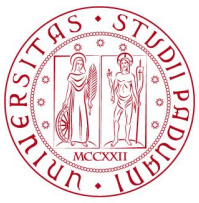

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

NETWORK SLICING: FUNDAMENTAL ELEMENTS, ARCHITECTURE AND TARGETS

Marco Rossanese 1177738
5G Systems - Prof. Stefano Tomasin
marco.rossanese@studenti.unipd.it

DEPARTMENT OF
INFORMATION
ENGINEERING
UNIVERSITY OF PADOVA

December 25, 2018

Contents

Acronyms

AMF Access and Mobility Management Function.

AN Access Network.

API Application Program Interface.

AUSF Authentication Server Functions.

BH Backthaul.

CN Core Network.

CP Control Plane.

CPF Control Plane Processing Functions.

E2E End to End.

EM Element Management.

eMBB enhanced mobile broadband.

EPC Evolved Packet Core.

FH Fronthaul.

HSS Home Subscriber System.

IaaS Infrastructure-as-a-Service.

IC Infrastructure SDN controller.

InP Infrastructure provider.

IoT Internet of Things.

KPI Key Performance Indicators.

MANO Management and Orchestration.

mMTC massive machine-type communications.

MNO Mobile Network Operators.

MTA Multi-Tenancy Application.

NAS Non-Access Stratum.

NBI Northbound Interface.

NF Network Functions.

NFV Network Functions Virtualization.

NFVI Network Functions Virtualization Infrastructure.

NGMN Next Generation Mobile Networks.

NMS Network Management System.

NS Network Services.

NSIL Network Service Instance Layer.

NSO Network Service Orchestrator.

ONF Open Networking Foundation.

OSS/BSS Operation/Business Support System.

PDCP acket Data Convergence Protocol.

QoE Quality of Experience.

QoS Quality of Service.

RAN Radio Access Network.

RLC Radio Link Control.

RO Resource Orchestrator.

RRC Radio Resource Control.

RRM Radio Resource Management.

SBI Southbound Interface.

SDN Software Defined Network.

SIL Service Instance Layer.

SLA Service Level Agreements.

SMF Session Management Function.

TC Tenant SDN controller.

UDM Unified Data Management.

UP User Plane.

UPF User Plane Processing Functions.

URLLC ultra-reliable low-latency communications.

VI Virtual Infrastructures.

VIM Virtual Infrastructure Manager.

VM Virtual Machines.

VNF Virtual Network Functions.

VNFM Virtual Network Function Manager.

1 Introduction

Mobile networks are a key element of today's society, allowing communication, access and information sharing. Moreover, traffic forecasts predict that the demand for capacity will grow exponentially over the next years, this is due mainly to video services.
Indeed, since cellular networks move from being voice-centric to data-centric, telecommunication operators are not able to keep pace with the predicted increase in traffic volume. Such pressure on operators has pushed research efforts toward designing 5G novel mobile network solutions able to open the frontier for new capabilities and revenue sources. In this context, the network slicing paradigm has emerged as a key 5G technology addressing this challenge.
Network slicing for 5G allows Mobile Network Operators (MNO) to open their physical network infrastructure platforms to the concurrent deployment of multiple logical self-contained networks, orchestrated according to their specific service requirements; such network slices are then (temporarily) owned by tenants. The availability of this vertical market[1] multiplies the money incoming opportunities of the network infrastructure as new opportunities come into play (e.g., automotive industry, e-health) and an higher infrastructure capacity utilization can be achieved by letting network slice requests and exploiting multiplexing gains.
With network slicing implementations, different services like automotive, mobile broadband, or tactile Internet can be provided by different network slice instances: each of these instances consist of a set of virtual network functions that run on the same infrastructure with a tailored orchestration. In this way, very different requirements can run on the same infrastructure, as different network slice instances can be orchestrated and configured separately according to their specific requirements. Moreover, 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:

[1] A vertical market is a group of companies that serve each other's specialized needs and do not serve a broader market, therefore it is mostly focused on meeting the needs of one specific industry.

1 Service Instance Layer (SIL);

2 Network Service Instance Layer (NSIL);

3 Resource Layer.

The SIL represents the business/end user services furnished by operators or the third party service providers, which are supported by the NISL. The NISL is set of necessary functions to run the instances and the resource layer, which consists of the resources such as compute, network, memory, storage.
The end goal of network slicing in 5G mobile networks is 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 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 service's requirements, for example, bandwidth, latency, processing, resiliency.
The two main slicing services needed to allow different degrees of control and management automation in NSIL and resource layers are: the Virtual Infrastructures (VI) under the control and operation of different tenants coherent with an Infrastructure-as-a-Service (IaaS) model[2] to split information from hardware; Network Services (NS) owned by tenants, that is actually creation of a service instance. The VIs can be operated by the tenant via different Software Defined Network (SDN), 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; both of them will be discussed later.
Multi-tenancy is an characteristic that it must be achieved to guarantee separation, isolation, and independence between different slices coupled with the efficient sharing of the underlying resources for both VI and NS concepts.

