Dr. Vivek Kachhatiya

Profile : Assistant Professor
Department : Electronics & Communication Engineering

College of Engineering and Technology
SRMIST-Chennai

Email : vivekdek@srmist.edu.in

18ECO127T
5G Technology – An Overview

OPEN ELECTIVE (by ECE)
SEM:7
B.TECH

MODULE 2:
5G Network Architecture

Introduction to 5G Network Architecture

5G Core Network (5GC

Radio Access Network

Network Slicing:
Concept and Implementation;

Virtualization in 5G

Software-Defined Networking (SDN) in 5G

Edge Computing

Mobile Edge Computing

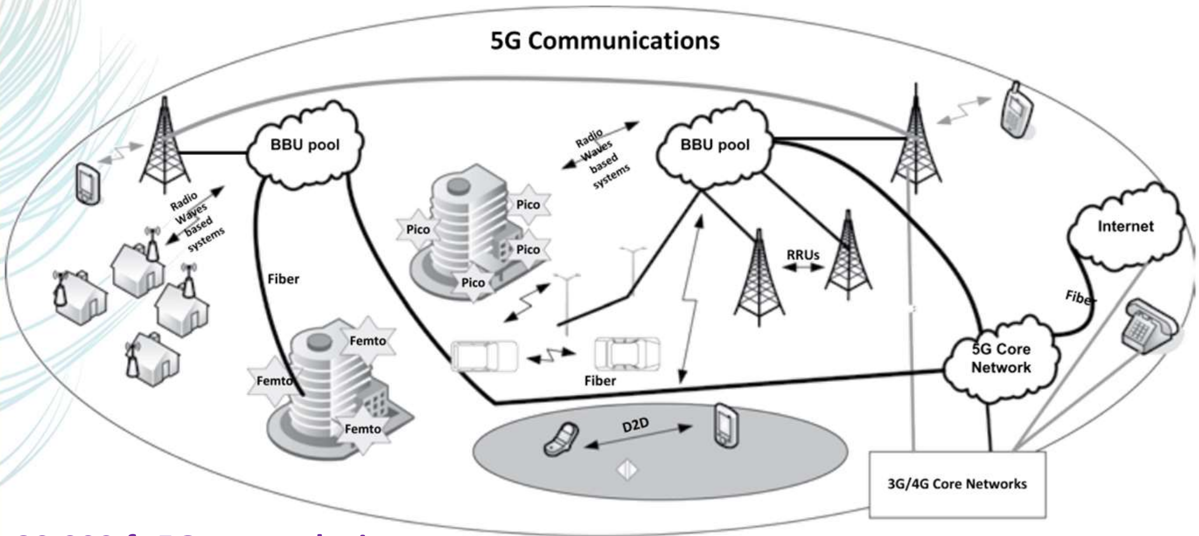
Quality of Service requirements.

MODULE 2:
5G Network Architecture

M2 S1

Introduction to 5G Network Architecture

Introduction to 5G Network Architecture



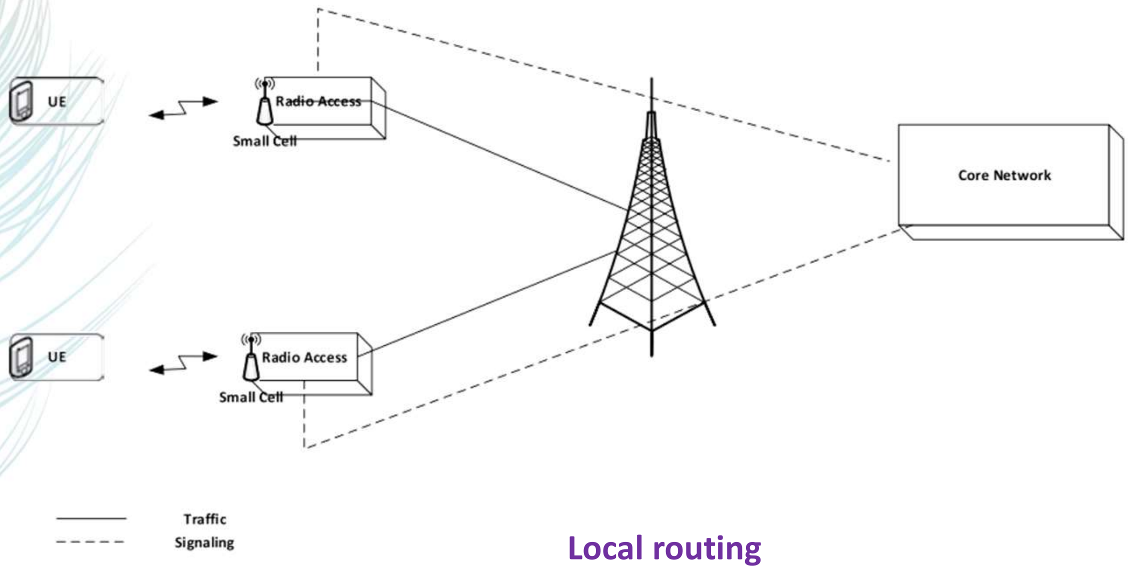
30,000 ft 5G network view

Introduction to 5G Network Architecture

30,000 ft 5G network view

- The design of 5G network is still evolving and the sector will take some time to thrash out the details.
- Nevertheless, it is expected that the architecture will utilize the cloud functionality and small cells to the deepest extent. A 30,000 ft bird’s eye view, shown in Figure, highlights the key segments of the architecture.
- It shows that small cells are connected with either view optical fiber cable or radio waves to the cloud radio access network (BBU pool).
- The pools of BBU are connected (which is called the fronthaul) to the RRUs (small cells or macro cells) and are backhauled to the core network.
- 5G core network in the form of cloud is integrated with existing 3G/4G core networks and provides connectivity to the Internet

Introduction to 5G Network Architecture

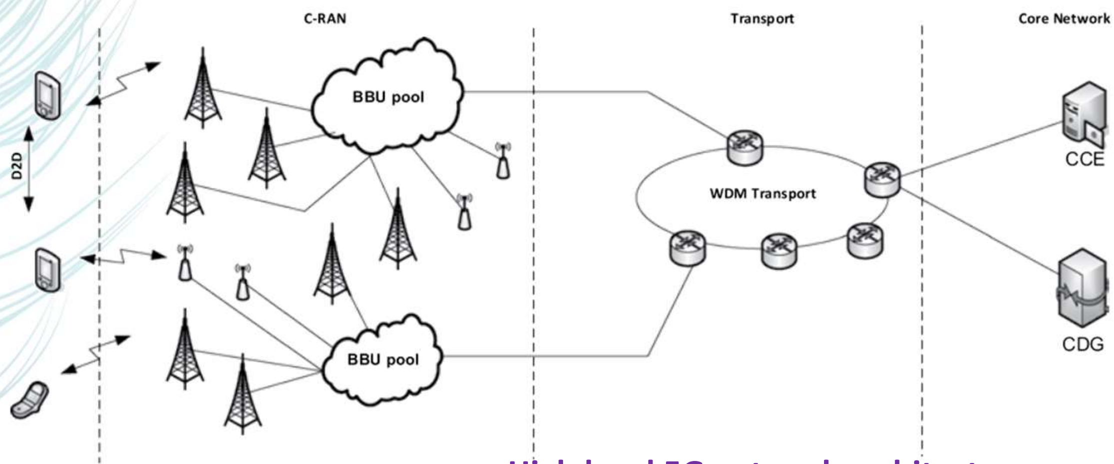


Local routing

Introduction to 5G Network Architecture

- 5G networks will consist of thousands of small cells making networks ultradense.
- This network densification is required to meet the latency and throughput requirements.
- Small cells are expected to carry a major portion of traffic with overall data volume expected to grow exponentially as predicted by many studies.
- Bringing small cells closer to the user will reduce latency and increase overall network efficiency by creating subnetworks.
- These subnetworks can have the functionality to route data traffic locally for the local users' calls who are communicating with each other while sending the signaling to the main network as shown in Figure.
- The net result from local switching is in the savings of network resources and cheaper data plans for the customers

