

2018

---

# 5G: Core Architecture

## Evolution || Use Cases || Services

---

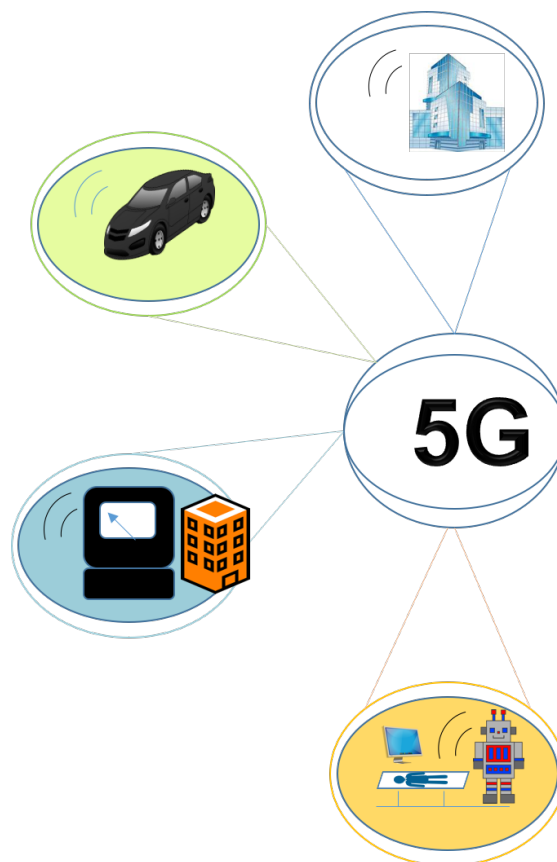
Dilip Krishna S & Nikhil Vyakaranam

---

This whitepaper explores the new 5G core architecture and how it solves the present-day problems of a closed framework through its programmability and service agility. It also delves into some of the use cases to see how other industry verticals are benefited by this new architecture thus creating new business models for the Service Providers.

## Introduction

The prime objectives of any cellular network architecture are to address the need of increased capacity, improved data rate, decreased latency and better quality of service. To meet these specific demands, there had to be a disruption in the mobility space and 5G is seen as the new normal. Different industry verticals like Automotive, Public Safety, Media Delivery, Railways and many more were already in search of an efficient transport system to manage their use cases. They saw this potential in 5G and strongly believed that it would help in solving their problems in service delivery. Their participation in the 5G standards discussion has led the ITU to classify these industry verticals under three different network services namely eMBB, URLLC, and mMTC. There have been many feasibility studies around these major 5G technology drivers and they have been captured under 3gpp technical requirements documents (TR) which forms the basis of the newly proposed 5G architecture.



*5G ecosystem*

---

## 5G Core Architecture

In the monolithic 4G architecture, there was a single piece of network designed primarily for the smart phones. Towards the later part of the generation, new use cases like NBIoT, mMTC and not to forget the ever-growing demand of speed due to proliferation of smartphones, started making waves. There was excitement around these varied nature of use cases but the 4G network could not provide any service or performance differentiation based on the type of use cases. A One-size-fits-All approach was followed. The 4G standard reference point architecture had different protocols running between most of the interfaces to support different services taking significant development cycles from vendors' release after release. The interoperability woes would add to the delay in the Service provider's GTM strategy.

Most of the initial deployments were on custom hardware from the vendors. However with the increased computational power and efficient capacity utilization delivered by x86 servers, Network Function Virtualization (NFV) along with Software Defined Networking (SDN) gained a lot of momentum in the service provider community in 4G deployments. The year 2017 saw many NFV based proof-of-concepts and trials, but the adoption has been on the slower side. The service providers who did make some progress on this front not only saw this as a financial benefit but also as a stepping stone to the 5G. NFV and SDN will continue to form the building blocks of 5G.

*NFV and SDN  
will form the  
foundation of  
5G architecture*

5G is moving from whiteboard to reality and there is a lot of interest and expectation around it. At the center of all the deliverables is the network architecture. With the varied nature of use cases, a fully agile and programmable network was expected to be delivered. The network should be developed as a framework to provide distributed cloud based services, context aware networking, low latency services and network capability in terms of coverage and access convergence.

The main concepts introduced in 5G include:

- Control & User Plane Separation (CUPS)
- Network Slicing
- Cloud Native and Microservices
- Service Based Architecture (SBA)

---

## Control & User Plane Separation (CUPS)

Software Defined Networking or SDN is defined as the physical separation between network control plane and user plane. This architecture decouples the network control and forwarding functions enabling the network control to become directly programmable and the underlying infrastructure to be abstracted for applications and network services. Taking inspiration from the advantages that SDN offers, Control and User Plane Separation or CUPS was suggested as a key core network feature to provide architecture enhancements for Evolved Packet Core nodes like SGW and PGW. The same concept has been extended to the 5G core architecture and will serve as a better feature than in 4G. In 5G SMF will act

*CUPS helps in achieving Low Latency, support for increase in Data Traffic along with independent evolution of Control Plane and User Plane*

