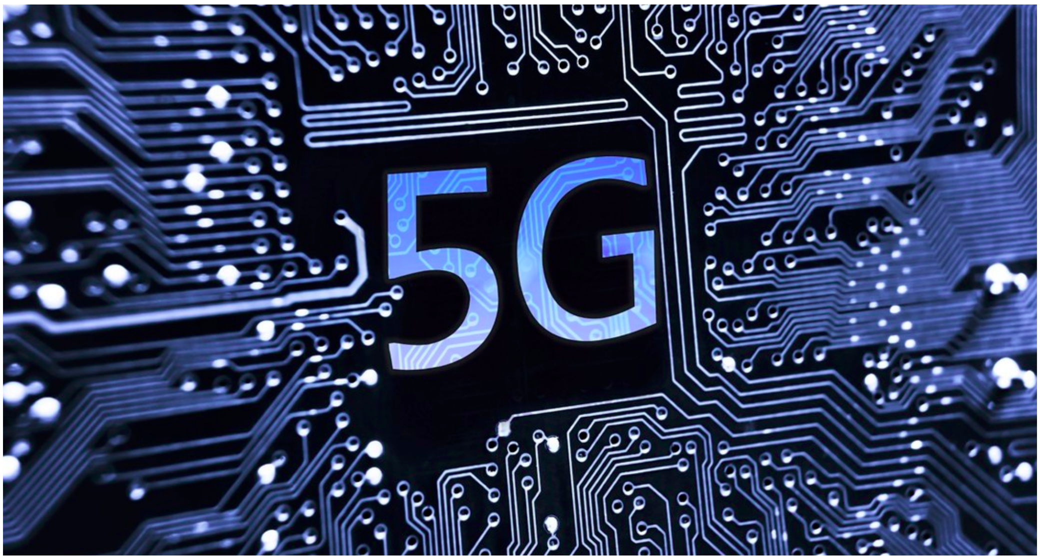


5G and Network Slicing for Layman

Published on June 10, 2018

Aayush Shrut
Consultant at Deloitte

Telecommunications is an evolving industry. From 1992 to 2016, GSM to 4G advanced, we have come a long way indeed evolving from Brick Mobiles to Smart Phones. However, in this 4th Industrial revolution, Telecommunications is evolving at much faster strides. To stay ahead, you have to move fast. Speed is what sells in this industry these days, and to keep the wheels churning at a faster rate, generations of researchers are creating standards to increase speed.

What is 5G? It's the evolution of Telecommunications to a whole new level, and at much faster speed. Why evolution? 5G aims to transcend connectivity from just the smartphones to **virtually almost everything in your home**. This digitalized and heavily connected world straight from Sci-Fi movies is dubbed as "Internet of Things" or IoT. 5G aims to serve IoT's demand of very high speed, connection autonomy and complex routing algorithms. Speed is just a by-product of 5G, as very low latency and consistent high throughput would be bread and butter for connecting such massive number of devices (20.8 billion by 2020 according to Gartner's report).

5G has not been rolled out yet, and standards are bein upon as of this writing. However, 5G is currently being feature tested behind back doors of almost all the major telecommunication giants such as Samsung and Qualcomm. Reportedly, Intel has also jumped in the 5G bandwagon. Samsung would field test it in Winter Olympics 2018, and by 2020 you can see the commercial deployment happening all around the world.

In this article, we try to establish the fundamentals of 5G, and why it is needed in the first place. We also explore the concept of Network Slicing, a key enabling feature which enables 5G.

Machines are Coming!

Currently, the actual download speed of LTE is fastest in Singapore at 37Mbps (India has 13 Mbps, better than the USA!). This is good enough speed to stream most of the HD content while on the go. So why is 5G needed in the first place? Going by the old saying, the thirst for money, power and speed never dies. And that's exactly the case here.

According to International Telecommunication Union's requirements for 5G, labelled as International Mobile Telecommunications - 2020 (or IMT 2020), there will be three distinct requirements for 5G. These three requirements will cover most of the use cases.

1.) **Massive Machine Communications**: A fancy term for the Internet of Things or IoT. It is a concept of connecting everyday objects with each other using the Internet, allowing them to constantly exchange and receive data. A smart house, smart car, smart window, smart "everyday objects", and the smartest city! Such a connected cybernetic augmented world would require one constant thing every single time, Fast Data! Lots and lots of data being exchanged in the cyberspace.

