NETWORK SLICING OVERVIEW

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

February 1, 2019

Contents

Acronyms

AF Application Function.

AN Access Network.

API Application Program Interface.

BH Backthaul.

CN Core Network.

CP Control Plane.

CSMF Communication Service Management Function.

E2E End to End.

EM Element Management.

eMBB enhanced mobile broadband.

EPC Evolved Packet Core.

FH Fronthaul.

IaaS Infrastructure-as-a-Service.

IC Infrastructure SDN controller.

InP Infrastructure provider.

KPI Key Performance Indicators.

MANO Management and Orchestration.

mMTC massive machine-type communications.

MNO Mobile Network Operators.

MTA Multi-Tenancy Application.

N3IWF Non-3GPP InterWorking Function.

NBI Northbound Interface.

NEF Network Exposure Function.

NF Network Functions.

NFV Network Functions Virtualization.

NFVI Network Functions Virtualization Infrastructure.

NGMN Next Generation Mobile Networks.

NMS Network Management System.

NRF NF Repository Function.

NS Network Services.

NSIL Network Service Instance Layer.

NSMF Network Slice Management Function.

NSO Network Service Orchestrator.

NSSF Network slice selection Function.

ONF Open Networking Foundation.

OSS/BSS Operation/Business Support System.

PCF Policy Control Function.

QoE Quality of Experience.

QoS Quality of Service.

RAN Radio Access Network.

RO Resource Orchestrator.

SBI Southbound Interface.

SDN Software Defined Network.

SIL Service Instance Layer.

SLA Service Level Agreements.

TC Tenant SDN controller.

UDR Unified Data Repository.

UP User Plane.

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 fundamental topic for modern society because they guarantee communications and access to information on a tap of our fingers. Furthermore we expect that the data flowing per month will be around 50 exabytes and this is caused mainly by video services. Indeed in the last years, mobile operators are moving from a voice-oriented infrastructures to data-oriented ones and they are not able to keep pace with the calculated traffic volume. Facing such new problems, operators have invested a lot in research efforts toward the designing of the Fifth Generation mobile architecture, in order to provide innovative solutions for new capabilities and revenue sources. Network slicing is one of proposed innovation to push forward the frontier of mobile communications and it will discussed in this overview.
5G Network slicing allows Mobile Network Operators (MNO) to share physical infrastructures they have to the simultaneous deployment of multiple independent logical networks, managed with respect to their specific service requirements. For example, different services like automotive, tactile internet or massive IoT can be put in practice by exploiting different network slice instances, that is a set of virtual network functions that run on the same infrastructure with customized policies. Thanks to that, heterogeneous requirements can run on the same infrastructure, since different instances can be orchestrated and set independently according to the specifics requested. Furthermore, this is executed in a costly efficient way because as the a network slice shares the same physical infrastructure with different slices.
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 is ideally composed by: the Service Instance Layer (SIL) that involves the business/end user services provided by operators or by the party which leases the services by the operators; the Network Service Instance Layer (NSIL) is set of functions to run the instances and the resource layer, which consists of the resources such as compute, network, memory, storage.
The target of slicing is to realize End to End (E2E) network slices from the mobile edge, through the mobile transport (Fronthaul (FH)/Backthaul (BH)) and up until Core Network (CN).
The two main tools needed to allow different degrees of control and automation in NSIL and resource layers are: the Virtual Infrastructures (VI) under the control and operation of different tenants in agreement with an

Infrastructure-as-a-Service (IaaS) model[1] to split information from used hardware; Network Services (NS) owned by tenants, that is actually creation of a service instance. The VIs can be operated by Software Defined Network (SDN), NS are instantiated directly over a shared infrastructure and as a set of Virtual Network Functions (VNF) connected through one or more VNF forwarding graphs are necessary to define the life cycle of the slice; both of them will be discussed later.
Multi-tenancy is an characteristic that must be achieved to guarantee separation, isolation, and independence between different slices decoupled from the shared underlying resources.