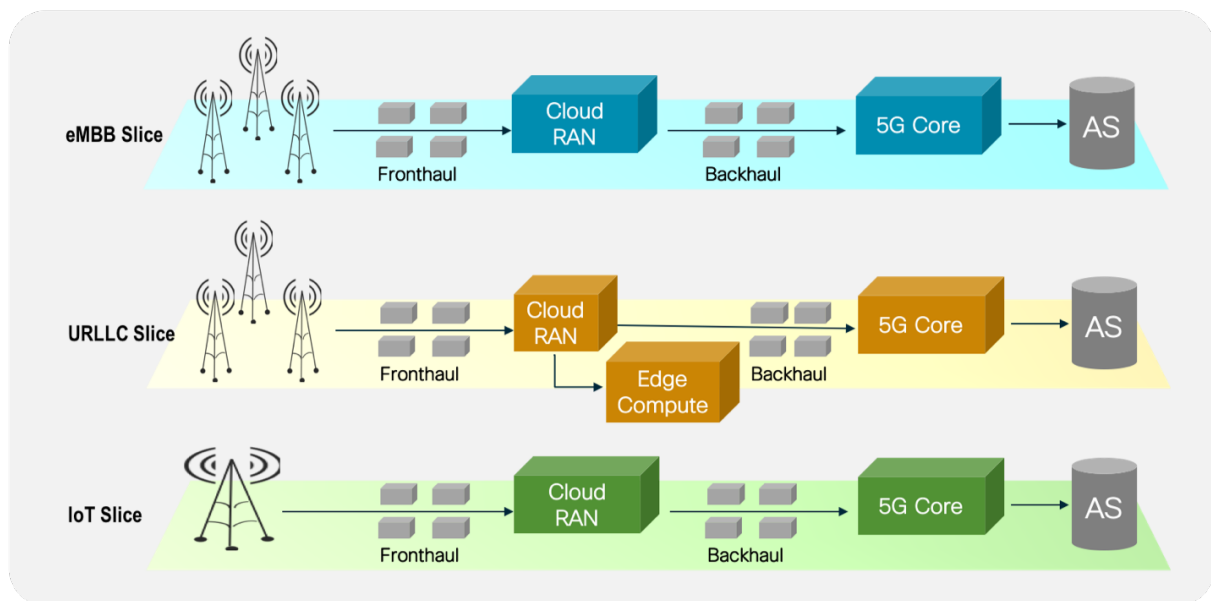
With the rapid increase in smart phones and applications like streaming video, the data traffic is set to accelerate. For effective delivery of these services with better user experience, lower latency is the criteria.

To meet the low latency requirements, the User Plane can be deployed at the edge cloud closer to the users. Mobile Edge Compute has a significant role to play in 5G deployments. Based on the increasing data traffic, the User Plane can be scaled independently. Upgrades can be performed without affecting rest of the ecosystem.

Packet Forwarding Control Plane (PFCP) is the native protocol between the Control Plane and the User Plane on the Sx reference point in case of 4G and is expected to be retained in 5G. A session is established on the User Plane, instructing the UP function on how to process a certain traffic.

## Network Slicing

End-to-End network slicing is the cornerstone of the 5G architecture and enables the support for diverse 5G services. It allows the operator to provide dedicated logical/virtual networks for specific requirement and functionality each having their own unique properties. Based on NFV and SDN the physical infrastructure is abstracted from the logical network architecture. Each network slice may have its own network architecture, protocols and security settings.



*Network Slicing*

As shown in the preceding figure, eMBB (Enhanced Mobile Broadband), URLLC (Ultra Reliable Low Latency Communication) and IoT (Internet of Things) slices can be supported independently on a single infrastructure. The eMBB slice has high requirement for bandwidth and will be supported by a physical infrastructure capable of high computations. URLLC is highly sensitive to network latency in application scenarios of self-driving, remote tele surgery equipment and virtual or augmented reality. Mobile Edge Compute should be placed at close proximity to the users to provide a short round trip time. The IoT slice sends very few data packets to the network but requires a large capacity to register millions of devices. This consequently allows low compute physical resources to be used for this slice thereby reducing the overall operating expenses.

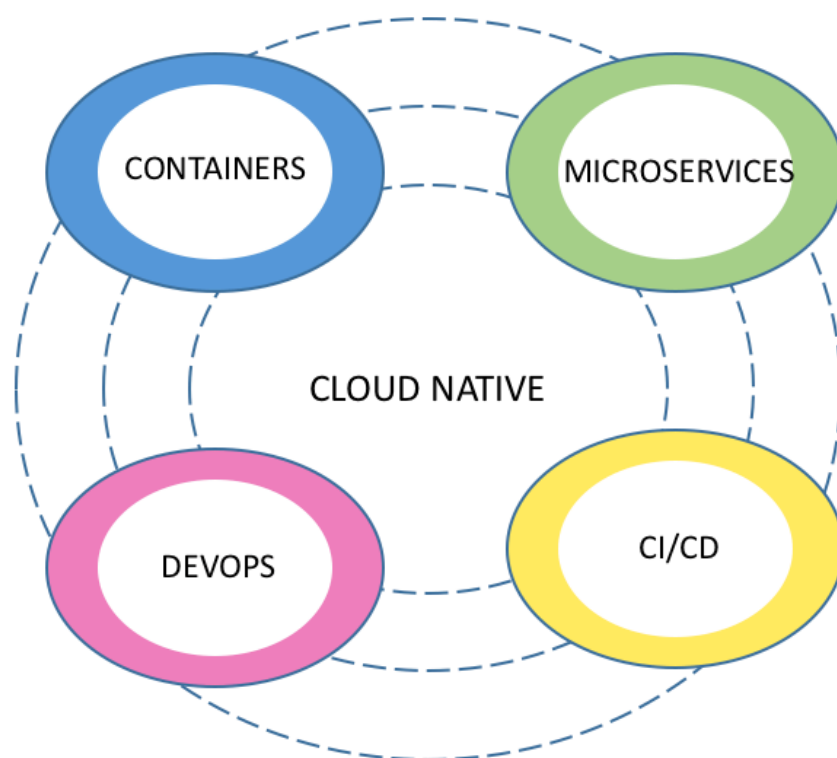
Network slicing enhances network security as the slices are isolated from each other and the traffic in one slice cannot interfere with the others. End-to-End network slice will include the Core, the RAN and possibly the transport network as well. Software Defined RAN or Cloud

---

RAN facilitates the RAN Slicing. The controller in the cloud RAN can allocate appropriate computational, network and radio resources to various network slices based on their service requirements. It's relatively easier to implement SDN and NFV on core network. Based on the service type, necessary VNFs can be provided to the network slice. These VNFs can be scaled on demand with changes in service and performance requirements. Considerable effort is being put to improve the transport to satisfy the 5G needs. Leading vendors have already demonstrated network routers that support backhaul slices with independent performance characteristics. SDN will play a crucial role in determining the best path for a packet to traverse through the fronthaul and backhaul elements, to meet the various 5G demands.