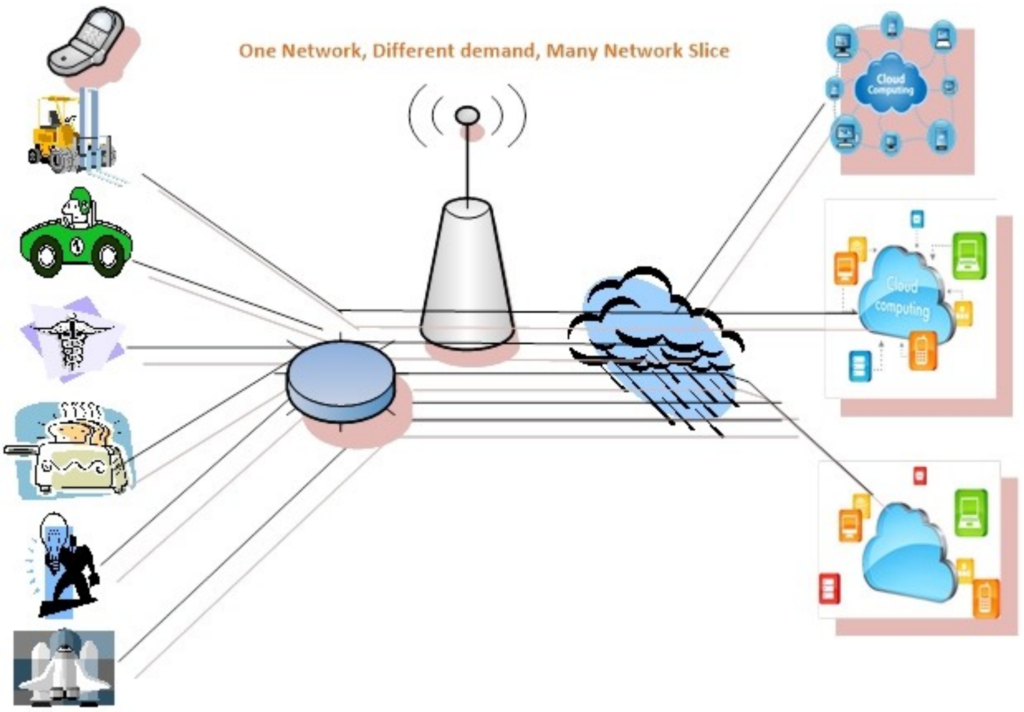
To keep up with such vast amount of data exchange, not only **high-speed low latency combo** is required, but also changes in the current network architecture is needed so as to accommodate the dense network topology that would be created by so many connections. Imagine Counter-Strike LAN party in your hostel room! Lots of LAN cables here and there, going snakeline through the room. Now imagine LAN party for your entire college! You possibly can't establish a messy network and would need to consider a sophisticated design so as to not trip anyone with the wires. The same thing goes for the demand of 5G.

2.) **Enhanced Mobile Broadband**: Your normal mobile broadband connection currently in 4G stage happily helps you stream all the kitten videos in ultra HD. But with the advent of 3D and Augmented reality, greater speed and more coverage would be required.

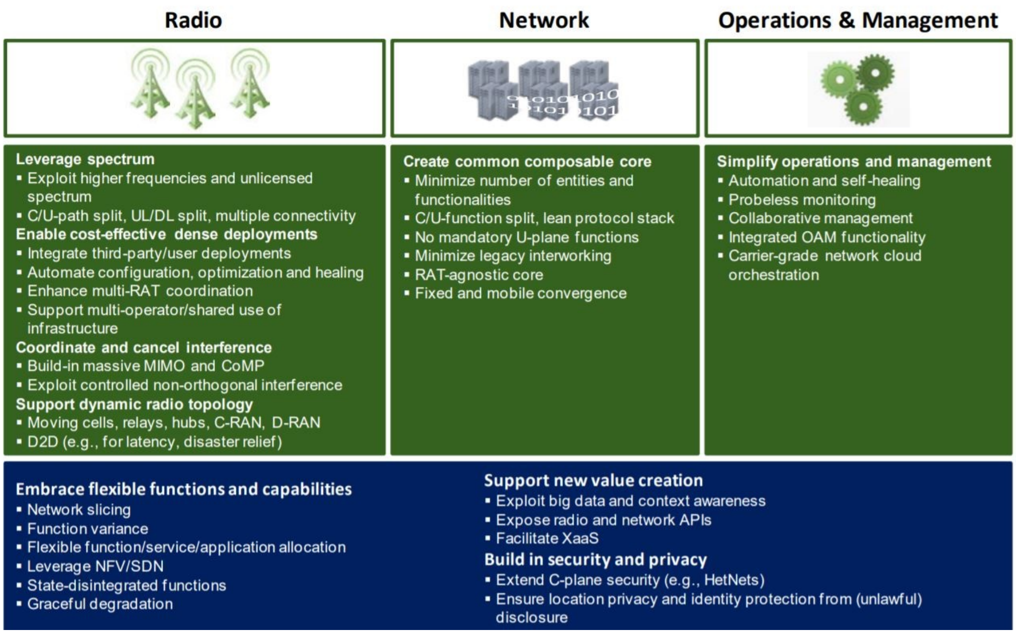
Speed can be achieved by following something called "**Demand Attentive Network**" (DAN) philosophy for broadband. DAN proposes an "on demand" service by undergoing changes in network regulation, infrastructure and technical standards. What it would do is that instead of always on high connectivity (which is constrained by infrastructure), it would do **load balancing** to give you the required bandwidth whenever you require. So the infrastructure of the guy who probably just streams some youtube videos before sleeping would be used to serve the demands of the torrent leecher who downloads 24/7 while watching movies in virtual reality, and vice versa! Theoretically, everyone will have an infinite bandwidth (1Gpbs+) and won't be constrained by infrastructure, a super efficient mobile network!