[2]Form of cloud computing that provides virtualized computing resources over the internet

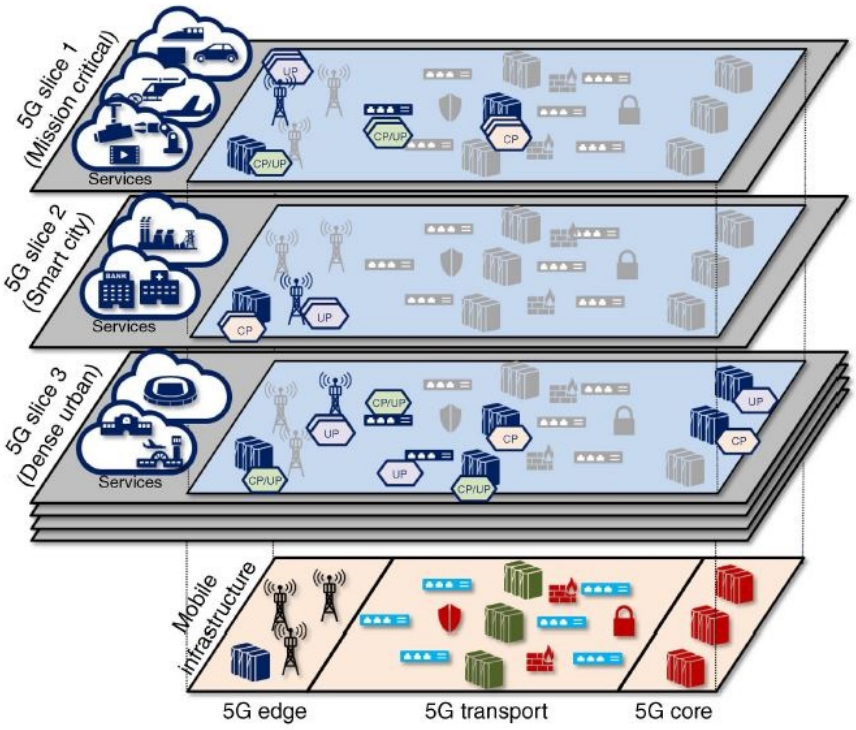

Figure 1: Example of network slicing in 5G. [1]

After this general overview this essay will treat accurately all the necessary fundamental components in order to fully understand how 5G network slicing is planned actually to be built.

2 Architecture for Network Slicing

Starting from how an architecture for network slicing is conceptually made, it will be explained what it should achieve and involve, that is, the aspects of modularization, resource virtualization, virtual infrastructure and network service management; they will be the main topics of this section. As antici-

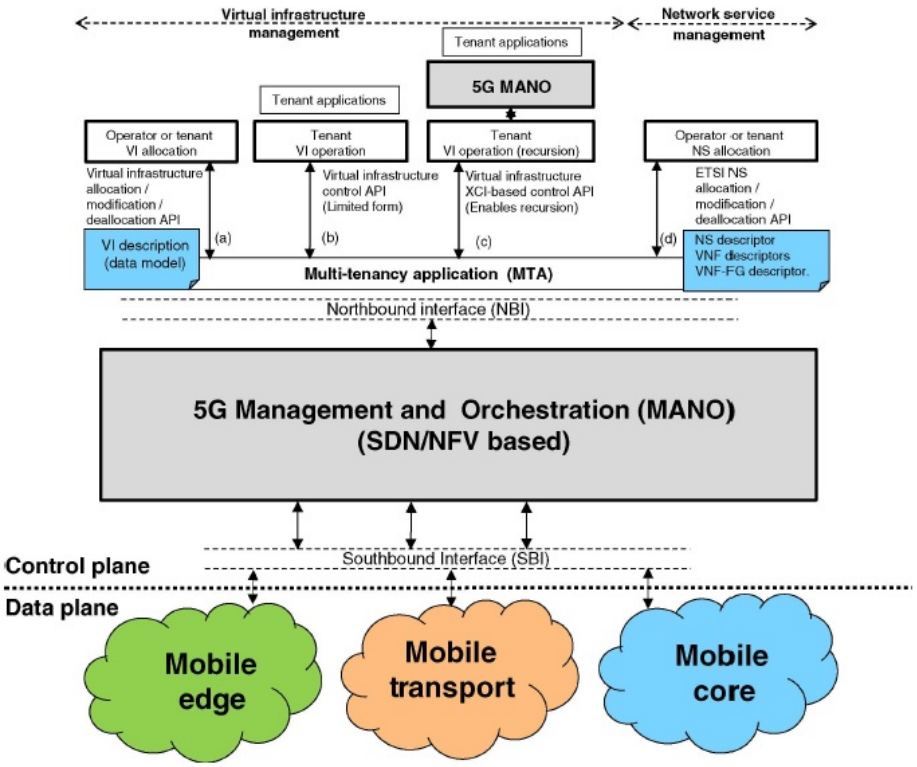

Figure 2: Architecture for network slicing. [1]

pated about virtualization of the infrastructure, the SDN design is proposed in Fig. 2 and follows the principles of:

- data and control plane fully decoupled;
- control logically centralized;
- applications having an abstracted view of resources and states.

The data plane is in practice the resource layer which includes mobile edge, mobile transport, and core. The infrastructure is composed of links, forwarding nodes as switches and routers, cloud nodes and 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 a view of available resources via a Northbound Interface (NBI), instead it is connected to the data plane elements via a Southbound Interface (SBI) to execute control and management functions (possible application to do the job are 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 Application Program Interface (API)[3] in cooperation with the data plane, remarking 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 novel concepts that were not concerned by the previous generation network architectures. The revolution provided by the introduction of SDN and Network Functions Virtualization (NFV)[4] 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 and Function Decomposition

The evolution of mobile communication systems towards 5G aims at achieving architecture flexibility, heterogeneous accesses and vertical business integration, exploiting NFV and SDN. The principle of architecture modularization and network function decomposition was proposed at the earliest 5G research stages in order to fulfill the above requirements.

