

4G/5G COMMUNICATION NETWORKS

UNIT – I EVOLUTION OF WIRELESS NETWORKS

Networks evolution: 2G, 3G, 4G, evolution of radio access networks, need for 5G, 4G versus 5G, Next Generation core (NG-core), visualized Evolved Packet core (vEPC)

1. Explain various Networks evolution: 2G, 3G, and 4G.

1.1.1 The First Generation 1G

- These phones were the first to be used, introduced in 1982 and completed in early 1990.
- It was used for voice services based on Advanced Mobile Phone System technology AMPS.
- The AMPS system used frequency and were frequency modulated. With a channel capacity of 30 kHz and a frequency band of 824-894MHz, frequency division multiple access FDMA is used. Its primary characteristics are as follows:
 1. 2.4 kbps speed enables voice calls in one nation.
 2. Use an analogue signal.
 3. Poor voice quality, insufficient battery life.
 4. Large phone size limited capacity poor handoff dependability inadequate security
 5. Spectrum efficiency was relatively low.
- It introduces mobile technologies such as MTS Mobile Telephone System, AMTS Advanced Mobile Telephone System, IMTS Improved Mobile Telephone Service, and Push to Talk Push to Talk PTT.
- It has low capacity, unpredictable handoff, poor voice connectivity, and no security because voice communications are replayed in radio towers, making them vulnerable to unwanted eavesdropping by other parties.

1.1.2 The Second Generation 2G

- The term 2G refers to the second Generation of GSM phones, which first appeared in the late 1980s.
- For voice transmission, it uses digital signals. This technology's primary focus was on digital signals, and it provides services for delivering text and picture messages at modest speeds in kbps.
- It has a frequency range of 30 KHz to 200 kHz. In addition to 2G, 2.5G systems use packet-switched and circuit-switched domains and can give data rates of up to 144 kbps, for example, GPRS, CDMA, and EDGE, the following are the primary characteristics of 2G and 2.5G:
 1. 2G stands for the Second Generation.
 2. Data transmission rates of up to 64kbps were achieved through digital signals.
 3. Enables services such as text messaging, picture messages, and MMS Multimedia message.
 4. Incapable of dealing with complicated data such as videos.
 5. Strong digital signals were required to enable mobile phones to function, digital signals will be weak if no network coverage in a particular area.
 6. GSM technology was constantly developed to give better services, which resulted in advanced technology between 2G and 3G.

7. Calls are made available.
8. Send and receive emails Web browsing Speed: 64-144 kbps Camera phones.
9. Take 6-9 minutes to download 3-minute MP3 music.

1.1.3 The Third Generation 3G

- 3G was introduced in 2000 and is based on GSM. The goal of this technique was to provide fast data.
- Using packet switching, the original design was upgraded to offer data rates of up to 14 Mbps and higher. It employs a Wide Band Wireless Network, which improves clarity.
- It also provides broadband services, access to television and video, and innovative services like Global Roaming.
- It operates at 2100MHz and has a bandwidth of 15-20MHz, utilized for high-speed internet access and video chatting, the primary characteristics of 3G are as follows:
 1. Speed 2 Mbps typically referred to as smartphones.
 2. Broadband and data transfer rates have been increased to handle web-based applications and music and video files.
 3. It allows for speedier communication.
 4. Sending/receiving huge email messages.
 5. High-speed web access, improved security, video conferencing, 3D gaming.
 6. Broadband capabilities and large capacities.
 7. Streaming TV, mobile TV, phone calls.
 8. It takes only 11 seconds to 1.5 minutes to download 3-minute MP3 music.
 9. Expensive license fees for 3G services.
 10. The infrastructure for 3G's high bandwidth requirement proved challenging to develop.
 11. Expensive 3G handsets.
 12. Large cell phones.
- 3G cellular services were launched in the year 2003. 3G was much more advanced, when compared to 2G/2.5G and offered up to 2 Mbps speed, supporting location-based services and multimedia services.
- It was ideal for web browsing. Apple, which was known to be a computer maker, got into the mobile equipment business by launching iPhone, with the advent of 3G. Android, the open source mobile operating system became popular with 3G.
- With 3G, the 3GPP group standardized UMTS. Universal Mobile Telecommunications System (UMTS) is a third- generation mobile cellular system for networks based on the GSM standard.
- Developed and maintained by the 3GPP (3rd Generation Partnership Project). UMTS uses wideband code division multiple access (W-CDMA) radio access technology to offer greater spectral efficiency and bandwidth to mobile network operators.

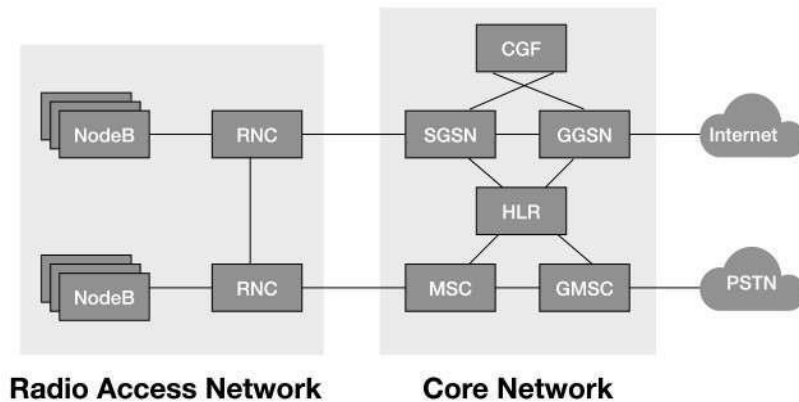


FIGURE 1.3 - 3G ARCHITECTURE

- UMTS specifies a complete network system, which includes the radio access network (UMTS Terrestrial Radio Access Network, or UTRAN), the core network (Mobile Application Part, or MAP) and the authentication of users via SIM (Subscriber Identity Module) cards.

3G network architecture has three distinct entities:

1. User Equipment (UE): In 2G, the handsets were called as mobile phones or cell phones, as they were predominantly used for making voice calls. However, in 3G, the handsets can support both voice and data services. Hence, the term User Equipment or UE is used to represent the end user device, which could be a mobile phone or a data terminal.

2. Radio Access Network (RAN): The RAN, also known as the UMTS Radio Access Network, UTRAN, is the equivalent of the previous Base Station Subsystem (BSS) in GSM. RAN includes the NodeB function and the Radio Network Controller (RNC) function. The NodeB function provides the air interface. The RNC manages the air interface for the overall network.

3. Core Network: The core network is the equivalent of Network Switching Subsystem or NSS in GSM and provides all the central processing and management for the system. Core network has both circuit switched and packet switched network elements.

3G core network architecture consists of the following functions:

Home Location register (HLR)

- HLR is a database that contains all information about the subscriber including their last known location.
- The HLR maintains a mapping between Mobile Station International Subscriber Directory Number (MSISDN) and International Mobile Subscriber Identity (IMSI). MSISDN is the mobile phone number used for making and receiving voice calls and SMS. IMSI is used for uniquely identifying a SIM card and the number is stored in the SIM card.
- Each network can have one or more physical or logical HLRs.
- User equipment periodically updates its location details to the HLR, so that calls can be routed appropriately to the user.
- Depending upon the implementation, the HLR may also have an in-built Equipment identity register (EIR) and Authentication Centre (AuC).

Equipment Identity Register (EIR)

- The EIR is the function that decides whether a user equipment is allowed onto the network or not. EIR is typically integrated with the HLR.
- EIR is used for blocking or monitoring calls from a stolen user equipment. Each user equipment is uniquely identified through a number known as the International Mobile Equipment Identity (IMEI).
- IMEI is exchanged by the user equipment at the time of registration with the network. Thus, EIR identifies a stolen equipment through its IMEI.

Authentication Centre (AuC)

- AuC is used for storing a shared secret key, which gets generated and burned in the SIM card at the time of manufacturing.
- AuC function is typically co-located with the HLR function. AuC does not exchange the shared secret key, but would run an algorithm on the International Mobile Subscriber Identity (IMSI), to generate data for authentication of a subscriber / user equipment. Each IMSI is unique and gets mapped to a SIM card.

Mobile switching center (MSC)

- MSC is responsible for functions such as routing calls and SMS messages. It interfaces with the HLR for keeping track of subscriber location and does call handovers, when the mobile subscriber moves from one location to another.
- Gateway MSC (GMSC) is a function that is present either within or outside of the MSC. A GMSC interfaces with the external networks such as the Public Switched Telephone Network (PSTN), which is our legacy land line network.

Serving GPRS Support Node (SGSN)

- SGSN is responsible for mobility management and authentication of subscribers / mobile devices in a GPRS network.
- It performs a role which is similar to the role played by the MSC for voice calls. The SGSN and MSC are often co-located in the network.

Gateway GPRS Support Node (GGSN)

- GGSN acts as a gateway to the Internet. It connects the GPRS network with the packet switched data network. GGSN receives data addressed to a given subscriber, checks if the subscriber is active and then forwards the data to the SGSN serving the particular subscriber.
- If the subscriber is inactive, the data is discarded. The GGSN keeps a record of active subscribers and the SGSN they are attached to. GGSN assigns a unique IP address to each subscriber.
- It also generates the call detail records (CDRs), which are processed by the Charging Gateway Function (CGF) or billing servers.

Charging Gateway Function (CGF)

- CGF handles Call Detail Records (CDRs) generated by the GGSN in a GPRS network. There are different types of CDRs processed by the CGF, based on the network node that generates the CDR.
- For example, when a SGSN generates CDRs, it is called S-CDR. When a GGSN generates CDRs, it is called G-CDR.
- One of the key differences between S-CDR and G-CDR is, G-CDR would have insights into the subscriber data transfers (for example, volume of data uploaded/downloaded by the subscriber).
- 3G technology evolved over a period of time to offer higher speeds by supporting a new standard called High Speed Packet Access (HSPA).

- Service providers who offered 3G services with HSPA support called their services as 3.5G or 3G+.
- The 3.5G networks that supported HSPA standards were able to offer speeds up to 7 Mbps. With the further evolution of HSPA standard (also called as Evolved HSPA), 3G networks were able to offer speeds up to 42 Mbps.

3.1.4 The Fourth Generation 4G

- 4G provides a download speed of 100Mbps, 4G offers the same features as 3G and new services such as Multi-Media Newspapers, the ability to watch TV programs with more outstanding quality, and transport data significantly quicker than previous generations.
- LTE Long Term Evolution technology is classified as 4G technology. 4G is being developed to meet the QoS and rate requirements of future applications such as wireless broadband access, Multimedia Messaging Service MMS, video chat, mobile TV, HDTV content, Digital Video Broadcasting DVB, minimal services such as voice and data, and other bandwidth-intensive services.

The significant characteristics of 4G are as follows:

1. Ability to give speeds ranging from 10Mbps to 1Gbps, high quality streaming video, a combination of Wi-Fi and Wi-Max and increased security.
 2. Provide any type of service at any time and from any location as needed by the user.
 3. Expansion of multimedia services.
 4. Low cost per bit, increased battery usage.
 5. Implementation is complex, intricate technology is required, and expensive equipment is necessary to build a next-generation network.
- In 2012, 4G services were launched, with speeds of up to 12 Mbps. 4G is an all-IP (Internet Protocol) network and it resulted in massive changes to the radio network and the core network architecture.

In 4G network,

1. The radio function is based on the Long Term Evolution (LTE) 3GPP standards and
2. The core network is based on the Evolved Packet Core (EPC) 3GPP standards

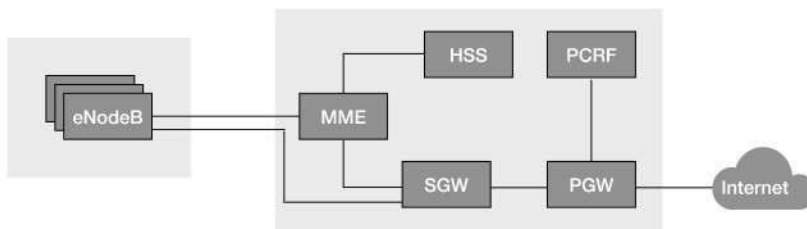


FIGURE 1.4 - 4G ARCHITECTURE

- One of the significant changes introduced by the Long Term Evolution (LTE) standards in 4G networks, is the changes to the Base station functionality.

- In 3G, the radio resources were centrally controlled by a node called Radio Network Controller (RNC). LTE introduces a new function called the Evolved NodeB (eNodeB), which manages the radio resource and mobility in the cell.
- In order to meet the 4G LTE requirements, functions of the eNodeB not only included the base station (NodeB) functions to terminate the radio interface but also the functions of the Radio Network Controller (RNC) to manage radio resources. This architecture is called Evolved UMTS Terrestrial RAN (E-UTRAN) architecture.
- In 3G, the RAN function included the base station (Node B) and the antennas. In 4G LTE architecture, the base station function is split into two key functions - Baseband Unit (BBU) and Remote Radio Head (RRH).
- RRH is connected to BBU through optical fiber. The BBU function is moved out from the cell site and hosted in a centralised location and called as Centralized RAN.
- The RRH function (i.e., the antenna function) is deployed closer to the users in a distributed fashion.
- The RAN architecture and the distribution of RRHs and BBUs is influenced by several factors such as quality of service, latency, throughput, user density and load demand.

The following are the key functional nodes/network elements in the LTE architecture:

Evolved Node B (eNB)

- eNodeB is the entity that supports air interface and performs radio resource management.
- It provides radio resource management functions such as IP header compression, user data encryption, and routing the user data to the Serving Gateway (SGW).
- The radio interface provided by eNodeB can be shared by several operators by having separate MME, SGW & PDN Gateway.

Home Subscriber Server (HSS)

- Home Subscriber Server (HSS) is a database for storing the Subscriber profile and authentication information.
- MME downloads subscriber profile information from the HSS, when a user equipment/mobile device attaches to the network. HSS also provides the subscriber profile information to the IP Multimedia Subsystem (IMS) Core function, at the time of the IMS registration.

Serving Gateway (SGW)

- SGW serves as the mobility anchor for the user plane. It takes care of inter-eNodeB handovers & User Equipment (UE) mobility between 3GPP networks.
- It is responsible for routing/forwarding data packets between the eNodeB & Packet Data Network Gateway (PDN GW).

Packet Data Network Gateway (PGW)

- PDN GW provides the UE with connectivity to the external packet data networks such as Internet.
- It serves as the anchor point for intra-3GPP network mobility, as well as mobility between 3GPP and non-3GPP networks.
- It takes care of Policy and Charging Enforcement Function (PCEF), which includes Quality of Service (QoS), online/offline flow- based charging data generation, deep-packet inspection, and lawful intercept.

Mobility Management Entity (MME)

- MME manages mobility, UE identities and security parameters.
- It operates in the Control plane and provides functions such as managing session states, authentication, mobility with 3GPP 2G/3G nodes, and roaming.

Policy and Charging Rules Function (PCRF)

- Policy and Charging Rules Function (PCRF) maintains the policy and charging related controls for all the subscribers. For example, a subscriber’s quality of service policy is stored in the PCRF server.
- The QoS policy can differ from service to service for each subscriber. The QoS for an IMS bearer may be different from the QoS for an Internet bearer for the same subscriber. Such differentiations in the QoS can be enforced by setting rules in the PCRF server.
- In addition, PCRF also helps the service providers in providing location-based services.
- PCRF allows a service provider in setting flow-based charging rules. For example, a service can be stopped, when the credit limit for the service is reached.
- With higher data speeds, 4G technology allowed users to watch high-definition video and play games online.
- Over a period of time, multiple enhancements were made to 4G technology - LTE-M (LTE Category M1 for Machines) allowed low powered IOT devices to connect to 4G networks and LTE-Advanced standards offer a network speed of up to 300 Mbps.
- Today, 4G offers adequate network speed for over the top services such as online video, gaming and social media.
- However, it does not support the bandwidth and latency needs of services such as Augmented Reality, Virtual Reality and Autonomous Cars. This paved the path for 5G technology research.

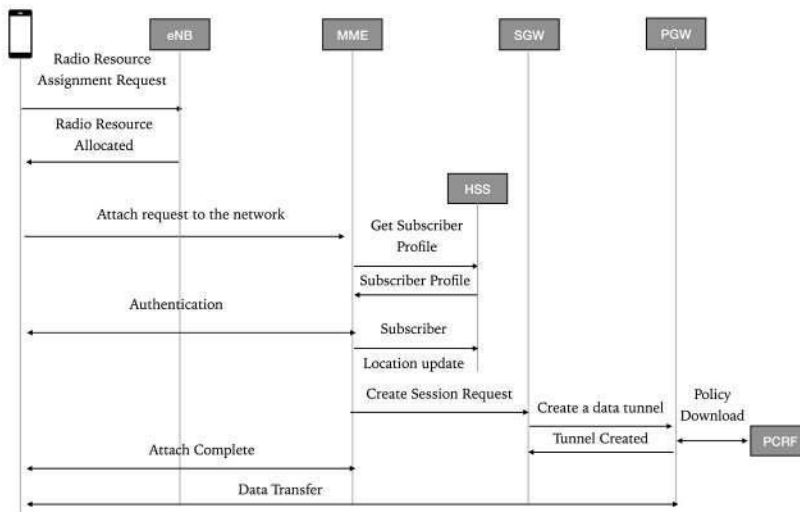


FIGURE 1.5 - 4G DATA CONNECTION ESTABLISHMENT

4G Data Connection Establishment

- There is a lot of similarities between how a data connection is established in a 3G and a 4G network. This section describes the procedures involved in establishing a data connection between the mobile equipment and the 4G network.