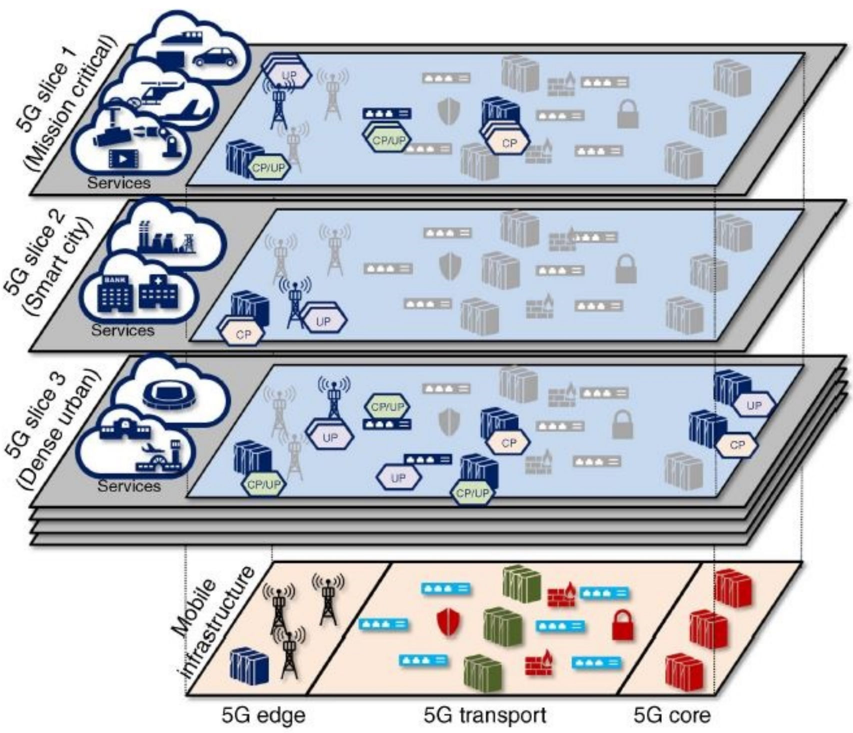

Figure 1: Example of network slicing in 5G. [1]

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. [1] [5]

[1]Form of cloud computing providing computed and virtualized resources over the internet

2 Structural components

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 NS management; they will be the main topics of this section. As anticipated about

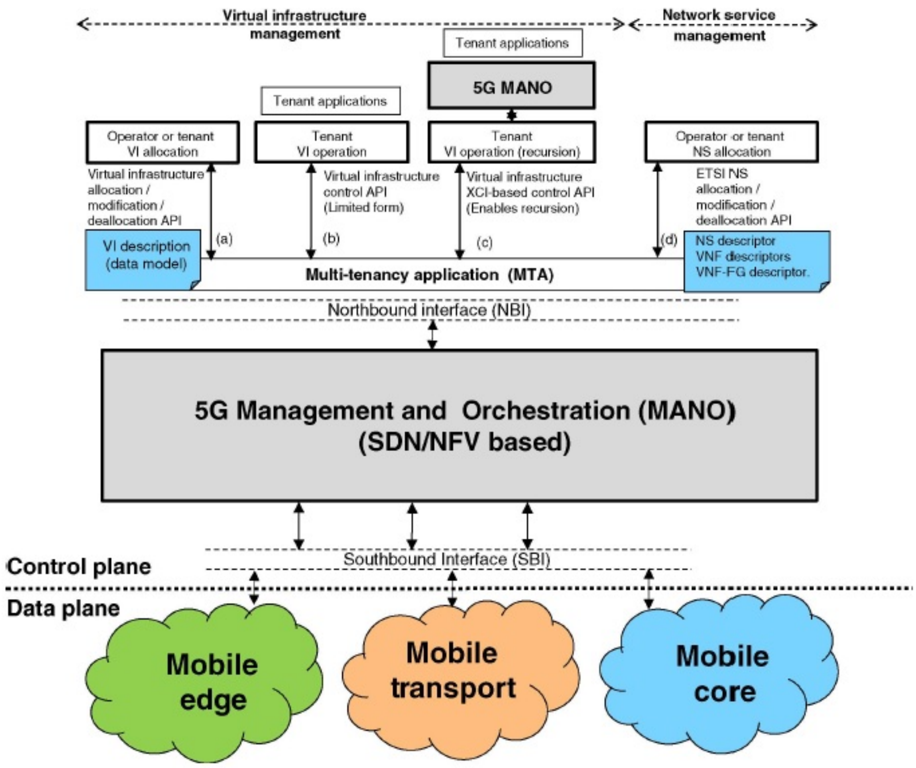

Figure 2: Architecture for network slicing. [1]