As for coverage, 5G would utilize "small cells" (or ultra small eNodeBs with a coverage area of 10m - 2kms). "Small cell" caters to providing continuous coverage to the mobile, especially in dense urban indoor areas (which according to research, 80% of time users stay inside). Such type of networks is also called **Ultra-dense Networks (UDN).**

3.) **Ultra reliable and low latency communications**: With 5G, comes a host of such use cases which require a constant high-speed connection that will never fail! Examples such as remote medical surgery, smart traffic controller, and other mission-critical services such as Rocket Launch! To support such services, a specialized network has to be created which would work dedicatedly for these services. You can't load balance them, and therefore a new type of network will have to be created. 5G will offer the ultra-low latency and reliable communication that such services will require. Such type of specialized networks is called **Ultra-reliable Networks (URN).**



So that's why we need to switch to 5G! With these three fundamental requirements, 5G evolves to more than just an upgrade to the speed of 4G. It would require massive changes and incorporation of all the latest technology. Design considerations for 5G is an ongoing research right now, but following lists the current design changes being proposed or discussed.



While the entire description and listing of the technologies that will enable 5G are beyond the scope of this article (we are particular technology should be dissected in detail. It's called "Network Slicing".

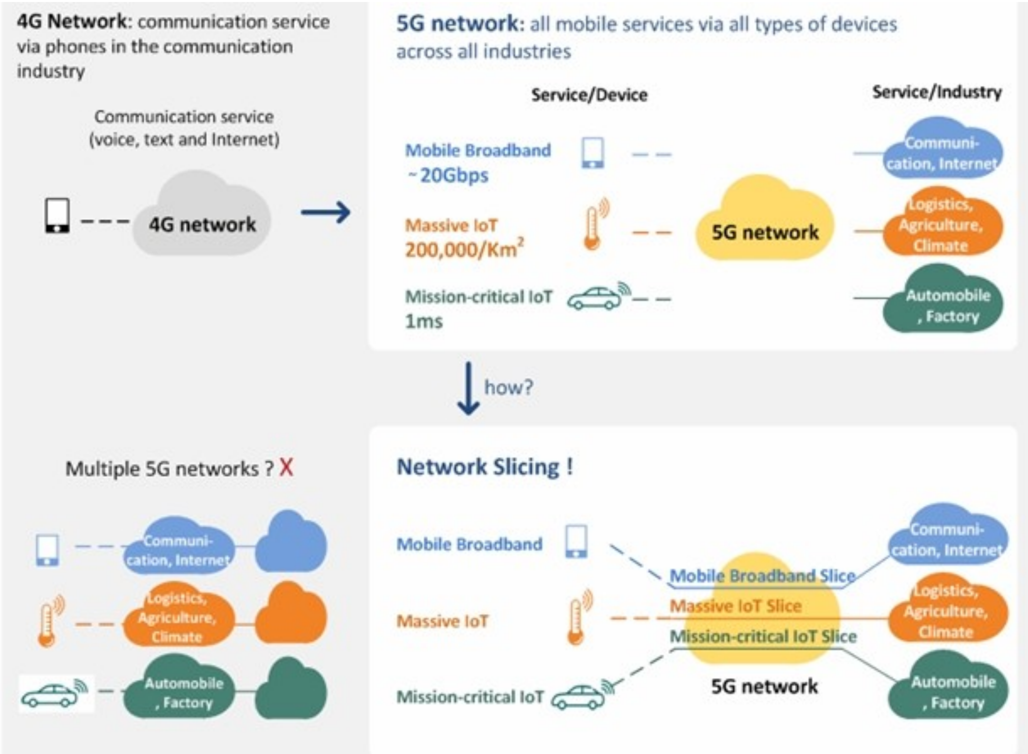
What is Network Slicing?

Network Slicing is a mechanism that could be used by operators to support multiple 'virtual' networks behind the air interface and across the fixed part of the mobile operator's network; both backhaul and core. This involves 'slicing' the network into multiple virtual networks to support different RANs (Radio Access Networks) or different service types running across a single RAN. This new technology lets operators slice one physical network into multiple, virtual, end-to-end (E2E) networks, **each logically isolated** including device, access, transport and core network (like partitioning an HDD into C and D drives) and dedicated for different types of services with different characteristics and requirements. For each network slice, dedicated resources (like resources within virtualized servers, network BW, QoS, etc.) are guaranteed. As slices are isolated from each other, an error or fault occurred in one slice does not cause any effect on communication in other slices.