- When a mobile phone is powered on, it looks for signals from the cellphone towers in the vicinity.
- Based on the International Mobile Subscriber Identity (IMSI) from the SIM card, the mobile phone picks the right service provider.
- The phone then requests for a radio resource from the eNodeB. The eNodeB allocates a radio resource for the mobile subscriber.
- The moment the mobile equipment gets the radio resource, it starts displaying the wireless 'signal bar' on the console.
- Then, the mobile device (also called the User Equipment or UE), sends an "Attach" request to the network.
- The "Attach Request" reaches the MME (Mobility Management Entity) in the Evolved Packet Core (EPC).
- The first step taken by the EPC is to authenticate the subscriber based on the SIM credentials.
- The MME retrieves the subscriber profile information from the HSS/HLR. The MME issues a challenge (which includes a set of encrypted keys) to the UE.
- The UE runs the challenge against the credentials stored in SIM card.
- The UE responds back to the challenge with an authentication response. The MME validates the authentication response based on the profile information retrieved from the HSS/HLR. The subscriber is now authenticated.
- Once the mobile subscriber is authenticated, the EPC proceeds with the session initiation process.
- The MME sends a "Create Session Request" to the Serving Gateway. The Serving Gateway sets up a tunnel with the PDN Gateway (PGW). As part of establishing the tunnel, the PGW downloads policy information from PCRF and applies them on the subscriber context.
- Once the tunnel, is created, the MME responds back to the UE with an "Attach Accept" response. The bearer / tunnel is setup based on the Internet Access Point Name (APN).
- APN will typically look like "internet.telco.com" and is configured in the UE by the service provider, as part of the initial configuration download to the mobile device.
- The moment, a tunnel is created (which means the data session is established), the mobile equipment starts displaying the '4G' symbol on the console.

Voice calls in 4G Network

- There are different mechanisms available to handle voice calls in a 4G network.
- The two popular mechanisms to handle a voice call are Circuit Switched Fall-Back (CSFB) and Voice over LTE (VoLTE).

Circuit Switched Fall-Back (CSFB)

- When LTE is just used for data transfer, voice calls are handled through the legacy circuit switched mechanisms - by falling back to a 3G or 2G network.
- Circuit Switched Fall- Back (CSFB) works only when the area covered by an LTE network is also covered by the 3G network. CSFB will be helpful for service providers when they are migrating from 2G/3G to a 4G network.
- In CSFB, the 4G MME talks to the 3G MSC through the new SGs interface, to setup the voice call.
- User Equipment (UE), initiates a "Combined Attach" procedure to both PS (Packet Switched) and Circuit Switched (CS) networks.

- MME receives the "Combined Attach" request and sets up the PS connection over the 4G Core, for data transfers.
- The newly introduced SGs interface between the MME and MSC is used for the CS connection setup over the 3G core, for voice calls.
- Once the UE is attached to both the 4G and the 3G networks, the eNodeB directs the UE to the 3G NodeB radio.
- The UE sets-up a voice call over the 3G NodeB. This circuit switched fallback to the 3G network, is equivalent of a handover from the 4G to 3G network, for voice calls.

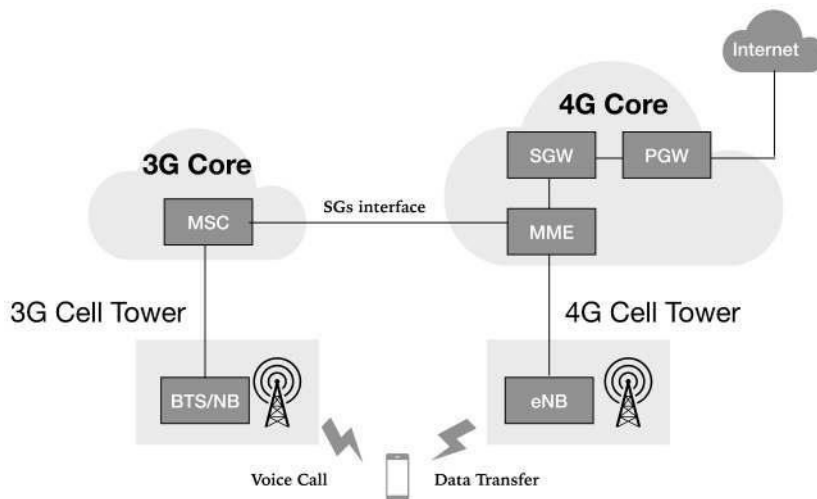


FIGURE 1.6 - CIRCUIT SWITCHED FALL-BACK

Voice over LTE (VoLTE)

- Voice over LTE is relatively a new concept, to support voice calls over the 4G network.
- While CSFB helped the service providers during the migration from 2G/3G to 4G networks, VoLTE runs completely on the 4G network.
- In case of VoLTE, the user equipment / mobile should be capable of initiating a VoLTE call and the network should support VoLTE.
- VoLTE calls are handled by the IP Multimedia Subsystem (IMS) core, in the 4G network.
- Unlike the OTT (Over the Top) calling services like Skype or Whatsapp, VoLTE service uses the same dialer application used by the CSFB service. It also provides reliability, when compared to the OTT calling services.
- For example, when the service provider is not able to establish the call via the VoLTE, the phone would automatically switch to the 2G/3G based circuit switched calls.
- This helps when a customer is making an emergency call.

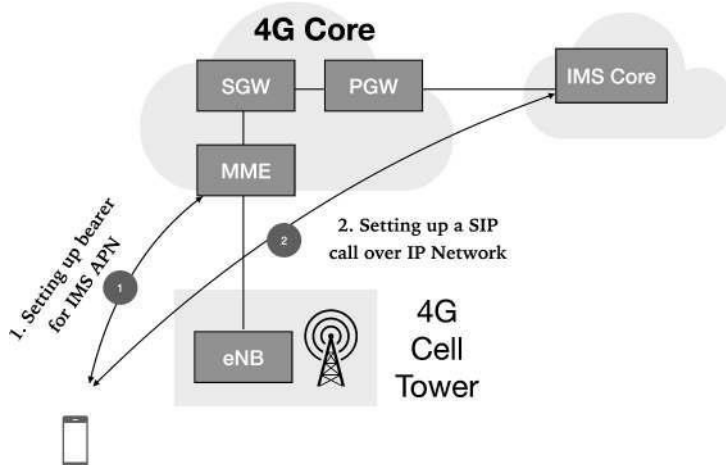


FIGURE 1.7 - VOICE OVER LTE

Setting up a VoLTE is a two-step process:

1. At first, the UE sets up a dedicated bearer/tunnel for IMS APN (Access Point Name). For example, the APN name will look like `ims.telco.com`. This is configured by the service provider on the UE. These settings automatically get downloaded to the mobile phone, as part of the service activation by the service provider. This bearer for IMS APN will be setup, in addition to the bearer setup for Internet APN (i.e., for data transfers).

The procedure for bearer setup is similar to the procedure outlined in the “4G data connection establishment” section.

2. Once the bearer is established, the UE sets up a Session Initiation Protocol (SIP) connection with the IMS Core. SIP is a popular protocol used for Voice over IP (VOIP) communications, over the Internet.

- Unlike the OTT VOIP dialer applications, the service provider guarantees reliability and security for the voice calls made over an LTE connection. Voice over Wi-Fi (VoWi-Fi) is also similar to VoLTE.
- However, the wireless service provider would not be able to guarantee the reliability of the voice calls made over the Wi-Fi connection.
- When the Internet speeds over Wi-Fi are high and reliable, VoWiFi calls help the service provider to offload the mobile wireless network for other applications/services. Hence, many service providers support VoWiFi capabilities.

1.1.5 The Fifth Generation 5G

- The term 5G refers to the Fifth Generation, which began from small developments in the late 2010s. 5G technology may provide significantly improved levels of connectivity and coverage.
- The primary focus of 5G will be on the global-wireless World Wide Web WWW. It is an entirely wireless communication system with no wires.
- 5G Technology is an abbreviation for 5th Generation Mobile Technology. 5G technology features exceptional data capabilities and the capacity to connect unlimited call volumes and complete data broadcast within the most current mobile operating system.

- 5G technology has a promising future because it can handle cutting-edge technologies and provide clients with priceless handsets.
- Perhaps 5G technology will take over the global market in the following days. 5G technologies offer a remarkable capacity to assist software and consulting. Router and switch technology is employed in the 5G network to provide a high connection.
- The 5G technology offers internet connectivity to building nodes and can be deployed with wired and wireless network connections.
- Shortly, 5G technology will deliver a cell phone that functions similarly to a PDA, and the entire office will be at our fingertips/on our phones.
- In a few years, we may be able to download a full-length HD movie in six seconds, but 4G takes seven minutes, and 3G takes more than an hour.
- In addition, video chats will be so immersive that we will feel as if we can reach out and touch the other person through the screen, 5G is a packet-switched wireless technology with high throughput and extended area coverage.
- 5G wireless employs OFDM and millimeter wireless, allowing for data rates of up to 20 Mbps and frequency bands ranging from 2 to 8 GHz. The 5G communication technology is envisioned as a virtual wireless network capable of wireless World Wide Web access WWW.
- The uncertainty around 5G stems from the fact that it is still primarily a concept, and the wireless industry has yet to agree on any standards for the new network, however, it is attempting to attain some success .
- Primary aims for 5G, significantly faster data speeds, and currently 4G networks may achieve peak download speeds of one gigabit per second however, this is never achieved in practice.
- This would increase to 10Gbps with 5G. Ultra-low latency, "Latency" refers to the time it takes for one device to transfer a data packet to another device. The latency rate in 4G is roughly 50 milliseconds; however, in 5G, it will be around one millisecond.
- This will be critical for industrial applications as well as driverless cars. A more "connected world", the Internet of Things intelligent home appliances, connected autos, etc., is predicted to increase enormously over the next decade, necessitating the development of a network capable of supporting billions of connected devices.
- 5G will provide users with the capacity and bandwidth they require. Technology is still a long way from being a reality, but it has the potential to drastically revolutionize how we interact with wireless gadgets, from our smartphones to the cars we drive.

The present 5G technology trend has the following feature.

1. 5G technology can provide connectivity speeds of up to 25 Mbps.
2. 5G technology provides cell phone users with high quality and bi-directional extensive bandwidth sharing.
3. 5G technology enables large-scale data transmission at gigabit speeds, with about 65,000 connections supported.
4. The uploading and downloading speeds of 5G technology have reached a new high.
5. Virtual private networks are also supported by 5G technology.
6. The 5G terminals will include software-defined radios and modulation Methods and new error-control systems that can be downloaded via the Internet.
7. The focus of 5G mobile networks is expected to shift toward user terminals.
8. 5G technology's improved billing interfaces make it more appealing and effective.

9. The 5G technology network provides expanded and widespread connectivity Worldwide.
10. The terminals will have simultaneous access to many wireless technologies and the terminal should be able to mix distinct flows from various technologies.
11. Vertical handovers should be avoided since they are impractical when there are numerous technologies, operators, and service providers.
12. In 5G, each network will handle user mobility, while the terminal will make the final decision among several wireless, mobile access network providers for a specific service.
13. Such a decision will be based on open, intelligent middleware in mobile phones.
14. The remote diagnosis is a fantastic function provided by 5G, allowing users to get better and faster solutions.

Table 1-1 shows a brief comparison of evolution of wireless communication technology.

Technology	1G	2G	3G	4G	5G
Start/Deployment	1970-80	1990-2004	2004-10	Now	Soon (probably by 2020)
Data Bandwidth	2Kbps	64 Kbps	2 Mbps	1 Gbps	Higher than 1 Gbps
Technology	Analog	Digital	CDMA 2000, UMTS,EDGE	Wi-Max, Wi-Fi, LTE	WWWW
Core Network	PSTN	PSTN	Packet N/W	Internet	Internet
Multiplexing	FDMA	TDMA/CDMA	CDMA	CDMA	CDMA
Switching	Circuit	Circuit,Packet	Packet	All Packet	All Packet
Primary Service	Analog Phone Calls	Digital Phone Calls and Messaging	Phone calls, Messaging, Data	All-IP Service (including Voice Messages)	High speed, High capacity and provide large broadcasting of data in Gbps
Key differentiator	Mobility	Secure, Mass adoption	Better Internet experience	Faster Broadband Internet, Lower Latency	Better coverage and no dropped calls, much lower latency, Better performance
Weakness	Poor spectral efficiency, major security issue	Limited data rates, difficult to support demand for internet and e-mail	Real performance fail to match type, failure of WAP for internet access	Battery use is more, Required complicated and expensive hardware	?

TABLE 1.1 EVOLUTION WIRELESS COMMUNICATION

7. Explain Evolution of Radio access networks/ What is RAN? And its functions (Nov/Dec 23)

- The Radio Access Network (RAN) architecture has evolved across the different generations of the wireless network, to support the bandwidth and scalability requirements.
- RAN has two distinct units – the Remote Radio Head (RRH) and the Baseband Unit (BBU). One end of the RRH is connected to the antenna and the other end to the BBU.
- RRH acts as a transceiver converting the analog signals to digital signals and vice versa.

Generation	Architecture / Technology	Base Station
2G	GSM	Base Transceiver Station (BTS)
3G	UMTS	NodeB
4G	LTE	Evolved NodeB (eNodeB)
5G	NR	Next Generation NodeB (gNodeB)

TABLE 1.2 - EVOLUTION OF RAN

- In addition, RRH also does filtering of noise and amplification of signals. The Baseband Unit (BBU) provides switching, traffic management, timing, baseband processing, and radio interfacing functions. BBU is typically connected to the RRH using a Fiber link.
- In the traditional 2.5G/3G networks, both the RRH and BBU functions remained in the cell site, as part of the Base Transceiver Station (BTS).
- In the 4G network, BBU function was moved out of the cell site to a centralized location. BBU function in a 4G network is hosted in the Central Office and is called as Centralized RAN.
- 4G architecture optionally supports the virtualization of BBUs and when the BBU function is virtualized, it is also called as Cloud RAN or Virtualized RAN. In a 5G network, virtualization of BBUs almost becomes mandatory as it helps the service providers to scale the network to support the various use cases.