Introduction to 5G Network Architecture

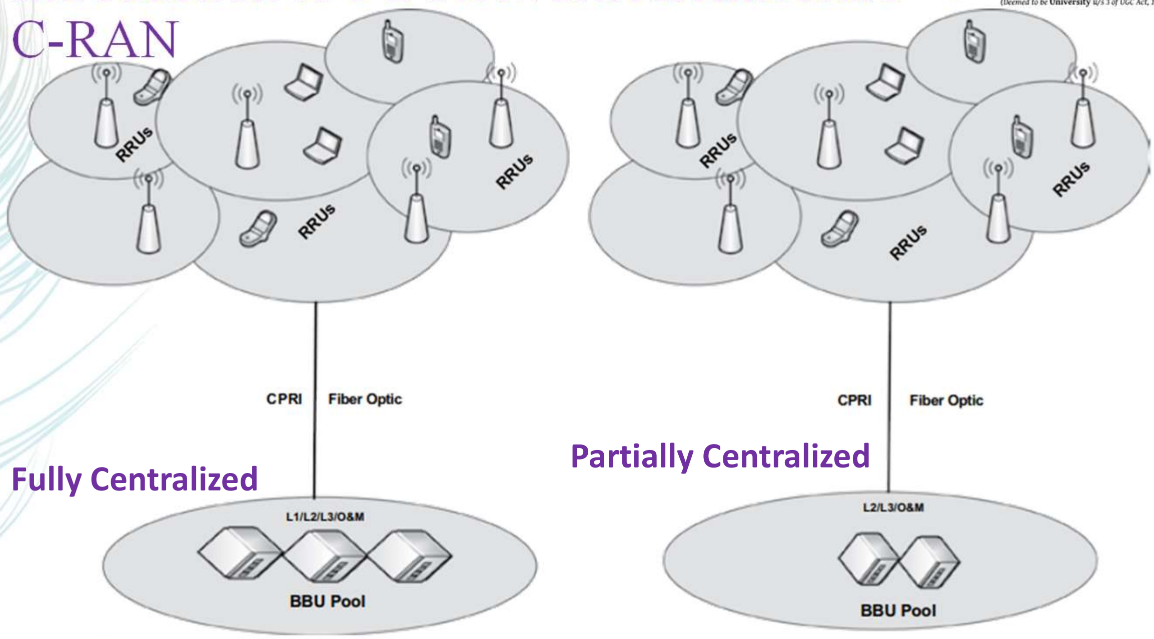


High level 5G network architecture

Introduction to 5G Network Architecture C-RAN

- Centralized RAN and Cloud RAN
- C-RAN refers to the non-static relationship between BBUs and RRUs.
- The name comes from the four “C”s, that is,
- Centralized processing,
- Clean (Green),
- Collaborative radio, and
- Cloud Radio Access Network.

Introduction to 5G Network Architecture
C-RAN

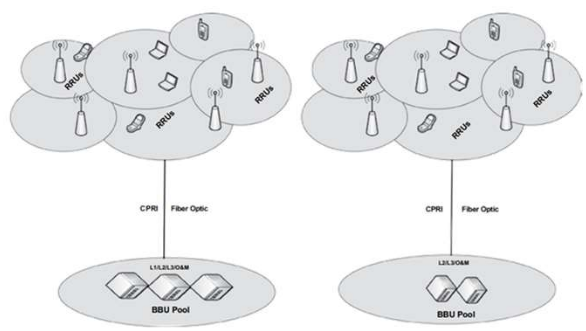


Fully Centralized **Partially Centralized**

Introduction to 5G Network Architecture
C-RAN

- Some key elements and functionalities of this architecture are

- 1. Transport
- 2. Interface
- 3. Virtualization



Fully Centralized **Partially Centralized**

Introduction to 5G Network Architecture C-RAN Architecture

- **Transport:**
- The connectivity between BBUs and RRUs in C-RAN is provided through fronthaul which is for the most part is supported via optical fiber cable, but in some cases, wireless links are also used.
- Wireless links are essential since it is very difficult to extend fiber to every RRU site.
- It is pertinent to note that as the distances for BBUs and RRUs run to 10s of kilometers, a mix of wired (fiber, Ethernet) and wireless technology may have to be utilized.
- Furthermore, in some cases, repeater sites may be placed between RRU sites and pools of BBUs to address terrain, right-of-way, and monetary challenges.

Introduction to 5G Network Architecture C-RAN Architecture

- **Interface(s):**
- C-RAN is primarily supported by CPRI (common protocol radio interface).
- The CPRI release 7.0 which was published in October 2015, specifies ten-line bit rate options with the lowest at 614.4 Mbps and the highest at 24.3 Gbps.
- It is essential for BBUs and RRUs to support at least one such bit rate.
- CPRI protocol is sensitive to latency and the synchronization performance of a transmission system, so in a number of cases, it is limited to a distance of about 40 km.

Introduction to 5G Network Architecture C-RAN Architecture

- **Virtualization:**
- The placement of a number of BBUs in a centralized pool while distributing RRUs according to targeted RF strategies means that operators employ virtualization technology that maps radio signals from/to one RRU to any BBU processing entity in the pool.
- The functions of BBUs may be realized through software instances making those virtual base stations.
- However, full virtualization is more of a long-term solution necessitating the use of virtualized BBUs running on commercial servers within an NFV (Network Function Virtualization) platform.

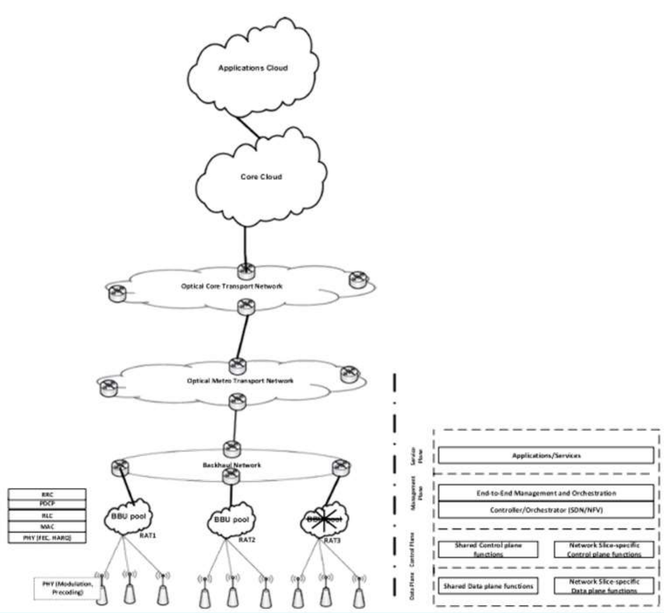
Introduction to 5G Network Architecture key benefits of C-RAN are

- **Energy Efficient Infrastructure:**
- C-RAN is an eco-friendly and energy efficient concept.
- The BBU pool in C-RAN is a shared resource, thus low power consumption and better load balancing can be achieved by dynamically allocating processing capability during a 24-hour period.
- During night time, several BBUs can either be turned off or put on low power to save energy without affecting the 24/7 service commitment.
- **Cost Savings:**
- The centralization allows placement of BBU pools in a few locations which will save O&M (operations and maintenance) costs.
- With C-RAN, transmission equipment can be shared as well, reducing CAPEX (Capital Expenditure) as well OPEX (Operational Expenditure) to some extent.

Introduction to 5G Network Architecture key benefits of C-RAN are

- **Improved Spectrum Utilization::**
- C-RAN allows implementation of joint processing and scheduling to mitigate intercell interference which improves spectral efficiency.
- For example, CoMP (Cooperative Multi-point Processing) technique of LTE-Advanced, which mitigates intercell interference, can be implemented under the C-RAN infrastructure.
- **New Business Models:**
- Cloud RAN may bring new as well as enhance existing business models such as base station (BS) pool resource rental system, cellular infrastructure sharing, intellectual property pooling (patent pooling), and so on.

Introduction to 5G Network Architecture



!!THANK YOU!!
!! Have a Nice Day!!

Today we learned about

Introduction to 5G Network Architecture

MODULE 2:
5G Network Architecture

M2 S2

5G Core Network (5GC)

5G Core Network (5GC)

- Core networks traditionally have been designed as a single architecture addressing a range of requirements and supporting backward compatibility.
- This one size fits all approach has been successful in keeping the costs down to a reasonable level and by supporting legacy circuit switched and today's packet switched functionalities.
- This core network, however, is rigid in the sense that it is not flexible enough to accommodate the customized and variable connectivity needs of individual users and businesses that are expected in the future.
- However, with virtualization, NFV, SDN, and network slicing, it is possible to make core networks more flexible and scalable.
- Thus, the next generation core network is expected to exist in a cloud-based environment with a high degree of virtualization and software-based networking.
- Such flexibility is needed to support a variety of access networks such as 3G, LTE, 4G, WiFi, and tomorrow's 5G.

5G Core Network (5GC)

Components of Core Network/High Level Architecture.

- The current EPC will further evolve to support virtualization and network slicing to become NGC applicable for 5G networks.
- Network slicing is often termed as logical instantiation of a network possibly due to virtualization technologies.
- The concept is seen as the natural extension/evolution of the current network sharing methodologies.
- Network slicing is one of the promising techniques that will likely exist in both radio access and core networks. It allows multiple logical networks to be created on top of a common physical infrastructure.

5G Core Network (5GC)

Components of Core Network/High Level Architecture.

- Either DCN (Dedicated Core Network) or a combination of NFV and SDN can be used as a technology to enable network slicing along with orchestration and analytics.
- DCN or Décor as defined in 3GPP TS 23.401 is a feature that enables an operator to deploy multiple logical mobile core networks connected to the same RAT or multiple RATs (e.g. GERAN, UTRAN, E-UTRAN, WB-E-UTRAN and NB-IoT).
- A DCN consists of one or more MME/SGSN and it may be comprised of one or more SGW/PGW/PCRF.
- This feature enables subscribers to be allocated to and served by a DCN based on subscription information (e.g., "UE Usage Type").

5G Core Network (5GC)

Components of Core Network/High Level Architecture.

- With 5G, a single terminal can use multiple services with different characteristics almost simultaneously.
- In such cases, a network slice can be created for each service, requiring all such slices to coordinate control for that particular single terminal.
- These slices can be mapped to respective radio and core network slices to provide end-to-end connectivity.
- The methodology is currently being specified for selecting radio/core networks particularities for supporting slicing in existing as well as in future 5G systems.

