5G Network Slicing

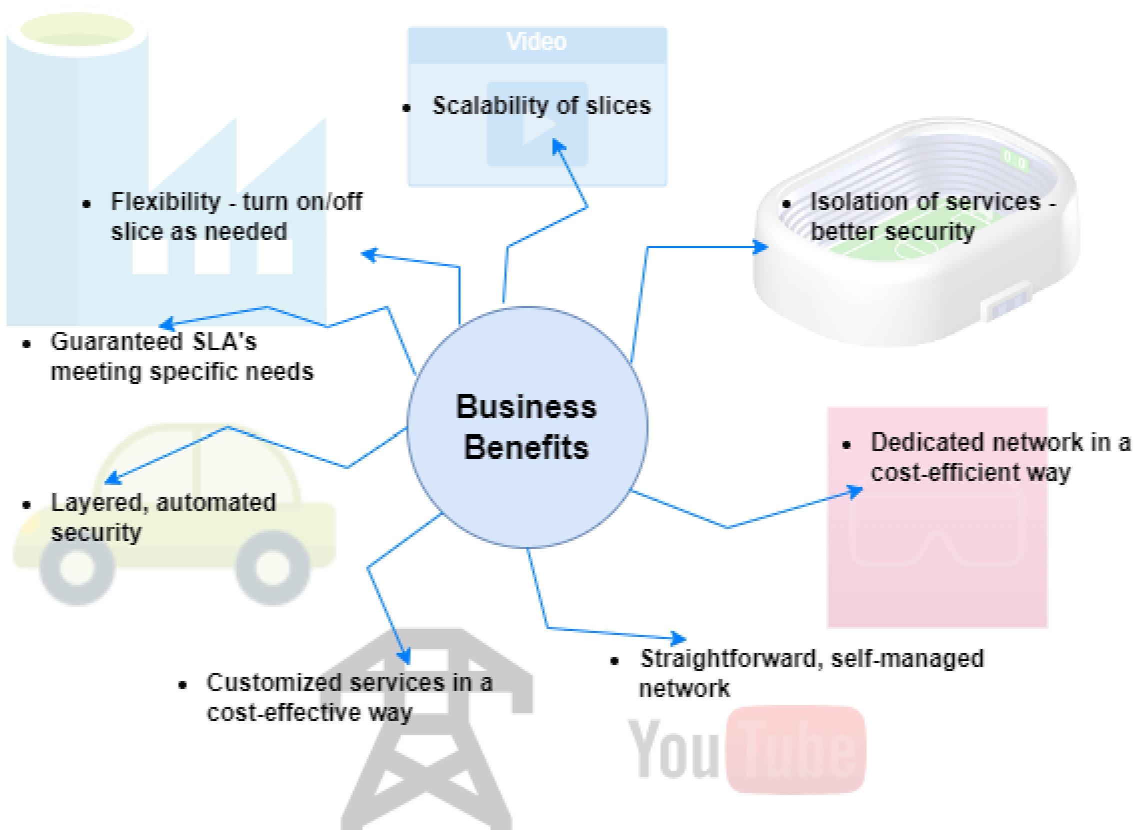
Contents

Chapter 1. Executive Summary

The following Executive Summary overviews the key information presented in each part of this report on Network Slicing in 5G mobile network. It introduces the basic **drivers and benefits of Network Slicing** as well as the **use cases** with the greatest potential to benefit from Network Slicing. This report also outlines the basic technological advancements enabling Network Slicing, that is **Recursive Model, Network Function Virtualization** and the **End-to-end Architecture of Network Slicing**with the emphasis on their role in delivering seamless, high-quality customer experience, in a cost-efficient way.

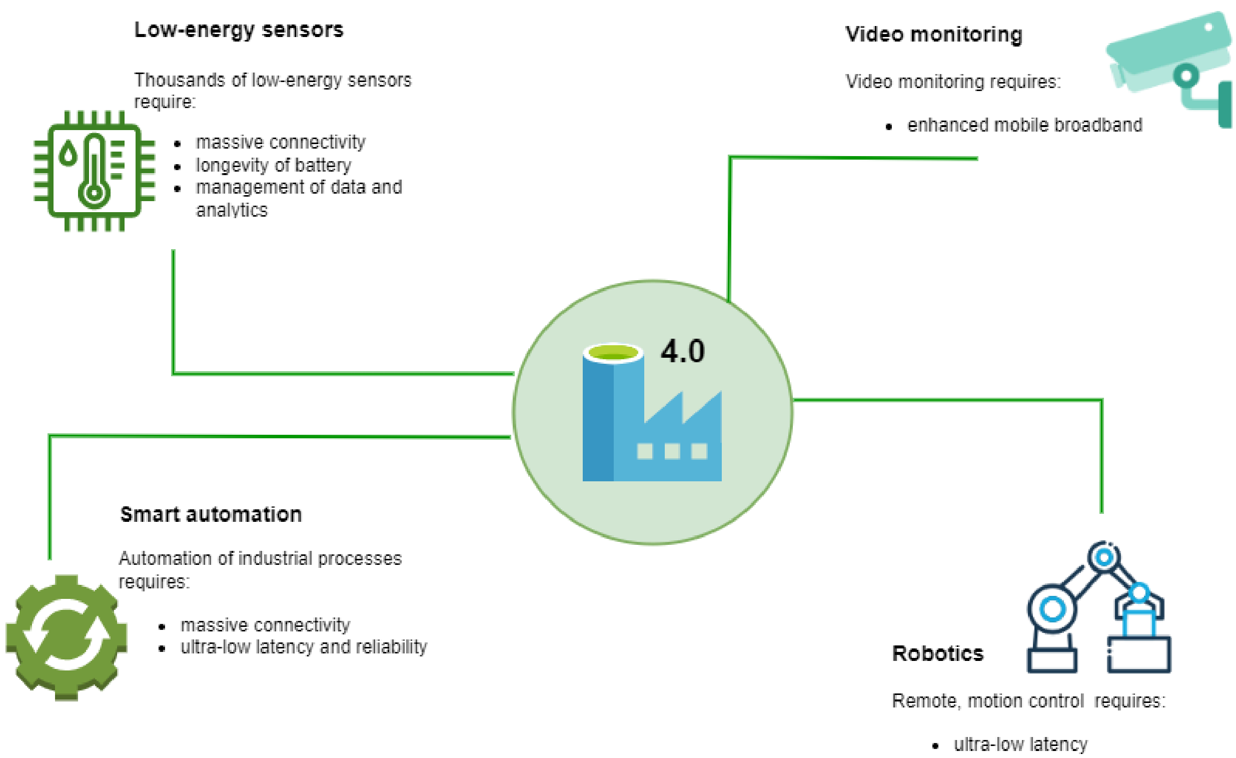
Business Drivers and Benefits:

- Network slicing is the most prominent technological advancement in 5G mobile network technology.
- Network slicing is **the solution to the rapidly growing digitalization of the world, emerging new use cases, increasing numer of smart devices and applications with specific requirements**.
- Network slicing gives operators capabilities of **creating specific, customized services in a cost-effective way**, introducing new services to old customers at the premium, address new customers and create new revenues.



Use Cases:

- Network slicing manages the **growing complexity of Smart Factory networks** by assigning each part of the factory network to a single, customized slice.
- Network slicing enables **massive connectivity of sensor in Smart Cities** and an uninterrupted and reliable data flow by isolation of network slices.
- Network slicing provides **high level of integrity and security of autonomous vehicles** by automated orchestration of slices.
- Each network slice is designed as **highly programmable and scalable**.
- Each slice characteristics is **separately customized** to serve the purposes of the specific service in terms of throughput, latency and connectivity.



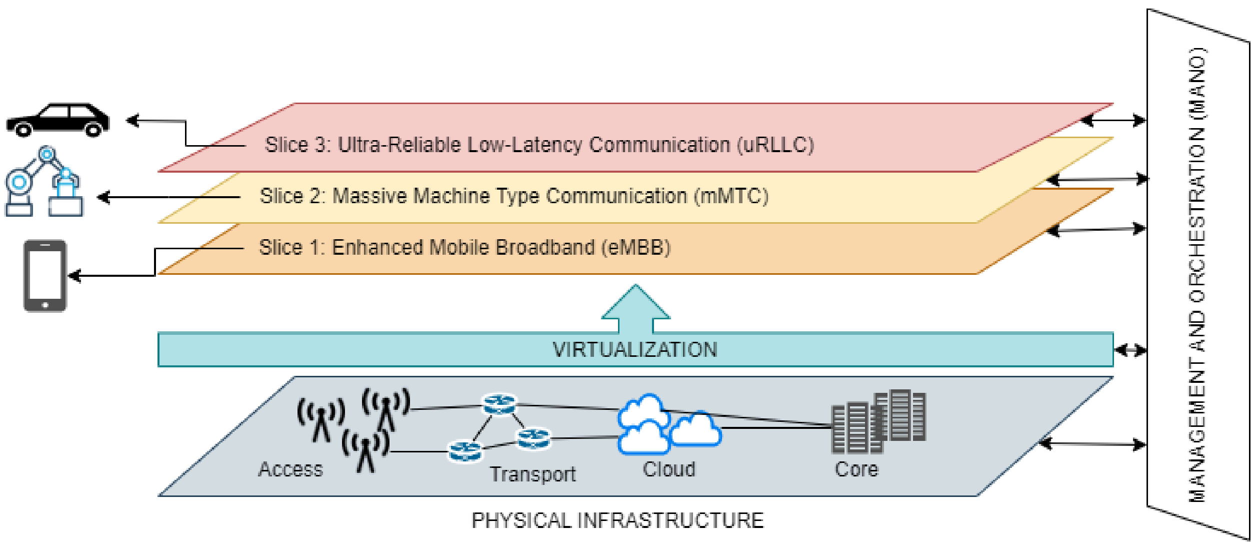
Network Slicing Architecture

Recursive Model

- The architecture of network slicing is primarily based on the Recursive Model, in which **each part of the network can be instantiated and placed in a different location.**
- In the virtualized infrastructure, the Recursive Model allows the creation of the **customized and independent instances of the network, i.e. "slices".**

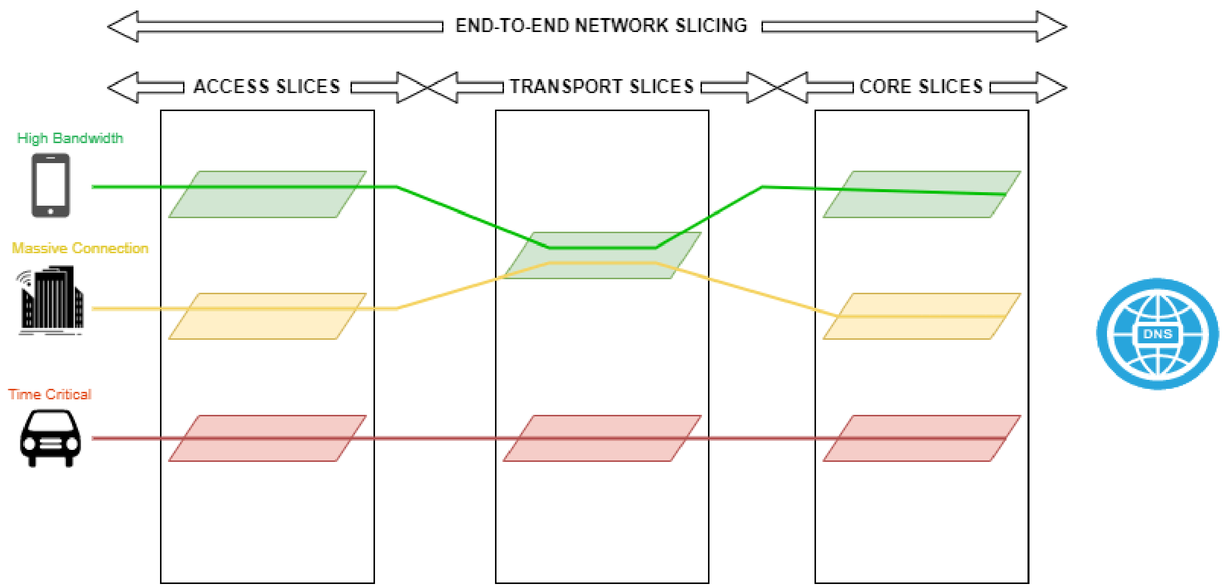
Network Function Virtualization

- **The functions and resources on network slices are virtualized** (in the form of a software), which provides the possibilty to dynamically configure them, turn them off/on, scale them up or down.
- The virtualized functions and resources are managed by **Software-Defined Networking (SDN)** ***(on page 42)***, which ensures more efficient use of resources, reduces costs and energy usage.
- Network slicing enables the creation of **multiple, virtualized, independent networks on the shared infrastructure.**



End-to-end Network Slicing

- Network slicing is end-to-end, i.e. all the way from user equipment (UE) *(on page 42)* to data center (DC) *(on page 39)*, **through separate and independent slices of access, transport and core.**
- The importance of end-to-end network slicing is mainly stressed for **extreme flexibility** in order to provide Service Level Agreement (SLA) *(on page 41)*-based quality of experience in a cost efficient way.