## Cloud Native and Microservices

Being Cloud Native is about building and running applications that fully exploits the advantages of a cloud delivery model. But the term Cloud Native can also mean getting services and products faster to the market at a lower cost. In the latter half of the 4G deployments, Service Providers opted for NFV based deployments. But these deployments were more about moving the workload into the cloud environment. With Cloud Native, the application itself will be developed by keeping the cloud characteristics in mind. Cloud Native Foundation (CNCF) recommends a cloud native system to have the following properties.



*Cloud Native Systems*

---

**Microservices** is an approach to develop an application as a collection of services. It is a set of small autonomous services that work together. A monolithic system will have a single large code base and it is difficult to know where a change needs to be made and there is a lot of dependency between the teams working on each of the modules as there isn't a distinct boundary defined. In most cases, the entire application needs to be developed using a single programming language due to the dependencies between them. With microservices, the application is broken down into small independent services. Each service will implement a given piece of functionality serving specific business capabilities. It will have its own processes and communicates to other services using REST APIs over a common message bus. Being independent services, they can be deployed and upgraded independently without affecting rest of the application.

**Containers** are the most suited to implement microservices due to the high level of resource isolation. Container technology allows the application and its dependent binaries and libraries to be packaged into a container with no extra baggage of an operating system. Multiple containers can be run on a host as it shares the Host OS kernel but runs in isolated user spaces each having its own unique filesystem, network and process table. Faster creation and deletion times due to the lightweight package and lesser overhead than a Virtual Machine makes it ideal for microservices.

**DevOps** is a concept where software development and operations teams work hand in hand to deliver applications and services at high velocity. A new philosophy adopted by some of the best companies in the world breaks the traditional barriers of software development and infrastructure management. This concept becomes an important part of any cloud native strategy bringing in business agility as the building, testing and releasing of software happens at a rapid pace.

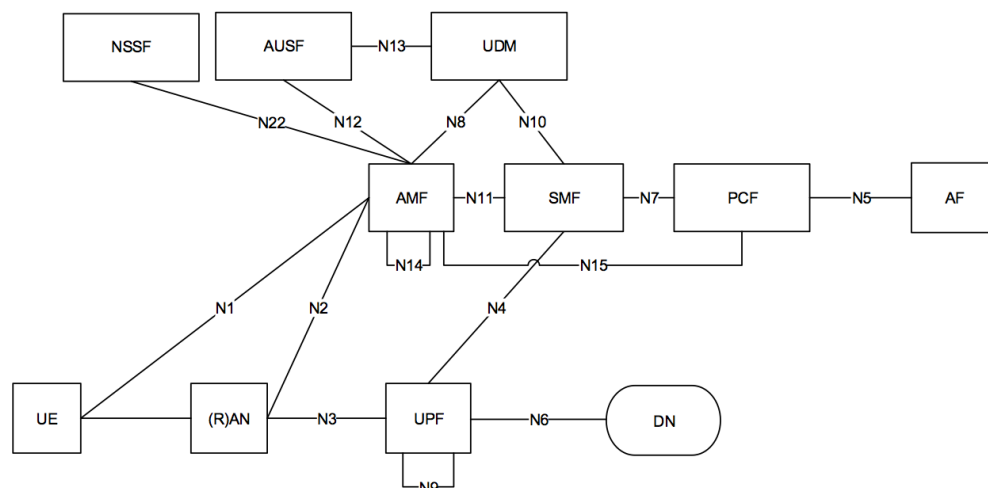
**Continuous Delivery** fueled by DevOps is about faster build, test and release of applications using automation tools. It aims at getting a faster feedback from the end user. A smallest of a change in the code can be configured to trigger the test and deployment of new releases using automated CI/CD pipelines.

*“5G is about agility, flexibility, programmability and microservices promises this evolution”*

## 5G Core Architecture

On the core side, 5G is conceptualized as one single architecture but has been represented in two different flavors - a Point-to-Point reference architecture and a Service Based Architecture (SBA).

The point-to-point architecture resembles the earlier generation models where discrete interfaces are defined between each of the nodes and running different protocols on most of the interfaces. This model has its drawbacks from being closed and not being programmable. With 5G, service providers are looking for operational agility and programmability in the cloud environment especially because of the varied nature of use cases waiting for them. All the initial 5G deployments are expected take this model as a reference due to the initial inhibition towards the new service-based model.



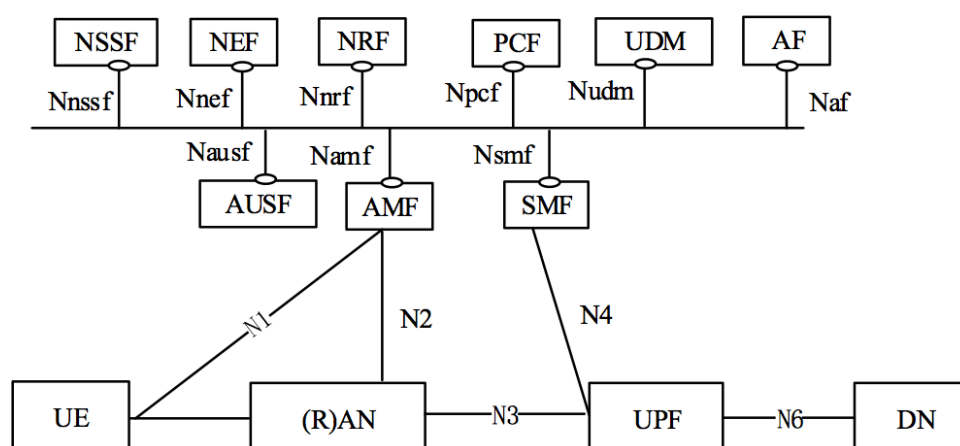
*Point-to-Point Reference Architecture*

Cloud Native Architecture comes as a solution to many of these problems. It is far from just moving the workload into the cloud as was in the case of 4G based NFV deployments. The application is expected to be baked in as cloud ready. It follows a microservices style of architecture where each self-contained service is created and maintained independently utilizing the container technologies like Docker which is easy to build, deploy and update. Each service will expose its capabilities and the others can interface with it using the APIs.