!!THANK YOU!!
!! Have a Nice Day!!

Today we learned about

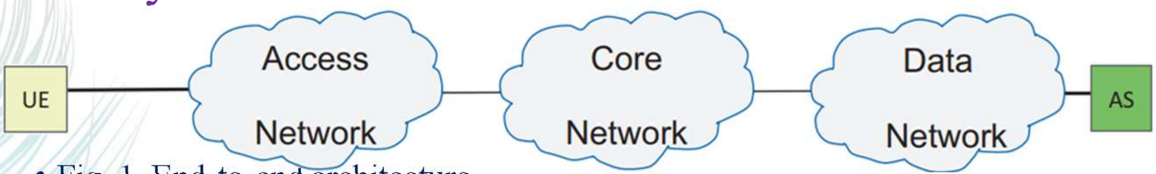
5G Core Network (5GC)

MODULE 2:
5G Network Architecture

M2 S3

5G System Architecture

5G System Architecture



- Fig. 1. End-to-end architecture
- The 3GPP 5G system defines the architecture for communication between a User Equipment (UE) and an end point, such as an Application Server (AS) in the Data Network (DN), or another UE.
- The interaction between the UE and the Data Network is via the Access Network and Core Network as defined by 3GPP Standards.
- Figure 1 depicts a simple representation of an end-to-end architecture.
- In this Unit we will focus on describing the 5G Core as defned by 3GPP 5G standards for Public land mobile network (PLMN).
- The Access Network in 3GPP is referred to as Radio Access Network (RAN).

5G System Architecture

- At a very high level, the Core and RAN consist of several Network Functions which are associated with Control Plane and User Plane functionalities.
- The actual data (also refer it as user data) is normally transported via a path in the User Plane, while the Control Plane is used to establish the path in the User Plane.
- **The Short Message Service (SMS) is an exception in which the data (short message) is communicated via the Control Plane**

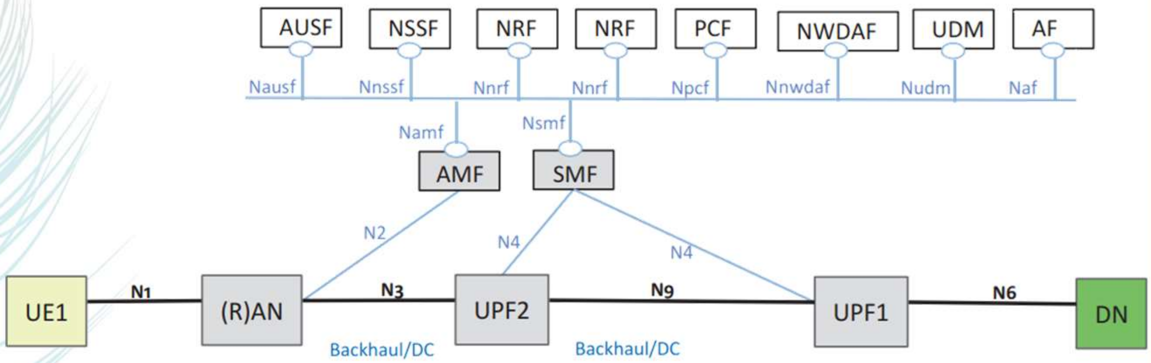
5G System Architecture

- The 5G System architecture (5GS) is represented in two ways in the 3GPP standards,
  one is a service-based representation
  in which the control plane network functions access each other's services,
  and
  the other is a reference point representation
  in which the interaction between the network functions is shown with point-to point reference points.
- In this unit we use the service-based representation since the 5G architecture is defind as service-based architecture

5G System Architecture

- In current specifications, the Service-based interfaces are defned within the Control Plane only.
- In 3GPP terminology, "a network function can be implemented either as a network element on a dedicated hardware, as a software instance running on a dedicated hardware, or as a virtualized function instantiated on an appropriate platform, e.g. on a cloud infrastructure."
- The Release 16 specifications add capability for Direct Communication and Indirect Communication between the Network Functions and Network Function services.
- The indirect communication is via a Service Communication Proxy (SCP).
- The SCP is not used for Direct Communication.

5G System Architecture



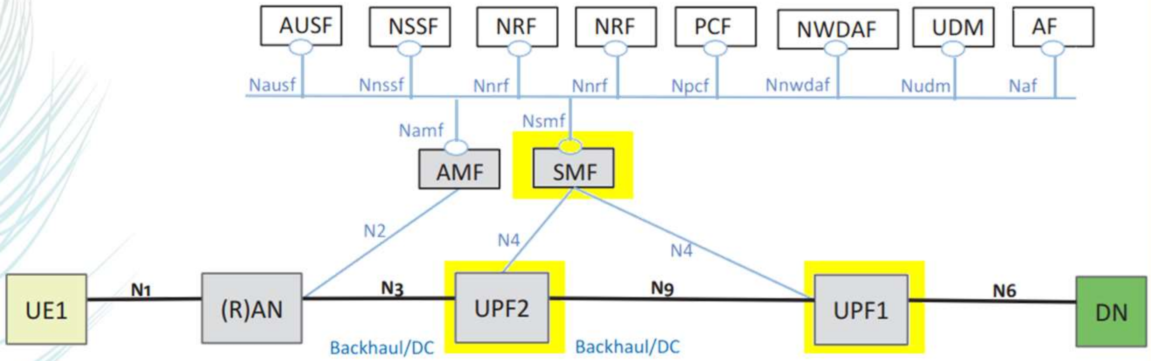
- Fig. 2. 5G System architecture

- The 3GPP 5GS service based non-roaming reference architecture in shown in Fig.

5G System Architecture

- in Release-14 (**4G**) was enhanced with an optional feature that allowed separation of control plane and user plane.
- In this feature, the Serving Gateway (SGW) and Packet Gateway (PGW) are divided into distinct control plane and user plane functions (e.g., SGW-C and SGW-U).
- This optional feature provided more flexibility and efficiency in network deployment.
- In 5G architecture, the separation of control plane and user plane is an inherent capability.
- The Session Management Function (SMF) handles the control plane functionality for setup and management of sessions, while the actual user data is routed through the User Plane Function (UPF).

5G System Architecture

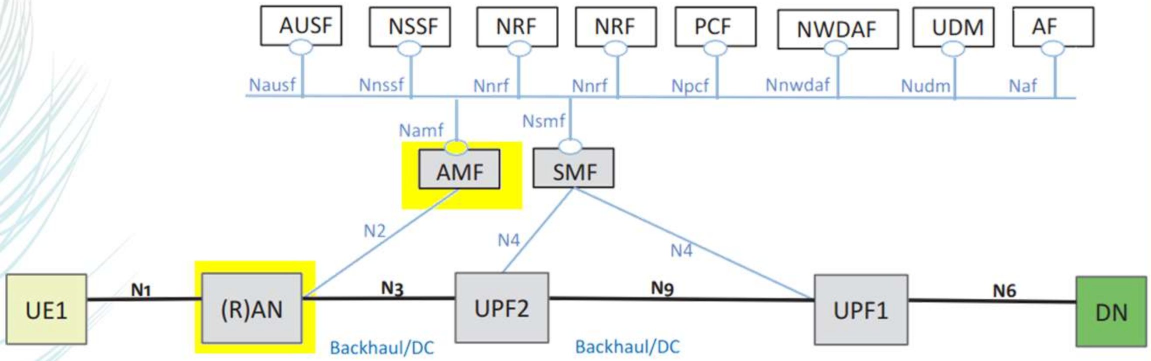


- Fig. 2. 5G System architecture
- The 3GPP 5GS service based non-roaming reference architecture in shown in Fig.

5G System Architecture

- In 5G architecture, the separation of control plane and user plane is an inherent capability.
- The Session Management Function (SMF) handles the control plane functionality for setup and management of sessions, while the actual user data is routed through the User Plane Function (UPF).
- The UPF selection (or re-selection) is handled by SMF.
- The deployment options allow for centrally located UPF and/or distributed UPF located close to or at the Access Network

5G System Architecture

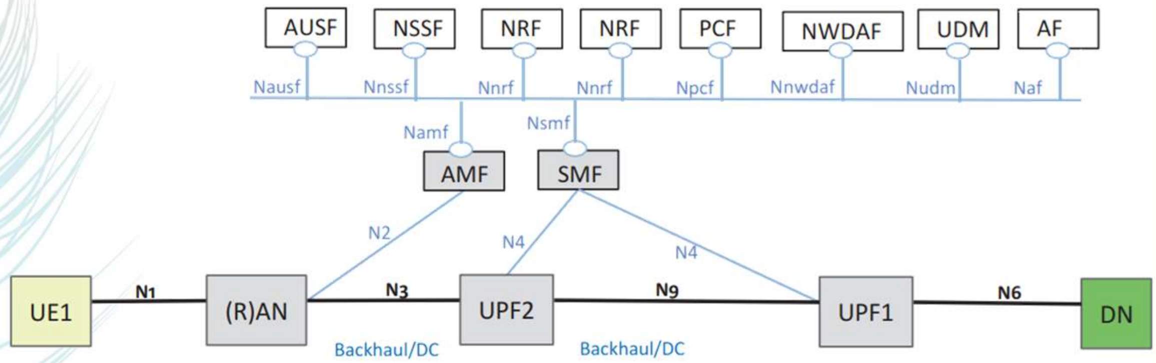


- Fig. 2. 5G System architecture
- The 3GPP 5GS service based non-roaming reference architecture in shown in Fig.