Access Part of Slicing

- Slicing in radio access network (RAN) *(on page 41)* is envisioned as **extremely flexible and adjustable.** The resources in RAN can be shared among different network slices.

Core Part of Slicing

- The most fundamental characteristics of 5G core *(on page 39)* architecture is **Control and User Plane Separation (CUPS)** ***(on page 39)***. User Plane Function, (UPF) *(on page 42)* is separated from other network functions in 5G core *(on page 39)* and located closer to UE *(on page 42)* for lower latency.
- **Network functions in 5G core can be either shared or dedicated**. In most use cases, session management function (SMF) *(on page 41)* and UPF *(on page 42)* are dedicated for a specific slice.

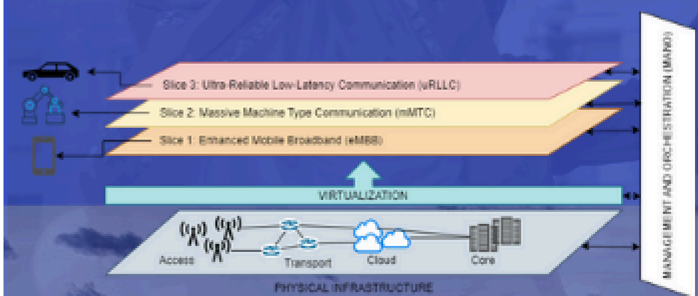
Transport Part of Slicing

- **Tranport network (from cell site to the core) is divided into 3 parts**, which provides higher level of flexibility, allocates the radio resources more efficiently, balances and optimizes the traffic load.
- Slicing at the transport layer mainly concerns **traffic routing and prioritization**, which is performed by the load balancer, such as software-defined networking (SDN) *(on page 42)*.

To discover more about each aspect of Network Slicing **click on the corresponding part on the image below**:

NETWORK SLICING

The most prominent technological advancement in 5G mobile network technology



DEFINITION

Network slicing is a network architecture that enables creation of multiple, virtualized, independent networks on the shared infrastructure.

DRIVERS

The development of Network Slicing has been driven by:

- rapid digitalization of the world
- growing number of smart devices
- new use cases and applications with specified requirements etc.

USE CASES

The verticals with the greatest potential to benefit from Network Slicing:

- Smart Factory

1. This link takes you to the definition of Network Slicing *(on page 9)*
2. This link takes you to the description of drivers of Network Slicing *(on page 13)*
3. This link takes you to the description of Use Cases in Network Slicing *(on page 19)*
4. This link takes you to the description of Business Benefits of Network Slicing *(on page 15)*
5. This link takes you the description of the Architecture of Network Slicing *(on page 25)*
6. This link takes you to the description of End-to-end Network Slicing *(on page 33)*

Chapter 2. Introduction to Network Slicing

Network slicing in 5G technology allows for the creation of multiple virtual networks on a shared infrastructure. Each network slice is customized independently **to serve the purposes of the specific service, business or application in terms of throughput *(on page)*, latency *(on page)* and connectivity *(on page)***. The characteristics of a slice is stated in Service Level Agreement (SLA) *(on page)* and agreed between a service provider and a tenant. Moreover, the fact that it is **virtualized, makes it more scalable and flexible.**

Definition of Network Slicing

Network slicing is based on the idea of creating multiple virtualized instances ("slices")of the shared physical network and customizing them in the specific way to fulfill the requirements of the service, business or application.

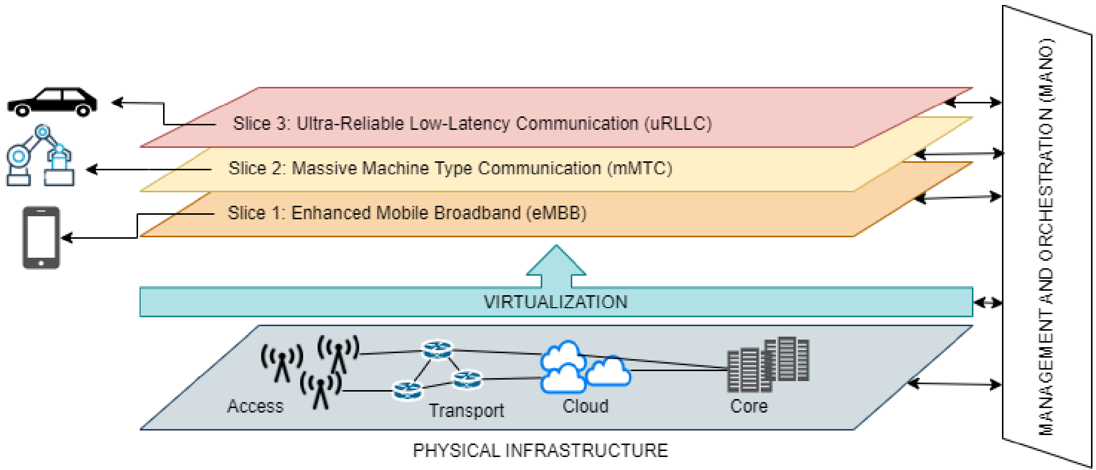
Definition od Network Slicing:

Network slicing is a network architecture in 5G technology that **enables the creation of multiple, virtualized, independent networks on a shared infrastructure**, as shown in Figure: Network Slicing Architecture.The basic idea behind network slicing is to **"slice" the existing infrastructure into multiple, virtual, end-to-end networks**, which are highly programmable and scalable.

Definition of Network Slice:

Thus, the network slice *(on page 41)*is a **separate, self-contained, and independently secured part of the underlying network infrastructure**. Each network slice is designed to **fulfill the specific requirements of the tenant *(on page 42)*** or a group of tenants, such as autonomous vehicle, smart factory or an application. Each network slice is highly programmable, scalable and has its own dedicated resources.

Figure: Network Slicing Architecture



Benefits:

Network slicing proves to be the most fundamental enabling technology to **fulfil thedivergent requirements of different types of use cases** in an efficient and **economically** sustainable way.

Network Slice Characteristics

Network slice is independently, **separately customized to serve the purposes of the specific servicein terms of throughput *(on page 42)*, latency *(on page 40)* and connectivity *(on page 40)***. The characteristics of a slice is stated in Service Level Agreement (SLA) *(on page 41)* and agreed between service provider and a tenant. Moreover, the fact that it is **virtualized, makes it more scalable and flexible.**

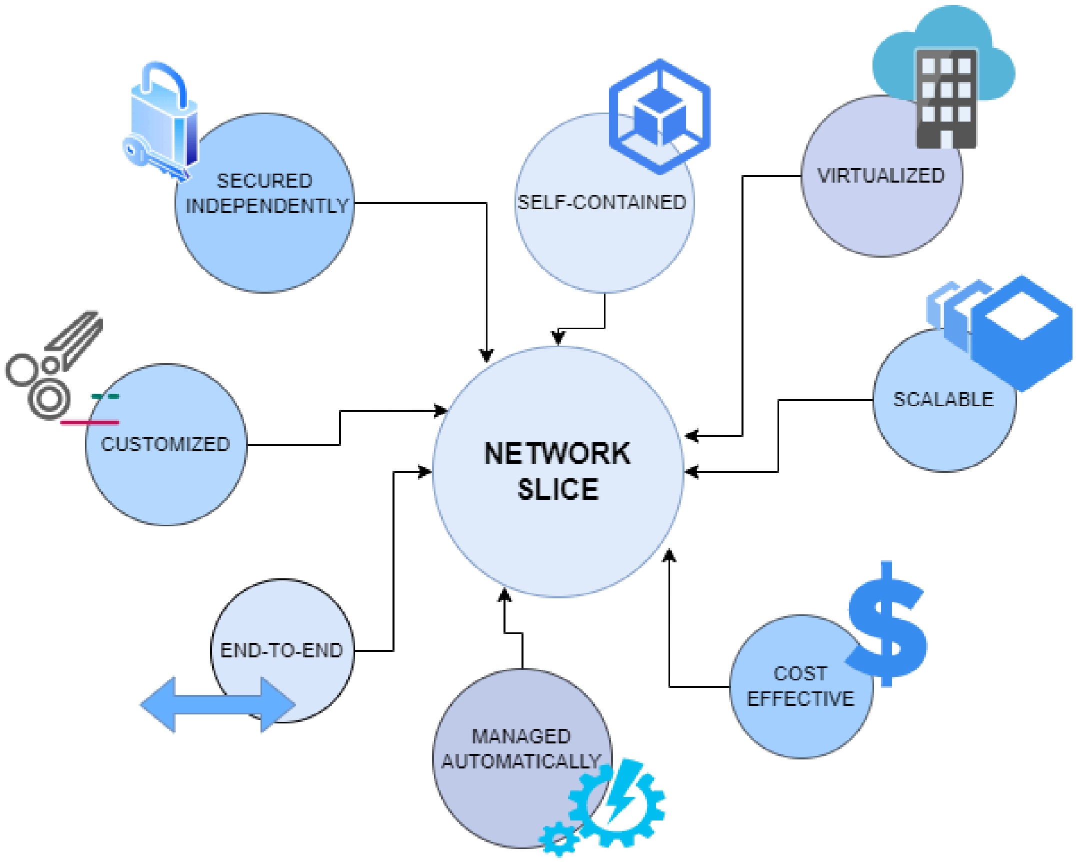
Features of Network Slice:

Since each network slice is a **virtualized** part of the physical network, it is:

- **highly scalable and flexible** – can be configured instantly, on demand
- **isolated** from other network slices – the breakdown or overload of one slice does not affect the functioning of the others
- **independently secured** – each slice has its own security
- **end-to-end** – from Data Centre to User Equipment through separate access, transport and core slices
- **automatized** – the lifecycle of a network slice is automatically managed, which **reduces the costs**.

Therefore, operators are provided with the exclusive capability to satisfy customer’s individual requirements and deliver them customized services in **a cost-effective way**.

Figure: Features of Network Slice



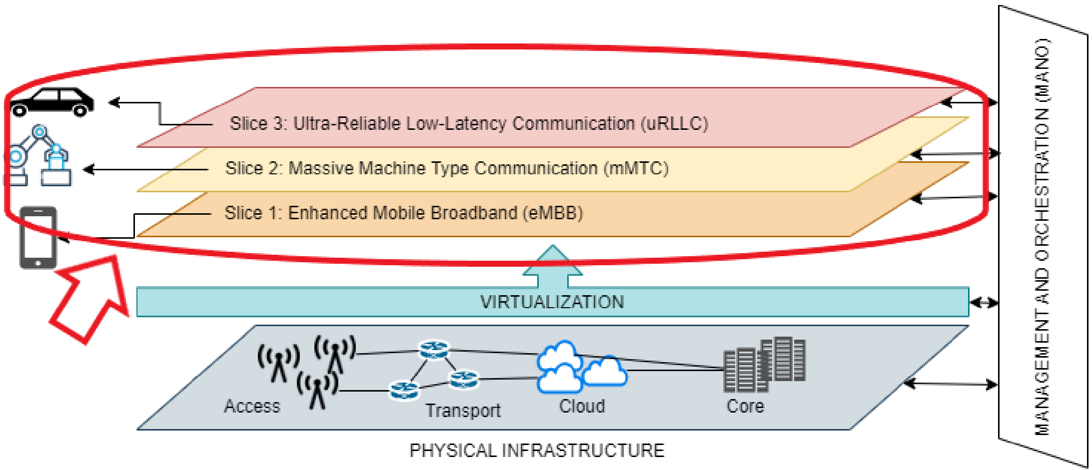
Characteristics of Network Slice:

The network slice can be **customized in a specific way to fulfil the requirements of a business or a service.** The requirements mostly concern throughput, latency and reliability or massive connectivity.