[3]An application program interface is code that allows two software programs to communicate with each other.

[4]Although NFV and VNF are often used interchangeably, for the sake of clarity NFV is an overarching concept, while a VNF is building block of a NFV framework.

Network Functions (NF)s are the functional blocks that provide specific network capabilities to support and realize particular services. Generally implemented as software instances running on infrastructure resources, NFs can be physical, that is a combination of specific hardware and software, or virtualized, that is function service is decoupled from the hardware it runs on. In particular, conventionally "monolithic" network functions are proposed to be split into basic modules, both for the Control Plane (CP) and User Plane (UP), allowing the definition of different logical architectures by means of 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 CN emerged. To minimize the dependency of the 5G core on the access and viceversa and achieve the definition of a convergent network[5], 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

Virtualization is a key process for network slicing as it permits effective resource sharing among slices and it actually is the abstraction of resources using appropriate techniques. Abstraction means the representation of a resource in terms of attributes that match predefined selection criteria in order to simplify the its use and management.
The resources to be virtualized can be physical or already virtualized as well, supporting a recursive system with different abstraction layers. Just like server virtualization makes Virtual Machines (VM)s independent of the underlying physical hardware, network virtualization enables the creation of multiple isolated virtual networks that are completely decoupled from the underlying physical network and can safely run on top of it. The framework consists of three kinds of actors:

- Infrastructure provider (InP): owns and manages a given physical network. It offers resources in the form of WANs or data centers that are

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

virtualized and handed in through programming interfaces to a single or multiple tenants.

- Tenant: leases virtual resources from one or more InPs in the form of a virtual network, where the tenant can realize, manage, and provide network services to its users. A network service is a composition of NFs, and it is defined in terms of the individual NFs and the mechanism used to connect them.
- End user: consumes the services supplied by the tenant.

2.1.3 Orchestration

Orchestration is also a key process for network slicing and can be defined as the concept of bringing together and coordinating different things into a coherent whole. In a slicing environment, where the players involved are so different, an orchestrator is needed to coordinate heterogeneous network processes for creating, managing, and delivering services.
According to the Open Networking Foundation (ONF), orchestration is defined as "*the continuing process of selecting resources to fulfill client service demands in an optimal manner*". This means to use a policy that governs the orchestrator behavior, which is expected to fulfill all the service level agreements Service Level Agreements (SLA)s associated with clients who request services remembering that the available resources, the service demands, and optimization criteria may change in time.
Noteworthy is that orchestration is also referred to as the defining characteristic of an SDN controller. However, in network slicing, orchestration cannot be performed by a single centralized entity, not only because of the complexity, but also because it is necessary to maintain management independence and support the possibility of recursion.
A framework in which each virtualization actor has an entity performing orchestration functions seems more suitable to satisfy the above requirements.

2.1.3.1 Isolation

Strong isolation is a major requirement that must be satisfied to operate parallel slices on a common shared underlying substrate. The isolation must be understood in terms of:

- Performance: each slice is defined to accomplish particular service requirements, usually expressed in the form of Key Performance Indicators (KPI)s. Performance isolation is an E2E issue and has to ensure that service-specific performance requirements are always met on each slice, regardless of the congestion and performance levels of other slices.

- Security and privacy: attacks or issues occurring in one slice must not affect other slices. Moreover, each slice must have independent security functions that prevent unauthorized entities to have read or write access to slice-specific configuration, management, accounting information.

- Management: each slice must be independently managed as a standalone network.

To achieve isolation, a set of appropriate, consistent policies and mechanisms have to be defined at each virtualization level, following the recursion principle discussed earlier. The policies contain lists of rules that describe how different entities must be isolated. To properly realize the required isolation level an team play of both virtualization and orchestration is actually needed.

2.1.4 SDN: Software-Defined Network

In a SDN a network engineer can manage traffic from a centralized control console without handling individual switches in the network. The centralized SDN controller directs the switches to deliver network services wherever they are needed, this is actually a move away from traditional network architecture, in which devices make traffic decisions based on their configured routing tables.
The SDN architecture comprises, as previously explained and shown in Fig. 1, an intermediate control plane is present to dynamically configure and to abstract the underneath forwarding plane resources so as to deliver custom services to clients located in the application plane. This is indeed well aligned with 5G network slicing requirements since needs to satisfy a wide range of service demands. Then, the SDN is claimed to be an is an appropriate tool for supporting the key principles of slicing architecture.
The main SDN components are resources and controllers. For SDN, a resource is anything that can be utilized to provide services in response to client requests, that is infrastructure resources and NFs, but also network services in application of the recursion principle described before. A controller is a logically centralized entity instantiated in the control plane which operates SDN resources at runtime to deliver services in an optimal way. Therefore, it finds out tradeoffs among clients and resources, acting simultaneously as server and client via client and server contexts, respectively. Both contexts are conceptual components of an SDN controller enabling the server-client relationships:

- Client context: represents all the information the controller needs to handle and transmit to a given client. It comprises a Resource Group

and a Client support function: the former contains an abstract and customized view of all the resources that the controller offers to the client, through its NBIs, in order to deliver its service demands and ease its interaction with the controller; the latter instead contains all the necessary to support client operations, like policies on what the client is allowed to see and do.