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 actually the resource layer where mobile edge, mobile transport, and core are present. The infrastructure is made of links, forwarding

nodes like a router or a switch, cloud nodes and a set of network, computing and storage resources.
The control plane is divided into two layers: an application layer at the top and at the bottom the Management and Orchestration (MANO).
The design of the MANO is based on the ETSI management and orchestration of the network by means of a framework exploiting SDN-based control. The MANO provides a panoramic of the resources available by means of Northbound Interface (NBI), instead it is linked to data plane entities by means of Southbound Interface (SBI) to execute control and management functions on the actual hardware components (possible application to do the job are OpenFlow, SNMP, OVSDB). For what concerns the Multi-Tenancy Application (MTA), it realizes the multi-tenancy support because it applies a coordination of the tenants accessing the infrastructure shared by the clients: so in order to do that it performs isolation of resource assigned to various instances of different tenants, and implements services related to allocation and operation of VIs using some dedicated Application Program Interface (API)[2].
As shown in Fig. 2 the choice of APIs depends on the kind of service. [1] [6]

2.1 Enablers and Design Principles

Future 5G networks will be built on concepts absolutely not related with the previous generation network architectures. In the Fifth generation a revolution is actually provided by the introduction of SDN and Network Functions Virtualization (NFV)[3] opening the doors to a new huge number of 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 idea to modularize the architecture and to decompose the network functions has been enunciated at the beginning stages of research in order to fulfill the above requirements.
Network Functions (NF)s are the entities that furnish specific network capabilities in order to realize and support the requested services. Generally they

[2]it is the code that permits softwares to create a communication among each other.
[3]Although NFV and VNF are often used without distinction, for the sake of clarity NFV is a general concept, instead the VNF is a building block of the NFV framework.

are software instances acting on infrastructure resources, but can be physical, a combination of them or virtualized, that is decoupled from the hardware it runs on; the last peculiarity will be fundamental. The general "monolithic" network functions (4g as well) 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 NFs for CP and UP.
To realize the highest level of decomposition possible, it is necessary to affirm a strong distinction between NFs of the Access Network (AN) and CN, in order to achieve what is called a convergent network[4]; AN/CN split is mandatory to reach the essentials to support network slicing. [1] [5]

2.1.2 Virtualization

Virtualization is a main process in the network slicing architecture, because it allows to share effectively the resource among the various slices and it operates the abstraction of resources. Abstraction means the representation of the characteristics of the underlying resources in order to recreate a virtual scenario with same peculiarities.
The resources to be virtualized can be physical or already virtualized, generating a recursive structure in the system counting different abstraction levels. Exactly like server virtualization makes Virtual Machines (VM)s make them free from the physical hardware, network virtualization allows to generate multiple isolated virtual networks, completely independent from the physical network and it can run over it. The framework consists of three actors:

- Infrastructure provider (InP): owner or a manager of the physical network. It offers resources to be virtualized and delivered to a single or multiple tenants.

- Tenant: leaser of the virtual resources from the InPs, which exploits to provide the necessary network services to the users.

- End user: consumes the services supplied by the tenant.

[3] [7] [5]

[4]It is the coexistence of different kind of information to transmit, actually voice, video and data, within a single network.

2.1.3 Orchestration

Orchestration is another key point to realize slicing and accounts the problem of allowing the coexistence and organizing all the constitutive elements of the architecture. In a such scenario, where the entities involved are so various, the so called orchestrator, is needed to coordinate the different services related to all the assigned requirements.
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 that a policy to handle the orchestrator behavior is required and which is expected to satisfy the service level agreements Service Level Agreements (SLA)s associated with clients requirements. This also has to remember that the available resources, the demands and optimization criteria may change in time.
Noteworthy is that orchestration is also referred to as the defining characteristic of an SDN controller. Orchestration process is not performed by a unique and central unity due to its complexity, but also because we want to maintain management independence and support recursion. Then a framework in which an entity performing orchestration functions is more suitable for the general architecture. [7] [5] [9]

2.1.3.1 Isolation

A effective isolation is, of course, an important requirement that must be fulfilled to let parallel slices run on a common underlying substrate. The isolation must be understood in terms of:

- Performance: each slice is defined to realize specific service requirements, expressed generally as Key Performance Indicators (KPI)s and it must ensure them always regardless of the congestion and performance levels on the others.
- Security and privacy: attacks or issues in a slice must not affect other slices. We say that we want each slice having independent security functions to block unauthorized entities to have access and so reading or writing capabilities on the configuration, management, accounting information.
- Management: each slice must be considered managed as a standalone network, that is as no other slices are present in parallel.