Automation services like Auto Scaling, Auto Healing, Auto Configuration and others which are still being explored in the 4G deployments today, will become an integral part of the 5G architecture. Upgradation procedures of the VNFs is a challenge today.



With a container based architecture, the traditional upgrade procedures vanish and it will be all about deleting the old version of the instance and bringing up the new one. Thanks to the CI/CD techniques and minimal start up times of the containers, each module/service can be upgraded independently. The upgrades should almost give us a zero down time now. The Service Based Architecture will provide the much needed agility to the service providers to respond to customer needs quickly.



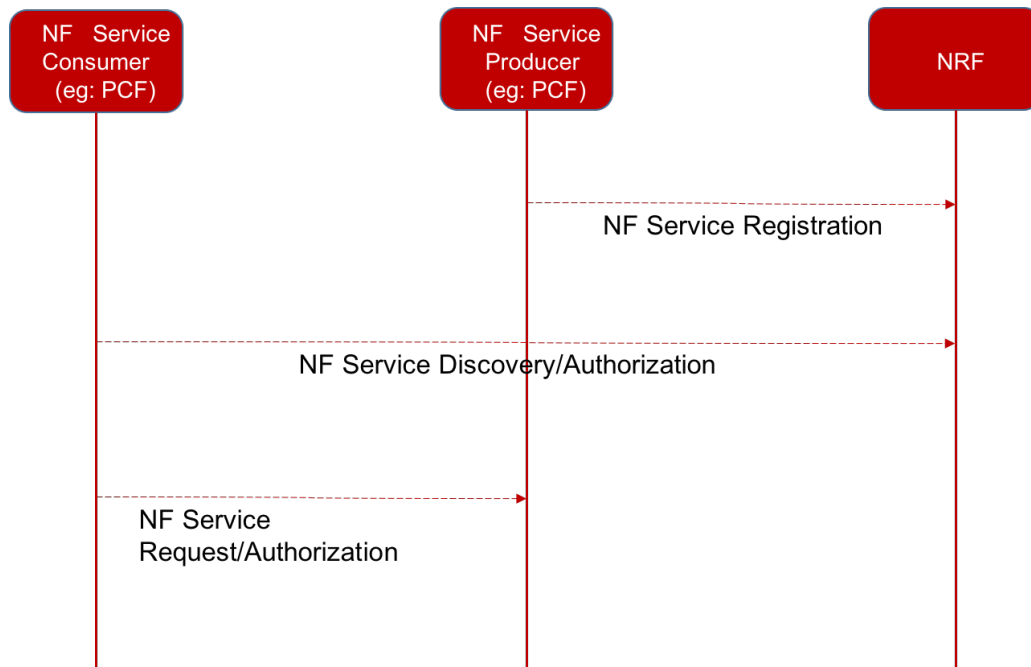
*Service Based Architecture (SBA)*

Looking at all the advantages of microservices, 3gpp decided to go with a Service Based Architecture (SBA) for 5G. All the Control Plane Network Functions are expected to be connected to a message bus and will expose their functionality over service based interfaces. The User Plane is expected to continue using GTPv1 with modifications. In this new model, it is expected that all the traditional 4G core control plane protocols like GTPv2, Diameter will make way for HTTP2.0 as a protocol on the service bus. Each Network Function Service will expose a set of REST APIs and the primitive model of request/response and subscribe/notify is set to be used for all interactions between the different services. QUIC, a protocol from Google, is being explored as the transport layer protocol replacing the TCP/TLS.

Due to the dynamic nature of microservices and features like Auto Scaling and Auto Healing, new service containers are spun up rapidly and it is demanding to keep track of these services. Service discovery becomes an important part of these dynamics. As a service comes

---

up, it registers itself with a DNS like service called Network Repository Function (NRF). NRF allows the network functions to discover the services offered by other network functions.



*Service Based Interaction*

The major components of 5G core and high level description is mentioned below

- **Access and Mobility Management Function (AMF)** – Like the MME, it takes care of access control and mobility. UE registration happens via AMF. It interfaces with AUSF for authentication, UDM for subscription information, talks to NSSF to get slice information and SMF for PDU establishment. The MM information from the UE is processed by AMF and the SM information is transparently passed to the SMF.
- **Session Management Function (SMF)** – Plays a role similar to S/PGW control plane elements in 4G. Takes care of Session Establishment and maintains the tunnel between UPF and Access Node. It manages the UPF over a Sx like interface. It also interfaces with PCF for network wide policies.
- **User Plane Function (UPF)** – It is a combination of SGW and PGW user planes in 4G. It interacts with the Access Node over a GTP-U interface. It does QoS handling and Policy

---

enforcement on the User plane part. They can be deployed anywhere in the network in various configurations.

- **Policy Control Function (PCF)** – From the same family as PCRF in 4G, it takes care of network policies to manage network behavior. It gets the subscription information from the UDM. It interfaces to both AMF to manage the mobility context and the SMF to manage the session contexts. It also plays a crucial role in providing a schema for network slicing and roaming. PCF triggers the UE route selection policy (URSP) which enables the UE to determine how a certain application should be handled in the context of an existing or new PDU session. This can be compared to the TFT or flow descriptors in the PCRF response. It also plays an important role in setting the Traffic steering policy for steering the subscriber traffic to appropriate service functions. PCF can get slice specific analytics from NWDAF, an analytics service which provides network wide analytics.
- **Unified Data Management (UDM)** – Like the HSS in 4G, it stores the subscriber data. It primarily interfaces with the AMF.
- **Authentication Server Function (AUSF)** – An authentication server which will also handle the EAP requests from the non-3gpp world.