In layman terms, Network Slicing enables operators to **create multiple networks using just one physical network**. With advances in virtualization and cloud technology, it is possible to create just one slice of the physical network, and dedicate that slice to realize the use case of one network (for eg: mobile broadband). Every actual physical hardware and resources allocated to this slice would be **logically isolated** from another slice of the physical network which would be serving another use case like IoT. The simplest analogy can be drawn to the concept of Virtualization. Just like you used one machine and created 10 different virtual machines, operators would be able to use one existing physical infrastructure and create 10 different networks from it!

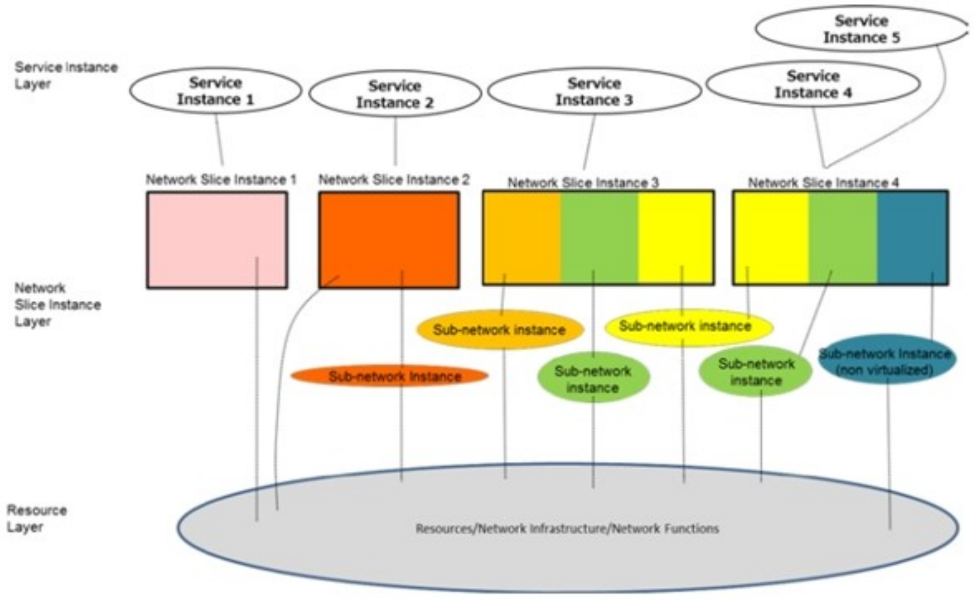
Why is network slicing necessary for 5G?

Why are some choices and options necessary in life than the other? Simple, it is cheaper! So far, mobile networks (of up to 4G), mainly serving mobile phones, have been optimized for phones only, at large. However, in the 5G era, they have to serve a variety of devices with different characteristics and needs. The 5G network, as discussed before, has to provide oodles of low-cost bandwidth at one end of the spectrum, as it were, while also providing low-power, l connections, video streaming optimized services and the oft-mentioned low latency, high speed, ultra-reliable premium services. We can't go on to create different physical networks for each of the mentioned use cases as the **cost implications would be substantial**. Therefore, Network Slicing comes into play to create different logical networks for different use cases using just one physical network.



Not all slices contain the same functions, and some functions that today seem essential for a mobile network might even be missing in some of the slices. The intention of a 5G slice is to provide **only the traffic treatment** that is necessary for the use case and **avoid all other unnecessary functionality**. The flexibility behind the slice concept is a key enabler to both expanding existing businesses and create new businesses. Third-party entities can be given permission to control certain aspects of slicing via a suitable API, in order to provide tailored service.

Key Concepts of Slicing:



As shown in the above figure, network slicing consists of three layers:

1.) Service Instance Layer: An instance of an end-user service or a business service that is realized within or by a Network slice. Typically services are provided by the network operator or by 3rd parties. Think of it as the layer that directly interacts with the user, the end user applications, the smartwatch, the smart car, the smart bathroom!

2.) Network Slice Instance: 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). It has following properties:

- A network slice instance may be fully or partly, logically and/or physically, isolated from another network slice instance.
- A Network Slice Instance may be composed of Sub-network Instances, which as a special case may be shared by multiple network slice instances. The Network Slice Instance is defined by a **Network Slice Blueprint.**
- **Instance-specific policies and configurations are required** when creating a Network Slice Instance.
- Network characteristics examples are ultra-low-latency, ultra-reliability etc.

The Network Slice Instance may be composed by none, one or more **Sub-network Instances**, which may be shared by another Network Slice Instance. A Sub-network instance has following properties:

- The Sub-network Instance is defined by a **Sub-network Blueprint**.
- A Sub-network Instance is not required to form a complete logical network.
- A Sub-network Instance may be shared by two or more Network Slices.

In Layman terms, A network slice instance is like a **computer program which defines what will be the characteristics of the slice**. So for example, if you a designing a IoT slice (one which interacts only with your smart bathroom), you might include network functions such as socket_open(), socket_close(), set_ultra_dedicated_channel() etc.