To get the wanted isolation then a set of policies and processes must be defined at each virtualization level, they are lists of rules and settings describing how the various entities have to be isolated; a team play of both virtualization and orchestration is actually needed.[7] [9]

2.1.4 SDN: Software-Defined Network

In a Software-Defined Network the admin or an engineer is able to handle the data traffic remotely exploiting a centralized control without putting hands on particular switches of the network. In this context, the SDN controller tunes the switches in order to deliver the NS wherever they are needed; this is a step away from the classical architecture, where devices take their traffic decisions based on the routing tables.
The SDN architecture is comprehensive of, as previously explained and shown in Fig. 1, an intermediate CP to set and abstract the underlying forwarding plane resources so that ocustom services are furnished to clients. After having said this, it is clear that SDN is a suitable tool for implementing slicing in the 5G architecture.
The main components in a SDN are resources and controllers. Resource is everything useful to provide services as answer to client requests that in this case are the infrastructure resources and NFs; also NS in application of the recursion principle are included. A controller, instead, is the centralized entity implemented at the control plane and it looks for trade-offs among clients and resources because it acts as server and client at the same time via client/server contexts respectively:

- Client context: all the information needed by a controller to manage the transmission to clients. It includes in turn: Resource Group, an abstraction of all the resources offered by the controller to the client via NBIs; the Client support function which contains the necessary to ease client operations, like policies.

- Server context: all the information needed by the controller to operate with a set of underlying resources grouped in a Resource Group via SBIs.

The idea is to transform a Resource Groups set instantiated in server contexts into those defined in client contexts, this indeed requires a SDN controller to virtualize and orchestrate the process.
During the virtualization function, the SDN controller first of all makes the abstraction and the merge or the partition of underneath resources. Then thanks to that, specific Resource Group provided by each client contexts are exploited by the client of that context to accomplish the wanted services. Afterwards, thanks to the orchestration, the SDN controller optimally gives the chosen resources to such separate Resource Groups. So the the services demanded by the clients are guaranteed and at the same time the isolation is maintained among them.

In the proposed architecture for a SDN an admin is present to accomplish the tasks of initialization and setting of controller, of the definition of server and client contexts and of selecting the related policies.
This provides a suitable solution for enabling slicing because the client context completely abstracts the set of resources into a Resource Group and furnishes the control logic that constitute a slice which includes the related client service attributes.
A noteworthy peculiarity that makes SDN ideal for slicing is recursion: thanks to the different abstraction layers that the recursion principle makes possible, the SDN CP may involve multiple controllers in a ranked ordered that enlarge the client-server relationships at more levels. Now it is clear that a SDN architecture does support a recursive composition of the slices and as consequence the resources that a controller brings to a client as a dedicated slice (client context), can in turn be virtualized and orchestrated by the client in the case of being an SDN controller. Hence, the new controller can exploit these resources it has accessed by means of its server contexts to define and bring new resources therefore guaranteeing effectively new slices to the clients. [7]

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.