The Service Based Architecture (SBA) is modeled around microservices and REST APIs. One of the properties of microservices is statelessness. But any mobile network is characterized around user data and user profiles which needs to be stateful. To accommodate the above requirements, a set of new nodes have been proposed as part of the 5G architecture.

- **NF Repository Function (NRF)** – The SBA architecture has all the nodes running as services and each of them exposing their own functionalities and there could be many such nodes in the network serving different set of features. The service discovery plays a key role here. NRF acts like DNS helping in service discovery. Every service registers itself with NRF as it comes up. The SBA interaction in the above Figure will explain it more.
- **Network Slice Selection Function (NSSF)** – Different use cases like IoT, eMBB, URLLC will be served by different set of nodes in the network based on the service and performance characteristics. AMF queries the NSSF based on the Slice details it gets from the UE during the PDU establishment.
- **Network Exposure Function (NEF)** – An API gateway to the mobile network like the SCEF in 4G. Third party applications can use the APIs to get information like network events, statistics for analytics, provisioning capability, policies and charging services exposed by the mobile network. NEF also interfaces with UDR to store and retrieve information.
- **Unified Data Repository (UDR)** – A central repository like UDR will help in storage and retrieval of subscription data, policy data and application data. NEF can reach out to UDR for specific information.

---

## Use Cases

The 5G innovation is built around higher throughput, low latency, security and programmability of the network. The new market segments and their requirements have outlined the newly proposed 5G architecture and the Service Based Architecture (SBA) is expected to provide the framework for this. The ability to support a wide range of such use cases and its abstraction from the complex mobile network makes the 5G architecture very flexible and industry friendly. We will take a look at some of these use cases to analyze how 5G helps in providing a differential service based on their requirements. This differentiation is achieved using Network Slicing.

### Use Case 1: Mobile broadband for Enterprise.

The Next generation enterprise or the startup culture is to have no strings attached in terms of the infrastructure. They tend to host all their applications on the cloud or subscribe to publicly available cloud applications with no single pie being spent on hardware. Similar is the case with the enterprise broadband access. The traditional DSL, Ethernet and Cable internet are all viable options but mobile broadband looks to be gaining a fair share in that space.

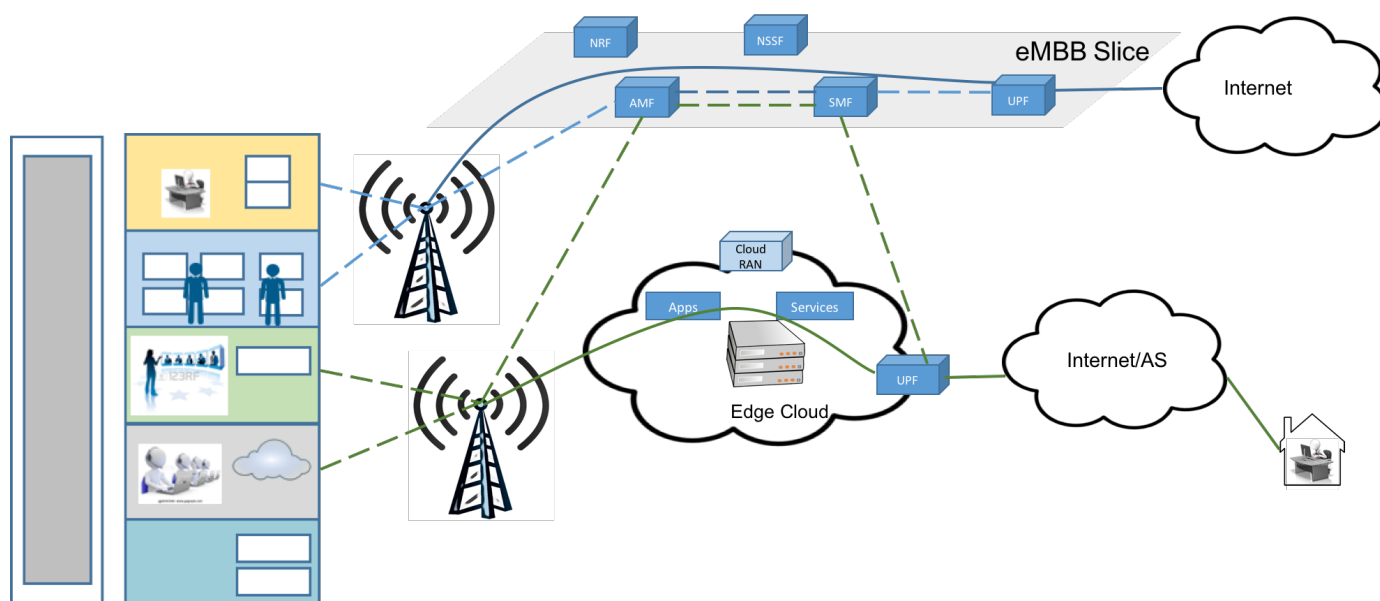
The arrival of 4G office routers connected to the fast LTE network on the egress changed the game for MNOs. The enterprise could now connect the router to their backend network or even access the internet through the Wifi provided by the router. The age old EPABX system could be replaced by the Fixed wireless terminals offered by many OpCos.

Portability is one of the compelling advantages of mobile broadband. The employees could be offered with bundled plans where they can access the office network on the go through a VPN or any cloud subscriptions at no extra cost. The pay-as-you-go models extended by many MNOs is definitely a better deal than a yearlong contract that some of the fixed line broadband companies offer. A research study shows that more than 80% of the enterprises in some of the European countries preferred mobile broadband over fixed broadband. This interest is astonishing as the maximum average speed across the globe in 4G was just around 26Mbps as per Akamai's State of the Internet Connectivity Report.