5G System Architecture

- The Access and Mobility Management function (AMF) handles the mobility management and procedures.
- AMF is termination point for control plane connection from (Radio) Access Network ((R)AN) and UE.
- The connection between UE and AMF (which traversed through RAN) is referred to as Non-Access Stratum (NAS).
- The Session Management Function (SMF) handles the session management procedures.
- The separation of the mobility and session management functionalities allows for one AMF to support different Access Networks (3GPP and non-3GPP), while SMF can be tailored for specific Accesses

5G System Architecture

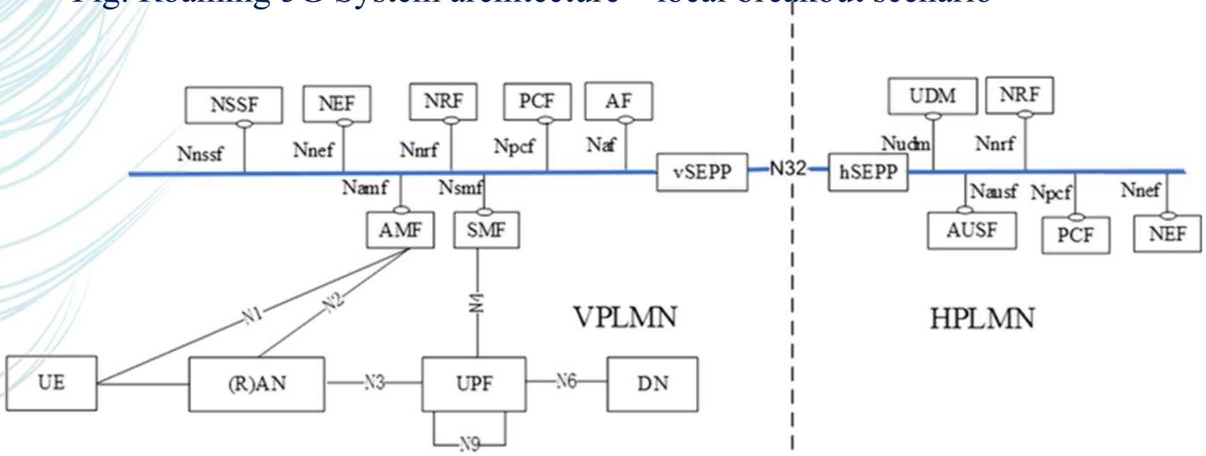


- Fig. 2. 5G System architecture

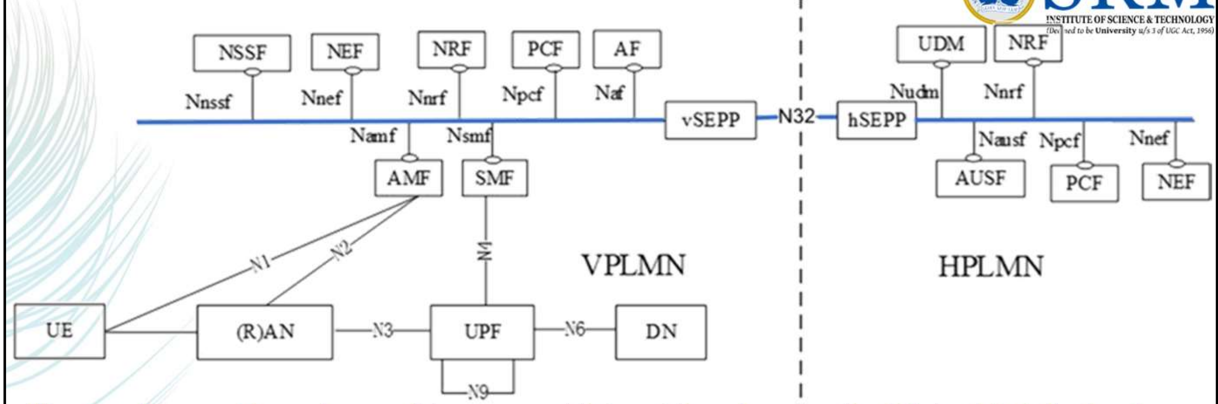
- The 3GPP 5GS service based non-roaming reference architecture in shown in Fig.

5G System Architecture

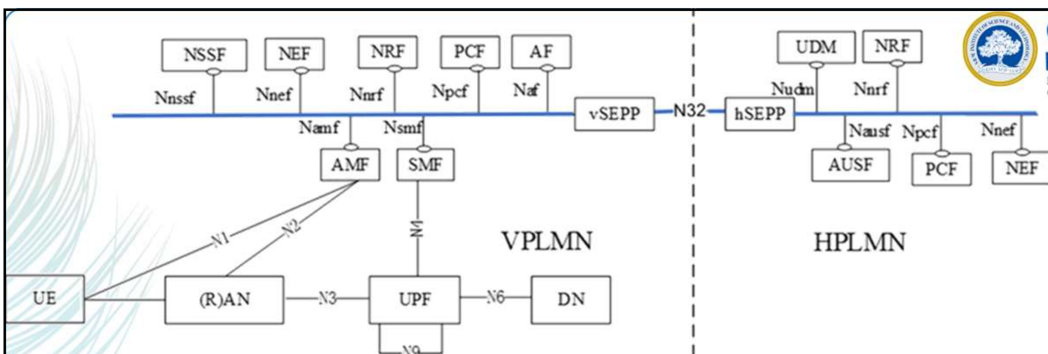
- Fig. Roaming 5G System architecture—local breakout scenario



- Fig. Roaming 5G System architecture—local breakout scenario



- Figure shows a Roaming architecture with local breakout at the Visited Public land mobile network (VPLMN).
- In this scenario the Unifed Data Management (UDM), which includes the subscription information, Authentication Server Function (AUSF), which includes authentication/authorization data, and Network Slice-Specifc Authentication and Authorization Function (NSSAAF), which support Network Slice-Specifc Authentication and Authorization, are located in the Home PLMN (HPLMN)



- There are Security Edge Protection Proxies (SEPP) that protect the communication between the Home and Visited PLMNS.
- UE communicates to Data Network (DN) via the User Plane Functions (UPF) in the VPLMN.
- The AMF and the Session Management Function (SMF) which handle the mobility and the session management for the UE are located in the VPLMN as well

!!THANK YOU!!
!! Have a Nice Day!!

Today we learned about

5G System Architecture

MODULE 2:
5G Network Architecture

M2 S4

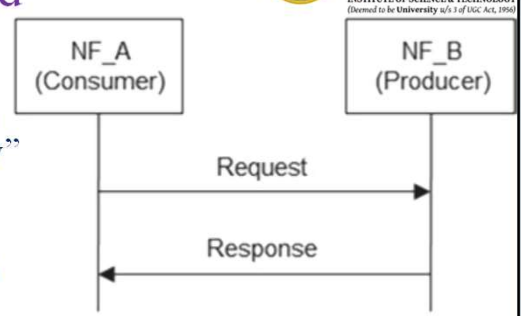
5G Core Network (5GC)

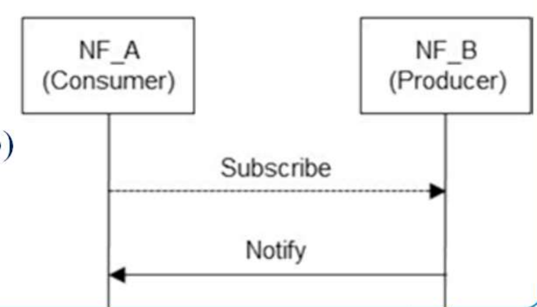
5G Core (5GC) Service-Based Architecture

- A major change in the **5G Core** (5GC) architecture compared to 4G and the previous generations is the introduction of the service-based architecture.
- In 4G architecture, the control plane functions communicate with each other via the direct interfaces (or reference points) with a standardized set of messages.
- In the service based architecture, the **Network Functions** (NF), using a common framework, expose their services for use by other network functions.
- In the 5GC architecture model, the interfaces between the networks functions are referred to as **Service Based Interfaces** (SBI).
- The Service Framework defines the interaction between the NFs over SBI using a Producer–Consumer model.
- As such a service offered by a NF (Producer) could be used by another NF (Consumer) that is authorized to use the service.
- The services are generally referred to as "NF Service" in 3GPP specifications.

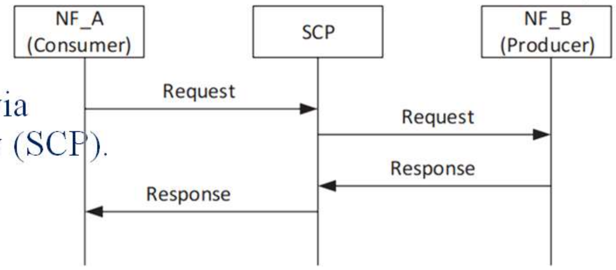
5G Core (5GC) Service-Based Architecture

- The interaction between the NFs may be a "Request-response" or a "Subscribe Notify" mechanism.
- In the "Request-response" model, an NF (consumer) request another NF (producer) to provide a service and/or perform a certain action (see Fig.a)
- In "Subscribe-Notify" model, an NF (consumer) subscribes to the services offered by another NF (producer) which notifes the subscriber of the result (Fig. b)
- The communication between the NF Service consumers and NF Service Producers maybe direct or indirect. .