As shown in Figure: Network Slices Characteristics, depending on the target use case, business vertical or a service, **network slice characteristics** can be:

- high bandwidth for a video streaming
- ultra-reliability, low-latency for autonomous vehicles
- massive connectivity for Internet of Things *(on page 40)*

Figure: Network Slices Characteristics



The specific characteristic of the slice is **agreed between a service provider and a tenant**, in a form of SLA (Service-Level Agreement) *(on page 41)*. It is also important to note that each device, such as a mobile phone, can use up to 8 network slices concurrently.

Chapter 3. Drivers of Network Slicing

The rapid digitalization of the world, growing number of smart devices, new use cases and applications with specified requirements have driven the deveopment of new era 5G mobile network with network slicing as a key technological enabler.

Why 5G Mobile Network?

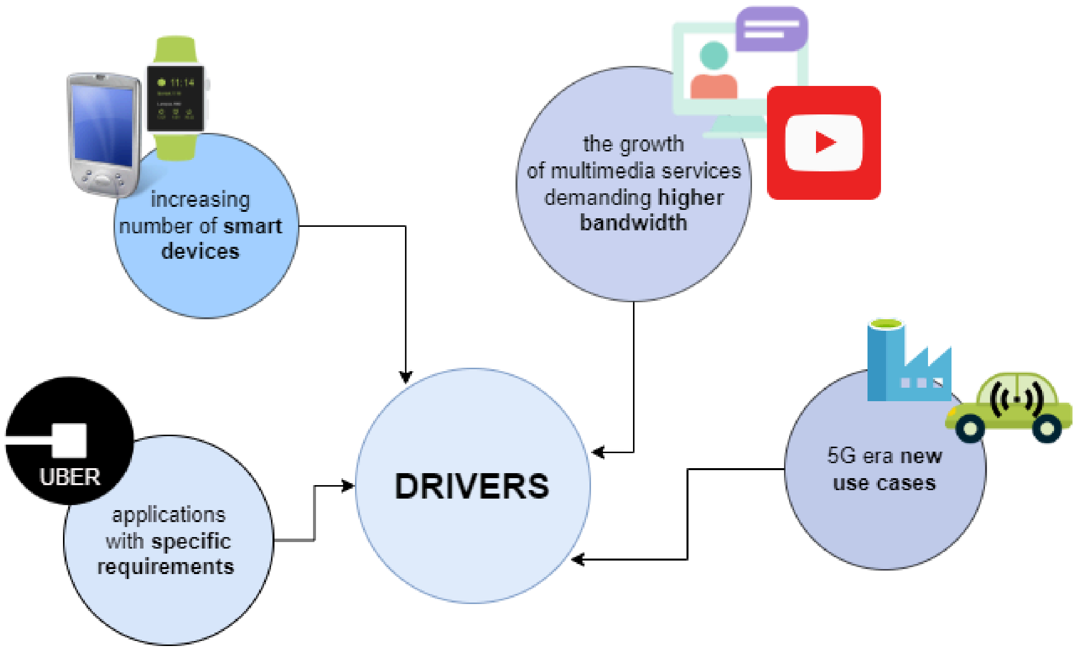
The previous monolithic network infrastructure (4G) proved to be largely inefficient in fulfilling the growing requirements of users or emerging use cases. The need for a new generation and technological advancements in mobile networks arose. It was envisioned to **overcome all the limitations of the previous mobile network generation in a cost-efficient way.** The 5G mobile network was designed with advanced technologies and has proved to be the most fundamental enabler of the digitalized world.

Drivers of Network Slicing:

Network slicing is believed to be the most prominent technological advancement in 5G mobile network. The design and development of network slicing has been triggered in the face of the **rapidly growing digitalization of the world**, that is:

- the increasing number of **smart devices**
- the growth of multimedia services demanding **higher data rates**
- the emergence of **applications with specific requirements** concerning throughput *(on page 42)*, latency *(on page 40)* and reliability
- the emerging **5G era new use cases**, such as autonomous vehicles, 4.0 industry, augmented reality

Figure: Drivers of Network Slicing



Benefits for Communication Service Providers:

Digitalization of the world provides incredible opportunities for Communication Service Providers *(on page 39)* to grow and increase their revenues. With the emergence of new 5G era use cases, Communication Service Providers have **new revenue opportunities**. They can offer new services to old customers at a premium or address new customers and create new revenues.

Chapter 4. Business Benefits

Network Slicing allows service providers to offer customized services, introduce new services, and create new revenues. Businesses benefit from network slicing by getting the same services as a private, dedicated network, but in a more cost-efficient manner.

Benefits for Service Providers *(on page 41)*:

Network Slicing gives service providers the following capabilities:

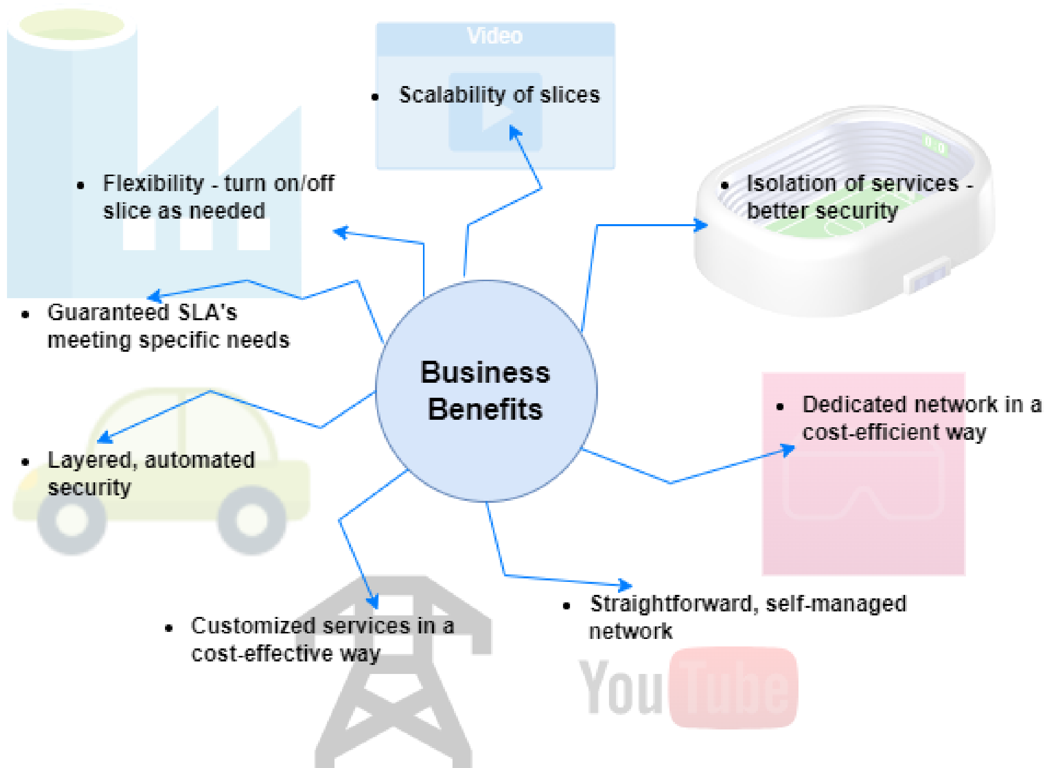
- **creating specific, customized services**
- offering SLA *(on page 41)*meeting specific needs, **in a cost-effective way**
- introducing **new services to old customers** at the premium
- **addressing new customers** and creating new revenues

However, the efficient management and growth of revenues can be achieved only by **the automation of all processes throughout the life cycle of the slice.** Service Providers need to have, at their disposal, ready templates of slices, that can be deployed off-the-shelf.

Benefits for Businesses:

Businesses are offered network slicing as a platform, which provides a variety of benefits. The Figure: Business Benefits below presents the most significant benefits of network slicing for businesses. Furthermore, network slicing offers the same services as a private, dedicated network, however, **in a more cost-efficient manner.**

Figure: Business Benefits

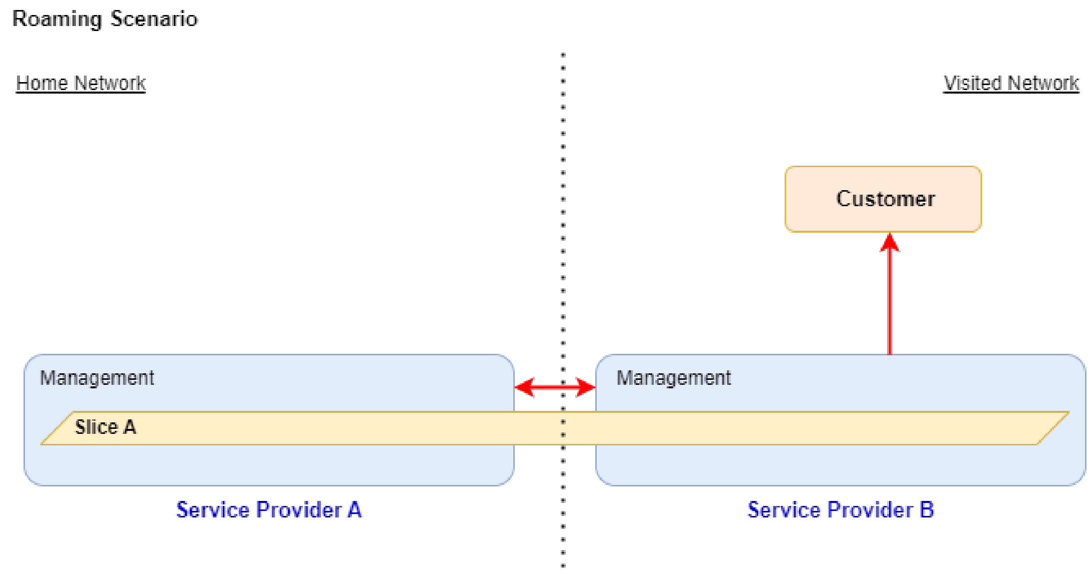


Cooperation of Service Providers

Multiple digital service providers *(on page 40)* can cooperate to fulfill customer's requirements, even if one provider cannot meet all of them. This cooperation is required to extend network capabilities and should be specified in the SLA. *(on page 41)*

Even if customer's requirements extend over the capabilities of one digital service provider, it is possible for **multiple digital service providers to cooperate in fulfilling customer's needs**. For instance, such cooperation is envisioned in roaming scenario as well as in case of specific business vertical's request, as shown in Figure: Cooperation of Service Providers. In both cases, the need of extending network capabilities from one provider to the other has to be specified in SLA *(on page 41)*.

