evolution (LTE) radio access and the Evolved Packet Core (EPC), future 5G networks should also support shared or dedicated logical architectures customized to the respective telco or vertical services, such as enhanced mobile broadband (eMBB), vehicular com- munications, ultra-reliable low-latency communications (URLLC), and massive machine-type com- munications (mMTC), as introduced in Section 2.2. These services need very different KPIs that are hard to be fulfilled by legacy systems, as they are characterized by monolithic network elements that have tightly coupled hardware, software, and functionality. In contrast, future architecture leverages the decoupling of software-based network functions from the underlying infrastructure resources by means of utilizing different resource abstraction technologies. Furthermore, as explained in the previous paragraph, modularization will play a fundamental role. For instance, well-known resource sharing technologies such as multiplexing and multitasking, e.g., wavelength division multiplexing (WDM) or radio scheduling, can be advantageously complemented by softwarization techniques such as NFV and SDN. Multitasking and multiplexing allow sharing physical infrastructure that is not virtualized. NFV and SDN allow different tenants to share the same general-purpose hardware, such as commercial off-the-shelf (COTS) servers. In combination, these technologies allow building fully decoupled E2E networks on top of a common, shared infra- structure. Consequently, multiplexing will not happen on the network level anymore, but on the infrastructure level, as depicted in Figure 5-1, yielding better quality of service (QoS) or quality of experience (QoE) for the subscriber (as different slices will have tailored orchestration for a given service) as well as improved levels of network operability for the mobile service provider or mobile network operator. In principle, a network slice is a logical network that provides specific network capabilities and network characteristics and comprises NFs, computing and networking resources to meet the performance requirements of the tenants, for instance verticals. This comprises both radio access network (RAN) and CN NFs and, depending on the degree of freedom that a tenant may have, also the management and orchestration (MANO) components. A network slice may be dedicated to a specific tenant or partially shared by several tenants that have the same performance requirements but different security or policy settings. The decoupling between the virtualized and the physical infrastructure allows for the efficient scaling in, out, up or down of the slices, hence suggesting the economic viability of this approach that can adapt the used resources on demand. Network slices are created mostly with a business purpose: Following the 5G vertical markets para-5G atom. The 5G use cases are in the center. The layers , from the center out, represent the requirements of the 5G use cases , the concepts that will allow network operators to satisfy the requirements , the technologies that enable the

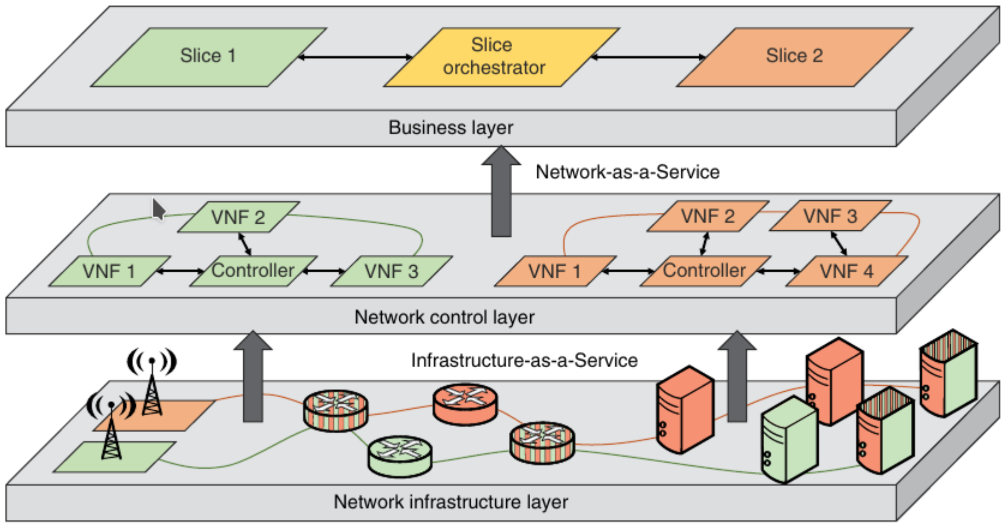

Figure 3: An example of a network-sliced architecture.

implementation of the concepts , and the novelties, that is, technologies that can be easily implemented due to softwarization and virtualization techniques.