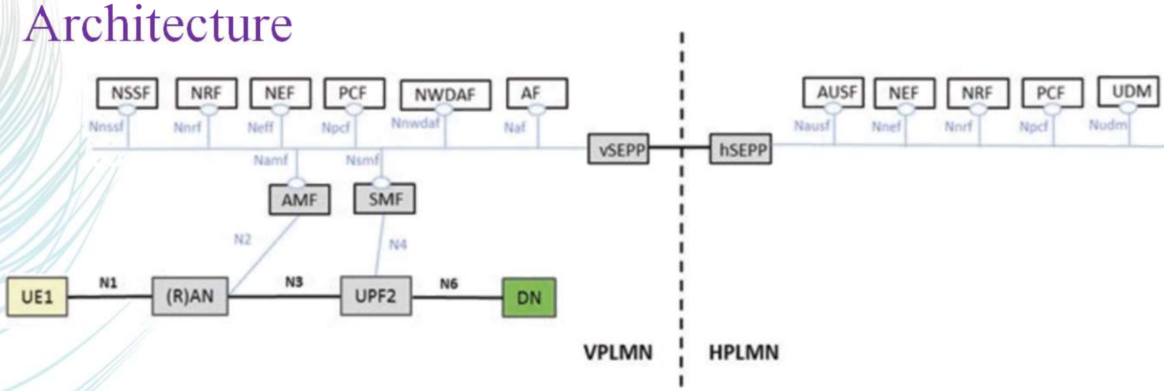


5G Core (5GC) Service-Based Architecture



- The indirect communication is via a Service Communication Proxy (SCP).
- The SCP is not used for direct Communication.
- The NF Service Producer/Consumer may include a Binding Indication in the Response/Request, which may be used by the NF Service Consumer to select a NF service producer instance for subsequent requests or used by the NF Service Producer to discover a suitable notification endpoint.
- For indirect communication, the binding indication is used by the SCP to discover a suitable NF service producer instance or a suitable notification target as appropriate

5G Core (5GC) Service-Based Architecture



- As can be seen in Fig. in the 5G System Architecture, each network function has an associated service-based interface designation.
- For example, "Namf" designates the services exhibited by the Access and Mobility Management function (AMF)

5G Core (5GC) Service-Based Architecture

- 3GPP specifications define a set of Services that are offered/supported by each Network Function. Access and Mobility Management function
- For example, the NF services specified for AMF are shown in Table

Service name	Description
Namf_ Communication	This service enables an NF to communicate with the UE through N1 NAS messages or with the AN. Enables an NF consumer to communicate with the UE and/or the AN through the AMF. This service enables SMF to request EBI allocation to support interworking with EPS
Namf_ EventExposure	Enables other NF consumers to subscribe or get notified of the mobility-related events, statistics, or other Event ID
Namf_MT	Enables an NF consumer to make sure UE is reachable to send MT signaling or data to a target UE
Namf_Location	Enables an NF consumer to request location information for a target UE

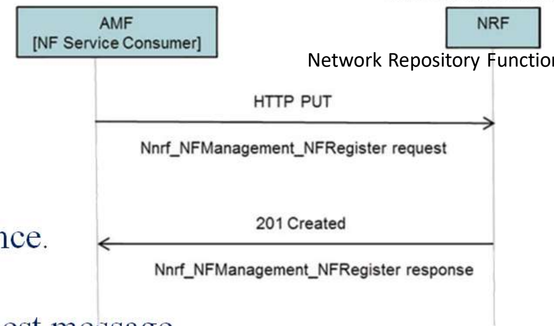
5G Core (5GC) Service-Based Architecture

There are **three main procedures** associated with the Service Framework as defined in 3GPP:

1. **NF service registration and de-registration**: to make the Network Repository Function (NRF) aware of the available NF instances and supported services.
2. **NF service discovery:** enables a NF (Consumer) to discover NF instance(s) (Producer) that provide the expected NF service(s). A NF typically performs a Services Discovery procedure with NRF for NF and NF service discovery.
3. **NF service authorization:** to ensure the NF Service Consumer is authorized to access the NF service provided by the NF Service Provider (Producer).

(Fully Qualified Domain Name)

5G Core (5GC) Service-Based Architecture

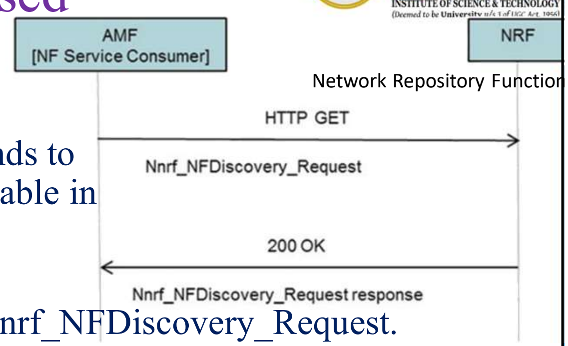


1. Example of NF Service Registration

- AMF as the NF service consumer sends a HTTP PUT request to NRF with the resource URI representing the NF Instance.
- The request contains Nnrf_NFManagement_NFRegister request message (the NF profile of NF service consumer) to NRF to inform the NRF of its NF profile.
- The NF profile of NF service consumer includes information such as NF type, FQDN or IP address of NF, names of supported services, etc.
- The NRF authorizes the request and upon success stores the NF profile of NF service consumer and marks the NF service consumer available.
- The NRF acknowledges the success of AMF Registration by returning a HTTP 201Created response containing the Nnrf_NFManagement_NF Register response (including the NF profle).

(Fully Qualified Domain Name)

5G Core (5GC) Service-Based Architecture



2. Example of NF Service Discovery

- The AMF as NF service consumer intends to discover NF instances or services available in the network for a targeted NF type.
- The AMF sends HTTP GET request to NRF in the same PLMN by invoking Nnrf_NFDiscovery_Request.
- This request contains Expected NF service Name, NF Type of the expected NF instance, NF type of the NF consumer, and may also include other information/parameters such as Subscription Permanent Identifier (SUPI), AMF Region ID, etc.
- The NRF authorizes the request, and if allowed the NRF determines the discovered NF instance(s) or NF service instance(s) and provides the search results to the NF service consumer (e.g., AMF) in a HTTP 200 OK.

!!THANK YOU!!
!! Have a Nice Day!!

Today we learned about

5G Core Network (5GC)

MODULE 2:
5G Network Architecture

M2 S5

Network Slicing:

Network Slicing

- From the 3GPP point of view, a 5G network slice is viewed as a logical network with specific functions/elements dedicated for a particular use case, service type, traffic type, or other business arrangements with agreed upon service-level agreement (SLA).
- It is important to note that 3GPP only defines network slicing for 3GPP defined system architecture and does not address transport network slicing or resource slicing of components.

Network Slicing

5G System architecture

- The most commonly discussed slice types in industry are enhanced Mobile Broadband (eMBB), Ultra-Reliable Low Latency Communications (URLLC), and massive IoT (mIoT).
- However, there could be many more network slices.
- In 4G systems there is an optional feature called eDecor to support Dedicated Core Networks (DCNs) to allow selection of the core networks based on UE's subscription and usage type.
- The network slicing in 5GS is more complete solution that provides capabilities for composing multiple dedicated end-to-end networks as slices..

Network Slicing

5G System architecture

- An end-to-end Network Slice includes the Core Network Control Plane and User Plane Network Functions as well as the Access Network (AN).
- The Access Network could be the Next Generation (NG) Radio Access Network described in 3GPP or the non-3GPP Access Network with the Non-3GPP Inter Working Function (N3IWF) or Trusted Non-3GPP Gateway Function TNGF, or the Trusted WLAN Interworking Function (TWIF) to the trusted WLAN (in case of interworking with Wire line Access Network, it will be Wire line Access Gateway Function (W-AGF)).
- To emphasis that there could be multiple instances of a network slice, the 3GPP 5GS specifications define the term 'Network Slice instance' as set of Network Function instances and resources (e.g., compute, storage, and networking resources) which form a Network Slice..

Network Slicing

5G System architecture

- In 5GS, the Network Slice Selection Assistance Information (NSSAI) is a collection of identifications for network slices.
- A network slice is identified by a term referred to as Single-NSSAI (S-NSSAI).
- The S-NSSAI signaled by the UE to the network assists the network in selecting a particular Network Slice instance.
- An S-NSSAI is comprised of a Slice/Service type (SST) and an optional Slice Differentiator (SD) which may be used to differentiate among multiple Network Slices of the same Slice/Service type..