Figure: Cooperation of Service Providers

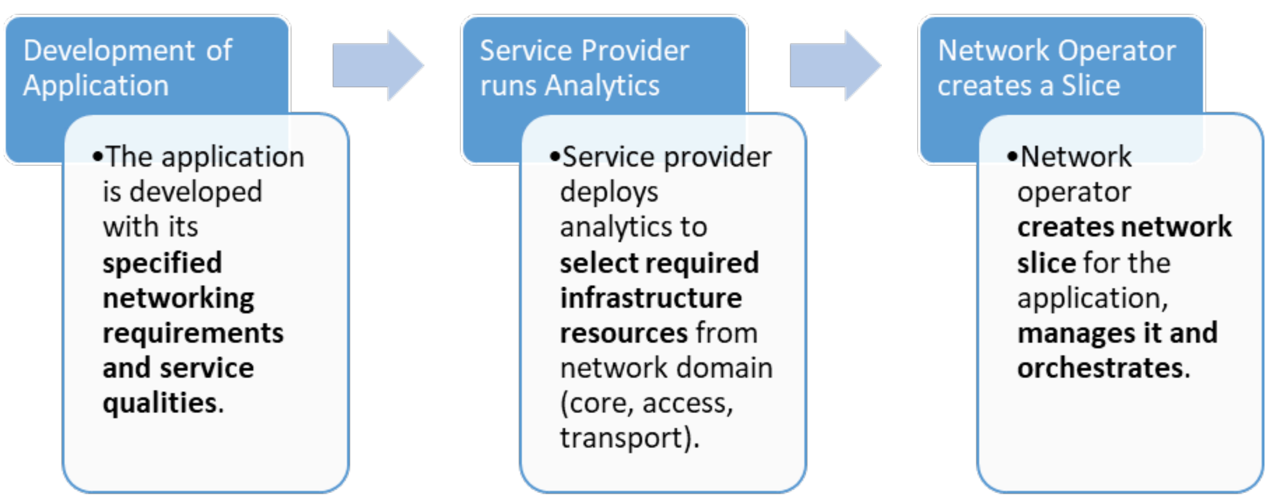


Cooperation of Network Operators and Service Providers

The need of efficient **cooperation between** network operator *(on page 41)* **and** service providers *(on page 40)* is evident in case of **creating application-aware network slices.**

Creating Application-aware Network Slices:

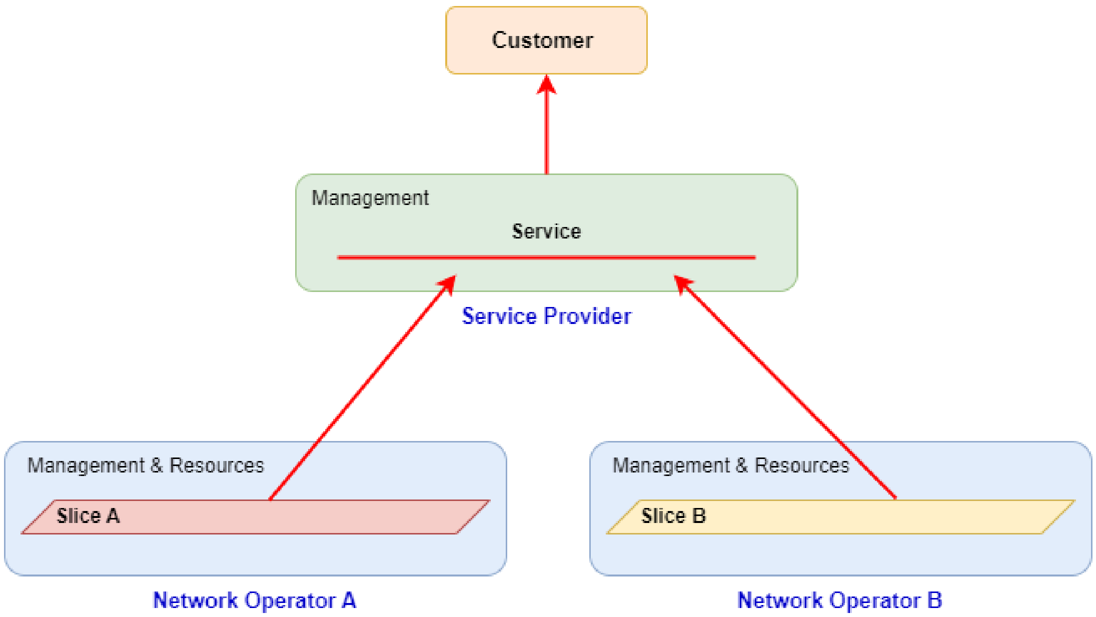
The 5G system introduces the following, revolutionary approach to creating application-aware slices:



Cooperation of Network Operators and Service Providers:

Digital service providers can browse and select form the wide range of network services provided by different network operators. As shown in Figure: Cooperation of Network Operators and Service Providers, digital service provider *(on page 40)* **combines network resources and functions from network operator A and B** to create a comprehensive digital service based on customer's SLA. The management and orchestration of this service is performed by the service provider.

Figure: Cooperation of Network Operators and Service Providers



Chapter 5. Use Cases

Network slicing plays the vital role in enabling mobile operators to offer new services to end consumers andbusiness verticals *(on page 39)* . The following section presents **three verticals with the greatest potential to benefit from network slicing**, that is Smart Factory, Smart City and Autonomous Vehicle.

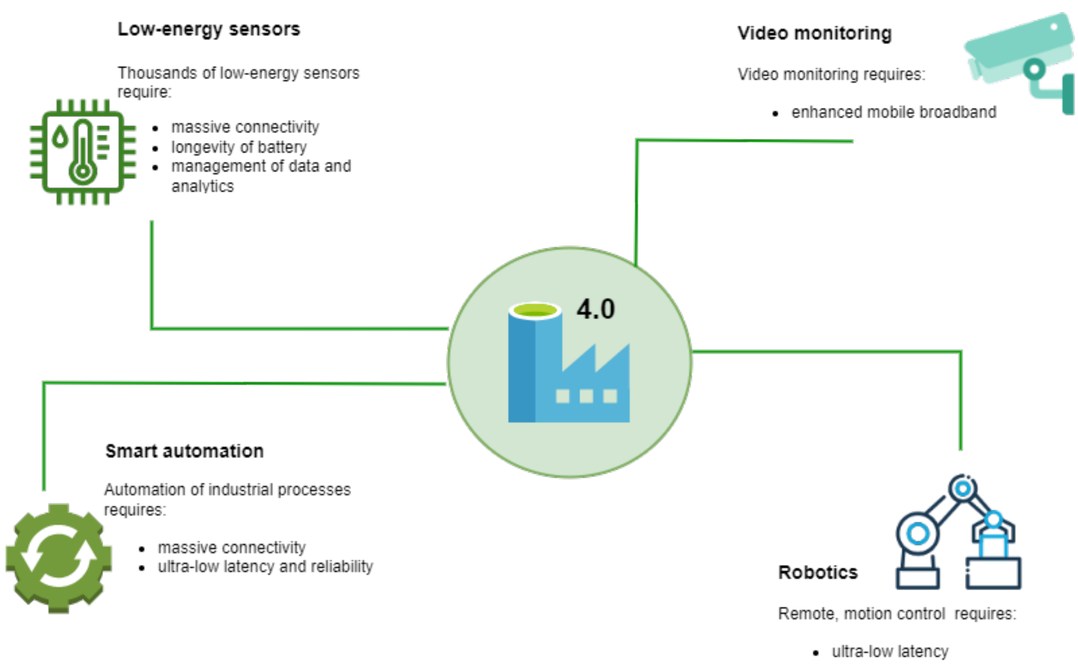
Smart Factory

Network slicing provides industrial businesses with the ability to quickly respond and adapt to changes, allowing for the addition of new slices without disrupting existing ones. This technology is capable of embracing the complexity and heterogeneity of industrial networks, enabling digitalization and speeding up the fourth Industrial Revolution.

Challenges of Smart Factory Network:

Recent **digitalization of factories** imposed a significant requirement on factory network to ensure high-performance and support critical process with a high speed and accuracy. As Figure: Smart Factory 4.0 shows, **each part of the factory** has been digitally transformed and **imposes divergent requirements** on a factory network.

Figure: Smart Factory 4.0



Capabilities of Network Slicing:

Network slicing is a way to manage the growing complexity of the factory networks and fulfil those requirements by introducing **rapid programmability and scalability**.Table: Smart Factory: Challenges and Capabilities presents the main problems new generation of factories faces and the solutions to them provided by Network Slicing.

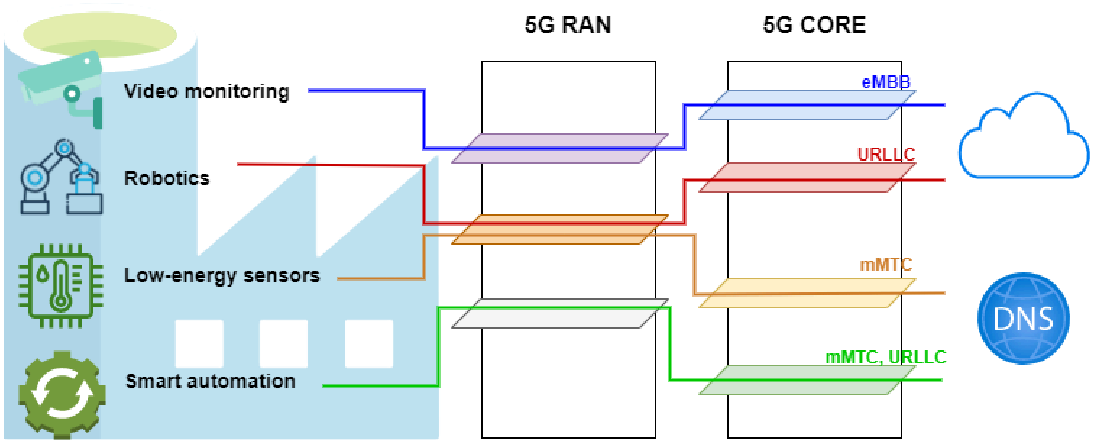
Table 1. Smart Factory: Challenges and Capabilities

Challenges	**Solutions provided by network slicing**
• thousands of sensors require **connectivity and longevity** • **massive amount of data** must be collected, analysed and managed • video monitoring requires **eMBB *(on page 40)*** • remote control requires **URLLC *(on page 42)*** • need for rapid **changeovers** • need for **flexible** network environment	• IoT *(on page 40)* analytics servers enable **storage and analysis of the data at the edge** • network slicing **offered as a service** • network slices tailored in line with **guaranteed SLA *(on page 41)*'s meeting needs** • **rapid network slice creation** and management • **automation** of slice lifecycle management • network slices **independently secured** • **flexible sharing of the resources** among slices

Network Slices in Smart Factory:

As the complexity of the factory network increases, **assigning each part of the network to a single, customized slice** seems to be the best solution to ensure efficiency and seamless performance of the factory. **Each part of the industrial site can be served by the specific, independent slice**, which is designed to satisfy the demands of a specific part of the factory network. For instance, one slice can provide high bandwidth and data collection of CCTV, another one can ensure seamless connectivity between multiple sensors, as shown in Figure: Smart Factory Network Slices.

Figure: Smart Factory Network Slices



Management of Slices in Smart Factory:

Network slicing is offered to industrial businesses as a service: it provides the capability to **instantly respond and adapt to the emerging changes** at the factory site.

For instance, in case of installation of new monitoring system, the management of a factory can on-demand **add a new, autonomous slice without the disruption to the already existing slices**. It can specify the slice's features in such a way to meet the requirements of a monitoring system and ensure its efficient functioning within the network. It can also rapidly turn off/on other slices if needed.

Smart City

A Smart City's network manages and analyzes data from millions of devices. It requires a consistent and reliable data flow. Network slicing enables the creation of separate network slices that can be managed independently, ensuring that an overload or problem in one slice will not affect the others.

Challenges of Smart City Network:

Smart Cities deploy **millions of devices that collect data** on weather, energy usage, waste management, and traffic. The majority of this **data is collected in real-time, analyzed**, and used to predict future events or instantly respond and adapt to changing circumstances.

Thus, Smart Cities must rely on a **consistent and dependable data flow** to support a range of services on a shared physical network.

Capabilities of Network Slicing

Table: Smart City: Challenges and Capabilities overviews the basic challenges Smart Cities face and solutions offered by network slicing.