The sub-network instance, though not necessary, can be thought of as **shared objects in object-oriented programs**. For example, in a C++ program, just like you share a Class across different programs, certain sub-network instances can be shared across different network instances. For example, for the smart bathroom, you can use a sub-network instance which has socket_open() and socket_close() functions only. This can be used for smart cars, smart watches, or even mobile phones as well.

Please note that Subnetwork instances **ARE NOT RESOURCES**, but just set of commonly shared functions.

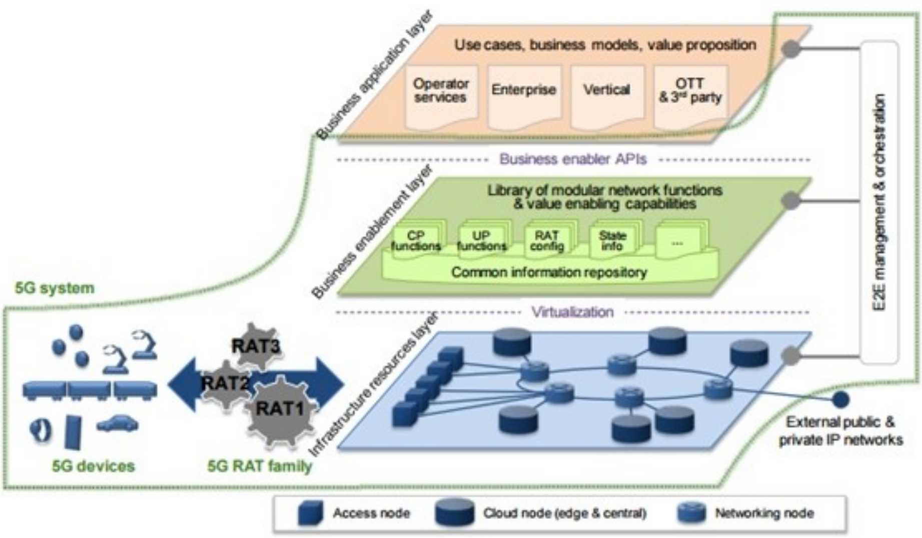
3.) Resource layer: It comprises of two types of resources:

- *Physical resource*: the physical asset for computation, storage or transport including radio access.
- *Logical resource*: Partition of a physical resource, or grouping of multiple physical resources dedicated to a Network Function or shared between a set of Network Functions.

To provide a way to provide these resources for different network slices, a **Network Slice Blueprint** is defined for network slice instance. Network Slice Blueprint provides the complete description of the structure, configuration and the plans/workflows for h and control the Network Slice Instance during its life cycle. A Network Slice Blueprint enables the instantiation of a Network Slice, which provides certain network characteristics (e.g. ultra-low latency, ultra-reliability, value-added services for enterprises, etc.). A Network Slice Blueprint refers to required physical and logical resources and/or to Sub-network Blueprint(s). Similarly, a **Sub-network Blueprint** provides the description of the structure (and contained components) and configuration of the Subnetwork Instances and the plans/workflows for how to instantiate it. A Sub-network Blueprint refers to Physical and logical resources and may refer to other Sub-network Blueprints.

In layman terms, all the necessary constructor functions in a Java program are provided by Resource layer and the blueprints. This layer's job is to provide the instantiation of all control variables, the access to all sockets, and opening up of dedicated channels for communication. Who will get what resources are detailed out in the blueprints.

Generic Architecture of 5G:



Given above is the generic architecture of the 5G system as realized by NGMN. Now that we have explored Network Slicing architecture, a quick overview of overall 5G architecture is warranted. The 5G architecture is a **native SDN/ NFV architecture** covering aspects ranging from devices, (mobile/ fixed) infrastructure, network functions, value enabling capabilities and all the management functions to orchestrate the 5G system.