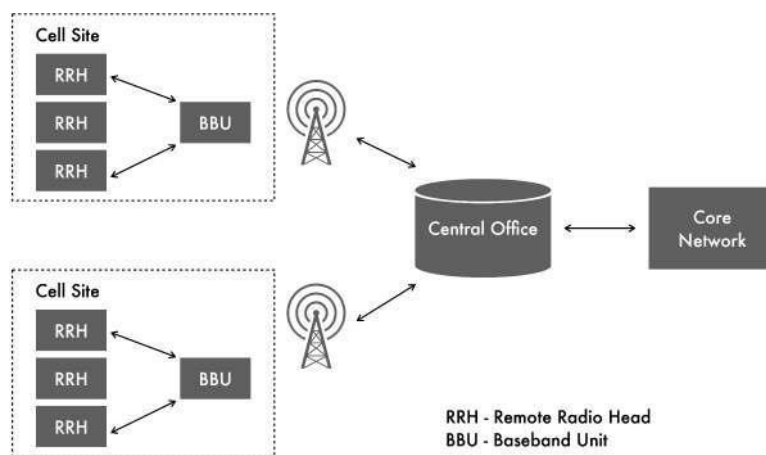


FIGURE 1.8 - TRADITIONAL RAN

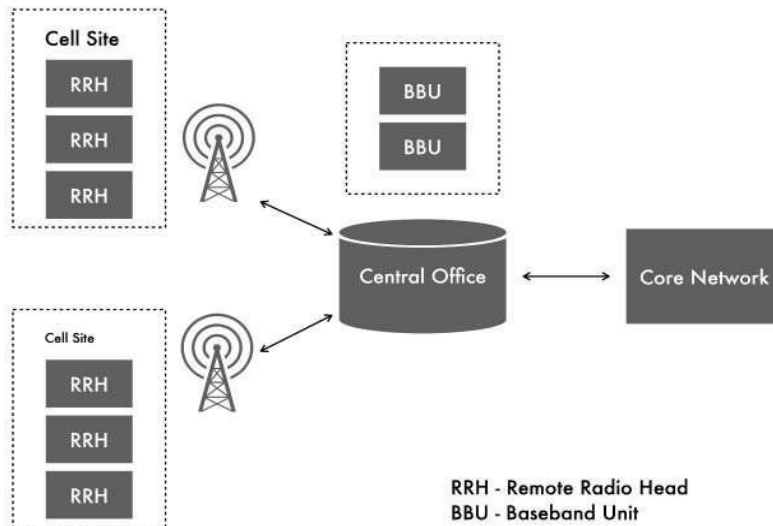


FIGURE 1.9 - CENTRALIZED RAN

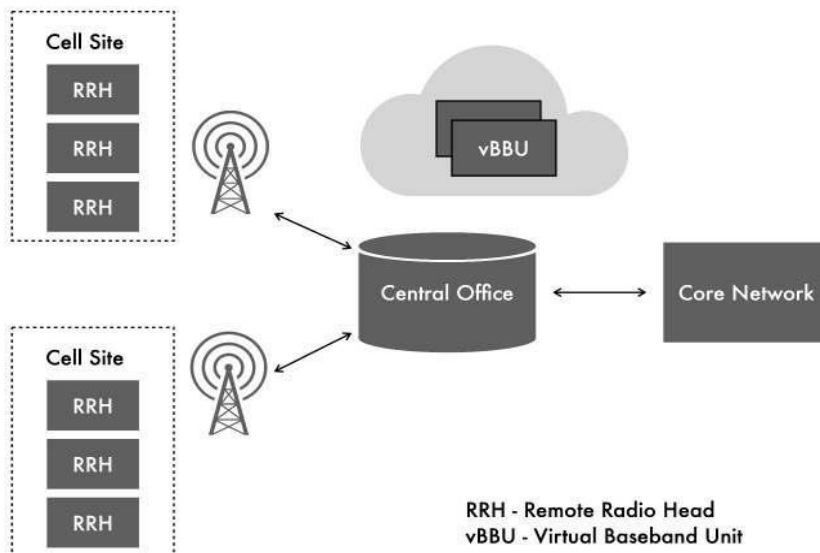


FIGURE 1.10 - VIRTUALIZED RAN

8. Discuss the Need for 5G.

- Most of the previous generation wireless technologies (such as 3G and 4G) were focused on increasing the speed of the wireless technology.
- 4G technology initially supported speeds of up to 12 Mbps - which was adequate for online video streaming and gaming services.
- However, 4G does not address the technology needs of some of the emerging use cases, in the areas of Internet of Things (IOT) and Virtual Reality. Here is the list of factors that drive the need for 5G technology:
- Internet of Things (IOT) will require an infrastructure that can handle several billions of network devices connecting to the wireless network and at the same time energy efficient.
- 3D video and Ultra High Definition Video streaming applications are hungry for additional bandwidth.

- Virtual Reality and Augmented Reality enabled gaming, video streaming and industrial applications require sub- millisecond latencies.
- Network operators have immense pressure to upgrade their networks continuously, to handle the growth in the mobile data traffic - and at the same time, reduce operational expenses.
- Enable new revenue streams for wireless service providers, by supporting new applications and use-cases

9. Discuss 4G versus 5G.

- 4G network infrastructure is based on Long Term Evolution (LTE) architecture. 5G network infrastructure is based on 5G Next Generation Core (5G NG-Core) architecture.
- There is a significant difference between both the technologies in terms of speed, latency, frequency ranges of the spectrum, use cases that are supported, support for network slicing, RAN architecture, and Core network architecture.

Criteria	4G	5G
Speed	300 - 400 Mbps (lab) 40 - 100 Mbps (real world)	1000 Mbps (lab) 300 - 400 Mbps (real world)
Latency	50 ms	1 - 2 ms
Frequency	2 - 8 GHz	Sub 6 GHz (5G macro optimized), 3-30 GHz (5G E small cells)
Use Cases	Voice over LTE Mobile Broadband Online Video Online Gaming	Enhanced Mobile Broadband Augmented Reality / Virtual Reality Internet of Things (IOT) Holographic Calls Fixed Wireless Autonomous Cars Robotic Surgeries
Network Slicing	No	Yes
Cell Towers	Large Towers in concentrated communities	Small Cells installed in almost every street corner, in addition to mobile towers
Service Architecture	Connection Oriented	Service Oriented
Architecture	Long Term Evolution (LTE)	Next Generation Core (NG-Core) New Radio (NR)

TABLE 1.3 4G VERSUS 5G

10. Explain Next Generation Core (NG-Core) (Nov/Dec 23)

- NG-Core for 5G is the equivalent of Evolved Packet Core (EPC) in a 4G network.
- 5G NG-Core architecture supports virtualization and allows the user plane functions to be deployed separately, from the control plane functions.
- In addition, the user plane and control plane functions can be scaled independently.
- 5G NG-Core supports both International Mobile Subscriber Identity (IMSI) based and non-IMSI based identities for authentication of services.
- NG-Core has support for capabilities such as network slicing, which allows the partition of network resources across different customers, services or use-cases.

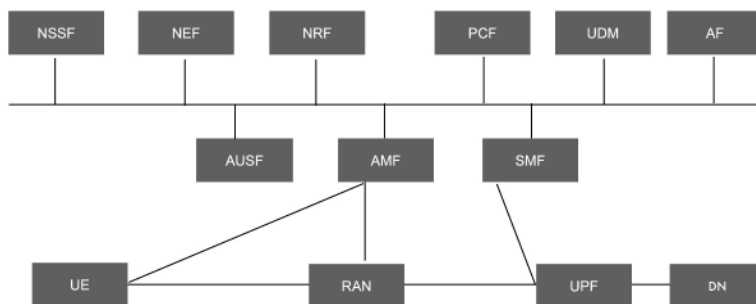


FIGURE 1.4: 5G DETAILED SYSTEM ARCHITECTURE

Network Functions in NG-Core

5G NG-Core architecture comprises of the following network functions:

1. Authentication Server Function (AUSF)
2. Access and Mobility Management Function (AMF)
3. Data Network (DN)
4. Network Exposure Function (NEF)
5. Network Repository Function (NRF)
6. Network Slice Selection Function (NSSF)
7. Policy Control Function (PCF)
8. Session Management Function (SMF)
9. Unified Data Management (UDM)
10. User Plane Function (UPF)
11. Application Function (AF)

Authentication Server Function (AUSF) - AUSF acts as an authentication server, performing UE authentication using Extensible Authentication Protocol (EAP).

- EAP is a popular protocol used in Wi-Fi networks for authenticating Wi-Fi clients. In the 4G network, AUSF function was part of the Home Subscriber Server (HSS) function.

Access and Mobility Management Function (AMF) - Responsible for connection management, registration management and mobility management (handling of reachability and idle/active mode mobility state).

- It also takes care of access authentication and authorization. AMF also supports Lawful intercept function for AMF events. In the 4G network, this function was part of the Mobility Management Entity (MME).

Data Network (DN) - DN offers operator services, internet access and third party services.

Network Exposure Function (NEF) - NEF is a proxy or API aggregation point for the core network and provides security when services or external application functions access the 5G Core nodes. This is a new function introduced in 5G architecture.

Network Repository Function (NRF) - NRF supports service discovery, and maintains/provides profiles network function instances. This is a new function introduced in 5G architecture.

Network Slice Selection Function (NSSF) - NSSF supports the selection of network slice instances to serve the User Equipment (UE), based on the Network Slice Selection Assignment Information (NSSAIs) configured or allowed for a given UE. This is a new function introduced in 5G architecture.

Policy Control Function (PCF) - PCF provides a unified policy framework and shares policy rules to control plane functions, to enforce them.

- It also accesses subscription information relevant for policy decisions from the Unified Data Repository (UDR). PCF was part of the PCRF function in the 4G network.

Session Management Function (SMF) - SMF provides session management, UE IP address allocation & management and DHCP functions.

- It also provides traffic steering configuration for User Plane Function (UPF) for proper traffic routing. SMF function was split between the MME and Packet Gateway (PGW) function in 4G network.

Unified Data Management (UDM) - UDM provides Authentication and Key Agreement (AKA) credentials, user identification handling, access authorization and subscription management functions. UDM was part of the HSS functionality in the 4G architecture.

User Plane Function (UPF) - UPF provides packet routing and forwarding functions. In addition, it also handles QoS services. UPF function was split between Serving Gateway (SGW) and PGW in the 4G architecture.

- Separating the user plane from the control plane in both SGW/PGW, enables the service providers to deploy the UPF closer to the network edge.
- In 5G, UPF function can be deployed at the network edge, in addition to the network core, to improve the network performance and to reduce latency.

Application Function (AF) - AF function is similar to the AF function in the 4G network.

- It interacts with the 5G core to provide services such as application influence on traffic routing, accessing Network Exposure Function (NEF) and interacting with policy framework for policy control.

11. Explain visualized Evolved Packet Core (vEPC)

- Control and User Plane Separation in 5G Core CUPS stands for Control and User Plane Separation.
- It was introduced by 3GPP, for Evolved Packet Core (EPC) as part of their Release 14 specifications.

Need for CUPS

- Service providers across the globe are seeing a jump in the mobile data growth, year-after-year, due to the growth in the consumption of video, online gaming and social media services.

- 5G is not only facing the challenge of supporting higher data speeds, but also has to reduce the network latency for customers.
- Network latency has a direct impact on the customer experience and almost a non-negotiable thing for the new 5G use-cases.
- The architects of 5G were looking at multiple ways of bringing down the network latency for customers, to meet the requirements of emerging 5G use cases such as Smart Cars, AR/VR and Holograms.
- 5G architecture tries to reduce the network latency through multiple mechanisms such as Network Slicing, Massive MIMO, Small Cells and Multi-access Edge Computing (MEC).
- MEC infrastructure, being closer to the user, plays a critical role in bringing down the network latency by providing a compute infrastructure for Over-The-Top (OTT) and Internet of Things (IOT) services. CUPS is another technique in 5G, that helps to bring down the network latency.
- The multiple deployment options supported by CUPS, provide great flexibility to the service providers, to deploy user-plane functions in one or more locations to meet the bandwidth and latency requirements of customer services.
- For example, a service provider may have to deploy more instances of the user plane function near a college dorm, where several 100s of students are watching video and playing online games.
- However, in a stadium, there will be several 1000s of mobile users who would be checking their emails, browsing Internet and uploading pictures. In such locations, the control plane has to scale to support several 1000s of customer sessions.
- So, the service provider may have to deploy more control plane functions in such geographies to support the 1000s of mobile users.

CUPS in 4G architecture

- CUPS was originally introduced in the 4G Evolved Packet Core (EPC) architecture. EPC with CUPS support separates the control plane function from the user plane function in the network.
 - Network functions within 4G EPC such as Packet Gateway (PGW), Serving Gateway (SGW) and Traffic Detection Function (TDF), were split into control plane and user plane functions.
 - EPC with CUPS support had PGW-U/ PGW-C, SGW-U/SGW-C and TDF-U/TDF-C.
- When EPC supports CUPS, service providers would have the option of
1. Deploying the control plane functions co-located with the user plane functions (i.e., in the same data center)
 2. Deploying the control plane functions and user-plane functions in a distributed fashion, across multiple locations
 3. Deploying the control plane function in a centralized location and deploy the user-plane functions in multiple locations
- 5G adopts CUPS based architecture for the 5G Core. 5G Core has a distinct User Plane Function (UPF) that handles all of the user-plane functions performed by SGW-U and PGW-U in 4G EPC.
 - 5G's control plane functions are distributed across different network functions such as Authentication Server Function (AUSF), User Data Management (UDM), Policy and Charging Function (PCF) and Session Management Function (SMF).
 - This gives a lot of flexibility for the service providers to decide the network functions which have to be deployed at the edge of the network versus the core of the network.
 - Since 5G supports cloud-native network services, it becomes easy for the vendors and service providers to implement CUPS in the 5G network architecture (when compared to the 4G network).

PART A- Two marks

1. What are the characteristics of first generation 1G?

1. 2.4 kbps speed enables voice calls in one nation.
2. Use an analogue signal.
3. Poor voice quality, insufficient battery life.
4. Large phone size limited capacity poor handoff dependability inadequate security
5. Spectrum efficiency was relatively low.

2. What are the characteristics of second generation 2G?

1. 2G stands for the Second Generation.
2. Data transmission rates of up to 64kbps were achieved through digital signals.
3. Enables services such as text messaging, picture messages, and MMS Multimedia message.
4. Incapable of dealing with complicated data such as videos.
5. Strong digital signals were required to enable mobile phones to function, digital signals will be weak if no network coverage in a particular area.

3. What are the characteristics of third generation 3G?

1. Speed 2 Mbps typically referred to as smartphones.
2. Broadband and data transfer rates have been increased to handle web-based applications and music and video files.
3. It allows for speedier communication.
4. Sending/receiving huge email messages.
5. High-speed web access, improved security, video conferencing, 3D gaming.

4. What are the characteristics of fourth generation 4G?

1. Ability to give speeds ranging from 10Mbps to 1Gbps, high quality streaming video, a combination of Wi-Fi and Wi-Max and increased security.
2. Provide any type of service at any time and from any location as needed by the user.
3. Expansion of multimedia services.
4. Low cost per bit, increased battery usage.
5. Implementation is complex, intricate technology is required, and expensive equipment is necessary to build a next-generation network.

5. List the features 5G.

The present 5G technology trend has the following feature.

1. 5G technology can provide connectivity speeds of up to 25 Mbps.
2. 5G technology provides cell phone users with high quality and bi-directional extensive bandwidth sharing.
3. 5G technology enables large-scale data transmission at gigabit speeds, with about 65,000 connections supported.
4. The uploading and downloading speeds of 5G technology have reached a new high.
5. Virtual private networks are also supported by 5G technology.

6. Define User Equipment (UE)

- In 2G, the handsets were called as mobile phones or cell phones, as they were predominantly used for making voice calls.
- However, in 3G, the handsets can support both voice and data services. Hence, the term User Equipment or UE is used to represent the end user device, which could be a mobile phone or a data terminal.

7. Define Radio Access Network (RAN)

- The RAN, also known as the UMTS Radio Access Network, UTRAN, is the equivalent of the previous Base Station Subsystem (BSS) in GSM.
- RAN includes the NodeB function and the Radio Network Controller (RNC) function. The NodeB function provides the air interface. The RNC manages the air interface for the overall network.

8. Define Core Network.

The core network is the equivalent of Network Switching Subsystem or NSS in GSM and provides all the central processing and management for the system. Core network has both circuit switched and packet switched network elements.

9. Define Home Location register (HLR)

- HLR is a database that contains all information about the subscriber including their last known location.
- The HLR maintains a mapping between Mobile Station International Subscriber Directory Number (MSISDN) and International Mobile Subscriber Identity (IMSI). MSISDN is the mobile phone number used for making and receiving voice calls and SMS.

10. Define Equipment Identity Register (EIR)

- The EIR is the function that decides whether a user equipment is allowed onto the network or not. EIR is typically integrated with the HLR.
- EIR is used for blocking or monitoring calls from a stolen user equipment. Each user equipment is uniquely identified through a number known as the International Mobile Equipment Identity (IMEI).

11. Define Authentication Centre (AuC)

- AuC is used for storing a shared secret key, which gets generated and burned in the SIM card at the time of manufacturing.
- AuC function is typically co-located with the HLR function. AuC does not exchange the shared secret key, but would run an algorithm on the International Mobile Subscriber Identity (IMSI), to generate data for authentication of a subscriber / user equipment. Each IMSI is unique and gets mapped to a SIM card.

12. Define Mobile switching center (MSC)

- MSC is responsible for functions such as routing calls and SMS messages. It interfaces with the HLR for keeping track of subscriber location and does call handovers, when the mobile subscriber moves from one location to another.
- Gateway MSC (GMSC) is a function that is present either within or outside of the MSC. A GMSC interfaces with the external networks such as the Public Switched Telephone Network (PSTN), which is our legacy land line network.

13. Define Serving GPRS Support Node (SGSN)

- SGSN is responsible for mobility management and authentication of subscribers / mobile devices in a GPRS network.
- It performs a role which is similar to the role played by the MSC for voice calls. The SGSN and MSC are often co-located in the network.

14. Define Gateway GPRS Support Node (GGSN)

GGSN acts as a gateway to the Internet. It connects the GPRS network with the packet switched data network. GGSN receives data addressed to a given subscriber, checks if the subscriber is active and then forwards the data to the SGSN serving the particular subscriber.

15. Define Charging Gateway Function (CGF)

- CGF handles Call Detail Records (CDRs) generated by the GGSN in a GPRS network.
- There are different types of CDRs processed by the CGF, based on the network node that generates the CDR. For example, when a SGSN generates CDRs, it is called S-CDR.