With 5G, the experience is going to get only better with significantly higher speeds and reliability. The service and performance differentiation can be done based on the use cases. ITU expects the key capabilities of 5G to include a peak data rate of 20Gbps and a user experienced data rate of 100Mbps. One of the largest operator in the world is already planning to launch 5G home fixed wireless. This is quite encouraging for the existing and the prospective enterprise customers looking to move the mobile broadband way.

Look at a scenario where an Enterprise is using Mobile broadband for their indoor network and how 5G helps achieving the best results. The serving radio in this case could be an in-building

solution or a small cell. This being a customized solution, it is expected to have a backhaul infrastructure with sufficient bandwidth. Frequent data transfer from/to the company's servers and real time video conferencing is expected to be their primary traffic pattern.



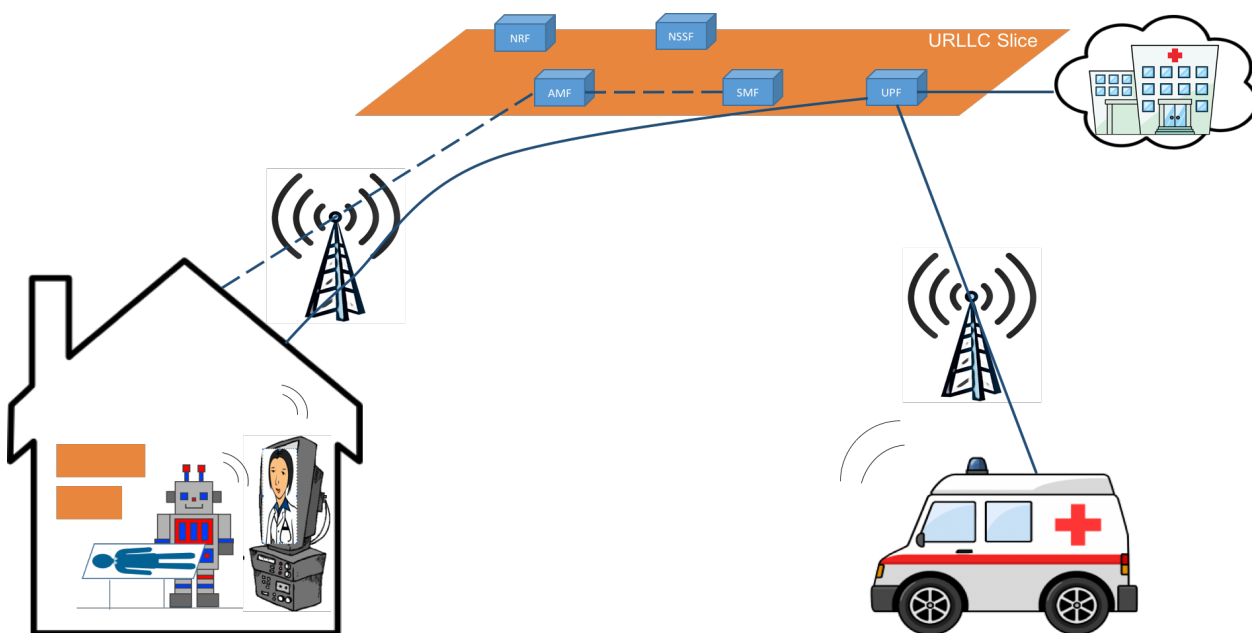
*Mobile broadband for Enterprise*

Based on these requirements, the mobile network is expected to handle high data rates and large volumes of data. The mobility aspect in this scenario is trivial but the network will have to support high density connections. Support for Low latency is expected to assist Real time video conference. The eMBB network slice should help in realizing this use case. For the locally hosted applications and video conferencing, usage of Mobile Edge Compute on the RAN network will benefit in low latency and saving in backhaul infrastructure. A dedicated network slice can be allocated to such an enterprise based on their requirements as shown in the below.

## Use Case 2: Telemedicine/Tele-surgery

Quality healthcare is one of the most important factors in a country's economic growth and national identity. All the developments in the health care sector has resulted in a improved life expectancy. As a result of this, we have begun to see a sharp rise in the population of people above the age of 65 in the recent times. As per the report published by WHO, by 2050 senior citizens are going to more than 16% of the total world population. Thanks to all the major inventions in the sector. It is the responsibility of every country to take care of their aging population but they are still gearing up to meet the challenges that is posed by aging. In a country like India, the number of doctors per 1000 patients is less than 1. The number of hospitals and hospital beds is yet another challenge.

Telemedicine/Tele-surgery or Remote diagnosis is seen as an answer to this. This allows healthcare practitioners to identify the nature of illnesses outside the conventional hospital settings such as at home or at a well-equipped clinic in the locality using sensors, cameras and other equipments to present the captured data. Telehealth which is providing health care facilities directly at home using telecommunications is primarily employed to manage chronic illness such as heart ailments. Tele robots are set to make an impact where surgeries can be now done in the comfort of the home without the need of any travel reducing the total healthcare delivery costs and time.



*Telemedicine/Tele-surgery*

---

Healthcare has some of the most important use cases for 5G. Right from providing robotic help in remote diagnosis or tele-surgery to handling medical services in rural areas, this vertical seems to be of primary interest to several countries. With 5G, the medical expertise is not bound to the location of the physicians any more. They can access their patients through tactile internet utilizing robots for diagnosis and tele-surgery. The real time scan videos and other vitals can be readily sent to a remote ambulance in emergency situations. The tele robot can also maintain health records of the patients.