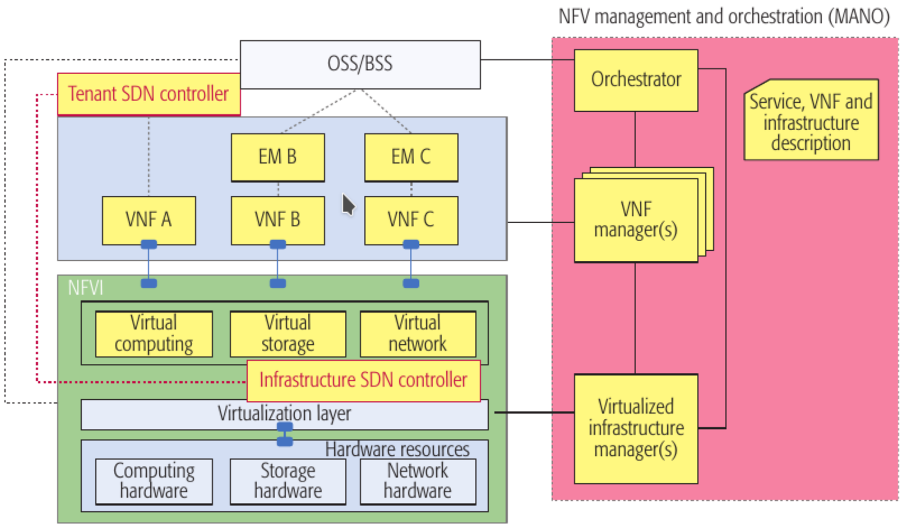

Figure 3: Integrating SDN controllers into the reference NFV architectural framework. [7]

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.

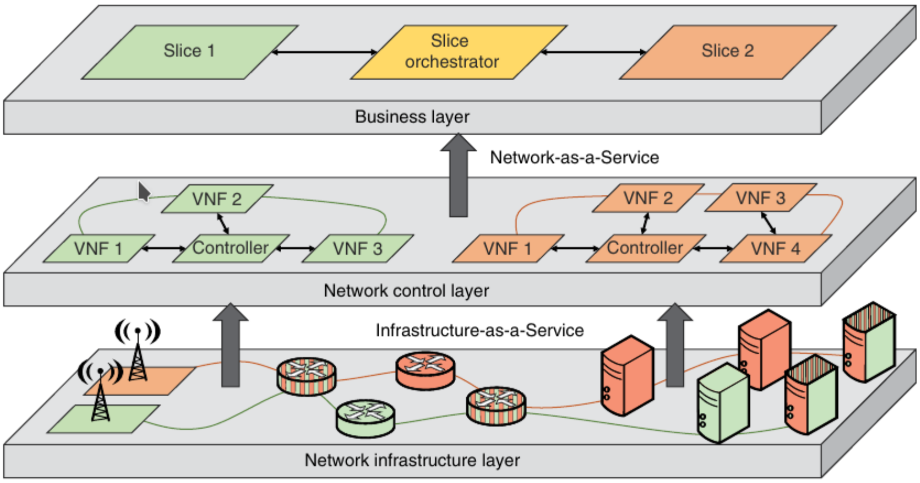

Figure 4: An example of a network-sliced architecture. [1]

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

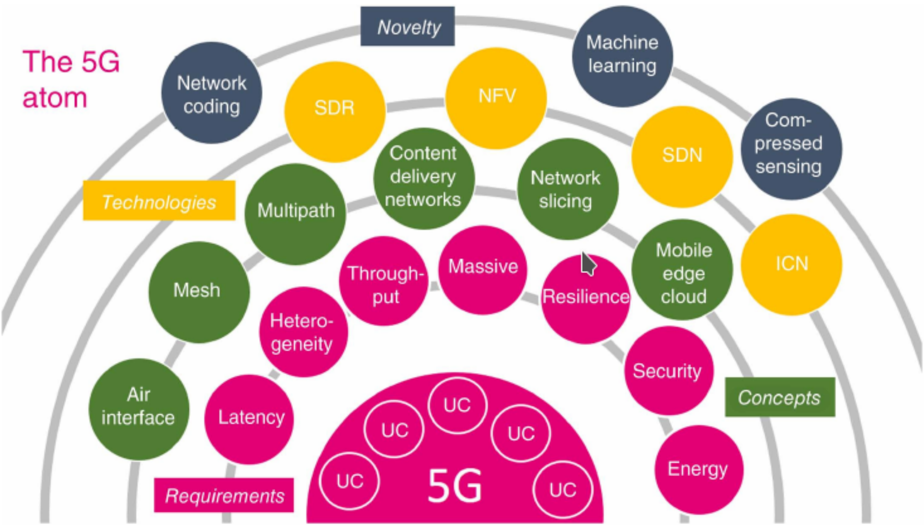

Figure 5: 5G atom representation. [5]

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 in 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 can be seen as a kind of minimum user experience in the coverage area. This requirement is set 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 requirements 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, defined only for the indoor hotspot case, with a target of 10 Mbps/m^2 for the downlink;

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

- Connection density, corresponding to the total number of connected per unit area, targeting 1000000 devices per km^2 for mMTC services;