16. Evolved Node B (eNB)

- eNodeB is the entity that supports air interface and performs radio resource management.
- It provides radio resource management functions such as IP header compression, user data encryption, and routing the user data to the Serving Gateway (SGW).

17. Define Serving Gateway (SGW)

- SGW serves as the mobility anchor for the user plane. It takes care of inter-eNodeB handovers & User Equipment (UE) mobility between 3GPP networks.
- It is responsible for routing/forwarding data packets between the eNodeB & Packet Data Network Gateway (PDN GW).

18. Define Packet Data Network Gateway (PGW)

- PDN GW provides the UE with connectivity to the external packet data networks such as Internet.
- It serves as the anchor point for intra-3GPP network mobility, as well as mobility between 3GPP and non-3GPP networks.
- It takes care of Policy and Charging Enforcement Function (PCEF), which includes Quality of Service (QoS), online/offline flow-based charging data generation, deep-packet inspection, and lawful intercept.

19. Define Mobility Management Entity (MME)

MME manages mobility, UE identities and security parameters. It operates in the Control plane and provides functions such as managing session states, authentication, mobility with 3GPP 2G/3G nodes, and roaming.

20. Define Policy and Charging Rules Function (PCRF)

Policy and Charging Rules Function (PCRF) maintains the policy and charging related controls for all the subscribers. For example, a subscriber's quality of service policy is stored in the PCRF server.

21. What is Circuit Switched Fall-Back (CSFB)?

- When LTE is just used for data transfer, voice calls are handled through the legacy circuit switched mechanisms - by falling back to a 3G or 2G network.
- Circuit Switched Fall- Back (CSFB) works only when the area covered by an LTE network is also covered by the 3G network.

22. Define Voice over LTE (VoLTE)

- Voice over LTE is relatively a new concept, to support voice calls over the 4G network.
- While CSFB helped the service providers during the migration from 2G/3G to 4G networks, VoLTE runs completely on the 4G network.
- In case of VoLTE, the user equipment / mobile should be capable of initiating a VoLTE call and the network should support VoLTE.

23. How do setting up a VOLTE?

Setting up a VoLTE is a two-step process:

1. At first, the UE sets up a dedicated bearer/tunnel for IMS APN (Access Point Name).
2. Once the bearer is established, the UE sets up a Session Initiation Protocol (SIP) connection with the IMS Core. SIP is a popular protocol used for Voice over IP (VOIP) communications, over the Internet.

24. What is Radio Access Network (RAN)?

The Radio Access Network (RAN) architecture has evolved across the different generations of the wireless network, to support the bandwidth and scalability requirements. RAN has two distinct units – the Remote Radio Head (RRH) and the Baseband Unit (BBU).

25. Draw the table for evolution of RAN.

Generation	Architecture / Technology	Base Station
2G	GSM	Base Transceiver Station (BTS)
3G	UMTS	NodeB
4G	LTE	Evolved NodeB (eNodeB)
5G	NR	Next Generation NodeB (gNodeB)

26. State the need for 5G. (Nov/Dec 23)

- Internet of Things (IOT) will require an infrastructure that can handle several billions of network devices connecting to the wireless network, and at the same time energy efficient.
- 3D video and Ultra High Definition Video streaming applications are hungry for additional bandwidth.
- Virtual Reality and Augmented Reality enabled gaming, video streaming and industrial applications require sub- millisecond latencies.

27. Define Next Generation Core (NG-core)

- NG-Core for 5G is the equivalent of Evolved Packet Core (EPC) in a 4G network.
- 5G NG-Core architecture supports virtualization and allows the user plane functions to be deployed separately, from the control plane functions.
- In addition, the user plane and control plane functions can be scaled independently.

28. State the functions of 5G NG –Core.

1. Authentication Server Function (AUSF)
2. Access and Mobility Management Function (AMF)
3. Data Network (DN)
4. Network Exposure Function (NEF)
5. Network Repository Function (NRF)

29. Define Authentication Server Function (AUSF)

AUSF acts as an authentication server, performing UE authentication using Extensible Authentication Protocol (EAP). EAP is a popular protocol used in Wi-Fi networks for authenticating Wi-Fi clients. In the 4G network, AUSF function was part of the Home Subscriber Server (HSS) function.

30. Define Access and Mobility Management Function (AMF)

- Responsible for connection management, registration management and mobility management (handling of reachability and idle/active mode mobility state).
- It also takes care of access authentication and authorization. AMF also supports Lawful intercept function for AMF events. In the 4G network, this function was part of the Mobility Management Entity (MME).

31. Define Data Network (DN)

DN offers operator services, internet access and third party services.

32. Define Network Exposure Function (NEF)

NEF is a proxy or API aggregation point for the core network and provides security when services or external application functions access the 5G Core nodes. This is a new function introduced in 5G architecture.

33. Define Network Repository Function (NRF)

NRF supports service discovery, and maintains/provides profiles network function instances. This is a new function introduced in 5G architecture.

34. Define Network Slice Selection Function (NSSF)

NSSF supports the selection of network slice instances to serve the User Equipment (UE), based on the Network Slice Selection Assignment Information (NSSAIs) configured or allowed for a given UE. This is a new function introduced in 5G architecture.

35. What is Policy Control Function (PCF)?

PCF provides a unified policy framework and shares policy rules to control plane functions, to enforce them. It also accesses subscription information relevant for policy decisions from the Unified Data Repository (UDR). PCF was part of the PCRF function in the 4G network.

36. What is Session Management Function (SMF)?

SMF provides session management, UE IP address allocation & management and DHCP functions. It also provides traffic steering configuration for User Plane Function (UPF) for proper traffic routing. SMF function was split between the MME and Packet Gateway (PGW) function in 4G network.

37. What is Unified Data Management (UDM)?

UDM provides Authentication and Key Agreement (AKA) credentials, user identification handling, access authorization and subscription management functions. UDM was part of the HSS functionality in the 4G architecture.

38. What is User Plane Function (UPF)?

- UPF provides packet routing and forwarding functions. In addition, it also handles QoS services. UPF function was split between Serving Gateway (SGW) and PGW in the 4G architecture.
- Separating the user plane from the control plane in both SGW/PGW, enables the service providers to deploy the UPF closer to the network edge.

39. What is Application Function (AF)?

AF function is similar to the AF function in the 4G network. It interacts with the 5G core to provide services such as application influence on traffic routing, accessing Network Exposure Function (NEF) and interacting with policy framework for policy control.

40. What is CUPS in 4G architecture?

- CUPS was originally introduced in the 4G Evolved Packet Core (EPC) architecture. EPC with CUPS support separates the control plane function from the user plane function in the network.
- Network functions within 4G EPC such as Packet Gateway (PGW), Serving Gateway (SGW) and Traffic Detection Function (TDF), were split into control plane and user plane functions.

4G/5G COMMUNICATION NETWORKS

UNIT – II 5G CONCEPTS AND CHALLENGES

Fundamentals of 5G technologies, overview of 5G core network architecture, new radio and cloud technologies, Radio Access Technologies (RATs), EPC for 5G.

1. Discuss the fundamentals of 5G technologies.

Drivers for 5G

1.1 Introduction

- The requirements on mobile and other types of communications networks have been growing significantly over the past decade.
- From humble beginnings just providing phone calls and text messages, these networks are now expected to form the underlying infrastructure for a truly digital economy—enabling new means of operation as the world transitions from 20th century operating models into ones that are designed for the challenges of the 21st.
- The drivers for 5G are far more, therefore, than merely the drive for a new core network but rather the result of intersecting requirements and demands— namely
 - (1) Business case demands from a broader set of economic actors, including industrial companies driving new use cases,
 - (2) New technologies for delivering core network components creating expectations of more efficient and flexible operations, and
 - (3) Shifts in how business, society and environmental needs are balanced to deliver services in a new way.

2.3 New technologies

Many new technologies have driven the development of 5G, in this section we very briefly discuss the main ones:

- (1) Virtualization,
- (2) Cloud native,
- (3) Containers,
- (4) Microservices, and
- (5) Automation

What is NVF? Explain its need and architecture (Nov/Dec 23)

2.3.1 Virtualization

- Traditionally Mobile core network element functional designs are distributed applications which scale horizontally and run on dedicated hardware such as processor blades in a chassis.
- The network element architecture is distributed internally onto specific types of blades that perform specific tasks.

- The first major step of virtualization was to migrate those application-specific blades to virtualized resources such as virtual machines (VMs) and later containers.
- ETSI NFV (Network Function Virtualisation) and OPNFV was created to facilitate and drive virtualization of the telecoms networks by harmonizing the approach across operators.
- The network element could then be realized as an application that is distributed among several virtual hosts.
- Because the application was no longer constrained by the resources and capacity of a physical chassis, this step allows much greater flexibility of deployment and for harmonization of the installed hardware.
- For example, the operator can deploy much larger (or even much smaller) instances of the network element. This first step was also mainly for proving that a virtualized host environment could scale appropriately to meet the subscriber and capacity demands of today's mobile core.
- However, most applications in this phase are like a 2-Tier application design wherein the second (Logic) tier the application itself was tightly coupled to state storage it required. The storage design to maintain state was ported from physical systems where individual blades had their own memories.

2.3.2 Cloud native

- Cloud Native architectures have gained a lot of interest over the past years and service operators attempt to emulate the efficiencies captured by so-called hyperscalers (e.g., Facebook, Google, Amazon) has led to a much heightened interest in this area.
- Simply put, the architectures and technologies (service-based interfaces, microservices, containers, etc.) used in web-scale applications bring benefits to networking infrastructure in elasticity, robustness and deployment flexibility.
- A cloud-native strategy therefore allows service providers to accelerate both the development and deployment of new services by enabling practices such as DevOps, while the ability to rapidly scale up or scale down services allows for resource utilization to be optimized in real-time, in response to traffic spikes and one-time events.
- There are several cloud-native design principles that hold for all installations, including:
 - **Infrastructure Agnostic:** Cloud-native applications are independent and agnostic of any underlying infrastructure and resources.
 - **Software decomposition and life cycle management:** Software is decomposed into smaller, more manageable pieces, utilizing microservice architectures. Each piece can be individually deployed, scaled, and upgraded using a CaaS (Container as a Service) environment.
 - **Resiliency:** In legacy applications, the MTBF (Mean Time between Failures) of hardware has been the base metric for resiliency. In the cloud, we instead rely on distribution and independence of software components that utilize auto-scaling and healing. This means that failures within an application should cause only temporary capacity loss and never escalate to a full restart and loss of service.

- **State-optimized design:** How we manage state depends on the type of state/data and the context of the state. Therefore, there is no “one size fits all” way of handling state and data, but there should be a balance between performance, resiliency, and flexibility.
- **Orchestration and automation:** A huge benefit of cloud-native applications is increased automation through, for example, a Kubernetes-based CaaS layer. A CaaS enables auto-scaling of microservices, auto-healing of failing containers, and software upgrades including canary testing (small-scale testing) before larger deployments.

2.3.3 Containers

- Virtualization has revolutionized IT infrastructure and enabled tech vendors to offer diverse IT-based services to consumers.
- From a simplistic perspective, system-level virtualization allows instances of an Operating System (OS) to run simultaneously on a single-server on top of something called a hypervisor.
- A hypervisor is a piece of computer software that creates and runs virtual machines. System-level virtualization allows multiple instances of OS on a single server on top of a hypervisor.
- Containers on the other hand are isolated from each other and share OS kernels among all containers.
- Containers are widely used in sectors where there is a need to optimize hardware resources to run multiple applications, and to improve flexibility and productivity.
- In addition, the eco systems and tooling for container based environment, e.g., Kubernetes are rapidly expanding.
- Containers are especially useful for telecommunications applications
- Where low-latency, resilience and portability are key requirements—e.g., in Edge computing environments.
- For implementing short-lived services, i.e., for highly agile application deployments.
- In machine learning or artificial intelligence when it is useful to split a problem up into a small set of tasks—it is expected therefore that containers will assist to some extent with automation.

2.3.4 Microservices

- Microservices are an architectural and organizational approach to software development where rather than be developed in a monolithic fashion, software is composed of small independent services that communicate over well-defined APIs.
- It is often considered a variant of the service-oriented architecture approach. The overall aim with microservices architectures is to make applications easier to scale and faster to develop, enabling innovation and accelerating time-to-market for new features. They also, however, come with some increased complexity including management, orchestration and create new data management methods. Microservice disaggregation has several benefits:
- Microservice instances have a much smaller scope of functionality and therefore changes can be developed more quickly.

- An individual feature is expected to apply to a small set of microservices rather than to the entire packet and 5GC function.
- Microservice instances can be added/removed on demand to increase/decrease the scalability of their functions. Microservices can have independent software upgrade cycles.

2.3.5 Automation

- One of the main drivers for the evolution of the core network is the vision to deliver networks that take advantage of automation technologies.
- Across the wider ICT domain, Machine Learning, Artificial Intelligence and Automation are driving greater efficiencies in how systems are built and operated.
- Within the 3GPP domains, automation within Release 15 and Release 16 refer mainly to Self-Organising Networks (SON), which provide Self-Configuration, Self-Optimisation and Self-Healing.
- These three concepts hold the promise of greater reliability for end-users and less downtime for service providers.
- These technologies minimize lifecycle costs of mobile networks through eliminating manual configuration of network elements as well as dynamic optimization and troubleshooting.

2. Explain overview of 5G core network architecture.

Architecture overview

3.1 Introduction

3.1.1 Balancing evolution and disruption

- Work on designing and specifying a Core network for 5G was done in parallel with and in close cooperation with the teams designing the 5G radio network.
- One key principle with the design of the 3GPP 5G Core architecture was not providing backwards compatibility for the previous generations of radio access networks, i.e., GSM, WCDMA and LTE.
- Previously, when new access network generations were developed, each one had a different functional split between the core network and the radio network, as well as new protocols for how to connect the radio and core networks.
- The 5G Core, however, brought a mindset shift aiming to define an “access-independent” interface to be used with any relevant access technology as well as technologies not specified by 3GPP such as fixed access.
- It is also, therefore, intended to be as future-proof as possible. The 5G Core architecture does not include support for interfaces or protocols towards legacy radio access networks (S1 for LTE, Iu-PS for WCDMA and Gb for GSM/GPRS).

- It instead comes with a new set of interfaces defined for the interaction between radio networks and the core network. These interfaces are referred to as N2 and N3 for the signaling and user data parts respectively.

3.1.2 3GPP architecture options

The outcome of the 3GPP work on the 5G network architecture was a number of architecture options, based on 3GPP making three important decisions:

- To specify LTE support for the new 5G architecture
- To specify support for combinations of LTE and NR access
- To specify an alternative 5G architecture based on an evolution of LTE/EPC

The fact that LTE access support is specified for the new 5G architecture means that an LTE access network in practice has two ways of connecting with a core network, potentially simultaneously and selected on a per device basis:

- Using S1 connectivity to an EPC core network
- Using N2/N3 to a 5GC core network
- When defining the 5G radio access network specifications, two variants of combining LTE and the new 5G radio access technology (NR) were discussed.
- Each one relies on the assumption that one of the technologies will have a larger geographical coverage and therefore be used for all signalling between devices and the network, while the other radio technology would be used to boost user traffic capacity inside geographical areas where both access technologies are present.

3.1.2.1 The non stand-alone (NSA) architecture

In conjunction with extending the new 5G architecture to not only include NR access but also LTE access, a parallel track was started in the 3GPP Release 15 work.

- This was driven by a widely established view in the telecom industry that there was a need for a more rapid and less disruptive way to launch early 5G services.
- Instead of relying on a new 5G architecture for radio and core networks, therefore, a solution was developed that maximizes the reuse of the 4G architecture.
- In practice it relies on LTE radio access for all signalling between the devices and the network, and on an EPC network enhanced with a few selected features to support 5G.
- The NR radio access is only used for user data transmission, and only when the device is in coverage. See [Fig. 3.1](#).



Fig. 3.1 The non stand-alone architecture.

- One drawback with this architecture is that NR can only be deployed where there is already LTE coverage. This is reflected in the name of the solution—the NR Non-Stand-Alone (NSA) architecture.
- Another drawback is that the available network features are limited to what is supported by LTE/EPC.
- The main differences in terms of capabilities are in the areas of Network slicing, Quality-of-Service handling, Edge computing flexibility and overall core network extensibility/flexibility for integrating towards applications in an IT-like environment.

In summary, there are four ways that LTE and/or NR can be deployed:

- Only LTE for all signalling and data traffic
- Only NR for all signalling and data traffic
- A combination of LTE and NR where LTE has the larger coverage and is used for signalling while both LTE and NR are used for data traffic
- A combination of LTE and NR where NR has the larger coverage and is used for signalling while both LTE and NR are used for data traffic

The common terminology around different variants of deploying radio access technologies are illustrated in Fig. 3.2.