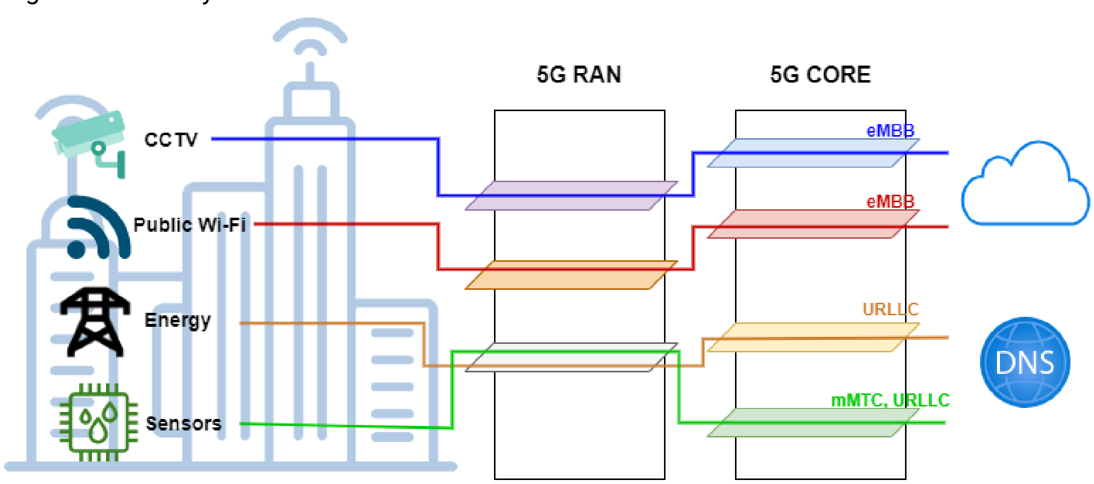
Table 2. Smart City: Challenges and Capabilities

Challenges	Solutions provided by network slicing
• millions of **connected devices** • massive amount of **data collected in re-al-time** • automatic **analysis of data** • uninterrupted, **reliable data flow** • **isolation** of services • **diverse requirements** of different services	• enabling **different levels of quality, latency, security and priority** for each network slice • IoT *(on page 40)* analytics servers en-abling storage and analysis of data • separate, **dedicated network slices** deprived of any interference from one another • **automatic selection** of the most appropriate slice to meet service level agreement (SLA) *(on page 41)* • network slices **independently secured**

Network Slices in Smart City:

Network slicing offers a solution by dividing a network into programmable slices that ensure optimal performance for various applications, as shown in Figure: Smart City Network Slices.

Figure: Smart City Network Slices



Management of Slices in Smart City:

More importantly, the slices can be managed, customized and secured independently, and **not affect one anotherif one slice is overloaded or down**.

For example, in the event of a disaster, Smart City residents can utilize their mobile phones to upload videos, photos, send messages, and make video calls. **Despite the significant increase in data flow,**

emergency calls are not impacted. The emergency calls are part of a separate network slice that has its own network connection, ensured traffic load, priority, and security. **Mobile operators can ensure a connection for emergency calls, even if the public network is overloaded.**

Autonomous Vehicle

Network slicing is crucial for autonomous vehicles, which demand ultra-reliable, low-latency communication, as well as a high level of security and the integration of multiple virtualized network slices.

Challenges of Autonomous Vehicle Network:

Autonomous vehicles constitute the greatest challenge for 5G network. It has to secure ultra-reliable communication between **thousands of sensors in the time-critical way**, provide **high bandwidth** ***(on page 39)*** for GPS and maintenance applications. as well as maintain **high level of integrity** and **security** of the whole vehicular network.

Capabilities of Network Slicing:

Table: Autonomous Vehicle: Challenges and Capabilities overviews the basic challenges faced by vehicles and the solutions provided by network slicing.

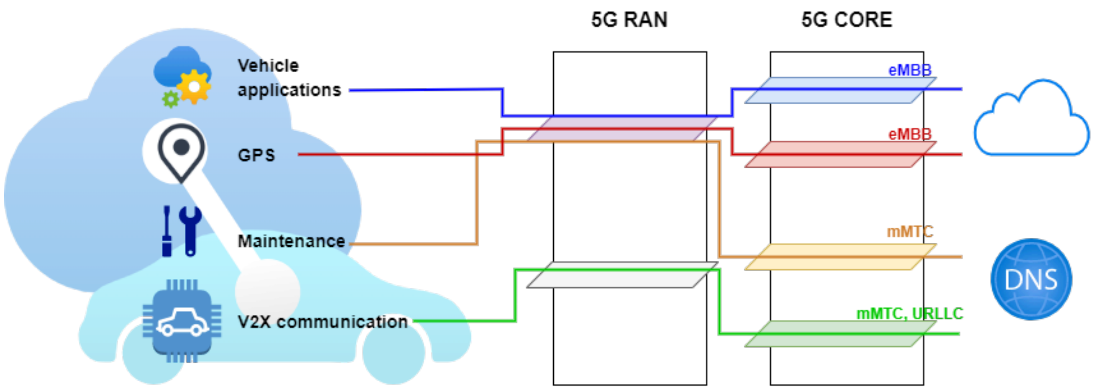
Table 3. Autonomous Vehicle: Challenges and Capabilities

Challenges	Solutions provided by network slicing
• thousands of sensors communicating in the **time-critical** way • **ultra-reliable** connection between sensors • **high bandwidth *(on page 39)*** for mainte-nance applications • **isolation** of services • separate security	• separate, **dedicated network slices for URLLC *(on page 42)* and eMBB *(on page 40)*** • **automated orchestration** of IoT *(on page 40)* slices • **IoT analytics servers** enabling storage and analysis of data at the edge • **Enhanced security** with **isolated network slices** supporting different services

Network Slices in Autonomous Vehicle:

One single, physical network is incapable of fulfilling all the above requirements. Instead, we need **multiple, virtualized network slices**, where, for instance, on slice can be for ultra-reliable, low-latency communication *(on page 42)* and separate one to support bandwidth-hungry applications, as shown in Figure: Autonomous Vehicle Network Slices. Thus, network slicing is the most fundamental enabler of autonomous vehicles.

Figure: Autonomous Vehicle Network Slices



Chapter 6. Network Slicing Architecture

Implementing all the 5G requirements, that is extreme throughput, ultra-low latency and reliability, massive connectivity, called for the design of the network architecture as a **dynamic, flexible and consistent framework of advanced technologies.** The architecture of network slicing is primarily based on the **Recursive Model,** in which each part of the network can be instantiated and placed in a different location. Still, the essential technological advancement enabling instantiation of network functions is **Network Function Virtualisation (NFV).** Both of these technological solutions are reviewed here, respectively.

Recursive Model of Network Slicing

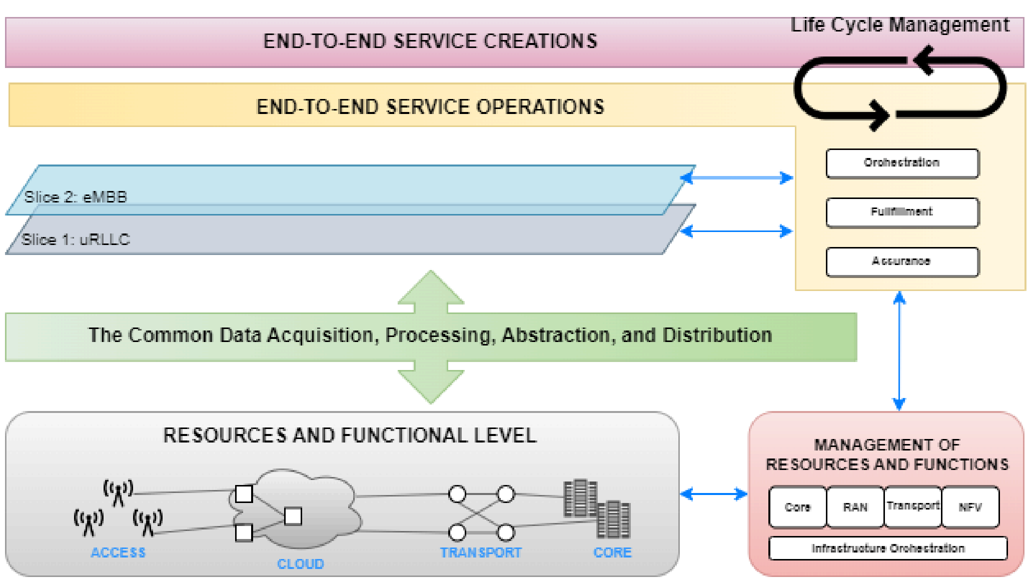
The 5G network architecture is based on a recursive model that allows for repeated instantiation and deployment of network services. This enables the creation of customized and independent network slices with dedicated resources and security.

Recursive Model:

The architecture of 5G network is based on the recursive model, presented in Figure: 5G Recursive Model. It allows **each part of the network service to be instantiated and deployed repeatedly at different places**, at the same time.

In this way **one service can be created out of the instances of already existing services or their parts**. In the virtualized infrastructure, the recursive model allows the creation of the customized and independent instances of the network, i.e. "slices", with dedicated resources and their own security.

Figure: 5G Recursive Model



The following section reviews the underlying parts of the Recursive Model:

1. At **Resources and Functional Level**, network functions and resources are **virtualized**.
2. **Management of Domain Resources & Functions** orchestrates virtualized resources and allocates them on network slices.
3. **The Common Data Acquisition, Processing, Abstraction, and Distribution** platform allows the access of data from all different levels.
4. At the **Service Level**, operators map the customer's requirements to the network infrastructure, to ensure that customers' needs are fulfilled..

The following section reviews the underlying parts of the Recursive Model.

1. At **Resources and Functional Level**, network functions and resources are virtualized.
2. **Management of Domain Resources & Functions** orchestrates virtualized resources and functions,ensures their fulfilment and assurance on each slice. It functions as an intelligence which has all the rules and policies. It allows the design of independent and customized network slices.
3. **The Common Data Acquisition, Processing, Abstraction, and Distribution** platform allows the access of data from all different levels. The data is connected to network resources, subscribers, network slices, applications or service instances. The platform ensures that the acquired data is processed, abstracted and distributed in network infrastructure.
4. The **Service Level** involves all the phases of network slice life cycle, i.e.: preparation, instantiation of resources and functions, configuration to meet customer's requirements, activation and decommission. At this level, operators perform the mapping of the customer's requirements and network infrastructure to ensure the fulfilment of customers' needs.

Network Function Virtualization

The 5G communication networks manifest a strong dependency on software. Thus, network functions virtualization (NFV) is envisaged as a fundamental technology enabling fulfilment of the 5G requirements.

Definition:

The major purpose of virtualization is to **abstract the resources from hardware and allocate them in a virtual environment**, which allows them to run in virtual machines or containers.

The virtualization of network functions *(on page 41)* offers the following benefits:

- ensures flexibility, scalability of network functions
- ensures rapidity of deployment of network functions
- provides efficient use of resources
- ensures rapid innovation and testing
- reduces cost and energy usage

Network Slicing and Network Function Virtualization:

The virtualization of functions and resources is the fundamental technological solution that enables network slicing. Virtualized network functions (NFs) are highly scalable and programmable. More importantly, they can be accommodated on each network slice, isolated, and secured independently.

Therefore, operators are provided with the unique capability to deliver different types of services on the common infrastructure, in a cost-effective way.

In the following sections, we present the virtualization of data, and network functions in RAN *(on page 41)*, respectively.