- Energy efficiency, refers 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. 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 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

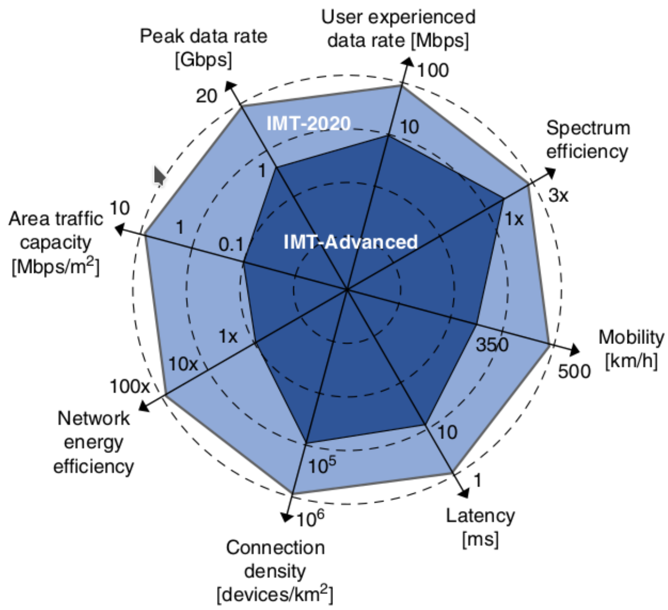

Figure 6: Capabilities to be achieved. [1]

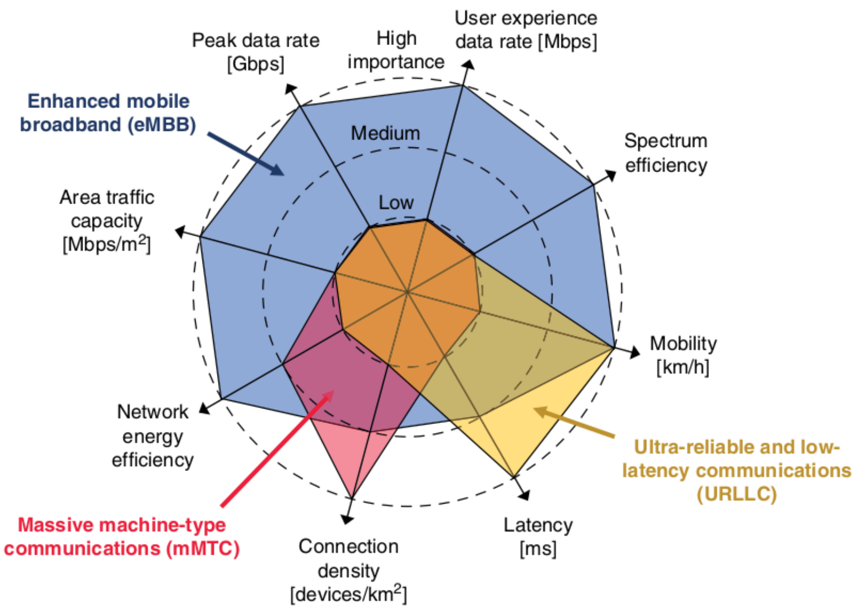

Figure 7: How capabilities are divided for each slice. [1]

4 5G Architecture

3GPP has split 5G specifications in two phases: the first called Release-15 and released on August 2018 addressing a urgent subset of commercial needs; the second, called Release-16 is planned to be completed by March 2020 addressing all identified use cases and requirements. Now, by giving a look at Release-15 a representation of 5G architecuture is illustrated in order to understand how the novel concept of netowork slicing is implemented; Fig. 8 Compared to the LTE architecture, 3GPP has decided to modularize