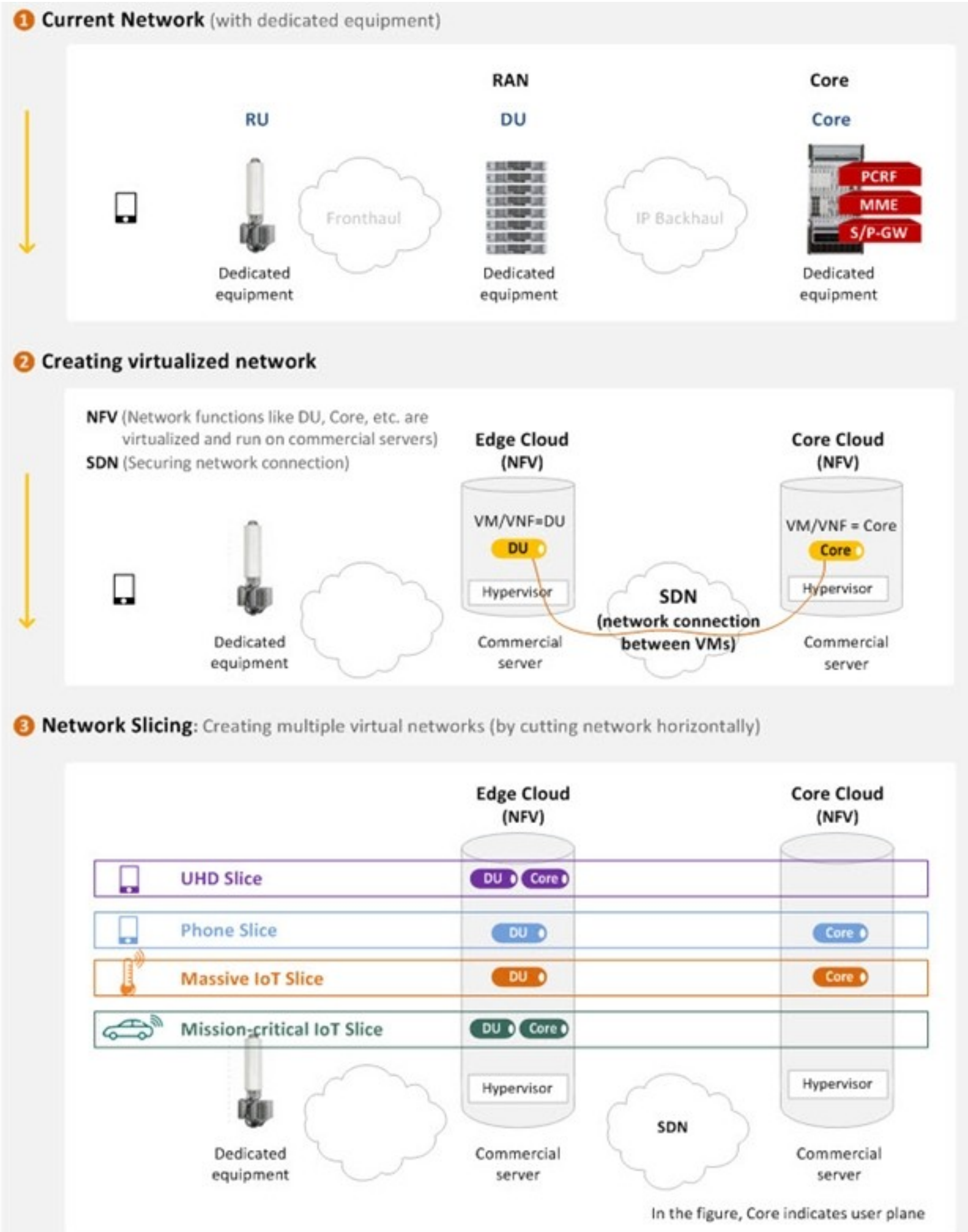
The **infrastructure resource layer** consists of the **ph** of a fixed-mobile converged network, comprising access nodes, cloud nodes (which can be processing or storage resources), 5G devices (in the form of (smart) phones, wearables, CPEs, machine type modules and others), networking nodes and associated links. 5G devices may have multiple configurable capabilities and may act as a relay/ hub or a computing/ storage resource, depending on the context. Hence, 5G devices are also considered as part of the configurable infrastructure resource. The resources are exposed to higher layers and to the end-to-end management and orchestration entity through relevant APIs. Performance and status monitoring, as well as configurations, are an intrinsic part of such an API.

The **business enablement layer** is a **library of all functions** required within a converged network in the form of modular architecture building blocks, including functions realized by software modules that can be retrieved from the repository to the desired location, and a set of configuration parameters for certain parts of the network, e.g., radio access. The functions and capabilities are **called upon request by the orchestration entity**, through relevant APIs. For certain functions, multiple variants might exist, e.g., different implementations of the same functionality which have different performance or characteristics. The different levels of performance and capabilities offered could be utilized to differentiate the network functionality much more than in today's networks (e.g., to offer as mobility function nomadic mobility, vehicular mobility, or aviation mobility, depending on specific needs).

The **business application layer** contains **specific applications and services of the operator**, enterprise, verticals or third parties that utilize the 5G network. The interface to the end-to-end management and orchestration entity allows, for example, to build dedicated network slices for an application, or to map an application to existing network slices.

The **E2E management and orchestration entity** is the **contact point to translate the use cases and business models into actual network functions and slices**. It defines the network slices for a given application scenario, chains the relevant modular network functions, assigns the relevant performance configurations, and finally maps all of this onto the infrastructure resources. It also manages to scale of the capacity of those functions as well as their geographic distribution. In certain business models, it could also possess capabilities to allow for third parties (e.g., MVNOs and verticals) to create and own network slices, through APIs and XaaS principles. Due to the various tasks of the management and orchestration entity, it will not be a monolithic piece of functionality. Rather it will be realized as a collection of modular functions that integrates advances made in different domains like NFV, SDN or SON. Furthermore, it will use data-aided intelligence to optimize all aspects of service composition and delivery