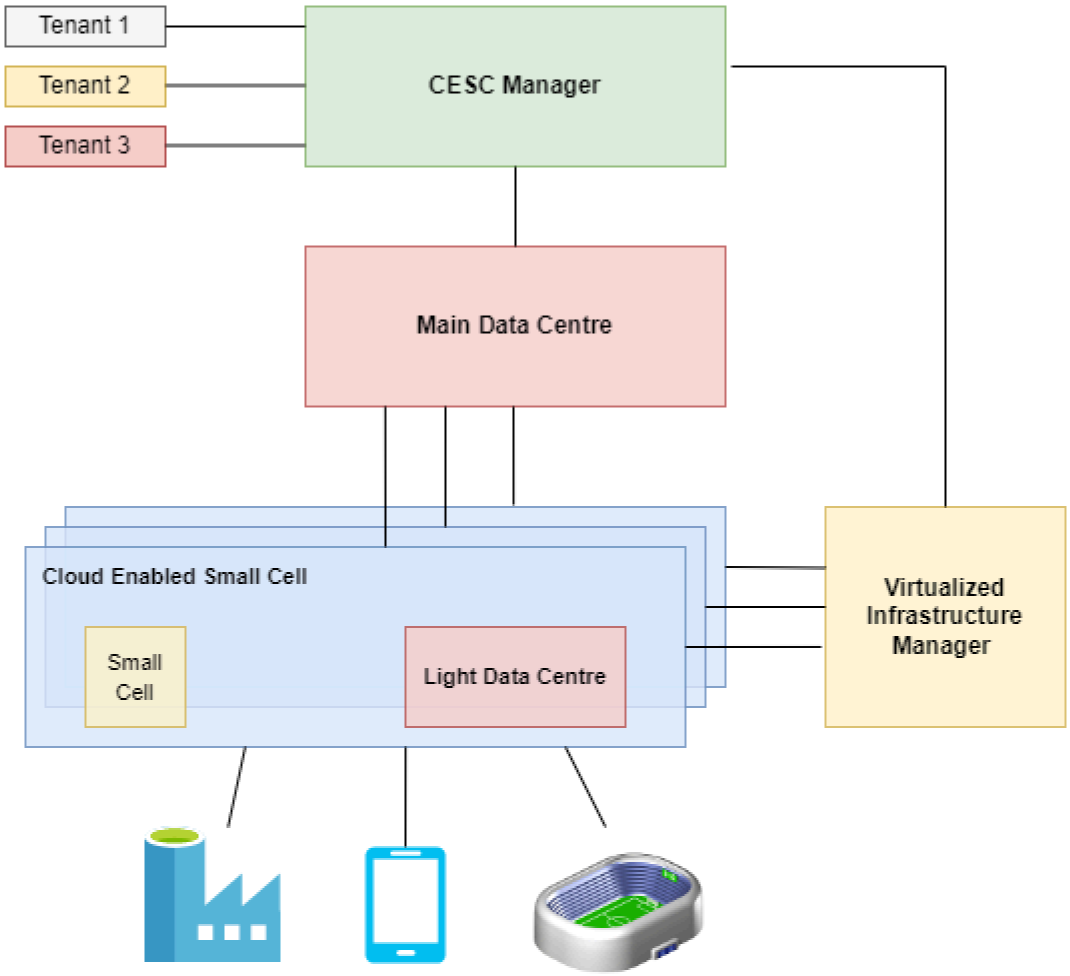
Virtualization of Data Plane

The software-defined networking virtualizes, allocates, and manages networking resources. In a Cloud Enabled Cell *(on page 39)*, this is done through the CESC Manager.

Data Plane Virtualization in Cloud Enabled Small Cell:

The data plane refers to the computing, storage and networking resources *(on page 41)* at the radio access node, edge cloud and data centre *(on page 39)*. **The virtualization of those resources, their management and orchestration is performed using software-defined networking (SDN)** ***(on page 42)*** **approach**. The example of this approach in cloud enabled small cell (CESC) is presented in Figure: Cloud Enabled Small Cell.

Figure: Cloud Enabled Small Cell



1. The **CESC Manager** initiates SDN operations.
2. The resources from **Main Data Centre** are virtualised and instantiated into separate **Light DC** and allocated at the **Small Cell** to form the **CESC**.
3. The **Virtualized Infrastructure Manager (VIM)** controls all clusters of CESC. It is responsible for analysing, monitoring and optimizing the resources from all distributed CESCs.
4. The resources from CESCs are used to design a network service in the SLA *(on page 41)*-complaint manner and offered to a **tenant.**

The **CESC Manager** not only supervises and orchestrates the infrastructure of CESC, but also acts as a middleware between the infrastructure and **tenants** ***(on page 42)***. More specifically CESC Manager:

- controls and orchestrates Light DC and small cell functions in the cloud environment
- organizes the allocation of radio resources

- coordinates the delivery of the services according to SLA *(on page 41)*
- ensures the service fulfils SLA requirements

Virtualization of Radio Access Network

The virtualization of RAN *(on page 41)* control functions as APPs in the Controller Layer allows for the cooperation of multiple links and low latency communication between vehicles in close proximity, making autonomous vehicles possible.

Requirements of 5G Radio Access Network:

The transport network domain in 5G system must implement such technological solutions which:

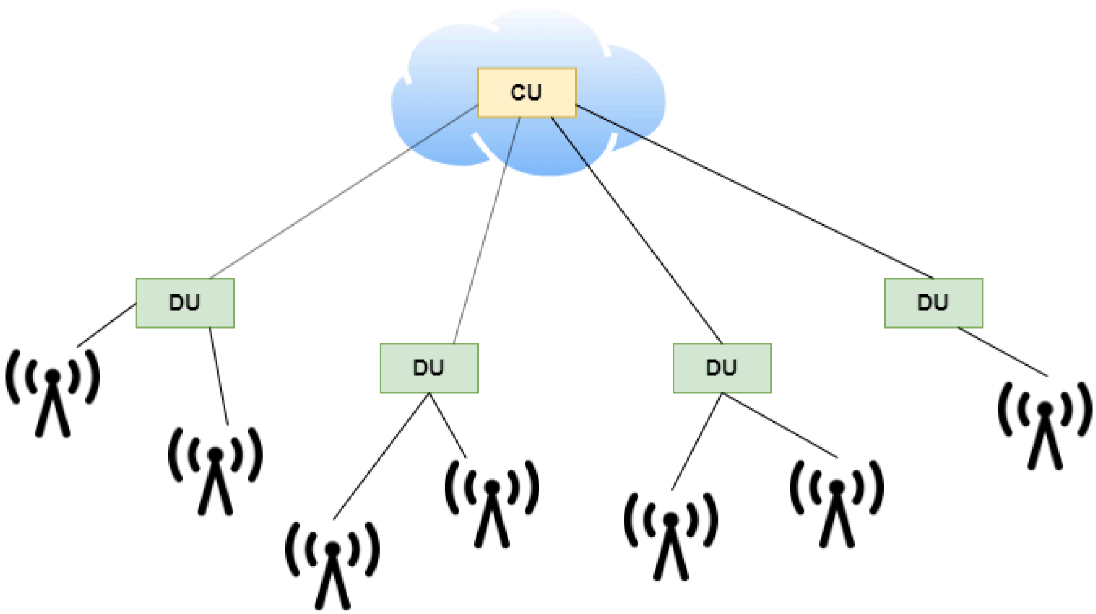
- provide high level of flexibility
- allocate the radio resources efficiently
- balance and optimize the traffic load

Functional Split:

The baseline architecture of 5G RAN is comprised of the **functional split *(on page 40)***, i.e. baseband unit (BBU) *(on page 39)* is split into distributed unit (DU) *(on page 40)* and centralized unit (CU) *(on page 39)*.

The significance of this architecture relies on the fact that **multiple DUs can be connected to one CU** located at next generation Node B (gNB) *(on page 40)*, as shown in Figure: Functional Split *(on page 40)*.

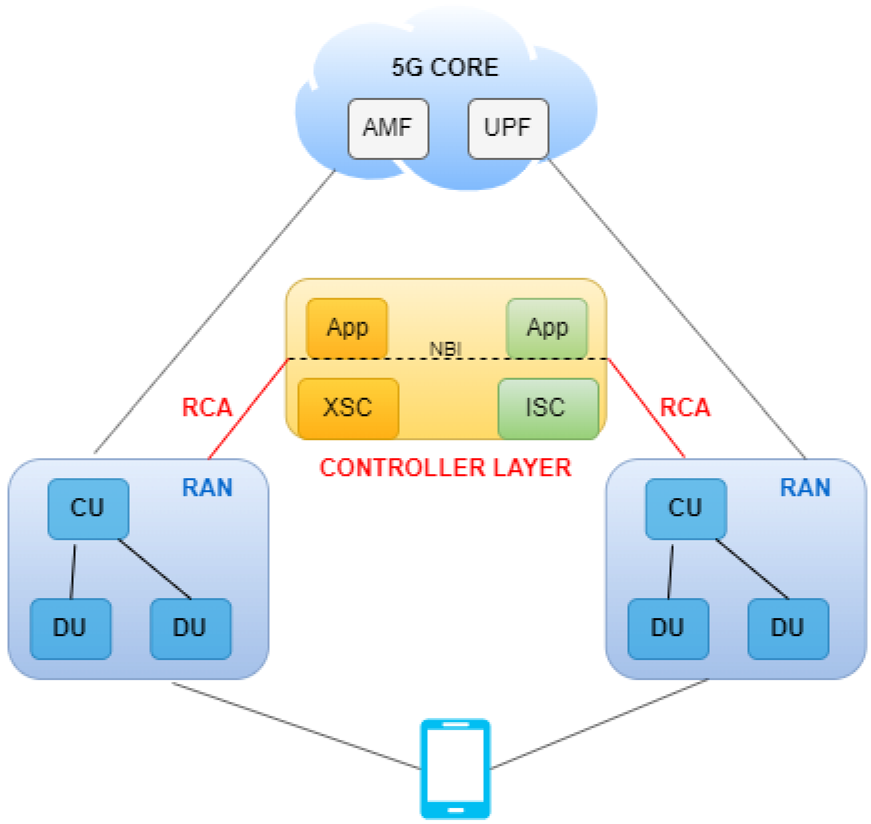
Figure: Functional Split



Architecture of RAN Virtualization:

The disaggregation of CU from DU allows certain **CU functionalities to be virtualized.** The **virtualization of RAN control functionalities** in CU, as applications (**APPs**), is provided by the **Controller Layer**, as it is shown in Figure: The Architecture of 5G RAN with the Controller Layer.

Figure: The Architecture of 5G RAN with the Controller Layer



The **APPs** can run on northbound interface (**NBI**) over corresponding cross-slice (**XSC**) and intra-slice controller (**ISC**).

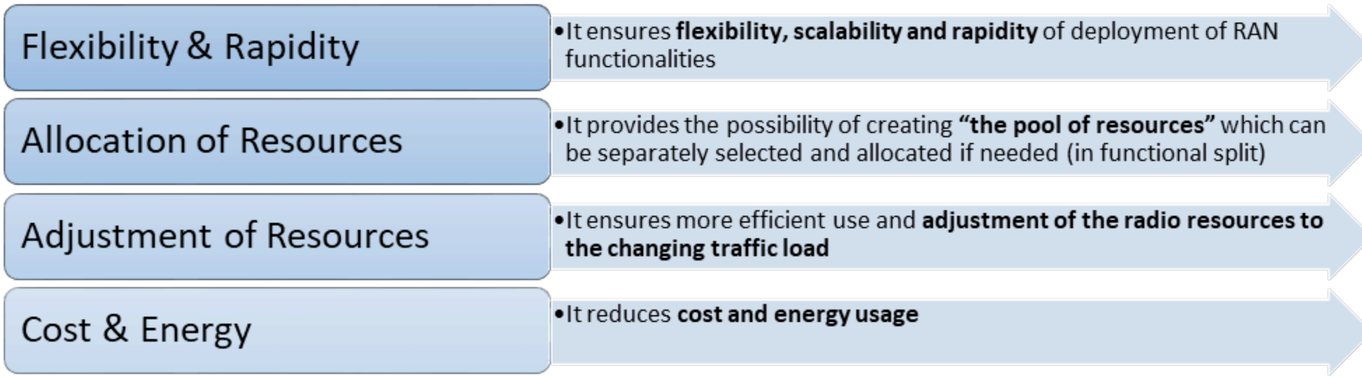
The **Controller Layer** is connected with gNB via the southbound interface (**SoBI**) and communicates with it through the **RAN Controller Agent** (RCA), which is situated in gNB and acts as a middleware.

In fact, each virtualized NF *(on page 41)* in CU or DU is connected to the Controller Layer via RCA. **The capabilities of RCA** involve the following:

- collects most recent information from NF *(on page 41)*s,
- exposes the data from NFs to the Controller Layer,
- assembles the analytics from UE *(on page 42)*s and RAN, regarding channel quality, power level, radio resource usage etc., and forwards it to the Controller Layer,
- interfaces the re-configuration information from the Controller Layer to NFs in CU and DU.

Benefits of RAN Virtualization:

The virtualization of RAN functionalities in CU introduces the following **benefits**:



More importantly, the virtualization of some RAN control functions as applications (APPs), in the Controller Layer, makes it possible to satisfy the most challenging use cases, such as autonomous vehicles. It is achieved by introducing the cooperation of multiple links as well as creating local paths for low latency communication between vehicles, which are close to each other.