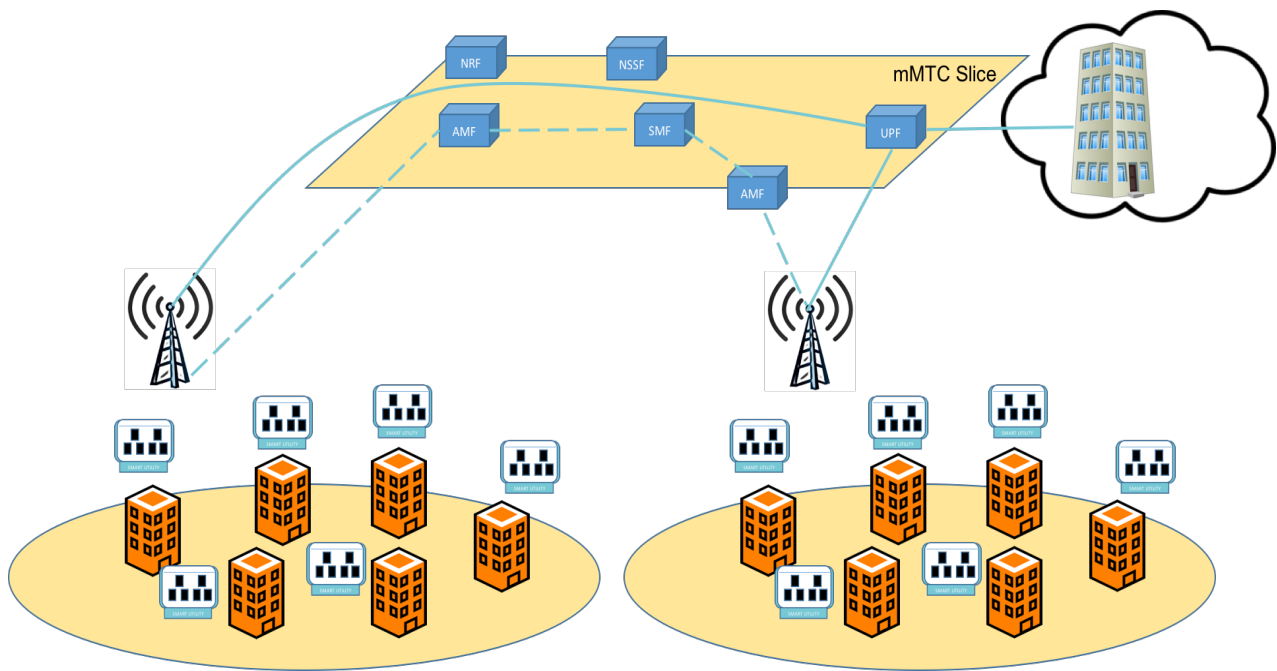
These requirements form probably one of the most important mission critical use cases. The nature of the traffic is real time and the network is expected to aid in low latency and ultra-high reliability. The packet loss is expected to be the least when compared to other slices. The URLLC slice will be the best fit in supporting this use case. The latency on this slice is expected to be <1ms. For reliability, the packets are marked with the highest priority DSCP marking and all possible HA is provided right from RAN to the Core. A part of the robotic computing can be offloaded to the Edge cloud as part of the cloud robotics which will make decision making faster. The Edge computing plays a major role in providing low latency due to the various applications it can handle. For e.g.: in cases where the physician is not residing in the same mobile network, the MEC can host a VPN service to which he can connect remotely and access his patients.

Some of the other use cases include Augmented Reality/Virtual Reality and Smart Grids.

### **Use Case 3: Smart Meters**

Today the Electric/Water utility companies in your locality get the consumption reports from the analog or digital meters installed in your premises. This manual method has a collection agent visiting you every month to take the readings on your meter. Bases on the readings, a manual or electronic bill is generated. The cost and time involved in this process is significant and is prone to manual error.

Smart Meters once installed at the consumption points will be capable to collecting the information in real time and is transmitted to the utility company. The information from remote areas can be gathered accurately with ease.



*Smart Utility*

A classic use case from the IoT stable where every smart meter in a particular area will keep uploading its data periodically to a centralized server. The server can reach out to these individual meters to initialize the device or to run some diagnostics as well. The amount of data sent is very small and is sent in bursts. Such IoT devices are generally low power sensors and a small area can have thousands of such devices.