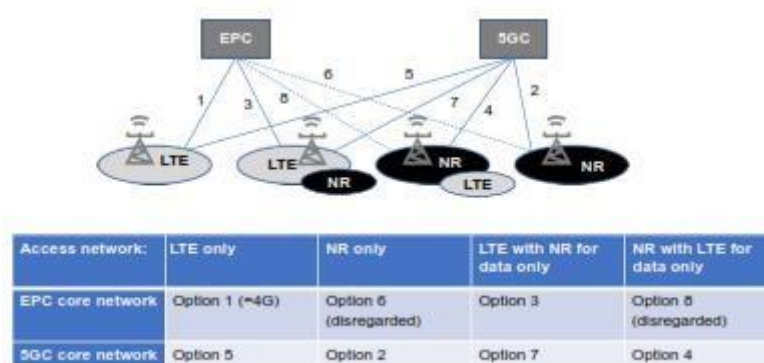


Fig. 3.2 The possible combinations of 5G radio and core networks.

- From a 5G Core network perspective, the four combinations of radio access technologies (options 2, 4, 5 and 7) all use more or less the same interface, protocols and logic.
- This is the first attempt to create an access independent interface between the core network and whatever access technology that is used.
- Option 3 is the popular name for Non Stand-Alone, or NSA, architecture described above.
- It was the first 5G network architecture to enter commercial services as it allows for expanding from the existing 4G LTE/EPC architecture, facilitating a smooth introduction of 5G, even if it is mainly addressing existing mobile broadband services.
- The formal name of the NSA radio network solution is EN-DC, short for “E-UTRAN-NR Dual Connectivity”

3.2 Two perspectives on 5G Core

- When comparing to the existing EPC architecture, the 5GC architecture is simultaneously very similar and very different.

- The user data processing parts, as well as the integration with 3GPP radio access networks, are quite similar between the new 5GC and the traditional EPC network architecture, originally defined for 4G/LTE.
- The part of the network that contains signalling-only functionality, is on the other hand very different.
- Another difference between the EPC architecture used for 4G and 5G NSA is that the architecture of 5G Core can be visualized and described in two different ways.
 1. The first visualization shows the way different network functions are connected. The major difference compared to previous 3GPP architectures in this visualization is the concept of Service-Based interfaces.
 2. It means that the network functions that include logic and functionality for processing of signalling flows are not interconnected through point-to-point interfaces but instead exposing and making available services to the other network functions.
 3. For each interaction between network functions, one of these acts as a “Service Consumer”, and the other as a “Service Producer”.

This representation of the architecture is shown in Fig. 3.3.

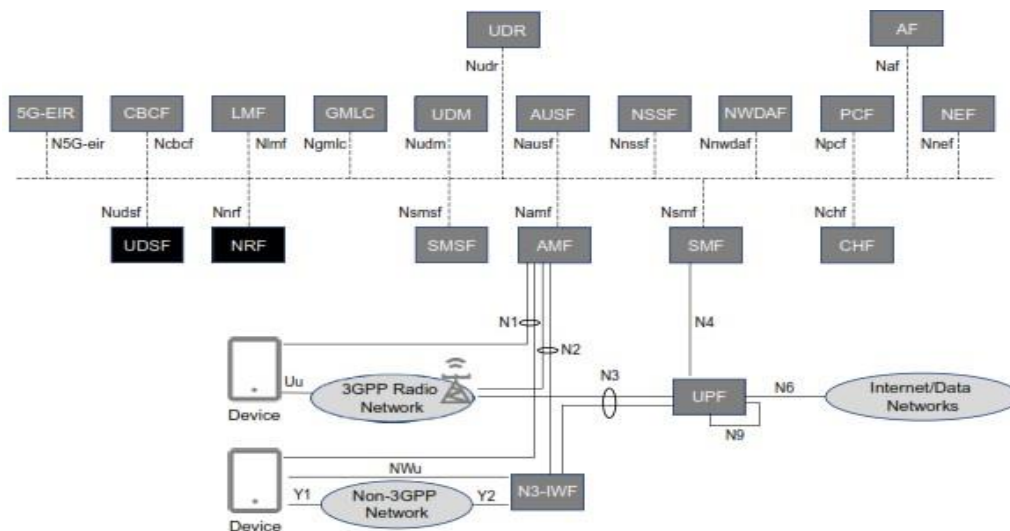


Fig. 3.3 5G Core architecture visualized with Service-Based interfaces.

- Firstly, however, let’s look at the other visualization of the architecture that illustrates how network functions interact with other network functions, represented by traditional point-to-point interfaces.
- Showing these interfaces can be useful to illustrate which of the network functions that utilize, or consume, the services of which other network functions.
- Even if all the network functions in theory could be connected in a full connectivity mesh, the actual call flows define which service combinations that apply in real operations.

- And these combinations are visualized as logical interfaces, or more correctly—reference points, in the view shown in Fig. 3.4.

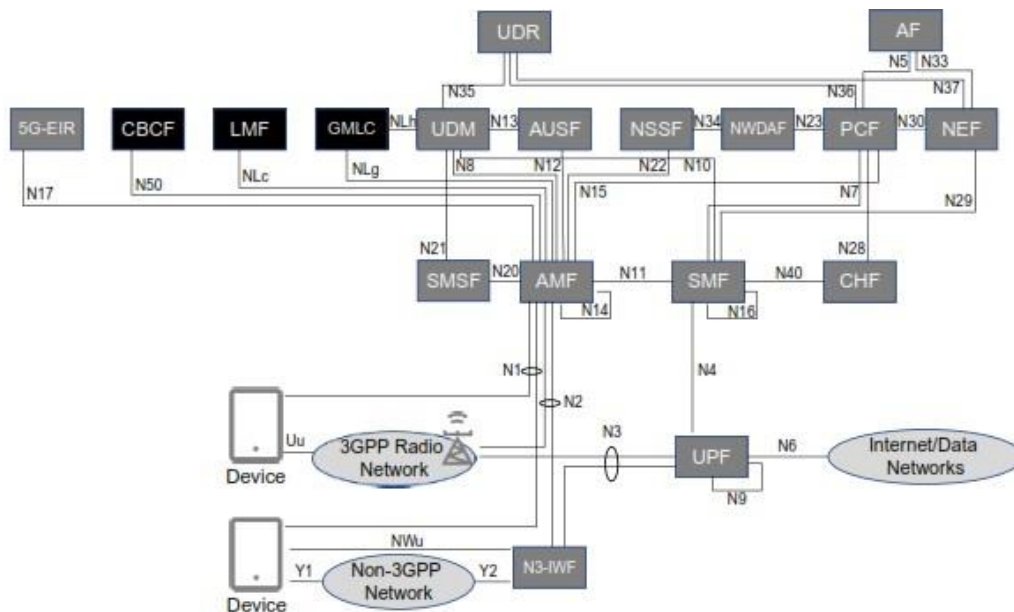


Fig. 3.4 The 5G Core architecture visualized with point-to-point interfaces.

3.3 Service-based architecture (SBA)

3.3.1 The concept of services

A major difference in 5G Core compared to previous generations of traditional network architectures represented by “nodes” or “network elements” connected by interfaces, is the usage of service-based interactions between Network Functions.

- This means that each Network Function offers one or more services to other Network Functions in the network.
- In the 5GC architecture, these services are made available over Network Function interfaces connected to the common Service-Based Architecture (SBA).
- In practice this means that functionality supported in a specific Network Function is made available and accessible over an API (Application Programming Interface).
- It shall be noted that this architecture applies to signalling functionality only, not to the transfer of user data.

3.3.2 HTTP REST interfaces

- The communication method defined for 5G Core relies on the widely used “HTTP REST paradigm” that are a set of rules or guidelines that define how web communication technologies access services from distributed applications using APIs. “REST” is short for “Representational State Transfer”
- And defines a set of design rules for how to implement the communication between different software modules in a networked architecture.
- This is the standard way of designing IT networking applications today, and it has been selected by 3GPP as a means of allowing for tighter integration between the mobile networks and surrounding IT systems, as well as for allowing for shorter and simplified service development efforts.

3.3.3 Service registration and discovery

- When two Network Functions communicate over the 3GPP SBA architecture, they take on two different roles.
- The Network Function that sends the request has the role of a Service Consumer, while the Network Function that offers a service and triggers some action based on the request has the role of a Service Producer.
- Upon completion of the requested action, the Service Producer responds back to the Service Consumer.

3.4 The core of the core

Having described the new mechanisms for how different Network Functions communicate, let’s now return to the functional view of the network.

- The core functionality of the network architecture includes functionality for establishing sessions in a secure way and to forward user data to and from mobile devices providing data connectivity.
- This is the part of the network that cannot be excluded from any 5G Core deployment.
 - AMF
 - SMF
 - UPF
 - AUSF
 - UDM
 - UDR

Fig. 3.10 illustrates the core components of any 5G network.

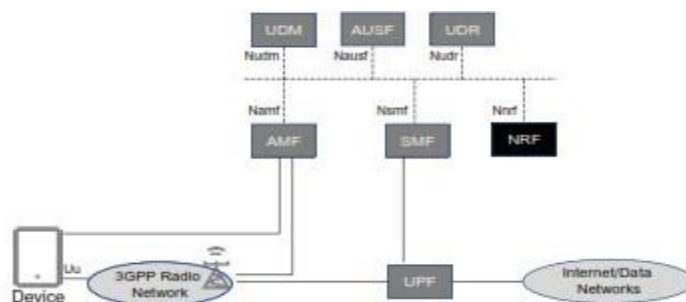


Fig. 3.10 Mandatory components of a 5G network architecture.

- The AMF is the “Access and Mobility Management Function”. It interacts with the radio network and the devices through signalling over the N2 and N1 interfaces respectively.
- Connections towards all other Network Functions are managed via service-based interfaces. The AMF is involved in most of the signalling call flows in a 5G network.
- It supports encrypted signalling connections towards devices, allowing these to register, be authenticated, and move between different radio cells in the network. The AMF also supports reaching and activating devices that are in idle mode.
- One difference to the EPC architecture is that the AMF (as opposed to the MME) does not handle session management.
- Instead the AMF forwards all session management-related signalling messages between the devices and the SMF Network Function. Another difference is that the AMF (as opposed to the MME) does not perform device authentication itself, instead the AMF orders this as a service from the AUSF Network Function.
- The SMF is the “Session Management Function”, meaning as the name suggests that the SMF manages the end user (or actually device) sessions.
- This includes establishment, modification and release of individual sessions, and allocation of IP addresses per session. The SMF indirectly communicates with the end user devices through that the AMF forwards session-related messages between the devices and the SMFs.
- The SMF interacts with other Network Functions through producing and consuming services over its service-based interface, but also selects and controls the different UPF Network Functions in the network over the N4 network interface.
- This control includes configuration of the traffic steering and traffic enforcement in the UPF for individual sessions.
- In addition to this, the SMF has a key role for all charging-related functionality in the network.
- It collects its own charging data, and controls the charging functionality in the UPF.
- The SMF supports both offline and online charging functionality. Furthermore, the SMF interacts with the PCF Network Function for Policy Control of user sessions.
- The “User Plane Function” (UPF) has as the main task to process and forward user data. The functionality of the UPF is controlled from the SMF.
- It connects with external IP networks and acts as a stable IP anchor point for the devices towards external networks, hiding the mobility.
- This means that IP packets with a destination address belonging to a specific device is always routable from the Internet to the specific UPF that is serving this device even as the device is moving around in the network.
- The UPF performs various types of processing of the forwarded data. It generates traffic usage reports to the SMF, which the SMF then includes in charging reports to other Network Functions.
- The UPF can also apply “packet inspection”, analyzing the content of the user data packets for usage either as input to policy decisions, or as basis for the traffic usage reporting.

- It also executes on various network or user policies, for example enforcing gating, redirection of traffic, or applying different data rate limitations.
- When a device is in idle state and not immediately reachable from the network, any traffic sent towards this device is buffered by the UPF which triggers a page from the network to force the device back to go back to connected state and receive its data.
- The UPF can also apply Quality-of-Service (QoS) marking of packets towards the radio network or towards external networks. This can be used by the transport network to handle each packet with the right priority in case of congestion in the network.
- The UDM is the “Unified Data Management Function”. It acts as a front-end for the user subscription data stored in the UDR (more on that further down) and executes several functions on request from the AMF.
- The UDM generates the authentication data used to authenticate attaching devices. It also authorizes access for specific users based on subscription data. This could for example mean applying different access rules for roaming subscribers and home subscribers.
- In case there are more than one instance of AMF and SMF in the network, the UDM keeps track of which instance that is serving a specific device.
- The UDR—the “Unified Data Repository”—is the database where various types of data is stored. Important data is of course the subscription data and data defining various types of network or user policies.
- Usage of UDR to store and access data is offered as services to other network functions, specifically UDM, PCF and NEF.
- The functionality of the “Authentication Server Function” (AUSF) is quite limited but very important. It provides the service of authenticating a specific device, in that process utilizing the authentication credentials created by the UDM.
- In addition, the AUSF provides services for generating cryptographical material to allow for secure updates of roaming information and other parameters in the device.

3. Explain new Radio and Cloud Technologies (Nov/Dec 23)

One of the significant changes to the 5G architecture is the radio specification. 5G introduces a new radio specification called 5G New Radio (5G NR).

Some of the key changes to the radio function in 5G are:

- ❖ **Spectrum:** 5G supports a wide range of spectrums from low bands below 1 GHz, to mid bands from 1 GHz to 6 GHz, to high bands 24 / 30 GHz to 300 GHz. This high band is called as millimetre wave.
- ❖ **Latency:** 5G NR supports lower latencies less than 10 milliseconds
- ❖ **Beamforming:** 5G NR supports a large number of multiple input multiple output (MIMO) antennas, which would allow it to work in a high interference environment through a technique called “beamforming”. It enables 5G radios to provide both coverage and capacity.
- ❖ **Interworking with 4G:** Coexistence with LTE (by supporting LTE NR), by bringing an overlay network, in cases where 5G coverage is not available.

5G Spectrum

- In order to achieve high speed and scalability, 5G supports a wide range of spectrums from low bands below 1 GHz, to mid bands from 1 GHz to 6 GHz, to high bands 24 / 30 GHz to 300 GHz (also, known as millimeter-wave or mmWave).
- Lower frequencies typically travel long enough and can propagate deeper than higher frequencies.
- Lower frequencies offer a better coverage. Higher frequencies will travel short distance, however can offer higher bandwidth.
- 5G has to offer both better coverage and higher speeds and hence, millimeter wave is very important for 5G to offer higher speeds. In addition to deploying large towers, 5G will also require service providers to deploy a large number of small cell towers to meet the speed and latency requirements.
- 5G promises ubiquitous wireless coverage by supporting both licensed and unlicensed spectrums.
- 5G NR-U is called Unlicensed Spectrum. 3GPP has categorized the use of 5GHz and 6GHz spectrum bands as 5G NR-U. The use of unlicensed spectrum significantly reduces the cost of deploying 5G networks.

There are two types of NR-U implementations:

1. Licensed Assisted Access (LAA) NR-U
2. Standalone NR-U

Licensed Assisted Access (LAA):

- In the LAA deployment, service providers aggregate the licensed spectrum with the unlicensed spectrum to provide wireless services to customers.
- The point of aggregation can be a small cell which supports the licensed and unlicensed frequencies.
- Alternatively, the mobile device can establish dual connectivity, one with the macrocell operating in the licensed spectrum and the other with the small-cell which is operating in the unlicensed spectrum.
- LAA in 5G is very similar to how LAA is implemented in 4G LTE network. LAA can help service providers in boosting the capacity of the network.

Standalone NR-U: 5G NR-U standalone implementation is unique and it may appear to be competing with existing WiFi technology (which is using the 5GHz frequency).

- 5G NR-U standalone implementations will help to accelerate the deployment of 5G in private networks such as enterprises and industries. This will eliminate the overhead for enterprises to maintain a separate WiFi network for wireless connectivity.
- 5G NR-U standalone deployments can support IOT use cases in enterprises, industries and communities.
- 5G NR supports Dynamic Spectrum Sharing (DSS), a new standard introduced in 5G, that helps service providers to migrate from 4G LTE to 5G NR, without a noticeable service quality degradation to customers.

- Dynamic Spectrum Sharing allows the service providers to share spectrum between their 4G and 5G users.
- For example, a 4G user can use a given spectrum for 10 milliseconds and the same spectrum can then be taken by a 5G user for 1 millisecond.
- Dynamic allocation of spectrum, improves the utilisation of the spectrum across 4G and 5G services.
- Traditionally, when service providers migrated from 3G to 4G, they carved out a chunk of the 3G spectrum and dedicated that to 4G.
- This may starve the users in the 3G network. However, Dynamic Spectrum Sharing enables a smoother migration from 4G to 5G.
- Initially, it is expected that more users will be in 4G and very few users will be in 5G.
- Over the next few years, the ratio between 4G and 5G users will change.
- More and more users are expected to migrate to 5G.
- Based on the capacity needs, the network can dynamically allocate spectrum to users. This would result in speedy services to both 4G and 5G users.