Chapter 7. End-to-end Network Slicing

Another significant feature of 5G network slicing is that it is **end-to-end, i.e. all the way from** user equipment (UE) *(on page 42)* **to** data center (DC) *(on page 39)*, **through separate and independent slices of access, transport and core.**

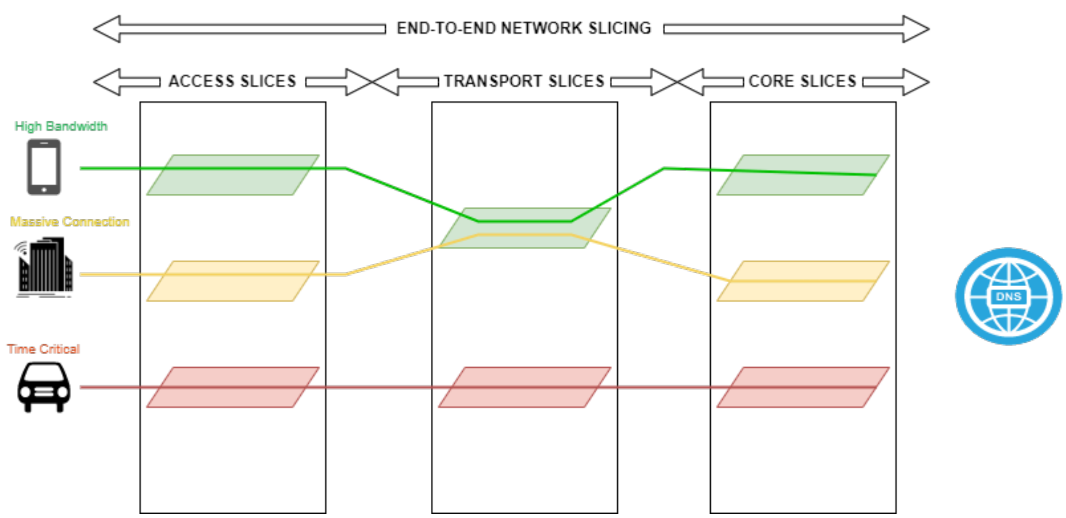
Definition:

End-to-end network slicing means that **each slice is composed of three independent slices: access, transport, and core**, as shown in Figure: End-to-end Network Slicing.

For instance, one slice may consist of a dedicated resources in access, shared resources in transport and dedicated resources in core. Whereas, another slice may require dedicated resources in all three parts of network slicing.

This enables efficient management and flexible allocation and scalability of the resources of each part of the network.

Figure: End-to-end Network Slicing



Benefits:

The importance of end-to-end network slicing is mainly stressed for **extreme flexibility** in order to provide Service Level Agreement (SLA) *(on page 41)***-based quality of experience in a cost efficient way**.

More importantly, automating the creation and registration of network slices significantly reduces Total Ownership Costs *(on page 42)*. With the increasing number of new use cases, manually creating, customizing, and managing the entire life cycle of slices would be expensive and time-consuming. Automating these processes essentially reduces TCO.

Access Part of Network Slicing

The 3GPP specification Release 15 for Next Generation-RAN *(on page 40)* allows for flexible and adjustable slicing in RAN *(on page 41)* to meet tenant requirements such as latency *(on page 40)*, data rate, and number of connections. Resources in RAN can be shared among different network slices while maintaining QoS *(on page 41)* requirements. One PDU *(on page 41)* session can be initiated for one slice instance, comprising of one or a number of QoS flows with mapped profiles and airtime.

Architecture of Network Slicing in RAN:

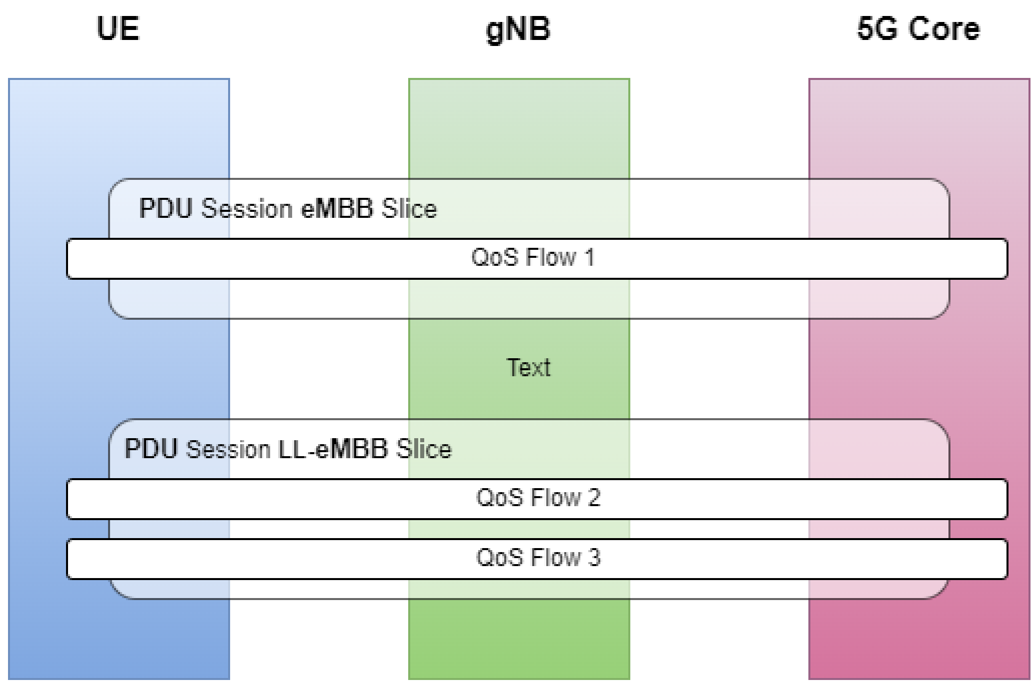
The slice awareness and slice selection is provided in radio access network (RAN) by the 3GPP specification: Release 15 for Next Generation-RAN. In line with this, slicing in RAN is envisioned as **extremely flexible and adjustable.** It foresees a number of scenarios to satisfy the requirements of tenants *(on page 42)*, such as: latency, data rate or number of connections. In some of these scenarios**, the resources in RAN can be shared among different network slices *(on page 41)*** provided that all the quality of service (QoS) requirements are maintained.

Slicing in RAN is highly flexible and adjustable. Depending on the situation, resources in RAN can either be shared among different network slices or dedicated to a specific slice.

As shown in Figure: Network slicing in RAN, **one PDU session can be initiated for one slice instance**, namely the enhanced mobile broadband (eMBB *(on page 40)*) and low-latency enhanced mobile broadband (LL-eMBB).

One PDU session can comprise of one or a number of QoS flows. **The incoming QoS flows, on the basis of their QoS profiles, are mapped to specific data radio bearers (DRBs) *(on page 40)* with assigned profiles and airtime.** It also provides mobility management and data encryption.

Figure: Network slicing in RAN



Core Part of Network Slicing

5G core architecture features **Control and User Plane Separation (CUPS)** ***(on page 39)*** **and Service Based Interface (SBI)** ***(on page 41)***. User Plane Function (UPF) *(on page 42)* is separated from other network functions and located closer to user equipment. Slice registration is performed by the cooperation of specific network functions in 5G core *(on page 39)*.**Network functions** ***(on page 41)*** **in 5G core can be either shared or dedicated** depending on the SLA *(on page 41)*-based requirements.

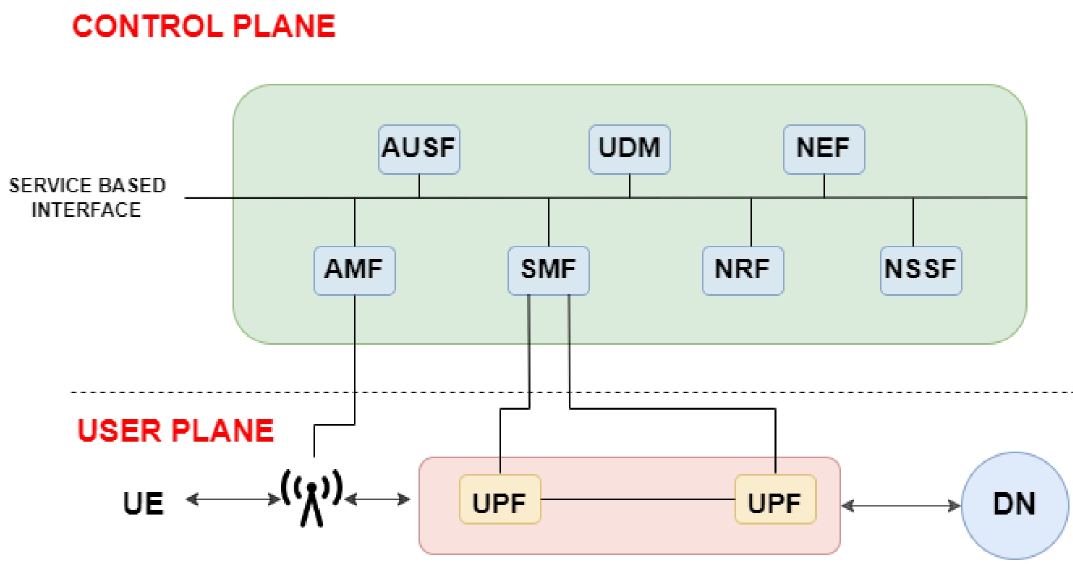
Architecture of 5G Core:

One of the most fundamental characteristics of 5G core architecture is **Control and User Plane Separation (CUPS)** and **Service Based Interface (SBI).**

More specifically, as shown in Figure: 5G Core Architecture, **User Plane Function**, (UPF) **is separated from other network functions in 5G core and located closer to UE** ***(on page 42)***. The significance of UPF's location and proximity to UE lies in its responsibility for initiating a protocol data unit (PDU) *(on page 41)* session for incoming QoS flows *(on page 41)*. The closer UPF is to the UE, the faster its reaction time will be for incoming QoS flows, resulting in lower latency.

The remaining network functions provide rules and policies for PDU session, while **freely communicating with each other**.

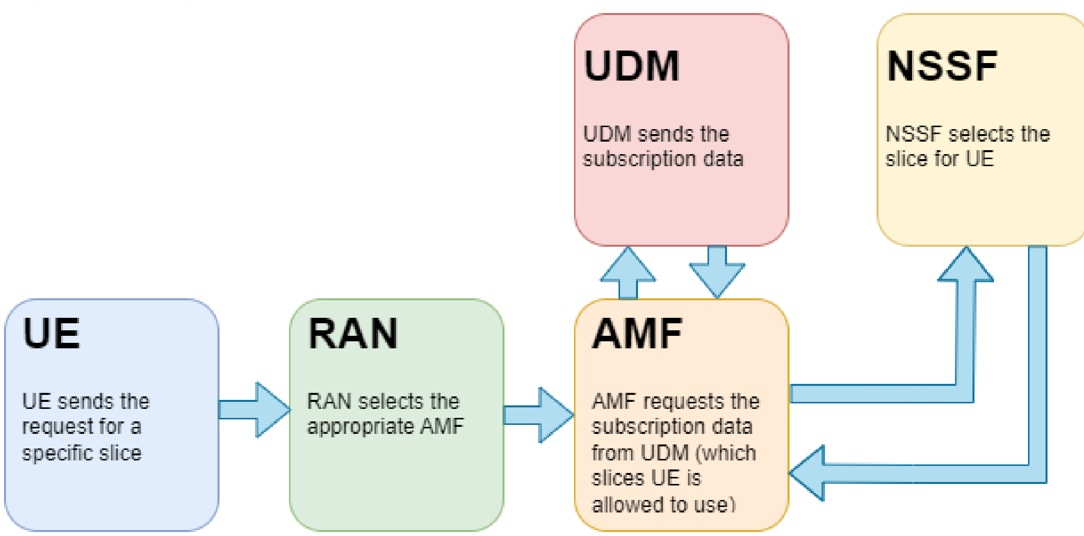
Figure: 5G Core Architecture



Network Slice Creation and Registration:

Slice creation and registration is mainly performed by the mutual cooperation of such network functions as: access and mobility function (AMF) *(on page 39)*, unified data management (UDM) and network slice selection function (NSSF), as shown in the Figure: Slice Registration.