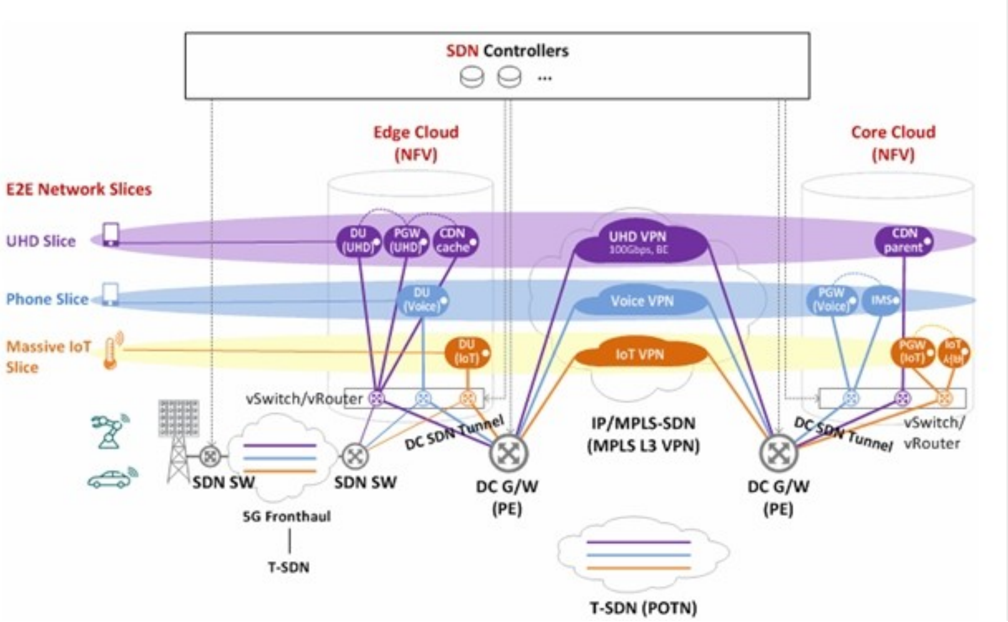
Deployment Architecture:



In the current mobile networks, main devices are phones, and RAN (DU and RU) and Core functions are built with dedicated n provided by RAN vendors. To implement network slices, **Network Function Virtualization (NFV) is a prerequisite**. The main idea of NFV is to install Network Function S/W (i.e., MME, S/P-GW and PCRF in Packet Core, and DU in RAN) all onto Virtual Machines (VMs) deployed on a virtualized commercial server (COTS; commercial off-the-shelf), **and not onto their dedicated network equipment individually.** This way, RAN works as edge cloud while Core works as core cloud. Connectivity among VMs located in edge and core clouds are provisioned using SDN (which is a native feature of 5G). Then, slices are created for each service (i.e., phone slice, massive IoT slice, mission-critical IoT slice, and so on).

In layman terms, instead of using dedicated hardware devices for MME, PCRF etc., vendors will use the technology of NFV to create **virtualized software**, and run the previously dedicated physical hardware, into virtualized software instance running over commercial VM dedicated hardware of a data centre. Think of it in this way: Currently, your wifi data is passed through a dedicated router. With 5G and network slicing, your wifi data will be passed through a software, which is running over a virtual machine, which is running over a commercial grade VM dedicated server (think Inception: a dream within a dream within a limbo!).

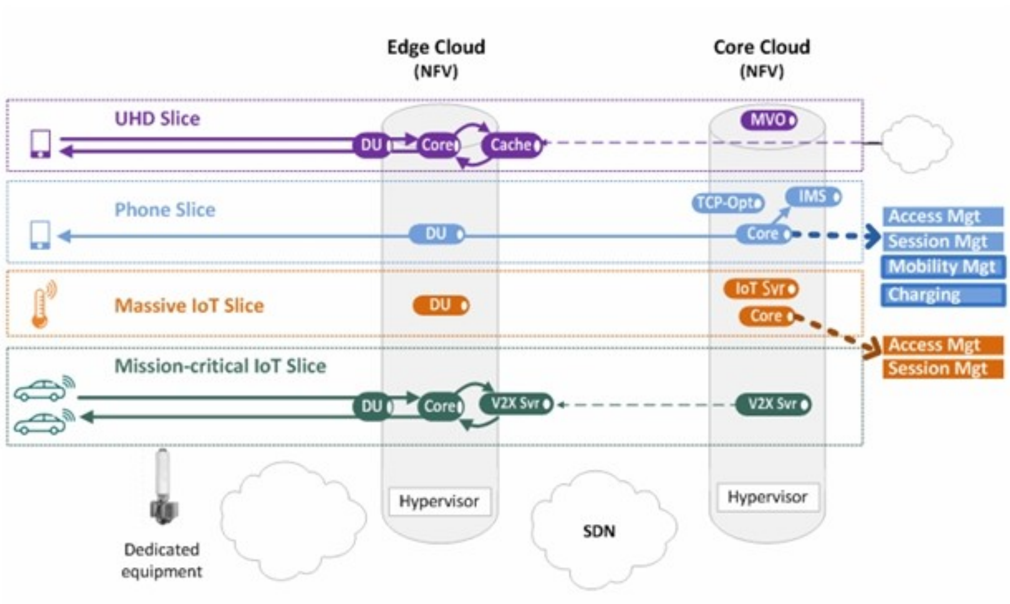
Now in 5G, like 4G, there will be Radio access network and Core network. Earlier, they were connected by the dedicated physical network because they were running on a physical infrastructure. Now since all of them will be running on the cloud as a virtual instance, they would be connected by a technology called **Software Defined Networking** (which is just a fancy term for virtual routers).