- Server context: represents all the information the controller needs to interact with a set of underlying resources grouped in a Resource Group through one of its SBIs.

The process of transformation of the Resource Groups set accessed by server contexts to those defined in separate client contexts is not easy and requires the SDN controller to perform virtualization and orchestration functions.
When performing the virtualization function, the SDN controller carries out the abstraction and the aggregation/partitioning of the underlying resources. Thanks to virtualization, each client context provides a specific Resource Group that can be used by the client associated with that context to realize its services. Through orchestration, the SDN controller distributes in an optimal way the selected resources to such separate Resource Groups. Thanks to these two functions the fulfillment of the service demands from all clients is guaranteed while preserving the isolation among them.
The SDN architecture also includes an admin whose tasks consist of initializing and configuring the entire controller, including the creation of both server/client contexts and the selection of their policies.
The SDN architecture also supports slicing because the client context provides the complete abstract set of resources as a Resource Group and the supporting control logic that constitute a slice, including the complete collection of related client service attributes.
Another key aspect that makes SDN architecture an ideal solution for 5G slicing is recursion. Because of the different abstraction layers that the recursion principle enables, the SDN control plane can involve multiple hierarchically arranged controllers that extend the client-server relationships at several levels. According to that, it is clear that SDN can support a recursive composition of slices. This implies as consequence that the resources a given controller delivers to one of its clients in the form of a dedicated slice, i.e. client context, can in turn be virtualized and orchestrated by such a client in the case of being an SDN controller. In this way, the new controller can utilize the resources it accesses via its server contexts to define, scale, and deliver new resources and hence new slices to its own clients, which might also be SDN controllers.

2.1.5 NFV: Network Functions Virtualization

Although the SDN architecture described above gives a comprehensive view of the control plane functionalities enabling slicing, it lacks capabilities to efficiently manage the life cycle of network slices and its resources.
VNFs move individual network functions out of dedicated hardware devices into software that runs on commodity hardware. These tasks, used by both network service providers and businesses, include firewalls, domain name system, caching or network address translation and can run as virtual machines. Then, the NFV architecture is ideal to play this role, as it manages the infrastructure resources and orchestrates the allocation of such resources needed to realize VNFs and network services.
To benefit from the management and orchestration functionalities of NFV, the right cooperation between SDN and NFV is required. However, embracing SDN and NFV architectures into a common reference framework is not an easy task. ETSI presents a framework to integrate SDN within the reference NFV architecture. This framework incorporates two SDN controllers, one logically placed at the tenant and another at the InP level. The NFV architecture comprises the following entities:

- Network Functions Virtualization Infrastructure (NFVI): a collection of resources used to host and connect the VNFs. While SDN makes resource a generic concept, the current resource definition in the NFV framework comprises only the infrastructure resources.
- VNFs: software-based implementations of NFs that run over the NFVI.
- MANO: Performs all the virtualization-specific management, coordination, and automation tasks in the NFV architecture. The MANO framework comprises three functional blocks:
  - Virtual Infrastructure Manager (VIM): responsible for controlling and managing the NFVI resources.
  - Virtual Network Function Manager (VNFM): performs configuration and life cycle management of the VNF(s) on its domain.
  - Orchestrator: according to ETSI, it has two set of functions performed by the Resource Resource Orchestrator (RO) and Network Service Orchestrator (NSO), respectively. The RO orchestrates the NFVI resources across (potentially different) VIMs. The NSO performs the life cycle management of network services using the

capabilities provided by the RO and the (potentially different) VNFMs.

- Network Management System (NMS): framework performing the general network management tasks. Although its functions are orthogonal to those defined in MANO, NMS is expected to interact with MANO entities by means of a clear separation of roles. NMS comprises:
  - Element Management (EM): responsible for the configuration, accounting, performance, and security of a VNF.
  - Operation/Business Support System (OSS/BSS): a collection of systems and management applications that network service providers use to provision and operate their network services. In terms of the roles we considered earlier, tenants would run these applications.

The ETSI proposal includes two SDN controllers in the architecture. Each controller centralizes the control plane functionalities and provides an abstract view of all the components connectivity related it manages. These controllers are:

- Infrastructure SDN controller (IC): sets up and manages the underlying networking resources to provide the required connectivity for communicating the VNFs. Managed by the VIM, this controller may change infrastructure behavior on demand according to VIM specifications adapted from tenant requests.
- Tenant SDN controller (TC): instantiated in the tenant domain as one of the VNFs or as part of the NMS, this second controller dynamically manages the pertinent VNFs used to realize the tenant's network service(s). The operation and management tasks that the TC carries out are triggered by the applications running on top of it (e.g. the OSS).

Both controllers manage and control their underlying resources via programmable SBI, implementing protocols like OpenFlow, NETCONF, and I2RS. However, each controller provides a different level of abstraction. While the IC provides an underlay to support the deployment and connectivity of VNFs, the TC provides an overlay comprising tenant VNFs that define the network service a tenant independently manages on its slices. Therefore, the IC is not aware of the number of slices that utilize the VNFs or about the tenants which operate in the slices. On the other side, for the TC the

network is abstracted in terms of VNFs, without a knowledge of how those VNFs are physically deployed.
Despite their different abstraction levels, both controllers have to coordinate and synchronize their actions. Note that the service and tenant concept mentioned here can be extended to higher abstraction layers by simply applying the recursion principle.
Finally an overall description of the system is given by Fig. 3.