Technologies accelerating 5G Radio

MIMO

MIMO stands for Multiple Input and Multiple Output, where several small antennas are grouped together at the transmitter and at the receiver.

- Before the introduction of MIMO, radio devices in wireless networks used a single antenna at the source and a single antenna at the destination for sending and receiving signals.
- When signals had to penetrate through walls and objects in between, they get scattered and lost before reaching the destination. This would result in errors and also affects the network performance, especially the speed.
- After the introduction of MIMO technology, radio devices were able to send or receive multiple signals at the same time - to minimize the errors and to improve the speed. MIMO uses a technique called multi-path propagation, where the radio signals are transmitted via two or more paths to reach the destination.
- Multi-path propagation approach increases the chances of signals reaching the receiver.
- When signals are sent via multiple paths, the overall strength of the signal received by the radio device is equal to the sum of all the signals received by the different antennas.
- At times, the overall signal strength would be higher when there are no interferences and at other times, the overall signal strength would be low, when there are interferences due to the presence of terrestrial objects in the path.
- Even a small difference in the space between the antennas would dramatically change the path traversed by a signal. So, in order to minimize the interferences, MIMO uses antennas that are spatially distributed.
- MIMO uses the additional paths to its advantage, by strengthening the signals to increase the data throughput.

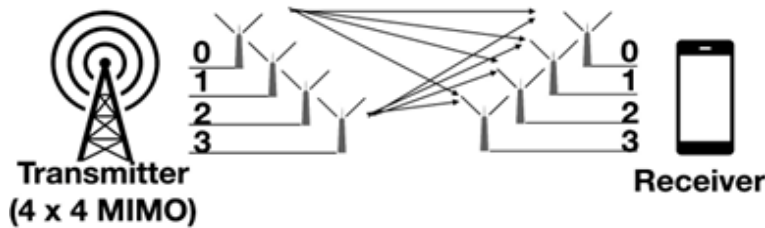


FIGURE 3.8 - HOW MIMO WORKS?

Massive MIMO

- 5G adopts massive MIMO technology where a few tens to several hundreds of antennas are used for propagating radio signals.
- The Massive MIMO enabled radio devices use algorithms to calculate the best possible transmission route through the air, to reach the receivers without any interference or with reduced interference.

Coordinated Multipoint (CoMP)

- Coordinated Multipoint (CoMP) is a technique used to improve radio performance, at cell edges by utilising the capability of more than one base station to enable the communication .
- Using CoMP, radio access nodes dynamically coordinate the transmission and reception of signals from the UE to multiple base stations.
- In the past, inter-cell interference used to be a challenge for wireless service providers. CoMP leverages inter-cell interference to improve the quality of service for the user, as well as improving the utilisation of the network resources.

CoMP has multiple advantages:

- ❖ Helps the service provider to improve the utilization of network resources. For example, data can be passed through a base station which is not loaded.
- ❖ User notices lesser call drops, as several cell sites come together to provide service to the user.
- ❖ User notices higher bandwidth, as multiple base stations serve the user, at a given time. The impact of the interference gets reduced and interference is used constructively.

CoMP falls under two categories:

1. **Joint Processing:** When there is coordination between multiple gNBs to simultaneously transmit or receive signals to or from a UE, it is called Joint processing.
2. **Beamforming:** In beamforming, instead of sending a signal in different directions, the radio device sends the signals concentrated in one direction towards a specific receiver. This reduces the loss of signals, and improves the efficiency.
 - Beamforming is very important when using milli-meter waves, as milli-meter waves get weakened when they get blocked or when they travel over long distances.
 - When radio signals are sent in a concentrated form in one direction, it increases the strength of the signals that get transmitted.

- This approach increases the chances of the receiver receiving the signals, intact and reduces interferences from other devices.

Small Cells

- 5G, because of the reliance on mmWave technology (which has a shorter wavelength) requires several thousands of small cells be deployed around the cities.
- Small Cells are mini base stations that handle the radio signals, like the mobile towers.
- Small Cells require very minimal power and space, unlike the large mobile towers in the neighbourhood.
- Small Cells can be installed on the top of a light pole or a building. Small Cells are not new to the wireless world.
- They were deployed in 4G networks too. However, in 4G networks, Small Cells were predominantly used to improve the indoor wireless coverage.

There are three categories of small cells that exist today:

1. **Femtocells:** Femtocells help the operators in quickly solving the network coverage problem. It takes a very long time for a service provider to install a mobile tower.
 - However, it takes only a few days to power up a femtocell in the network. So, if there is a residential or an enterprise customer that requires extended wireless coverage, the operator can quickly deploy a Femtocell.
 - Femtocell provides multiple benefits such as helping the service provider to offload the network congestion, provide better coverage and increase the data transfer rates.
 - A Femtocell typically caters to 16 users and can provide coverage for a distance between 10 and 50 meters.
2. **Picocells :** Picocells are similar to Femtocells in terms of functionality. However, their coverage area is 100 to 500 meters. Also, they can support 32 to 64 users at a time. Picocells are well suited for small enterprise applications.
3. **Microcells:** Microcells are designed for providing wireless coverage to a large geographic area and suited for providing network coverage to outdoor applications such as smart communities. Microcells provide coverage for a distance between 500 meters and 2.5 kilometers. Microcells can support nearly 200 users at a time.

11.1.1 Information on Cloud technologies

- Cloud computing is a commercial technology that many businesses depend on to access the latest solutions in the IT industry.
- With cloud computing, businesses can access the best IT solutions without necessarily buying any expensive and space-consuming hardware.
- With the help of 5G technology, service providers of cloud computing will be able to provide a more efficient solution for businesses.

The following are the benefits of 5G to cloud-hosted solutions:

- **Faster Transfer of Data** – One of the benefits of deploying the 5G technology is its fast data transfer rate. With the 5G technology, cloud computing helps accelerate the data streaming process.
- **Uninterrupted Storage** – Cloud computing supported by uninterrupted storage will support many businesses as they perform complex operations requiring space-consuming hardware. Rather than choose on premise solutions (which can be expensive), many businesses prefer cloud solutions.
- **Reliability with Large Data** – 5G for cloud computing is important, especially in big data. Many companies deal with large sets of data all the time. They will prefer to transfer all of these data on time. This way, large quantities of data can be transferred easily and in real-time.
- **Enhanced Productivity** – Due to the efficiency and effectiveness of the 5G technology, businesses will be more productive.

Impact of cloud on 5G

5G systems and cloud-hosted solutions are some of the changing technologies in the IT industry. Using both technologies together will lead to a bigger world full of opportunities and innovations.

There are many benefits of using both technologies together, such as –

- **Increased Accessibility** – Using a 5G network with cloud computing will close any bandwidth gap across different regions. It will also enhance the availability of cloud computing solutions in remote locations.
- **IoT Connectivity** – Blending both technologies will lead to a high-speed network with low latency. This way, IoT devices, and systems can easily access cloud solutions. This connectivity will lead to improved business digitization and automated machines.
- **Better Work Connectivity** – When cloud computing is driven by the 5G technology, remote working systems are enhanced. There will be no delays or technical difficulties, giving organizations the ability to work from any location.
- **Improved Security Systems** – Hacking is always a threat to cloud solutions. When powered by 5G, this will upgrade the system's security protocol. With 5G, attacks can be identified easily as hybrid cloud environments are much safer/.
- **Supports Edge Computing** – Driving cloud computing solutions using 5G network systems will enhance edge computing. With edge computing, remote systems are enhanced while a lesser bandwidth is consumed.
- **Easy Access to Customers** – With cloud technology and 5G, it becomes possible for machines to communicate easily with each other. This way, communications will be more reliable as businesses have faster customer access.

(1). RAN Virtualization

- Novel heterogeneous networks can be realized by radio access network (RAN) virtualization and softwarization in the 5G mobile network.
- The RAN virtualization is an effective approach in a fabric of Cloud-RAN structure for example, since the mobile network needs to support flexible capabilities in terms of frequency bands, transmission schemes, antenna configuration, multiplex access attributes.

- Given the technical trends above, future RAN may have a capability of intelligent control on radio configurations, front/back-haul transmissions, and radio resources of a number of small cells which are organized virtually from the central controller.
- A concept of RAN virtualization is given in Fig. 11.1, it includes several slices are configured in association with the service profile to achieve the required quality and reliability by radio network arrangement with RATs, bandwidth, antenna configuration, transmission power, latency, mobility, and so on.

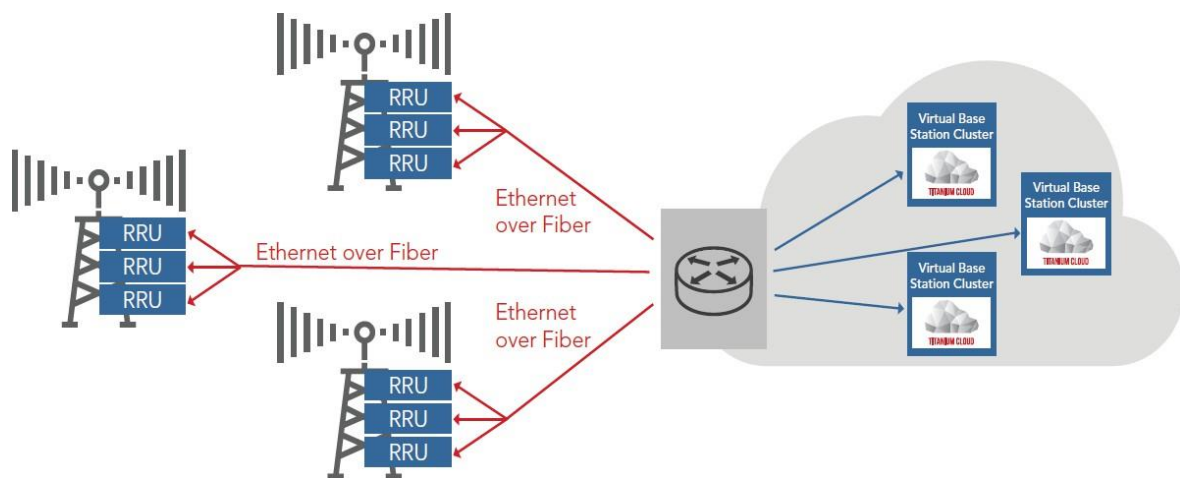


Fig. 11.1 RAN Virtualization

- The technology would be useful when applied to : eMBB, URLLC, mMTC:
- In the future, more diverse services will come out in various usage scenarios in some working environments with wide ranges of data rate, latency, connection density, data size, mobility, reliability, etc. for the associated service profiles.
- Flexible programming on the softwarized virtualization can help handling those service data transmissions under the virtual control function.
- It will be beneficial to the network operators, service providers and end users.
- Expected performance/ features when applied:
- RAN controller integrates overall network control, scheduling, and data transport control throughout the user devices, remote radio unit (RRU), radio access schemes, fronthaul, backhaul, and radio resources such as transport bandwidth, RAT attributes, signaling on BBU.
- Upper controller of cloud network can potentially cover the RAN in the end-to-end network coordination.

- Depending on the requirements for service application, the network slices are flexibly arranged and scaled up or down in a configuration set with appropriate network resources, virtual network functionalities (VNFs) in the virtual network topology in dynamic way.
- The RAN has a capability of orchestrating the VNF chain by arranging and scheduling the virtual machines, storage memories, processing units, and so on. In consequence, all the data processing functions and the transport lines become programmable in software.

4. Explain Radio Access Technologies (RATs).

5G Radio Access Technologies

3.1. Overview of 5G radio access network

- The radio access network (RAN) and aggregated backhauls support the capabilities of data transport, radio transmission and reception.
- In the 5G era, these capabilities shall be enhanced to accommodate massive traffic capacity and device connectivity while providing enhanced quality of user experience.
- As has been mentioned in the previous chapters, ‘5G’ communications system should serve wide range of use cases.
- Depending on each of these use case, range of required capabilities to radio access technologies would be extremely different.
- Consequently ‘5G’ communications system should be an intrinsic and genuine-type heterogeneous network which utilizes every proper radio access technologies according to required capabilities of the use case concerned.
- Intrinsic and genuine-type heterogeneous network would not be a simple ‘overlaid cellular networks’ aiming improved communication capacity anymore but it should be a consolidated communication system consist of functional elements tailored to each of the use cases and serve them in a suitable manner.

➤ Some of these technologies are illustrated in Fig. 5.2-1

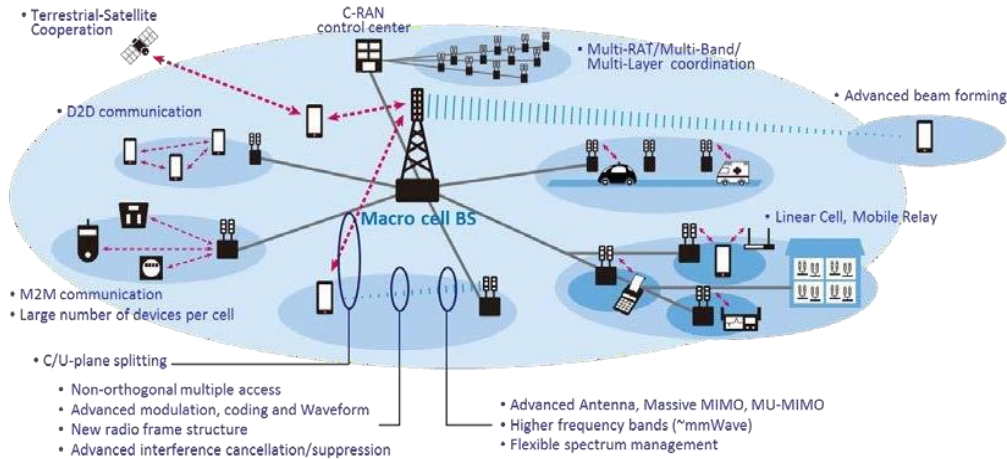


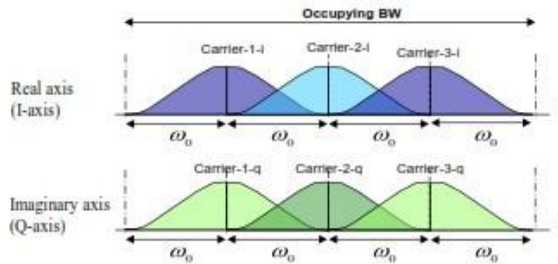
Fig. 5.2-1 Overview of ‘5G’ RAN technologies

5.1.2 Information on modulation or coding scheme

(1). OFDM-SSB-QAM

This method belongs to orthogonal multiple modulation/ demodulation technologies, which is based on the analytic frequency form using the Hilbert transform. While the current OFDM uses DSB (double side band) carriers, this method uses a SSB (single side band) which separates one DSB into two of SSB carriers, so that the spectral efficiency is twice that of LTE/OFDM.

(a) OFDM



(b) OFDM-SSB-QAM

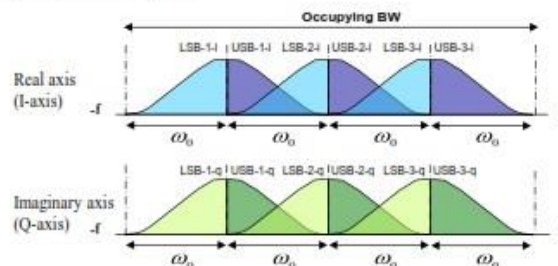


Fig. 5.2.2 Spectral structure of OFDM and OFDM-SSB-

- SSB forming is carried out using a forming data table. Generating multicarrier and integration of demodulation are carried out using FFTs. These elements are very common in LTE/OFDM systems.

5.1.2.1 The technology would be useful when applied to : eMBB

5.1.2.2 Expected performance/features when applied:

- This method not only provides double the spectral efficiency of LTE, but also takes over the whole access function built by LTE.
- Furthermore, this method is signal-processed in the baseband part closely, so that the spatial multiplication technologies; MIMO or NOMA can be adopted easily.

5.1.2.3 Preconditions when applied:

- SSB is said to be weak against frequency fluctuations like the Doppler shift effect, but its tolerance is the same when compared to OFDM. This solution method has the frequency synchronization as OFDM, as well.
- Because both OFDM and this method are multicarrier systems, it is suitable to adopt this transformation into SC-FDMA when using millimeter wave bands.