Figure: Slice Registration



Network functions in 5G core can be either shared or dedicated depending on the SLA-based requirements. In most use cases, session management function (SMF) *(on page 41)* and UPF are dedicated for a specific slice.

Transport Part of Network Slicing

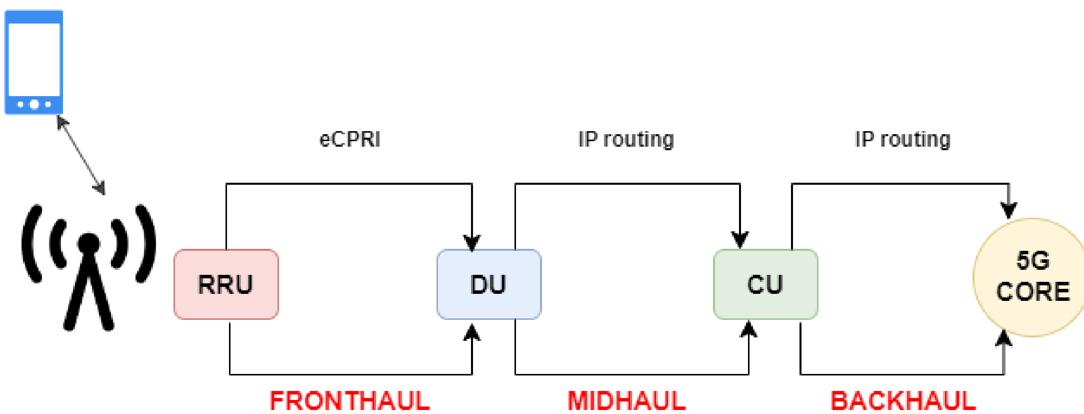
The transport network's high modularity and intelligent traffic steering system allow for efficient transport slice creation and management.

Architecture of 5G Transport Network:

The transport network domain in 5G system must implement such technological solutions which **provide high level of flexibility, allocate the radio resources efficiently, balance and optimize the traffic load**. The proposed solution involves **functional split in RAN *(on page 40)***, which means dividing the baseband unit (BBU) *(on page 39)* into a distributed unit (DU) *(on page 40)* and a centralized unit (CU) *(on page 39)*. This technological solution is the response to the densification of the traffic and need for the greater flexibility to serve diverse users.

The transport network becomes **more modular and scalable**. The single **backhaul is divided into fronthaul, midhaul and backhaul**, as shown in Figure: Transport Network Architecture, which enables greater flexibility and diversification.

Figure: Transport Network Architecture



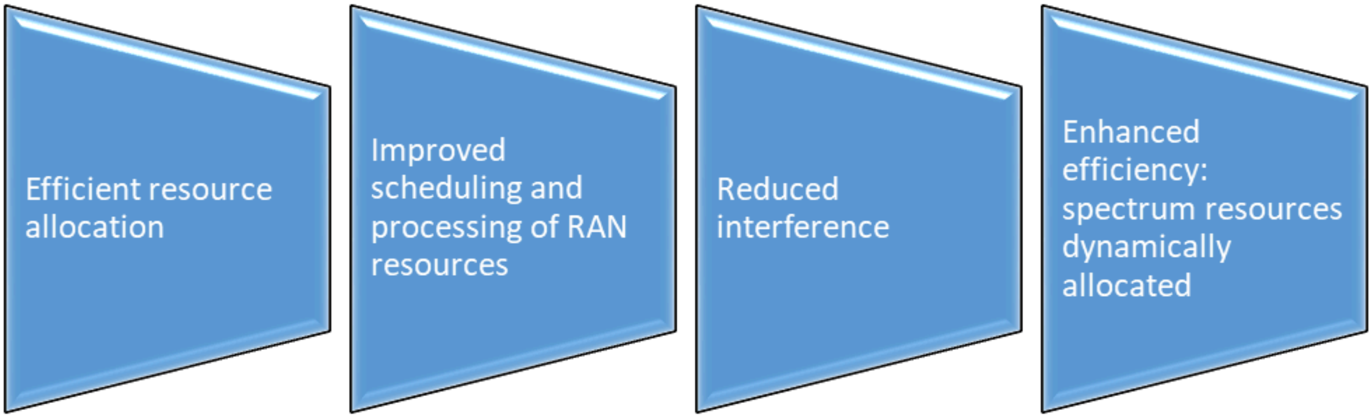
The connection between RRU and DU (i.e. **fronthaul *(on page 40)***) is standardised as common public radio interface (CPRI *(on page 39)*) or its enhanced version (eCPRI).

The interfaces between DU and CU (**midhaul *(on page 40)***) as well as CU and 5G core *(on page 39)* (**backhaul *(on page 39)***) are based on Internet protocol (IP) routing *(on page 40)*.

The primary responsibility of transport network is to provide connectivity between radio remote unit (RRU) *(on page 41)*, DU, CU and Core *(on page 39)*.

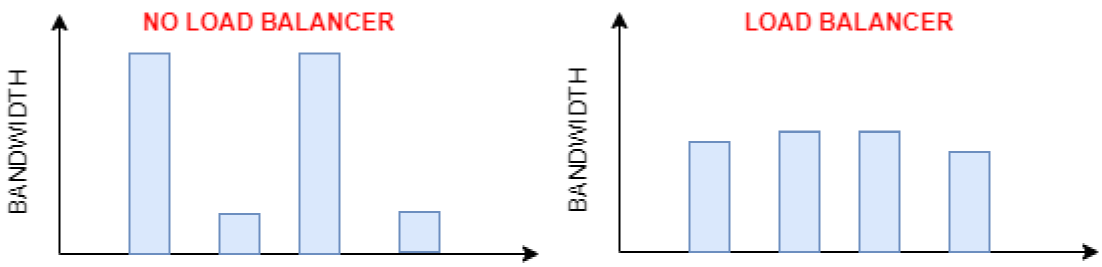
Benefits:

Higher level of granularity of the transport architecture offers the following benefits:



Network Slicing in Transport Network:

Since transport network management is primarily based on IP routing, **transport layer slicing** mainly involves traffic routing and prioritization. This is accomplished by the **load balancer**, such as software-defined networking (SDN) *(on page 42)*.



Software-defined networking (SDN) ***(on page 42)*** not only intelligently steers, maintains and balances the traffic, but it is also responsible for:

- creating and managing the transport slices
- optimizing transport slices
- orchestrating services and allocating resources on transport slices

Higher level of modularity of transport network along with the **intelligent traffic steering system** enable efficient transport **slice creation and management**.

Chapter 8. Glossary: Network Slicing Important Terms

5G Core - the heart of a 5G network, controlling data and control plane operations. The 5G core aggregates data traffic, communicates with UE, delivers essential network services and provides extra layers of security, among other functions.

Access Mobility Function (AMF) - a control plane function in 5G core network

Backhaul - portion of the network that connects a cell site with the core network

Bandwidth - a range of frequencies within a given band, in particular that used for transmitting a signal

Baseband Unit (BBU) - It is a device that interprets Baseband frequencies in telecom systems including computer networks, the internet, phone networks and radio broadcasting systems

Business vertical - are business niches where vendors serve a specific audience and their set of needs. Some broad examples of business verticals include insurance, banking, hospitals, retail, real estate, government, and more. Verticals can also be subcategorized into narrower niche markets.

Centralized Unit - It is a centralized unit connected between the 5G core network and the DUs. It provides the access control and coordinates high-level protocols including RRC (Radio Resource Control) from the control plane and IP, SDAP (Service Data Adaptation Protocol), PDCP (Packet Data Convergence Protocol) from the user plane

Cloud Enabled Small Cell - a miniature radio access point or wireless network base station with a low radio frequency power output, footprint and range.

Common Public Radio Interface – standard that defines an interface between Radio Equipment Control (REC) and Radio Equipment

Communication Service Provider - offers telecommunications services or some combination of information and media services, content, entertainment and application services over networks, leveraging the network infrastructure as a rich, functional platform.

Control and User Plane Separation - the complete separation between control plane functions (which take care of the user connection management, as well as defining QoS policies, performing user authentication, etc.) and user plane functions (which deal with data traffic forwarding)

Data Centre - a facility composed of networked computers, storage systems and computing infrastructure.

Data Radio Bearer - It defines how the UE data/signaling are treated when it travels across the network

Digital Service Provider - provides an online search engine, online marketplace or cloud computing service (either alone or in combination)

Distributed Unit - the main processing unit in Radio Access Network that is responsible for the High Physical, MAC, and RLC protocols in the RAN protocol stack

Enhanced Mobile Broadband - a concept that focuses on the speed, capacity and mobility to allow for new mobile uses such as high-definition video streaming and immersive augmented reality (AR) and virtual reality (VR) on the go

Fronthaul - the optical network that connects Remote Radio Units (RRUs) and Baseband Units (BBUs).

Functional Split - a new, flexible architecture for the 5G RAN, where the base station or gNodeB (gNB) is split into three logical nodes: the Central Unit, (CU), the Distributed Unit, (DU) and the Radio Unit, (RU), each capable of hosting different functions of the 5G NR stack

gNodeB (gNB) - a node in a cellular network that provides connectivity between user equipment (UE) and the evolved packet core (EPC). A gNodeB is the functional equivalent of a base station in a traditional cellular network

Internet Protocol Routing - a set of protocols that determine the path that data follows to travel across multiple networks from its source to its destination

Internet of Things - describes physical objects (or groups of such objects) with sensors, processing ability, software and other technologies that connect and exchange data with other devices and systems over the Internet or other communications networks.

Latency - the delay before a transfer of data begins following an instruction for its transfer

Massive Machine Type Communication (mMTC) - service in environment which has very high density of connected devices. mMTC focuses on applications that require a huge number of low-powered (IoT) devices operating over long-range with little to no human intervention. These autonomously communicating machine-to-machine (M2M) devices require low bandwidths and a long life.

Midhaul – portion of a network that connects DU and CU

Network Function - a functional building block within a network infrastructure which has well-defined external interfaces and behaviour

Network Function Virtualization - the replacement of network appliance hardware with virtual machines.

Network Operator - a company that provides its subscribers with internet and telecommunications services

Network Resources - any kind of device, information, or service available across a network. A network resource could be a set of files, an application or service of some kind, or a network accessible peripheral device

Network Slice - is a logically separated, self-contained, independent and secured part of the network, targeting different services with different requirements on speed, latency and reliability

Protocol Data Unit - a single unit of information transmitted among peer entities of a computer network

Quality of Service - any technology that manages data traffic to reduce packet loss, latency and jitter on a network, to guarantee its ability to dependably run high-priority applications and traffic under limited network capacity

Radio Access Network - is a major component of a wireless telecommunications system that connects individual devices to other parts of a network through a radio link.

Radio Unit - a mobile, stationary, or portable radio communications device, which communicates at certain air wave frequencies

Service provider - a vendor that provides IT solutions and/or services to end users and organizations

Service-based Interface - the control plane functionality and common data repositories of a 5G network are delivered by way of a set of interconnected Network Functions (NFs), each with authorization to access each other's services

Session Management Function (SMF) - the control function that manages the user sessions including establishment, modification and release of sessions, and it can allocate IP addresses for IP PDU sessions

Service Level Agreement - sets the expectations between the service provider and the customer and describes the products or services to be delivered

Software-defined networking - an approach to network management that enables dynamic, programmatically efficient network configuration in order to improve network performance and monitoring, in a manner more akin to cloud computing than to traditional network management

Tenant - a group of user in software and virtualization context

Throughput - the measurement of how much data is exchanged between the user and the server, over a given period.

Total Cost of Ownership - an estimation of the expenses associated with purchasing, deploying, using and retiring a product or piece of equipment

Ultra-Reliable Low-Latency Communication (URLLC) - a subset of the 5G network architecture, ensures more efficient scheduling of data transfers, achieving shorter transmissions through a larger subcarrier, and even scheduling overlapping transmissions. It supports highly important data transfer that requires low latency, such as self-driving cars and remote surgery

User Equipment - any device used directly by an end-user

User Plane Function - the function that does all of the work to connect the actual data coming over the Radio Area Network (RAN) to the Internet