Network Slicing

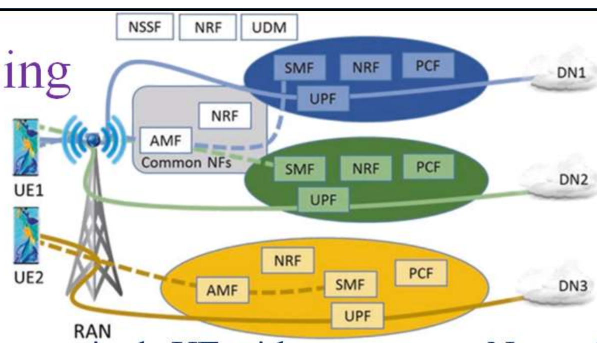
5G System architecture

- An S-NSSAI can have standard values or non-standard values.
- The S-NSSAI with standard value means that it is comprised of an SST with a standardized SST value.
- An S-NSSAI with a non-standard value identifies a single Network Slice within the PLMN with which it is associated..

Slice/service type	SST value	Characteristics
eMBB	1	Slice suitable for the handling of 5G enhanced Mobile Broadband
URLLC	2	Slice suitable for the handling of ultra-reliable low latency communications
MIoT	3	Slice suitable for the handling of massive IoT
V2X	4	Slice suitable for the handling of V2X services

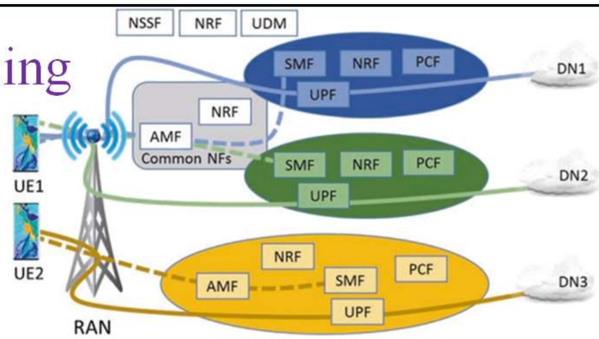
Network Slicing

5G System architecture



- A network may serve a single UE with one or more Network Slice instances simultaneously via a 5G-AN.
- Figure shows an example of three Network Slices in 5GS.
- For Slice 1 and Slice 2, the Access and Mobility Management Function (AMF) instance that is serving the UE1 and UE2 is common (or logically belongs) to all the Network Slice instances that are serving them.
- The UE in Slice 3 is served by another AMF.
- Other network functions, such as the Session Management Function (SMF) or the User Plan Function (UPF), may be specific to each Network Slice..

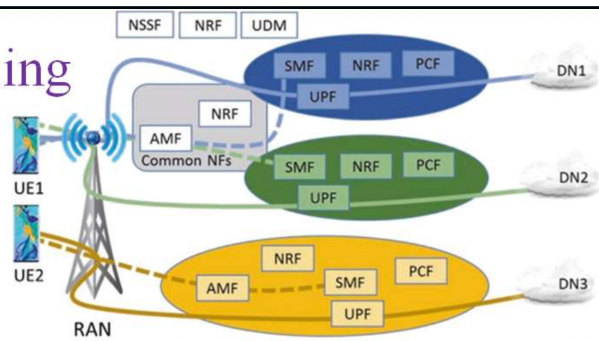
Network Slicing



5G System architecture

- The Network Slice instance selection for a UE is normally triggered as part of the registration procedure by the first AMF that receives the registration request from the UE.
- The AMF retrieves the slices that are allowed by the user subscription and may interact with the Network Slice Selection Function (NSSF) to select the appropriate Network Slice instance (e.g., based on Allowed S-NSSAIs, PLMN ID, etc.).
- The NSSF contains the Operators' policies for slice selection. Alternatively, the slice selection policies may be configured in the AMF. It is also possible for the Home PLMN to provision the UE with Network Slice selection policy (NSSP) as part of the URSP (UE Route Selection Policy) rules...

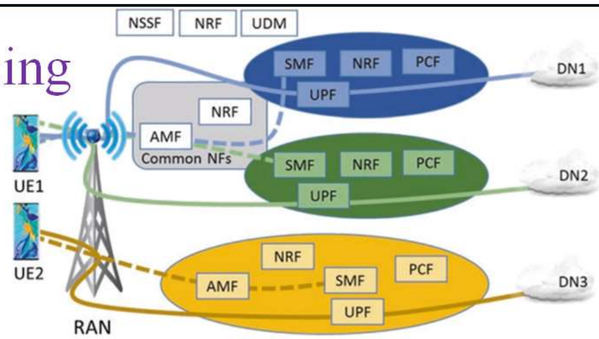
Network Slicing



5G System architecture

- The data connection between the UE and Data Network (DN) is referred to as PDU session in 5GS.
- In 3GPP standards, a PDU Session is associated to an S-NSSAI and a DNN (Data Network Name).
- The establishment of a PDU session is triggered when the AMF receives a Session Management message from UE.
- The AMF discovers candidate Session Management Functions (SMF) using multiple parameters (including the S-NSSAI provided in the UE request) and selects the appropriate SMF....

Network Slicing

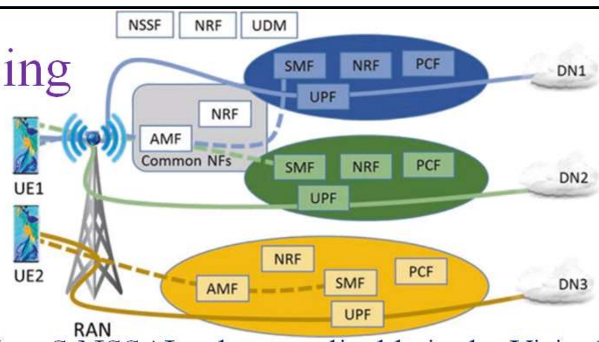


SRM INSTITUTE OF SCIENCE & TECHNOLOGY (Deemed to be University u/s 3 of UGC Act, 1956)

5G System architecture

- The selection of the User Plane Function (UPF) is performed by the SMF.
- The Network Repository Function (NRF) is used for the discovery of the required Network Functions using the selected Network Slice instance The data transmission can take place after a PDU session to a Data Network is established in a Network Slice.
- The S-NSSAI associated with a PDU Session is provided to the (R)AN, and to the policy and charging entities, to apply slice-specific policies.....

Network Slicing



SRM INSTITUTE OF SCIENCE & TECHNOLOGY (Deemed to be University u/s 3 of UGC Act, 1956)

5G System architecture

- For roaming scenarios, S-NSSAI values applicable in the Visited PLMN (VPLMN) depend on if the UE only uses standard S-NSSAI values or not.
- If UE is using the standard S-NSSAI values, then Home PLMN (HPLMN) S-NSSAI values will be used in Visited PLMN (VPLMN).
- If there is an SLA between the Visited and Home networks to support non-standard S-NSSAI values, then the NSSF in VPLMN maps the Subscribed S-NSSAIs values to the S-NSSAI values that are to be used in the VPLMN.
- It is also possible for the visited operator to set policy and the configuration in the AMF, in which case the AMF may decide the S-NSSAI values to be used in the VPLMN and the mapping to the Subscribed S-NSSAIs......

!!THANK YOU!!
!! Have a Nice Day!!

Today we learned about

Network Slicing:

MODULE 2:
5G Network Architecture

M2 S5

Virtualization in 5G & Edge Computing

Virtualization in 5G

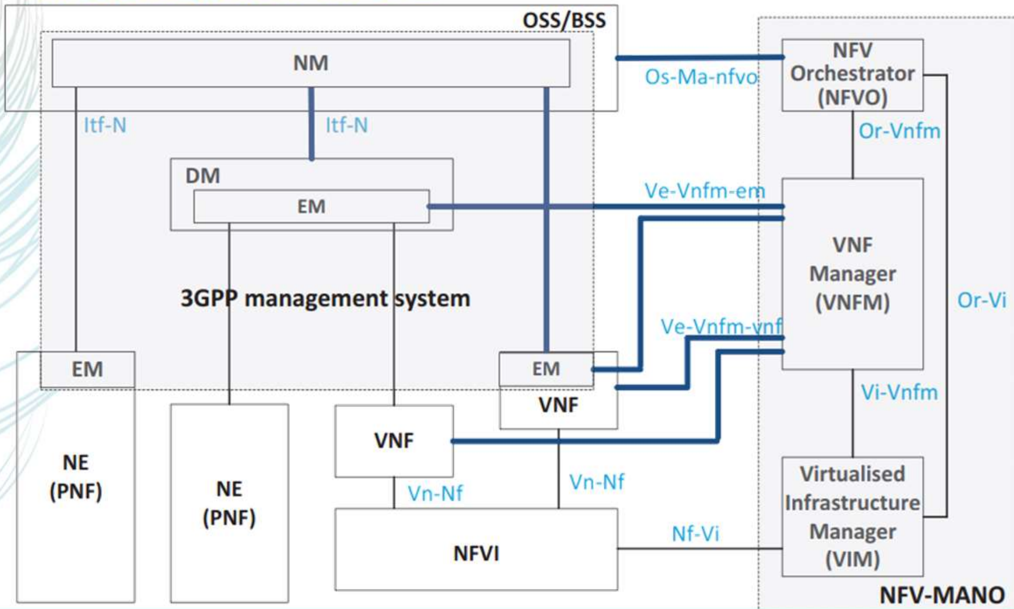
- The 5GS embraces Network Function virtualization (NFV) and cloudification for its architecture design.
- The 5GS supports different virtualization deployment scenarios, such a
  - – A Network Function instance can be deployed as fully distributed, fully redundant, stateless, and fully scalable NF instance that provides the services from several locations and several execution instances in each location.
  - – A Network Function instance can also be deployed such that several network function instances that are present within a NF set are fully distributed, fully redundant, stateless, and scalable as a set of NF instances.
- The network slicing feature supported by 5GS is also enabled by the virtualization.
- Network function instances can be created, instantiated, and isolated with each other in the virtualized environment, into different network slices in order to serve different services

Virtualization in 5G

- In order to manage the life cycle of the virtualized 5GS functions and its instances as well as the virtual resource of network slice,
  5G OAM provides means to integrate with virtualized network function management and
  orchestration capability,
  as well as
  providing standardized life cycle management interfaces with other virtualized function management and
  orchestration system which are defined by other standards, such as ETSI ISG NFV, and other open source project, such as ONAP.