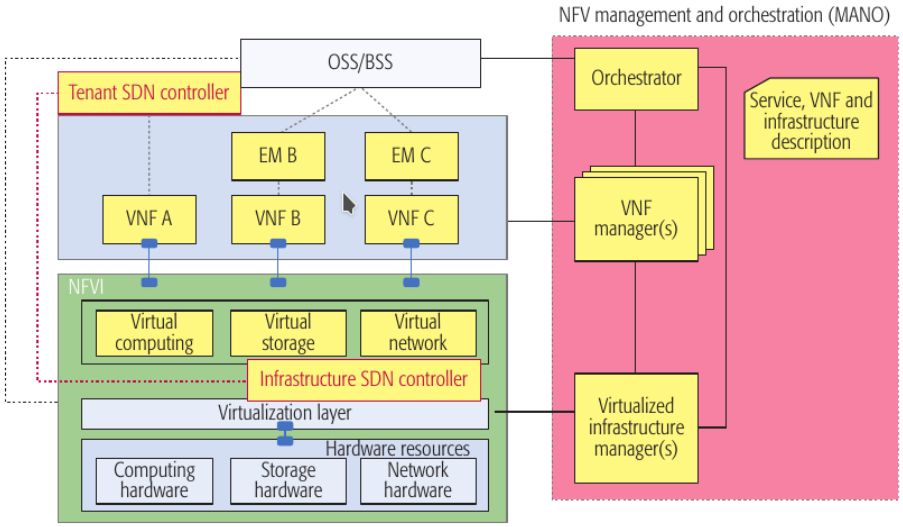

Figure 3: Integrating SDN controllers into the reference NFV architectural framework. [7]

3 Network Slicing

Now we have all the players to look inside what NGMN has proposed as the concept of network slicing: while legacy systems host multiple telecommunication services as mobile broadband, voice, SMS on the same mobile network architecture, for instance composed of Long Term Evolution 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 as enhanced mobile broadband (eMBB), vehicular communications, ultra-reliable low-latency communications (URLLC), and massive machine-type communications (mMTC). These services need different KPIs that are hard to be fulfilled by legacy systems which are characterized by monolithic network elements that have tightly coupled hardware, software, and functionality.
Future architectures must leverage on the decoupling of software-based network functions from the underlying infrastructure resources by means of utilizing different resource abstraction technologies. Furthermore exploiting modularization, resource sharing technologies such as multiplexing and multitasking (for instance wavelength division multiplexing or radio scheduling), can be complemented by softwarization techniques. Multitasking and multiplexing allow sharing physical infrastructure that is not virtualized. NFV and SDN allow different tenants to share the same general purpose hardwares.

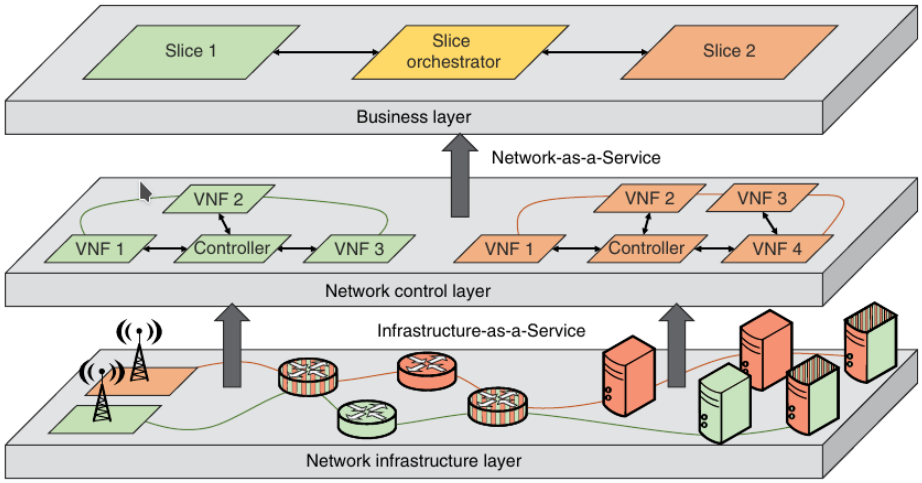

Figure 4: An example of a network-sliced architecture. [1]

In combination, these technologies allow building fully decoupled E2E networks on top of a common, shared infrastructure and consequently, mul-

tiplexing will not be done on the network level anymore, but on the infrastructure level, as depicted in Fig. 4, yielding better Quality of Service (QoS) or Quality of Experience (QoE) for the subscriber since different slices will have tailored orchestration for a given servic.
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. This comprises both Radio Access Network (RAN) and CN NFs that depending on the degree of freedom a tenant may have, also the MANO components. A network slice may be dedicated to a specific tenant or partially shared by several tenants in order to 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 of the slices, hence suggesting the economic viability of this approach that can adapt the used resources on demand.
The 5G atom proposed in Fig. 5 summarizes the discussion: 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 implementation of the concepts, and the novelties, that is, technologies that can be easily implemented due to softwarization and virtualization techniques.