(2). Time and frequency localized single carrier technology

- Insertion of zeros or a static sequence before DFT operation in DFT-s-OFDM can reduce out of band emission compared with the conventional DFT-s-OFDM.
- Fig. 5.3-3 shows a comparison between DFT-s-OFDM and DFT-s-OFDM with zero or static sequence.
- Fig. 5.3-4 demonstrates out of band suppression performance of DFT-s-OFDM with zero or static sequence. Maintaining the low peak to average power ratio of SC-FDMA, which is the standardized uplink waveform in LTE, the aforementioned technologies can reduce out of band emission compared to DFT-s-OFDM waveforms.
- The inserted zeros or static sequence can be used as a cyclic prefix, providing robustness against frequency selectivity in channels.

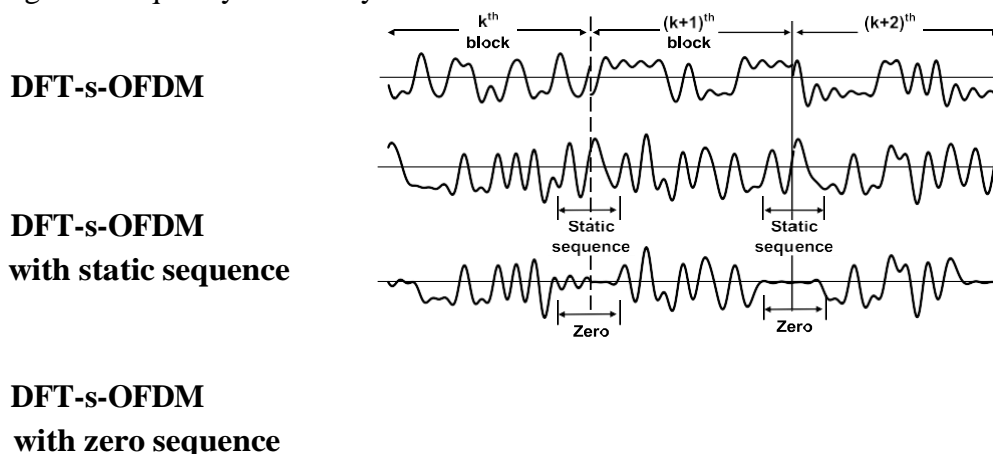


Fig. 5.3-3 DFT-s-OFDM with zero or static sequence insertion

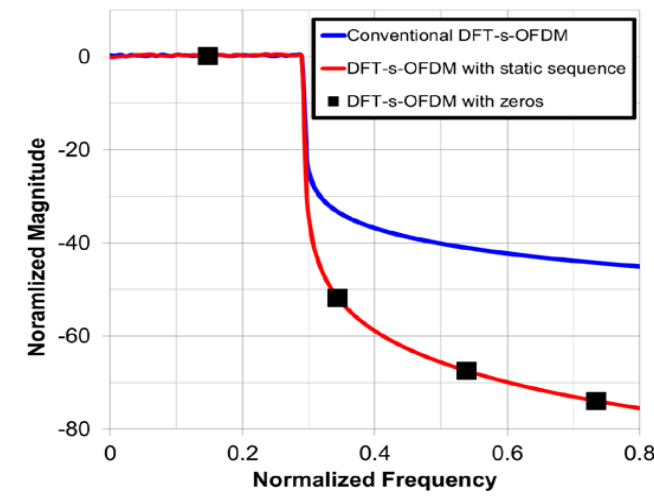


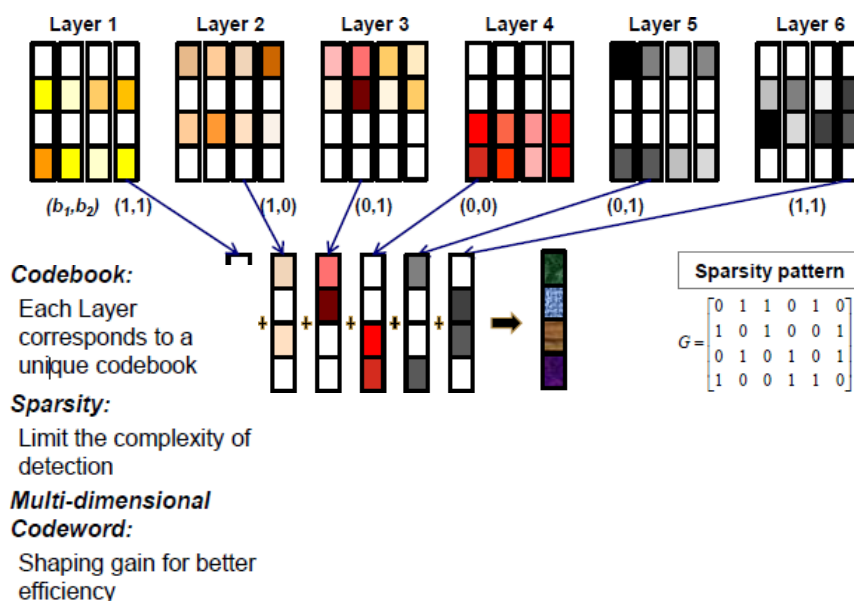
Fig. 5.3-4 Out of band suppression performance

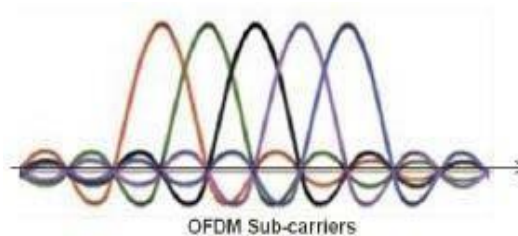
- The technology would be useful when applied to : eMBB, mMTC, URLLC
- Expected performance/features when applied:
- Low PAPR, low out-of-band emission, coverage expansion, saving cost for amplifiers.

5.1.3 Information on multiple access scheme, duplex scheme

(1). Sparse Code Multiple Access (SCMA)

- SCMA is introduced as a new multiple access scheme. In SCMA, different incoming data streams are directly mapped to code words of different multi-dimensional cookbooks, where each code word represents a spread transmission layer.
- Since the multiple SCMA layers are not fully separated in a non-orthogonal multiple access system, a non-linear receiver is required to detect the intended layers of every user.
- The sparsity of SCMA code words takes advantage of the low complexity message passing algorithm (MPA) detector which achieves ML-like performance.





Overlaid code words with code domain resource sharing
 Fig. 5.3-6 SCMA codebook mapping, encoding, and multiplexing

The technology would be useful when applied to : eMBB, mMTC, URLLC

- Massive connections with user devices become available via SCMA introduction.
- Long and short burst data packets on the devices are carried smoothly. It is also beneficial to achieve higher data throughput, compared with conventional OFDMA, under the same level of channel resource utilization with a smaller packet drop rate in small latency processing.

Following improvements are expected compared with OFDMA:

- Multiplexing gain for massive connections.
- Grant-free multiple access that eliminates the dynamic request and grant signaling overhead, which is an attractive solution for small packets transmission in low latency connection.
- Robust with lower packet drop late, better BLER in link budget, higher throughput in loaded conditions.
- Some adaptive parameters can compromise among spectral efficiency, coverage, detection complexity, connectivity, and link budget, to adapt to different application scenarios.

(2). Non-orthogonal multiple access (NOMA)

- In non-orthogonal multiple access (NOMA) with advanced receiver, multiple users can use the same time and frequency resource.
- In downlink NOMA, a base station multiplexes signals for users in power domain. In uplink NOMA, which is grant free access or scheduled access, multiple users' signals are spatially multiplexed at the base station.
- The technology would be useful when applied to : mMTC:

- NOMA can increase the number of users who simultaneously transmit or receive data at the same resource.

Expected performance/features when applied:

- This technique can improve spectral efficiency since more users can transmit or receive data at the same resource compared to orthogonal multiple access, e.g. OFDMA or SC-FDMA.
- The number of users, which transmit or receive data simultaneously, also increases. As a result, a base station with NOMA can accommodate more users than orthogonal multiple access.

(3). Space Division Full Duplex

- Full duplex or STR (Simultaneous transmission and reception) is extremely challenging since very large TX/RX isolation is required.
- Space division full duplex utilizes spatially separated small transmission points (STPs) alongside with macro transmission points (MTPs).
- While the MTP serves DL to one or some terminals, the STP serves UL to other terminals, or vice versa simultaneously.
- MTP and/or STP may employ adaptive beamforming and successive interference cancellation (SIC) in order to reduce interference to acceptable level for receiving operation.
- Smart algorithms have to be developed since the selection of combination of STPs and terminals being served will have impact on the system performance.
- The technology would be useful when applied to : eMBB
- Following improvements are expected compared with OFDMA:
- Ideally, cell capacity will increase by the factor of 2 compared to conventional duplex scheme.

4. Discuss EPC for 5G.

4.1 Introduction

The key baseline functions for the EPC-based system include support of multiple 3GPP RATs (i.e., GERAN, UTRAN, and E-UTRAN), support for non-3GPP accesses such as W-LAN, and support of fixed wireline access.

- All integrated with functions as Mobility management, Session management, Network sharing, Control and User plane separation, Policy control and Charging, Subscription management, and Security.
- Over the years, EPC has grown with additional features such as Machine Type Communications and Cellular Internet of Things (MTC and CIoT), support for Proximity Services with Device to Device communication and Vehicle to Anything communications support (V2X), Dedicated Core Network selection (DECOR) and Control and User Plane Separation for the GWs (CUPS).
- DECOR and CUPS are two key enablers for the base core network architecture that enhances EPC for 5G based on EN-DC due to the flexibility and versatility they provide for the operators for deployment of differentiated core networks towards specific targeted users. Figs. 4.1 and 4.2 illustrate the key EPS architecture and the simplified architecture of EPC for 5G, respectively.
- As the radio network increases its throughput and bandwidth capacity for 4G and enhanced 4G Radio, operators seek more flexibility and different grades of requirements from the user plane functions provided by the GWs.
- Basic EPC provided separation of control and user plane to some extent, in particular by separating the session management, user plane functions, and external data connectivity into separate GWs but these GWs (e.g., Serving GW and PDN GW) still hold session management control plane functions.
- The CUPS work was driven by operator requirements to scale control and user plane functions independently of one another and the ability to deploy user plane functions in a flexible manner independently of control plane functions.

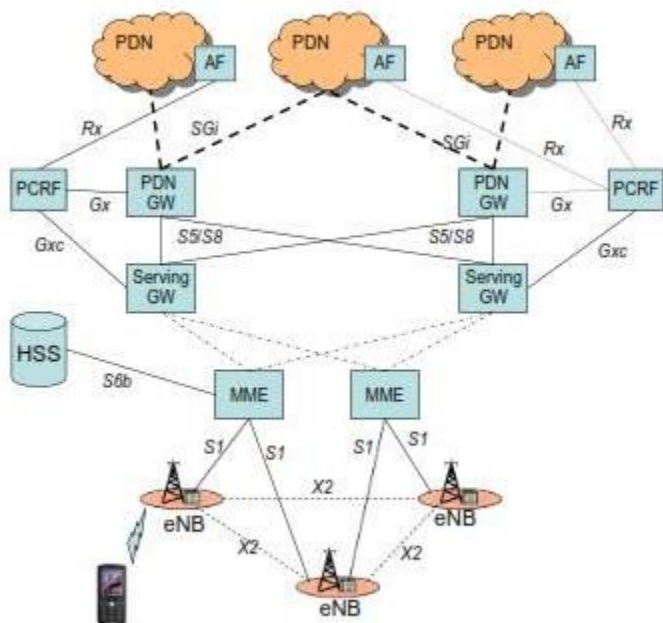


Fig. 4.1 Core EPS architecture for LTE.

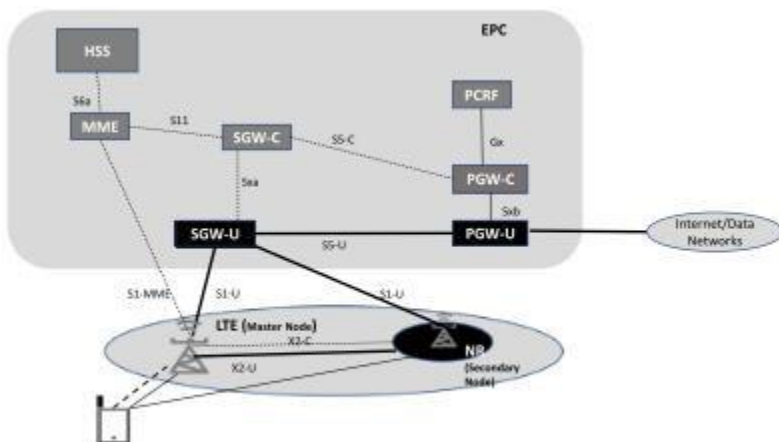


Fig. 4.2 Simplified EPC for 5G architecture.

- The result enables the separation of the SGW and PGW (as well as TDF) control and user plane functions and the flexibility to have a single control plane function to control multiple user plane functions.
- This ability to scale the control and user plane independently allows increased user plane capacity in the network without affecting control plane components.
- DEDicated CORE networks (DECOR and enhanced DECOR), meanwhile, enables operators to partition their core networks into separate dedicated core networks with potentially dedicated MME, SGW, and PGWs used for specific purposes such as dedicated core for CIoT and MBB.
- Together with the Dual connectivity function in the Radio Access network (for more details, see Chapter 12), where RAN can boost the throughput of the UE by adding a secondary RAT using NR 5G Radio for the UE, an operator is able to create the early 5G system using EPC.
- One key aspect for the two features in EPS (DECOR and CUPS) is that both features were developed to minimize UE impacts (or have no UE impacts) and as CUPS functions developed it did not impact existing peripheral nodes such as the MME and PCRF.
- Let us consider an example deployment use case where an operator plans to deploy NB-IoT and MBB. Some MBB users have IMS services, while others are using data services with high data volume requirements.
- An operator may decide to separate its core network components using DECOR principles into two core networks, one for NB-IoT and one for MBB.
- Within the MBB part of the core network, the operator additionally decides to deploy User Plane GWs for high data volume for the MBB APN and use another set of User Plane GWs for IMS services, both being controlled by a single Control Plane function.
- The operator may also decide to deploy DC to boost the radio. The combination of this functionality provides an EPC for 5G enabling early NR deployment, which also continues to support all 4G EPS features without any additional impacts to existing installations.

What are the key EPC functions? (Nov/Dec 23)

4.2 Key EPC functions

- The key entities are HSS, MME, Serving GW, PDN GW, and PCRF. Fig. 4.4 illustrates a simplified architecture for EPS, including only the key components relevant for EPC and specifically for EPC for 5G.

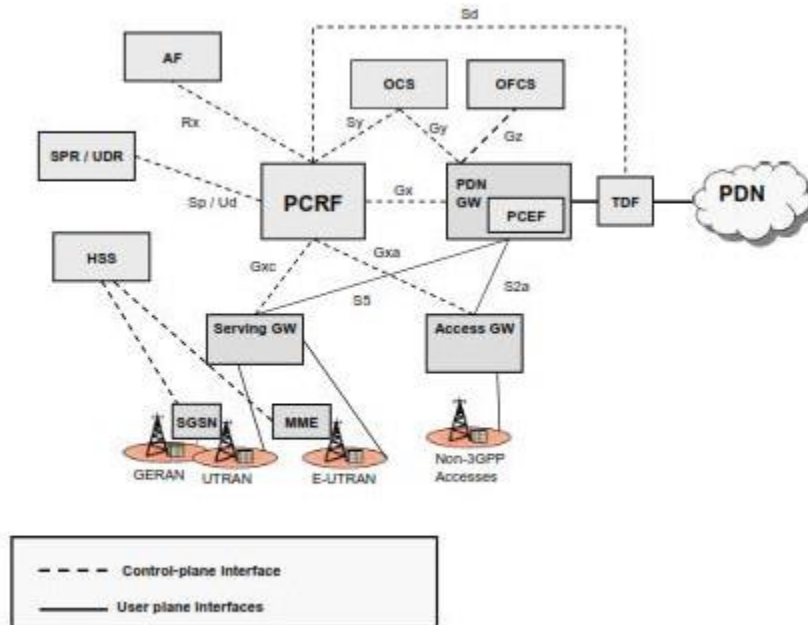


Fig. 4.4 Simplified EPS architecture.

- The 3GPP Radio access networks are GERAN, UTRAN, and E-UTRAN, with our focus being on E-UTRAN (LTE) access only.
- MME is the Mobility Management Entity, responsible for connectivity of the control plane signalling with the eNB, the E-UTRAN component responsible for UE connectivity.
- This node is also responsible for NAS termination, Registration and Tracking Area management, Paging and Authentication, and Authorization with support from HSS (and AUC) for the users and the UEs connecting to the EPS.
- The Serving GW (S-GW) is the user plane termination point from eNB and provides connectivity towards the PDN GW which is the anchor point towards a Packet Data Network (PDN). Each UE is normally served by a single S-GW.