SRM INSTITUTE OF SCIENCE & TECHNOLOGY (Deemed to be University u/s 3 of UGC Act, 1956)

Virtualization in 5G



SRM INSTITUTE OF SCIENCE & TECHNOLOGY (Deemed to be University u/s 3 of UGC Act, 1956)

Virtualization in 5G

- Figure illustrates the integration of 5G OAM system with ETSI NFV management and orchestration (MANO) system.
- In this illustration, 5G management system provides the 5G service and functionality management, such as NM plays one of the roles of OSS/BSS and provides the functions for the management of mobile network which includes virtualized network functions; and EM/DM is responsible for FCAPS management functionality for a VNF on an application level and physical NE on a domain and element level.
- ETSI ISG NFV MANO provides the virtualized resource and life cycle management for those virtualized 5G functions and the network service and network slice composed by those functions

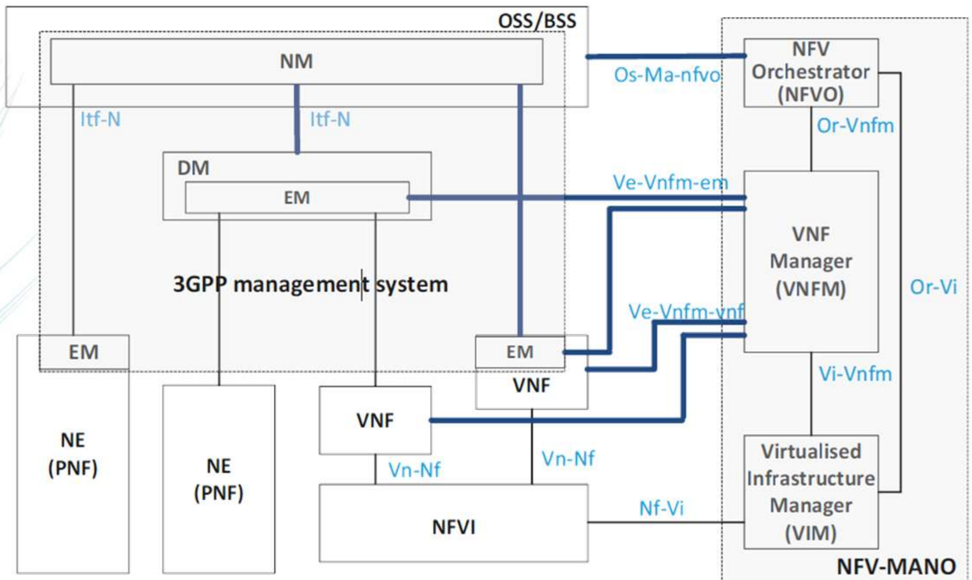
Edge Computing

- Edge computing is considered a key technology for efficient routing to application servers hosted by the operator or a third party to achieve low latency as well as efficient use of the transport network.
- In the context of 3GPP, the Edge Computing refers to the scenarios where the services need to be hosted close to the access network of the UE (e.g., at or close to the RAN).
- As described earlier, in 5GS the routing of the data (or user) traffic is done via UPF interface to a Data Network.
- The 5G core network supports the capability to select a UPF that allows routing of traffic to a local Data Network that is close to the UE's access network.
- This includes the local breakout scenarios for roaming UEs as well as non-roaming scenarios.

Edge Computing

- The decision for selection (or reselection) of UPF for local routing may be based on the information from an Edge Computing Application Function (AF) and/or to other criteria such as the subscription, location, and policies.
- Depending on the operator's policy and arrangements with the third parties, an AF may access the 5G core directly or indirectly via the Network Exposure Function (NEF).
- For example,
  an external AF at the edge data center could influence the routing of the traffic by altering the SMF routing decisions via its interaction with the Policy Control Function (PCF).

Edge Computing



Edge Computing

- The figure shows local and central data networks (DN) that a UE can reach for edge compute services.
- In this example, each of the UEs (UE1–3) requests the SMF to establish PDU sessions with a central and local anchor.
- The SMF establishes UPF anchor (PSA-0) as the central anchor for each of the PDU sessions, but they each have different UPF PSA for session breakout since the aim is to breakout to a local N6 network that is proximate to the edge application server (EAS).

Edge Computing

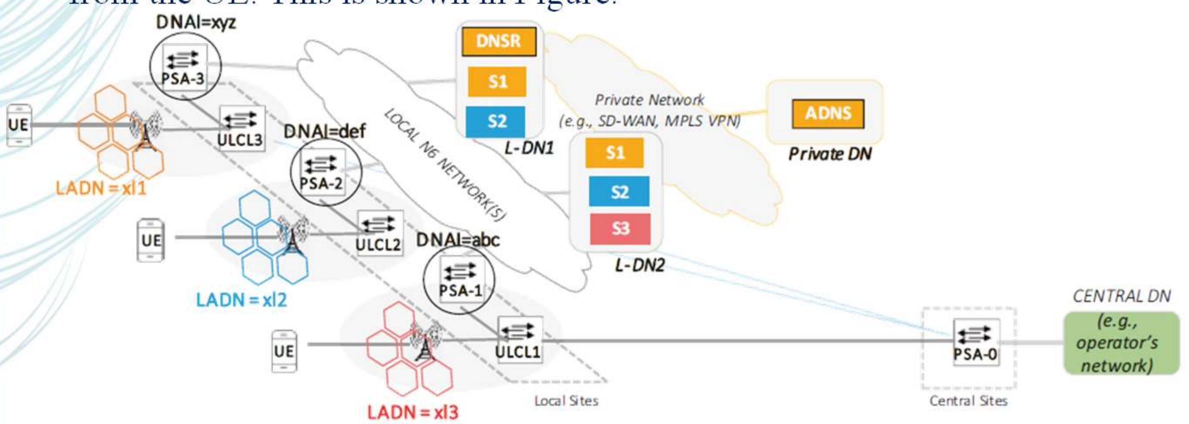
- Edge computing is considered a key technology for efficient routing to application servers hosted by the operator or a third party to achieve low latency as well as efficient use of the transport network.
- In the context of 3GPP, the Edge Computing refers to the scenarios where the services need to be hosted close to the access network of the UE (e.g., at or close to the RAN).
- As described earlier, in 5GS the routing of the data (or user) traffic is done via UPF interface to a Data Network.
- The 5G core network supports the capability to select a UPF that allows routing of traffic to a local Data Network that is close to the UE's access network.
- This includes the local breakout scenarios for roaming UEs as well as non-roaming scenarios.

Edge Computing

- For end-to-end QoS and other service guarantees (reliability, protection, etc.), the PDU session segment and the N6 segment to the application server are important to manage.
- 3GPP-specified QoS with URLLC considers the PDU session segment.
- For the N6 segment, IP routing and transport network underlays with SD-WAN, MPLS, TSN, or other means of providing reliable service guarantees is necessary.
- If the PDU session segment and the N6 segments are operated by different providers (e.g., MNO, IP transport provider, application service provider) there needs to be coordination to support end-to-end guarantees

Edge Computing

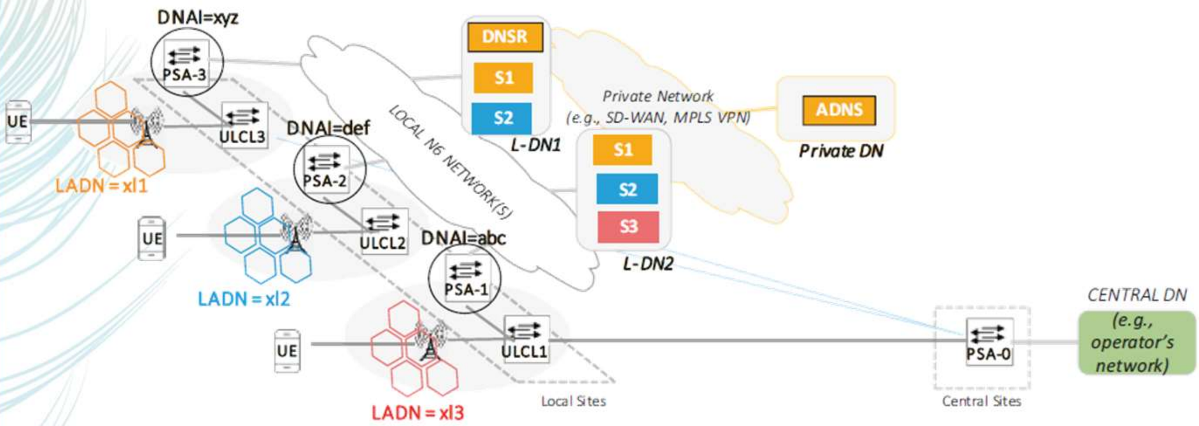
- Another concept introduced in 5G standards to support routing to application data networks (DN) based on the origin/location of the request from the UE. This is shown in Figure.