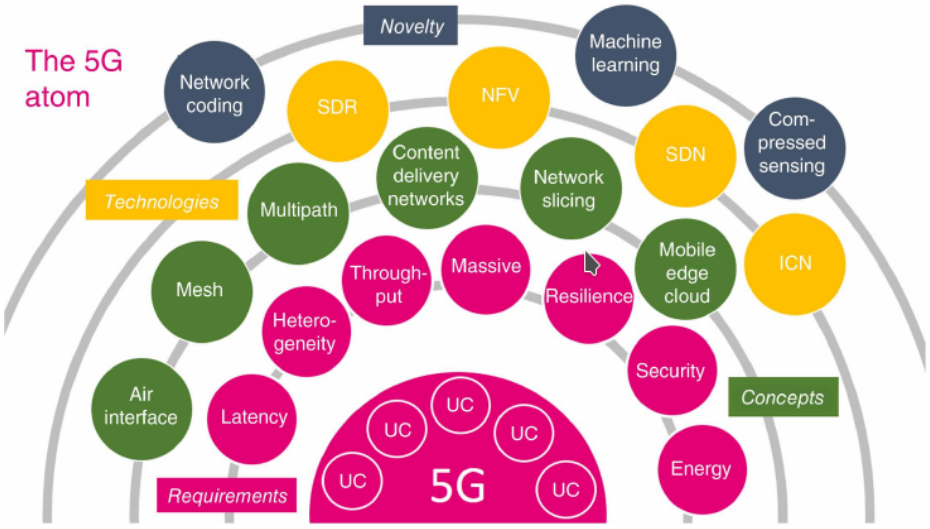

Figure 5: 5G atom representation. [5]

3.1 Main Services Types

The main 5G service types typically considered are:

- Enhanced mobile broadband (eMBB): related to human operations and to an enhanced access to multimedia content, services and data with improved performance with an increasing user experience. This service type, which can be seen as an evolution of the services today provided by 4G networks, covers use cases with very different requirements, ranging from hotspot characterized by a high user density and very high traffic capacity and low user mobility, to wide area coverage cases with medium to high user mobility; besides the need for seamless radio coverage practically anywhere and anytime with visibly is requested to improve user data rates;

- Ultra-reliable and low-latency communications (URLLC): related to use cases with tight requirements for capabilities such as latency, reliability and availability. Examples include the wireless control of industrial manufacturing or production processes, remote medical surgery, distribution automation in a smart grid, transportation safety, etc. It is expected that URLLC services will provide a main part of the foundations for the 4th industrial revolution (often referred to as Industry 4.0) and have a substantial impact on industries far beyond the information and communication technology industry;

- Massive machine-type communications (mMTC): capturing services that are characterized by a very large number of connected devices typically transmitting a relatively low volume of non delay sensitive data. However, the key challenge is here that devices are usually required to be low-cost, and have a very long battery lifetime. Key examples for this service type would be logistics applications, smart metering, or for instance agricultural applications where small, low-cost and low-power sensors are sprinkled over large areas to measure ground humidity, fertility and what concerns Intern of Things in general.

The concept of network slicing will be implemented at the beginning of the 5G era in order to realize these main services as slices and their 8 most important KPI are the following:

- Peak data rate, referring to the maximum achievable data rate under ideal conditions per user or device in bits per second. The minimum 5G requirements for peak data rate are 20 Gbps in the downlink and 10 Gbps in the uplink;

- User experienced data rate, referring to the achievable data rate that is available ubiquitously across the coverage area to a mobile user or device in bits per second. This KPI corresponds to the 5% point of the cumulative distribution function of the user throughput, and represents a kind of minimum user experience in the coverage area. This requirement is set by ITU-R to 100 Mbps in the DL and 50 Mbps in the UL;

- Average spectral efficiency, also known as spectrum efficiency and defined as the average data throughput per unit of spectrum resource and per cell in bps/Hz/cell. Again, the minimum require- ments depend on the test environments as follows:

  - Indoor Hotspot: 9 bps/Hz/cell in the DL, 6.75 bps/Hz/cell in the UL;

  - Dense Urban: 7.8 bps/Hz/cell in the DL, 5.4 bps/Hz/cell in the UL;

  - Rural: 3.3 bps/Hz/cell in the DL, 1.6 bps/Hz/cell in the UL.

- Area traffic capacity, defined as the total traffic throughput served per geographic area in Mbps/m2. ITU-R has defined this objective only for the indoor hotspot case, with a target of 10 Mbps/m2 for the downlink;

- User plane latency, given as the contribution of the radio network to the time from when the source sends a packet to when the destination receives it. The one-way end-to-end latency requirement is set to 4 ms for eMBB services and 1 ms for URLLC;

- Connection density, corresponding to the total number of connected and/or accessible devices per unit area. ITU-R has specified a target of 1 000 000 devices per km^2 for mMTC services;

- Energy efficiency, on the network side referring to the quantity of information bits transmitted to or received from users, per unit of energy consumption of the RAN, and on the device side to the quantity of information bits per unit of energy consumption of the communication module,both cases in bits/Joule. The specification given by ITU-R in this respect is that IMT-2020 air interfaces must have the capability to support a high sleep ratio and long sleep duration;

- Mobility, here defined as the maximum speed at which a defined QoS and seamless transfer between radio nodes which may belong to different layers and/or radio access technologies can be achieved. For the rural test environment, the normalized traffic channel link data rate at 500 km/h, reflecting the average user spectral efficiency, must be larger than 0.45 bps/Hz in the uplink.