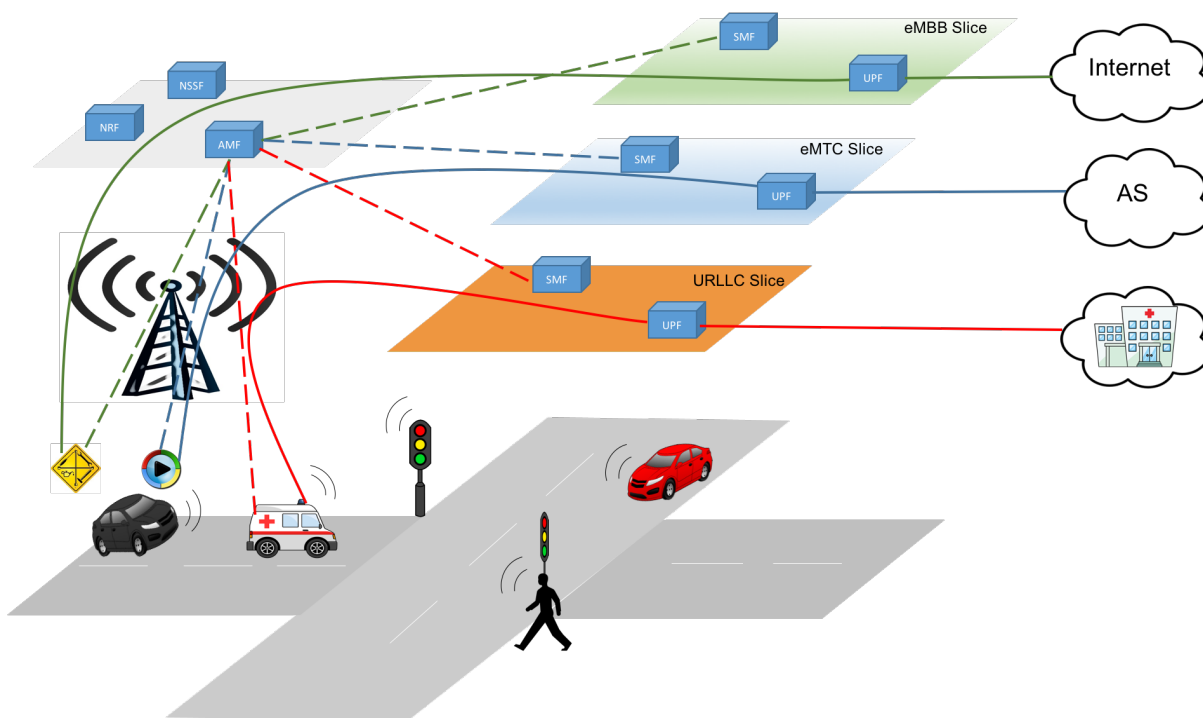
This scenario demands efficient resource and signaling support for low power devices. The network should be able to handle the burst traffic from a highly dense IoT device ecosystem. More AMFs might be needed to handle the traffic. The latency requirements are not stringent to none. Mobility as a parameter would vary across different IoT use cases. To satisfy all the above requirements, the mMTC or the MIoT slice. Since the throughput requirements are unsubstantial, the MNOs can save on the infrastructure cost in terms of using cheaper hardware to host VNFs/Network Services. Even though the latency impact is considered as minimal, MEC seems to be the way forward for many of the IoT services. This is mainly due to the savings they see in the backhaul expenditure.



## Use Case 4: Connected Vehicles

According to WHO, more than 1.25 million people die each year as result of road traffic crashes and another 20 to 30 million getting injured. This is even said to have an impact on the economy of certain countries. These accidents are attributed to unsafe road infrastructure including improper traffic laws, distracted driving, over speeding and many such risk factors. There is a constant effort from the automobile companies and their associations to use latest technologies to be integrated into vehicles to bring down these alarming number of road accidents.

With 5G discussions in background in 2016, 5GAA (5G Automotive Association) was formed with its members ranging from major players in the automobile industry and many telecom vendors. Their motto was to build intelligent transport systems where vehicles are allowed to share information to make transportation safer. An interactive domain called C-ITS (Cooperative Intelligent Transport Systems) was chosen which will allow road users and traffic managers to share information and use it to coordinate their actions. This domain would cover vehicle-to-vehicle (V2V) communication, vehicle-to-infrastructure (V2I), vehicle-to-network (V2N) and vehicle-to-pedestrian (V2P) communication which can be collectively called as V2X communication (vehicle-to-everything). Due to the reliability, robustness and features like network slicing, 5G was seen as the network which can deliver V2X.



*V2X Communication*

---

V2X demands high reliability, low latency, high mobility, higher positioning and medium traffic density. It also means higher bandwidth when the vehicles try to reach a streaming content or to use AR/VR applications. Due to this varied nature of requirements, V2X cannot be mapped to any reference slices like eMBB, URLLC or mMTC. Vehicles will carry multiple slices. Edge computing as for any use case, is seen as one of the key technologies serving V2X services. In this new age of driverless cars and autonomous driving, V2X is looked up to as one of the key drivers by the automobile community.

## Summary

The world is growing digital thus driving new business ecosystems. The technological advancements in every industry vertical is aimed at building new business models. 5G guarantees to be at the middle of this digital revolution. With the extent of consumer coverage along with higher speeds that it can provide and the flexibility that the newly proposed service-based architecture brings in, there is no other technology which can match its service delivery capabilities. 5G is set to loom large in the next few years and there is immense excitement around it.

## Reference

- 3gpp TS 23.501 - [http://www.3gpp.org/ftp/Specs/archive/23\\_series/23.501/](http://www.3gpp.org/ftp/Specs/archive/23_series/23.501/)
- 3gpp SBA - [http://www.3gpp.org/news-events/3gpp-news/1930-sys\\_architecture](http://www.3gpp.org/news-events/3gpp-news/1930-sys_architecture)
- Akamai Internet Connectivity report - <https://www.akamai.com/fr/fr/multimedia/documents/state-of-the-internet/q1-2017-state-of-the-internet-connectivity-report.pdf>
- Connected Vehicles - <http://5gaa.org/news/c-v2x-edge-computing-the-winning-technologies-for-connected-vehicles-and-autonomous-driving/>
- WHO - <http://www.who.int/mediacentre/factsheets/fs358/en/>
- Tactile Internet - [https://www.itu.int/dms\\_pub/itu-t/oth/23/01/T23010000230001PDFE.pdf](https://www.itu.int/dms_pub/itu-t/oth/23/01/T23010000230001PDFE.pdf)
- Cloud-Native Apps - <https://pivotal.io/cloud-native>
- 3GPP 5G CoreNetwork Status - [http://www.3gpp.org/ftp/Information/presentations/Presentations\\_2017/webinar-ct-status-11-2017.pdf](http://www.3gpp.org/ftp/Information/presentations/Presentations_2017/webinar-ct-status-11-2017.pdf)
- 3GPP eMTC TR - [http://www.3gpp.org/ftp/specs/archive/22\\_series/22.861/](http://www.3gpp.org/ftp/specs/archive/22_series/22.861/)
- 3GPP Critical Communications TR - [http://www.3gpp.org/ftp/specs/archive/22\\_series/22.862/](http://www.3gpp.org/ftp/specs/archive/22_series/22.862/)
- 3GPP eMBB TR - [http://www.3gpp.org/ftp/specs/archive/22\\_series/22.863/](http://www.3gpp.org/ftp/specs/archive/22_series/22.863/)