Edge Computing

- The UE is configured with LADN (Local Area Data Network) information on a per PLMN basis for 3GPP accesses.
- LADN Service Area applies to a set of tracking areas (TA) of PLMN and is configured in AMF and is only accessible by the UE at those specific locations.
- An LADN DNN (Data Network Name) corresponds to an LADN service that the UE is configured with during registration.
- Location-based selection and direction of application data packets is facilitated by LADN.
- For example, in the figure above, there are three LADN (xl1, xl2, xl3) comprising of a set of TAs (some of which may overlap).
- A UE that registers and is configured to use these sets of services (S1, S2, S3).
- Each of these sets may consist of applications that can only be accessed from a DNAI (Data Network Access Identifier).

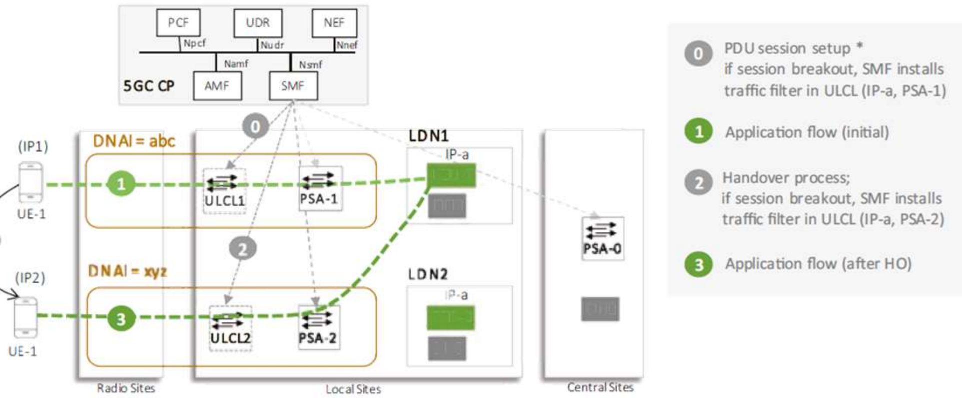
Edge Computing



Mobile Edge Computing

- A data network (DN) where applications are located is identified by a DNAI.
- For example,
  UEs with LADN xl1 is allowed to access the set of services in S1 via DNAI = xyz.
- Route filters in ULCL perform access filtering to limit access to only authorized sessions.

Mobile Edge Computing

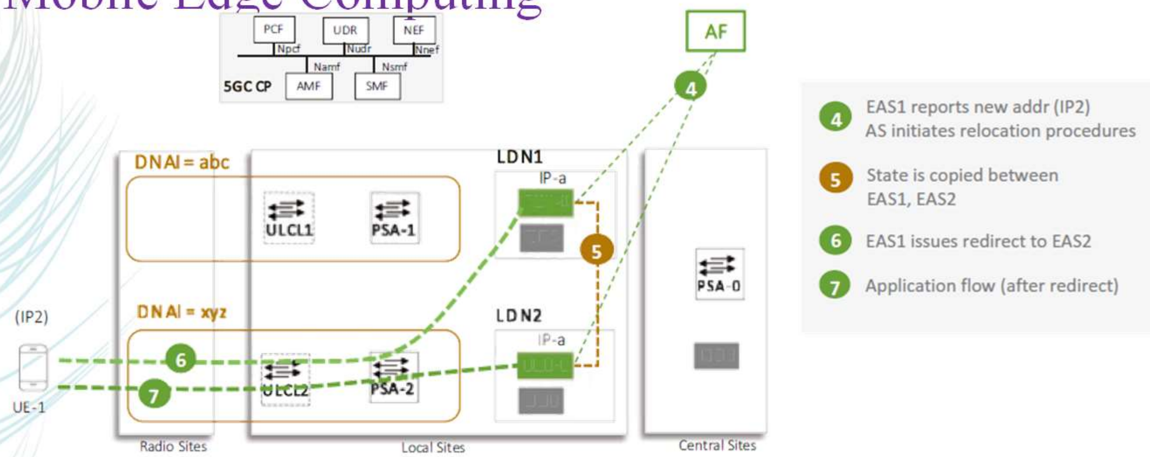


- For stateful applications and UE mobility (new IP access/DNAI), it results in selecting a new LDN (Local Data Network).
- The UE continues to be served by the old LDN location where its data is stored.
  In the figure, after the PDU session is set up (step 0),
  the UE is served by EAS-1 in the local data network (step 1).
  After handover (step 2), the UE continues to be served by EAS-1 (step 3).

Mobile Edge Computing

- The rationale is that not all UE mobility requires application server relocation.
- It may be that the current application server is close enough, or that closer application servers may be overloaded.
- The application domain orchestrator determines the appropriate change of application server based on UE mobility, but also managing congestion, site outage, mitigation of DDoS attacks, and other administrative policy.
- The application domain may also decide to move the serving application server for a variety of reasons.
- The trigger may be that following handover, the UE is better served by a closer server, but the change of application server may be initiated to balance application load, handle server failure, or comply with domain policies on transit cost.

Mobile Edge Computing

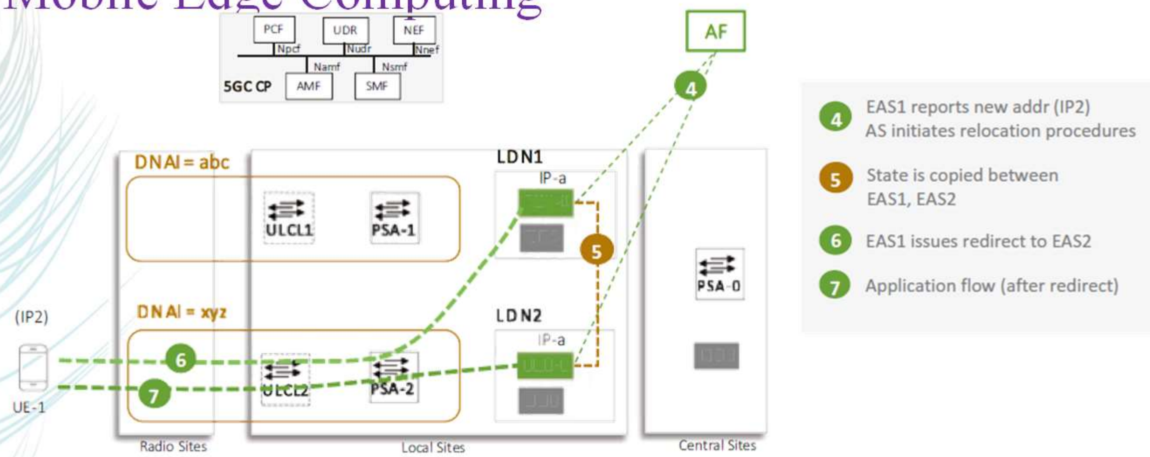


- Figure shows an overview of possible steps to achieve the relocation of application server.
- 3GPP procedures also describe application server relocation using AF influenced routing.
- In this case, the AF coordinates both PDU session and application server relocation.

Mobile Edge Computing

- When the application domain orchestrator (AF in figure) determines that the server should be relocated based on IP address change or other criteria (step 4), it selects the new server (EAS2).
- The old server (EAS1) and new server (EAS2) replicate state (step 5) for the UE during this transition.
- The old server (EAS1) then sends an application layer redirect request, e.g., HTTPS redirect (step 6) to the new server (EAS2).
- The UE has the URL translated and initiates application message to EAS2 (step 7), and the replication of state across EAS1/EAS2 is stopped.
- Thus, PDU session relocation and application relocation can both be accomplished with minimum coordination.

Mobile Edge Computing



- Engineered backhauls and IP transport along with new protocols such as QUIC enable a good end-to-end QoS during this transition by using zero RTT application session establishment capability.

Quality of Service Requirements

- Quality of service (QoS) requirements are technical specifications that specify the system quality of features such as performance, availability, scalability, and serviceability.

System Quality	Description
Performance	The measurement of response time and throughput with respect to user load conditions.
Availability	A measure of how often a system's resources and services are accessible to end users, often expressed as the uptime of a system.
Scalability	The ability to add capacity (and users) to a deployed system over time. Scalability typically involves adding resources to the system but should not require changes to the deployment architecture
Security	A complex combination of factors that describe the integrity of a system and its users. Security includes authentication and authorization of users, security of data, and secure access to a deployed system
Latent capacity	The ability of a system to handle unusual peak loads without additional resources. Latent capacity is a factor in availability, performance, and scalability qualities.
Serviceability	The ease by which a deployed system can be maintained, including monitoring the system, repairing problems that arise, and upgrading hardware and software components.

https://docs.oracle.com/cd/E19636-01/819-2326/gaxqg/index.html#:~:text=Quality%20of%20service%20(QoS)%20requirements,specified%20in%20the%20business%20requirements.

!!THANK YOU!!
!! Have a Nice Day!!

Today we learned about

Virtualization in 5G
Edge Computing
Mobile Edge Computing
Quality of Service Requirements

OLD Slides