The following web-spider diagrams in Fig. 6-7, sum up optimally these capabilities and how they have to be split to realize each particular slice

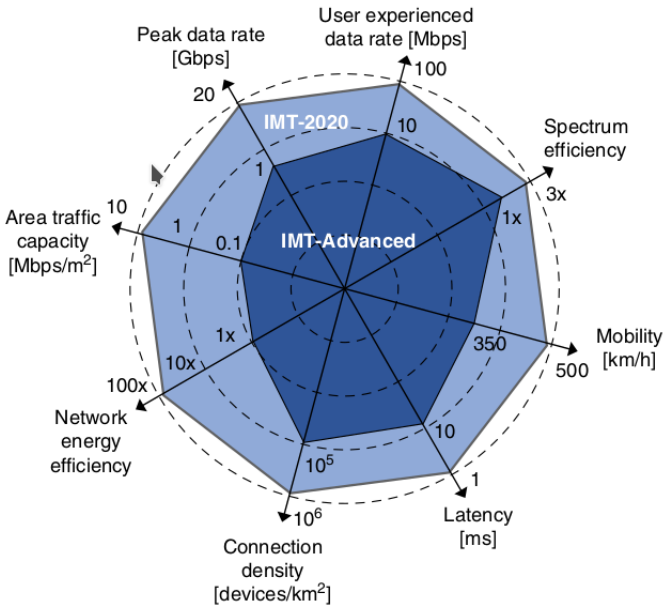

Figure 6: Capabilities to be achieved. [1]

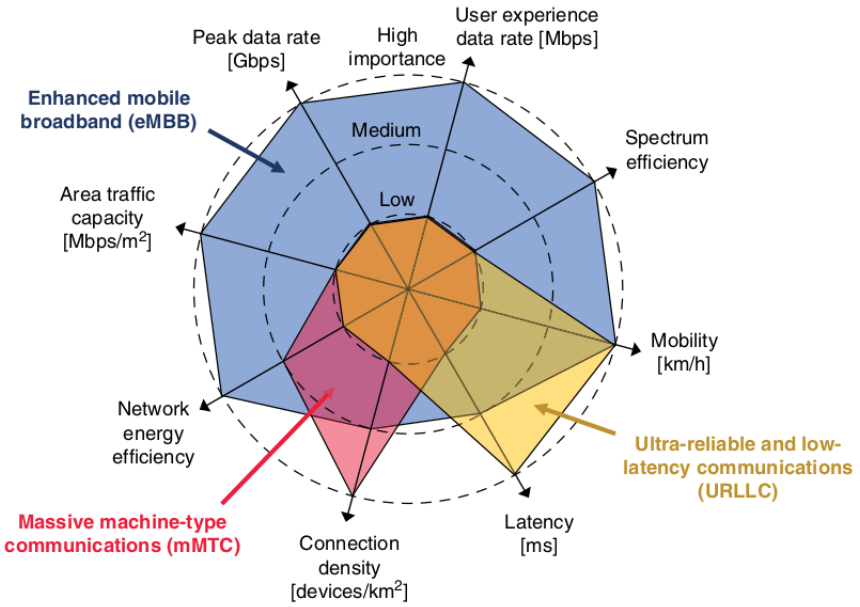

Figure 7: How capabilities are divided for each slice. [1]

3.2 Example

An illustrative example for deploying multiple network slice instances in a mixed environment consisting of public, i.e. mobile network operator owned, networks and private network infrastructure owned by a vertical enterprise. The infrastructure is used to commission an mMTC network slice and two eMBB network slices.

3.2.1 5G Services on Factory Premises

A process automation use case from industrial manufacturing is considered. Traffic is composed of sensor readings, actuator control signaling, and eMBB services providing access to local as well as remote applications (e.g. augmented reality for machine maintenance).
In a process automation environment, Internet of Things (IoT) devices include actuators, such as pumps, valves and sensors for capturing heterogeneous physical and logical quantities. The latter may for instance include sensors for supporting maintenance processes or for critical safety applications, aiming to improve the overall operational efficiency and safety of the factory. When connecting such IoT devices, latency and bandwidth requirements can be very diverse. Such a setup requires two types of network slices,

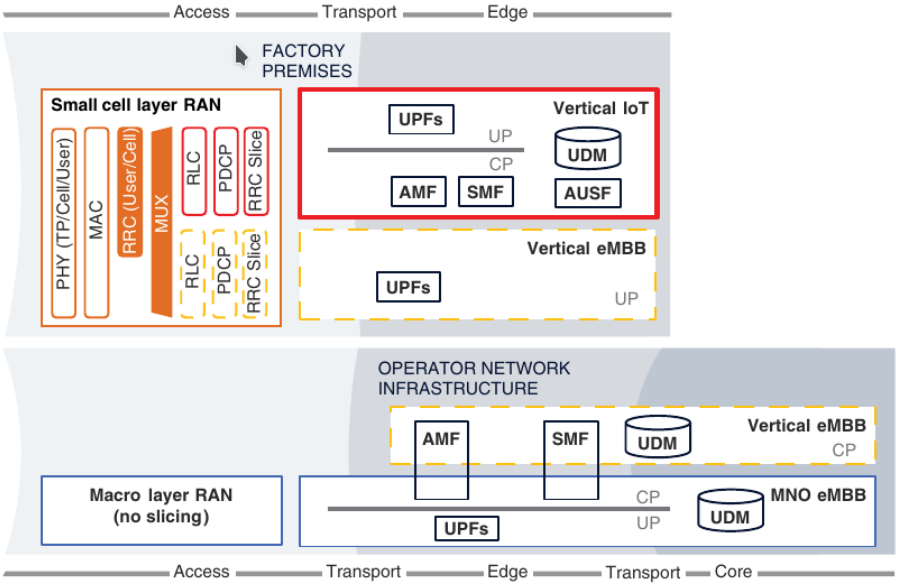

Figure 8: E2E network slice example for 5G services on factory premises. [1]