Furthermore, each VM of both Edge Cloud and Core Cloud are connected to the vendor supplied SDN enabled IP/MPLS external gateway routers using **SDN tunnelling**. In the hypervisors of each VM lies a switch/vRouter, which is provisioned by **SDN controller to provide this connectivity between the VMs and the gateway routers**. Additionally, SDN controller also provisions for mapping between these tunnels and respective VPNs created for network slices such as IoT slice.

Resources for different use cases of network slices

Following resource distribution can be considered for Network Slicing for 5G:



- **UHD slice**: All virtualized DU, 5G Core (UP), and Cache server in Edge cloud, and virtualized 5G Core (CP) and MVO server in Core cloud.

- **Phone slice**: 5G Core (UP and CP) **with full mobility features**, and IMS server, all virtualized in Core cloud.

- **Massive IoT slice** (e.g., sensor network): Simpler, light-duty 5G Core **WITHOUT mobility management** feature in Core cloud.

- **Mission-critical IoT slice**: 5G Core (UP) and associated servers (e.g., V2X server) all down in Edge cloud for minimized transmission delay.

Radio scheduler would be an example of a **shared function across multiple slices**. The scheduler of a RAT will typically be shared among multiple slices and be critically used to ease network congestion by multiple scheduling. However, the entire Telecommunications industry is based on proprietary scheduling algorithms. So it remains to be seen how open and collaborative these technology giants become in the future to develop shared scheduling algorithms.

Conclusion:

In Cloud, we have heard of Infrastructure as a service (IaaS), Platform as a service (PaaS), Software as a Service (SaaS), with service (AaaS) and function as a service (FaaS). Now, to realize the challenging demands of 5G such as mission-critical ultra-low latency networks, ultra HD streaming, and your typical VoIP calls, **connectivity as a service** (CaaS?) must be realized using network slicing.

Network slices are the key to delivering 5G use cases, as they can provide connectivity that is adapted and optimized for each and every use case, application and user in such a way that **uses network resources efficiently**. Also, the benefit of slicing networks is not just the capability to deliver a wide range of connectivity services to any industry, but also to ensure that the **usage can be billed accordingly**. Slicing networks provide greater insight into network resource utilization, as each slice is customized to match the level of delivery complexity required by the service or services using the slice.

The concept of network slices requires a number of technologies such as SDN, NFV and cloud technologies. These will enable networks to be broken out from their underlying physical infrastructures so that they can, programmatically, provide connectivity as a service and usher in the new revolution of high-speed on-demand connectivity.

Writer's Note: *If this post helped you in any way possible, please do take a minute to like, share or comment. You can also read some of my previous work here*. As always, *Sharing is Caring!*

References:

https://www.ngmn.org/fileadmin/user_upload/160113_Network_Slicing_v1_0.pdf

http://www.netmanias.com/en/post/blog/8325/5g-iot-nfv-network-slicing-sdn/e2e-network-slicing-key-5g-technology-what-is-it-why-do-we-need-it-how-do-we-implement-it

https://ngmn.org/5g-white-paper/5g-white-paper.html

http://www.eurecom.fr/en/publication/4641/download/cm-publi-4641.pdf

http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7169508

Aayush Shrut
Consultant at Deloitte
Published • 4yr

11 articles + Follow

#network #5G #cloud #connectivity #highspeed #telecommunications #nfv #sdn #4G #slicing #technology #latest

Like Comment Share 30 · 1 Comment

Reactions

+18

1 Comment

Most Relevant

Add a comment…

Michael Elling • 3rd+ 12mo ···
TMT/ICT Strategy & Business Development

Aayush, very good overview. Thank you.

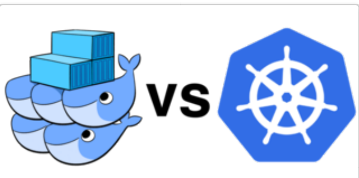
The carriers, much like with wholesale 4G will be loathe to price all the elements at or near marginal cost; like we see in other internet ecosystems: ad-tech, content streaming, VPN, etc.... As such, anyone who can piece together a vertically complete hybrid solution based on fiber, ultra-high mmw frequencies an ...see more

Aayush Shrut
Consultant at Deloitte

+ Follow

n Aayush Shrut

Layman
ut on LinkedIn



Docker and Kubernetes for Layman
Aayush Shrut on LinkedIn

Containerization for Layman!
Aayush Shrut on LinkedIn

LTE for Layman (part 3) - The Complete Picture!
Aayush Shrut on LinkedIn

ırticles

Linked in

Messaging