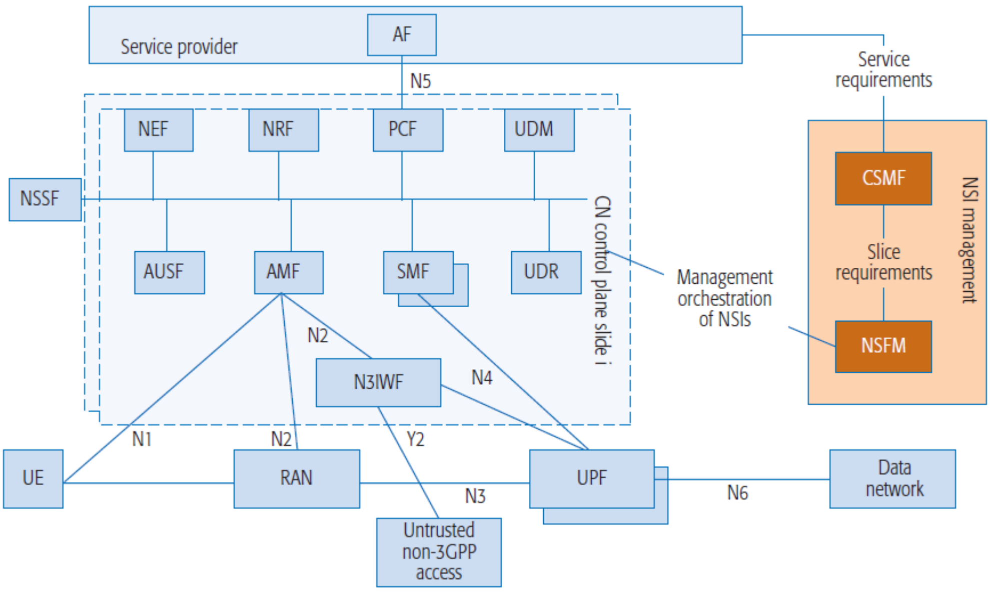

Figure 8: 5G System architecture. [2] [4]

mainly the CN and allocate slice-related NFs inside it. The CN control plane functions are now described:

- Core access and mobility management function (AMF) necessaru mobility management, access authentication and authorization, security anchor functions and context management;
- Session management function (SMF) handles session management, IP address allocation, traffic steering, selection and control of UP functions, control part of policy enforcement, and QoS;
- Policy Control Function (PCF) specifies an unified policy framework to govern network behavior and policy rules to control plane functions;

- Network Exposure Function (NEF) provides to securely expose the services and capabilities provided by NFs;
- NF Repository Function (NRF) is a service discovery function used by NF instances;
- Unified data management (UDM) concerns the authentication credential repository and processing function, user identification handling, subscription management;
- Authentication server function (AUSF): supports the authentication server;
- Non-3GPP InterWorking Function (N3IWF) is necessary to support: IPsec tunnel establishment, relay signaling between user equipment and AMF, relay user plane packets between UE and UPF, anchor point within non-3GPP access networks;
- Unified Data Repository (UDR) stores subscription and policy data;

The architecture also contains the following functions:

- User plane function (UPF) is an anchor point for inter/intra-radio access technology mobility and necessary for packet routing and forwarding, QoS handling for the user plane;
- Network slice selection Function (NSSF) supports the functionality to bind a user equipment with a specific slice;
- Application Function (AF) influences traffic routing, accesses the NEF, interacts with PCF;

5G allows user equipments to concurrently access multiple network slices concurrently, but in such a case, only a single AMF will be used for all slices. It has been agreed that they cannot use more than eight slices in parallel. To identify slices, the network slice selection assistance information is used and consists of the slice/service type, an entity which identifies the NFs a slice provides. 3GPP has standardized 3 slice/service type values,one for eMBB, 2 for URLLC and 3 for MIoT. Therefore, CN is responsible to authorize the attachment of UE to a slice. The management functions for network slices are called Communication Service Management Function (CSMF) and Network Slice Management Function (NSMF). The CSMF is used to translate the communication service requirements to network slice requirements (capacity, throughput, delay, number of users, geographical identifications, authentication level). The NSMF is responsible for managing and orchestrating the life cycle of a network slice. This life cycle consists of the following phases:

- Preparation phase;
- Instantiation, configuration, and activation phase;
- Runtime phase;
- Decommissioning phase;

The first phase involves the creation and verification of a slice template. During the second phase, all shared or dedicated NFs and resources are allocated and configured. During the third phase, a network slice is essentially a fully operational network, and monitoring and optimization functions are also performed. The last phase is related to the deactivation of a slice and the decommissioning of previously allocated resources.

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] ETSI. 3gpp ts 23.501 v15.2.0, 2018.

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

[4] Alexandros Kaloxylos. A survey and an analysis of network slicing in 5g networks. *IEEE Communications Standards Magazine*, 2(1):60–65, 2018.

[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.