one covering machine-type communications for IoT devices and one covering eMBB traffic from smartphones, tablets and similar terminals. Figure 8 shows an example scenario with two E2E network slices (mMTC and vertical eMBB) for the factory owner (also referred to as "vertical") as well as an eMBB network slice for the MNO. The network slices run MNO infrastructure as well as the vertical's telecommunication infrastructure on the factory premises.

3.2.2 Ownership of Infrastructure, Spectrum, and Subscriber Data

In the given scenario, the vertical provides the small cell layer RAN equipment for all network slices that require coverage on the factory premises, i.e., both the IoT slice and vertical eMBB slice. Beyond the RAN, this includes the transport network and edge cloud resources, such as general purpose hardware for computing and storage. For operating the small cell layer RAN, the vertical rents dedicated spectrum resources from the MNO, such as higher frequency spectrum (e.g., above 6 GHz) with coverage strictly limited to the factory premises. For the vertical eMBB service, both small cell layer RAN and, if required, the MNO-provided macro layer RAN are utilized. Radio Resource Management (RRM) functions for the macro layer strictly remain under control of the MNO, which further owns a 5G-compatible network infrastructure consisting of centralized datacenters and distributed edge clouds

comprising general-purpose as well as application-specific hardware.
Subscriber information data including long-term security credentials are in possession of the vertical for the IoT devices. This assures full isolation of the vertical's IoT subscriber information from the MNO. For the eMBB subscribers, the MNO holds the corresponding data for own eMBB subscribers as well as vertical eMBB subscribers.

3.2.3 Domain-specific Network Slice Deployment

Regarding the deployment of the individual network slices, the mMTC network slice deploys all functions in the domain of the vertical, and it is only used by IoT devices that are registered in the vertical's Home Subscriber System (HSS) or Unified Data Management (UDM) and Authentication Server Functions (AUSF). These devices are mostly stationary and never leave the factory premises. The small cell layer RAN as well as the IoT-specific User Plane Processing Functions (UPF)s and Control Plane Processing Functions (CPF)s, in particular Access and Mobility Management Function (AMF) and Session Management Function (SMF), are operated locally under full control of the vertical. Since also the security mechanisms are strictly realized locally, the entire network slice operates in the shielded factory environment without exposure of any data to the MNO. In contrast, the vertical eMBB slice is deployed in an inter-domain manner: the CN control plane is shared with the MNO eMBB network slice and operated by the MNO outside the factory, including AMF and SMF as well as AUSF and UDM for authentication towards the core network and Non-Access Stratum (NAS) ciphering and integrity protection, respectively.
Regarding the transport network, independent slices with guaranteed levels of isolation and security are used by both the vertical and the MNO. In the vertical's small cell layer RAN, on the factory premises, physical and MAC layer in the UP and Radio Resource Control (RRC) in the CP are shared by both slices. This approach limits the complexity because resource multiplexing is implemented across all network slices on MAC level, forcing each network slice to make use of the same efficient flexible RAN implementation. On the other hand, each network slice may still customize the operation through configuration and parameterization of Radio Link Control (RLC), acket Data Convergence Protocol (PDCP), and RRC-Slice functions, RRM for the small cell layer realizes resource allocation according to the defined SLAs for the mMTC and eMBB slices of the vertical.

3.3 Actual Implementations

Some early trials have been conducted demonstrating network slicing with cross-industry collaboration among operators, vendors, and vertical industries. Two specific examples are given here.

- Deutsche Telekom and Huawei demonstration of E2E autonomous network slicing. In this demo, eMBB, mMTC and uRLLC are envisaged as network classes that could be built as slices. E2E network slicing included not only the core network and RAN, but also interconnecting transport networks. The demo implements E2E network slicing automation based on service oriented network auto creation. It uses software-defined topology, software-defined protocol, and software-defined resource allocation to ensure the automatic implementation of slice management, service deployment, resource scheduling, and fault recovery.

- SK Telecom, Deutsche Telekom and Ericsson have jointly built and demonstrated a trial network on federated network slicing for roaming, making SK Telecom and DT network slices available in each operators footprint, connecting South Korea and Germany. The demonstration was hosted at Deutsche Telekom's corporate R&D center in Bonn, Germany and Sk Telecoms 5G testbed at Yeongjongdo (the BMW driving center) in Korea. The demo featured an industrial maintenance use case involving a repair worker communicating via augmented reality with support colleagues in a visited network. The scenario used local breakout and edge cloud to enable the best service experience in terms of latency and throughput for the augmented reality repairman.

Bibliography

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

[4] NFVISG ETSI. Network functions virtualisation (nfv) release 3; evolution and ecosystem; report on network slicing support with etsi nfv architecture framework, 2017.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.