3.1 Services

3.2 Example

3.3 Actual realizations

bbbbb

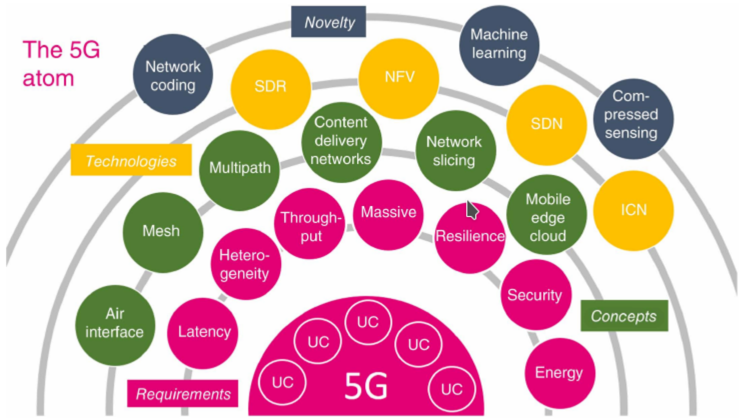

Figure 4: 5G atom rappresentation.

References

[1] Anwer Al-Dulaimi, Xianbin Wang, and I Chih-Lin. *5G Networks: Fundamental Requirements, Enabling Technologies, and Operations Management.* John Wiley & Sons, 2018.

[2] Carsten Bockelmann, Nuno Pratas, Hosein Nikopour, Kelvin Au, Tommy Svensson, Cedomir Stefanovic, Petar Popovski, and Armin Dekorsy. Massive machine-type communications in 5g: Physical and mac-layer solutions. *IEEE Communications Magazine*, 54(9):59–65, 2016.

[3] M Condoluci, R Trivisonno, T Mahmoodi, and X An. miot connectivity solutions for enhanced 5g systems.

[4] Jim Doherty. *SDN and NFV simplified: a visual guide to understanding software defined networks and network function virtualization.* Addison-Wesley Professional, 2016.

[5] Mohammad Asif Habibi, Bin Han, and Hans D Schotten. Network slicing in 5g mobile communication architecture, profit modeling, and challenges. *arXiv preprint arXiv:1707.00852*, 2017.

[6] Patrick Marsch, Ömer Bulakci, Olav Queseth, and Mauro Boldi. *5G System Design: Architectural and Functional Considerations and Long Term Research.* John Wiley & Sons, 2018.

[7] Peter Öhlén, Björn Skubic, Ahmad Rostami, Matteo Fiorani, Paolo Monti, Zere Ghebretensaé, Jonas Mårtensson, Kun Wang, and Lena Wosinska. Data plane and control architectures for 5g transport networks. *Journal of Lightwave Technology*, 34(6):1501–1508, 2016.

[8] Jose Ordonez-Lucena, Pablo Ameigeiras, Diego Lopez, Juan J Ramos-Munoz, Javier Lorca, and Jesus Folgueira. Network slicing for 5g with sdn/nfv: concepts, architectures and challenges. *arXiv preprint arXiv:1703.04676*, 2017.

[9] Petar Popovski, Kasper F Trillingsgaard, Osvaldo Simeone, and Giuseppe Durisi. 5g wireless network slicing for embb, urllc, and mmtc: A communication-theoretic view. *arXiv preprint arXiv:1804.05057*, 2018.

[10] Ahmad Rostami, Peter Ohlen, Kun Wang, Zere Ghebretensae, Bjorn Skubic, Mateus Santos, and Allan Vidal. Orchestration of ran and transport networks for 5g: an sdn approach. *IEEE Communications Magazine*, 55(4):64–70, 2017.