- The PDN GW (P-GW) is the anchor point for a specific UE accessing certain PDN(s), which is represented by the SGi interface. P-GW provides support for all packet data related enforcement functions related to policies and QoS.
- The Policy and Charging Rules Function (PCRF) is the central policy handling entity in EPS. This node is responsible for functions such as QoS control, bearer binding, IP-CAN gating, and policy-related rules instructions towards the PCEF.
- The Policy and Charging Enforcement Function (PCEF) is the entity part of the P-GW function responsible for enforcing policies and rules installed by PCRF.

4.2.1 Subscription and mobility management

In mobile networks there are many functions and processes that require subscription-related information.

- The most obvious example of user subscription data that is used in LTE/EPC networks may be the user identity and security credentials that are required when an end-user device connects to an LTE/EPC network and performs authentication.
- The user identity (IMSI) and the security keys are stored in the USIM card in the device and the same information is also stored for each user in the operator's core network, in the Home Subscriber Server (HSS).
- The subscriber data and functionality of the HSS are used for many functions in 3GPP networks.

The functionality of the HSS includes:

User security support: The HSS supports authentication and security procedures for network access by providing credentials and keys towards network entities such as SGSN, MME, and 3GPP AAA Server.

Mobility management: The HSS supports user mobility by, for example, storing information about what MME is currently serving the user.

User identification handling: The HSS provides the appropriate relations among all the identifiers uniquely determining the user in the system.

Access authorization: The HSS authorizes the user for mobile access when requested by the MME, or 3GPP AAA Server (for PS access), by checking that the user is allowed to roam to a particular visited network.

Service authorization support: The HSS provides basic authorization for mobile terminated call/session establishment and service invocation.

Service provision support: The HSS provides access to the service profile data for use within the CS domain, PS domain, and/or IMS. For the PS domain, the HSS provides the APN profiles that include what APNs the user is authorized to use. The HSS also communicates with IMS entities to support Application Services.

4.2.2 Mobility management

- Mobility management in LTE/EPC involves keeping track of the UE as it may move among cells, via what are known as Tracking Areas.
- UEs are tracked while connected to the core network by the MME, which also where needed and possible, change the serving MME along the way through Tracking Area update and handover procedures.
- As part of the mobility procedure, the UE initially attaches to the network via a registration procedure.

In EPS the registration areas are called Tracking Areas (TAs). In order to distribute the registration update signalling, the concept of tracking area lists was introduced in EPS.

A summary of the idle mobility procedure in EPS is:

- A TA consists of a set of cells.
- The registration area in EPS is a list of one or more TAs.
- The UE performs TA Update when moving outside its TA list.
- The UE also performs TA Update when the periodic TA Update timer expires.

When the UE reselects a new cell and realizes that the broadcast TA ID is not in their list of TAs, the UE initiates a TAU procedure to the network.

4.2.3 Session management

- Session management is how a 3GPP system provides connectivity between the UEs and the service network the UEs are trying to communicate with.
- In EPS, this connectivity is achieved via establishing one or more PDN connections, which connects a UE through the RAN to the PDN GW, which is the 3GPP entry/exit point towards the external networks outside the EPC.
- One of the main goals during session management is establishing PDN connectivity and initial EPS was mainly providing PDN Type IP (IPv4 and IPv6).
- As different types of services demand arose for EPS, the need to support other PDN types also became important.
- Current EPS systems support two additional PDN types known as Non-IP and Ethernet PDN.

- The Non-IP and Ethernet PDN types are useful tools, for example, for low complexity and low throughput UEs specifically designed for Cellular IoT services.

4.2.4 Control-plane aspects

There are several procedures available in EPS to control the bearers. These procedures are used to activate, modify, and deactivate bearers, as well as to assign QoS parameters, packet filters, etc., to the bearer.

4.2.5 QoS

- The EPS only covers QoS requirements for the traffic within the EPS—that is, between UE and PDN GW. If the service extends beyond that, QoS is maintained by other mechanisms that, for example, depend on operator deployments and service level agreements (SLAs) between network operators.
- The EPS bearer represents the level of granularity for QoS control in E-UTRAN/EPS and provides a logical transmission path with well-defined QoS properties between UE and the network.
- The QoS concepts of the EPS bearer are then mapped to the QoS concepts of the underlying transport. We distinguish between two types of bearers: GBR bearers and non-GBR bearers, where GBR is short for Guaranteed Bit Rate.
- A GBR bearer has, in addition to the QoS parameters discussed above, associated bit rate allocations: the GBR and the Maximum Bit Rate (MBR). A non-GBR bearer does not have associated bit rate parameters.
- A bearer with an associated GBR means that a certain amount of bandwidth is reserved for this bearer, independently of whether it is utilized or not.
- The GBR bearer thus always takes up resources over the radio link, even if no traffic is sent. The GBR bearer should not in normal cases experience any packet losses due to congestion in the network or radio link.
- This is ensured since GBR bearers are subject to admission control when they are set up. A GBR bearer is only allowed by the network if there are enough resources available.
- The MBR limits the bit rate that can be expected to be provided by a GBR bearer. Any traffic in excess of the MBR may be discarded by a rate shaping function.

4.2.6 The EPS bearer for E-UTRAN access

- For E-UTRAN access in EPS, one basic tool to handle QoS is the “EPS bearer”. In fact, the PDN connectivity service described above is always provided by one or more EPS bearers (also denoted as “bearer” for simplicity).
- The EPS bearer provides a logical transport channel between the UE and the PDN for transporting IP traffic. Each EPS bearer is associated with a set of QoS parameters that describe the properties of the transport channel, for example, bit rates, delay and bit error rate, scheduling policy in the radio base station, etc.

4.2.7 Default and dedicated bearers

- A PDN connection has at least one EPS bearer but it may also have multiple EPS bearers in order to provide QoS differentiation to the transported IP traffic.
- The first EPS bearer that is activated when a PDN connection is established in LTE is called the “default bearer.”
- This bearer remains established during the lifetime of the PDN connection.
- Even though it is possible to have an enhanced QoS for this bearer, in most cases the default bearer will be associated with a default type of QoS and will be used for IP traffic that does not require any specific QoS treatment.

4.2.8 User-plane aspects

- The UE and the PDN GW use packet filters to map IP traffic onto the different bearers. Each EPS bearer is associated with a so-called Traffic Flow Template (TFT) that includes the packet filters for the bearer.
- These TFTs may contain packet filters for uplink traffic (UL TFT) and/or downlink traffic (DL TFT). The TFTs are typically created when a new EPS bearer is established, and they can then be modified during the lifetime of the EPS bearer.
- When an EPS bearer is established, a bearer context is created in all EPS nodes that need to handle the user plane and identify each bearer.
- For E-UTRAN and a GTP-based S5/S8 interface between Serving GW and PDN GW, the UE, eNodeB, MME, Serving GW, and PDN GW will all have bearer context.
- The exact details of the bearer context will differ somewhat between the nodes since the same bearer parameters are not relevant in all nodes.
- Between the core network nodes in EPC, the user-plane traffic belonging to a bearer is transported using an encapsulation header (tunnel header) that identifies the bearer.

- The encapsulation protocol is GTP-U. When E-UTRAN is used, GTP-U is used on S1-U and can also be used on S5/S8.

4.2.9 Policy control and charging

- Policy control is a very generic term and in a network there are many different policies that can be implemented, for example, policies related to security, mobility, use of access technologies, etc.

When discussing policies, it is thus important to understand the context of those policies. When it comes to PCC for EPC, policy control refers to the two functions gating control and QoS control:

1. Gating control is the capability to block or to allow IP packets belonging to IP flows for a certain service. The PCRF makes the gating decisions that are then enforced by the PCEF. The PCRF could, for example, make gating decisions based on session events (start/stop of service) reported by an Application Function (AF) via the Rx reference point.
 2. QoS control allows the PCRF to provide the PCEF with the authorized QoS for the IP flows.
- With online charging, the charging information can affect, in real time, the services being used and therefore a direct interaction of the charging mechanism with the control of network resource usage is required.
 - Online credit management allows an operator to control access to services based on credit status. For example, there has to be enough credit left with the subscription in order for the service session to start or an ongoing service session to continue.
 - The OCS may authorize access to individual services or to a group of services by granting credits for authorized IP flows. Usage of resources is granted in different forms.

4.3 (Enhanced) Dedicated Core Networks ((e)DECOR)

- (e)DECOR was inspired by the desire and flexibility for the operators to deploy within an operator's network (designated by PLMN ID(s)) multiple core networks and directing users towards specific core networks and thus allowing partitioning off the full core networks.
- This feature as such enables an operator to deploy such multiple Dedicated Core Networks (DCN) within a PLMN with each DCN consisting of one or multiple CN nodes (e.g., MME only, MME with GWs, MME, GWs, and PCRF).
- Each DCN may be dedicated to serve specific type(s) of subscriber and the difference between DECOR and (e)DECOR is that the latter requires the UE to provide specific information (i.e., DCN) to facilitate faster and optimal selection of the core network preferred.
- EPS already makes it possible to direct the UEs towards specific PLMNs, which in turn takes the UE to the specific network (including Core network).
- Usage of the concept of APN allows the PLMN to direct the UEs towards specific service networks in a differentiated manner via selecting different user plane entities (i.e., PDN GWs).
- Fig. 4.1 in Chapter 4.1 shows interconnects of different network elements within a PLMN.
- Figs. 4.5 and 4.6 illustrate how the routing differentiation works in pre-DECOR networks. The EPS system already allowed for directing the UEs towards different CN based on PLMN ID separation within an operator's network (i.e., PLMN).
- But this is quite static and did not allow separation of a deeper granularity and UEs stayed within the PLMN/CN chosen. Then within a CN, using the concept of APN, which allows for user plane separation towards different GWs (PDN GWs) leading to service differentiation within the CN, such as an APN for MBB allowed UEs to be connected to GWs that support MBB services and that allows operators flexibility to isolate them accordingly. A UE may be connected to multiple APNs within a single CN.
- (e)DECOR enables the EPC to "slice" the core network into components that can be tailored to serve specific group of UEs (users), based on their subscription and optionally configured for specific UEs in the UE (only for enhanced DECOR) and in the subscription (HSS).

- This provides operators additional flexibility to separate users into different core network types (e.g., MBB, IoT) as appropriate for the intended usage.
- Prior to DECOR being introduced, users (UE) could access different data services using concepts like Access Point Name (APN), which led to the selection of a different edge GW in the core towards a different data network or selecting a PLMN based on supported PLMN-ID which allowed routing the UEs to a specific CN.
- DECOR allows operators to separate certain types of traffic into specific core network node(s) and if needed, scale them differently than rest of the core network nodes.
- In this way, the operator is also able to segregate specific users more efficiently and control this from the subscription data.
- Whereas, if preferred, with enhanced DECOR, which requires update of the UEs as well, users are also able to choose when registering to the network, which type of DCN network it preferred for the connection.
- Where it may come in handy, as an example, a user coming back into the factory, may choose to use the Dedicated Core Network (DCN) that enables connectivity specific to the factory floor giving access to specific services provided only in that location.
- In 5GS, support of the network slicing function (a more elaborate version of DECOR) in the UEs is enabled from day one, thus eliminating any issues of different types of devices requiring different network behaviour on how to trigger the dedicated core network (aka slicing) separation.
- Some of the key principles of DECOR is that millions of already deployed devices must be able to benefit from this feature. That means network entities need to be able to (re)route UEs within the network and using existing system procedures that the UEs already support.
- At the same time, DECOR must not force an operator to have to handle every UE with DECOR and for that reason the existing core network that is common for all users prior to DECOR must also coexist in a DECOR deployment.

Part – A Two marks

1. What are the drivers for 5G? /List the use cases enabled by 5G (Nov/Dec 23)

- (1) Business case demands from a broader set of economic actors, including industrial companies driving new use cases,
- (2) New technologies for delivering core network components creating expectations of more efficient and flexible operations, and
- (3) Shifts in how business, society and environmental needs are balanced to deliver services in a new way.

2. List the new technologies have driven the development of 5G.

- (1) Virtualization,
- (2) Cloud native,
- (3) Containers,
- (4) Microservices
- (5) Automation

3. Define virtualization.

- Traditionally Mobile core network element functional designs are distributed applications which scale horizontally and run on dedicated hardware such as processor blades in a chassis.
- The network element architecture is distributed internally onto specific types of blades that perform specific tasks.
- The first major step of virtualization was to migrate those application-specific blades to virtualized resources such as virtual machines (VMs) and later containers.

4. Define cloud native.

Cloud Native architectures have gained a lot of interest over the past years and service operators attempt to emulate the efficiencies captured by so-called hyperscalers (e.g., Facebook, Google, Amazon) has led to a much heightened interest in this area.

5. What are several cloud-native design principles?

- ✓ Infrastructure Agnostic
- ✓ Software decomposition and life cycle management
- ✓ Resiliency
- ✓ State-optimized design
- ✓ Orchestration and automation

6. What is infrastructure agnostic?

Cloud-native applications are independent and agnostic of any underlying infrastructure and resources.

7. What is Software decomposition and life cycle management?

Software is decomposed into smaller, more manageable pieces, utilizing microservice architectures. Each piece can be individually deployed, scaled, and upgraded using a CaaS (Container as a Service) environment.

8. Define Resiliency.

In legacy applications, the MTBF (Mean Time between Failures) of hard-ware has been the base metric for resiliency. In the cloud, we instead rely on distribution and independence of software components that utilize auto-scaling and healing. This means that failures within an application should cause only temporary capacity loss and never escalate to a full restart and loss of service.

9. Define State-optimized design.

How we manage state depends on the type of state/data and the context of the state. Therefore, there is no “one size fits all” way of handling state and data, but there should be a balance between performance, resiliency, and flexibility.

10. What is Orchestration and automation?

A huge benefit of cloud-native applications is increased automation through, for example, a Kubernetes-based CaaS layer. A CaaS enables auto-scaling of microservices, auto-healing of failing containers, and software upgrades including canary testing (small-scale testing) before larger deployments.

11. Define container.

- From a simplistic perspective, system-level virtualization allows instances of an Operating System (OS) to run simultaneously on a single-server on top of something called a hypervisor.
- A hypervisor is a piece of computer software that creates and runs virtual machines. System-level virtualization allows multiple instances of OS on a single server on top of a hypervisor.

12. What is microservices?

Microservices are an architectural and organizational approach to software development where rather than be developed in a monolithic fashion, software is composed of small independent services that communicate over well-defined APIs.

13. What are the benefits of microservices?

- Microservice instances have a much smaller scope of functionality and therefore changes can be developed more quickly.
- An individual feature is expected to apply to a small set of microservices rather than to the entire packet and 5GC function.
- Microservice instances can be added/removed on demand to increase/decrease the scalability of their functions. Microservices can have independent software upgrade cycles.

14. What is the key principle of 3GPP 5G Core architecture?

One key principle with the design of the 3GPP 5G Core architecture was not providing backwards compatibility for the previous generations of radio access networks, i.e., GSM, WCDMA and LTE.

15. List the options of 3GPP 5G architecture.

- To specify LTE support for the new 5G architecture
- To specify support for combinations of LTE and NR access
- To specify an alternative 5G architecture based on an evolution of LTE/EPC

16. Mention the ways that LTE or NR can be deployed.

In summary, there are four ways that LTE and/or NR can be deployed:

- Only LTE for all signalling and data traffic
- Only NR for all signalling and data traffic
- A combination of LTE and NR where LTE has the larger coverage and is used for signalling while both LTE and NR are used for data traffic
- A combination of LTE and NR where NR has the larger coverage and is used for signalling while both LTE and NR are used for data traffic.

17. What is AMF?

The AMF is the “Access and Mobility Management Function”. It interacts with the radio network and the devices through signalling over the N2 and N1 interfaces respectively.

18. What is SMF?

The SMF is the “Session Management Function”, meaning as the name suggests that the SMF manages the end user (or actually device) sessions.

19. What is UPF?

The “User Plane Function” (UPF) has as the main task to process and forward user data. The functionality of the UPF is controlled from the SMF.

It connects with external IP networks and acts as a stable IP anchor point for the devices towards external networks, hiding the mobility.

20. What is UDM?

The UDM is the “Unified Data Management Function”. It acts as a front-end for the user subscription data stored in the UDR (more on that further down) and executes several functions on request from the AMF.

21. What is UDR?

The UDR—the “Unified Data Repository”—is the database where various types of data is stored. Important data is of course the subscription data and data defining various types of network or user policies.

22. What is AUSF?

The functionality of the “Authentication Server Function” (AUSF) is quite limited but very important. It provides the service of authenticating a specific device, in that process utilizing the authentication credentials created by the UDM. In addition, the AUSF provides services for generating cryptographical material to allow for secure updates of roaming information and other parameters in the device.

23. What is Beamforming?

Beamforming: 5G NR supports a large number of multiple input multiple output (MIMO) antennas, which would allow it to work in a high interference environment through a technique called “beamforming”. It enables 5G radios to provide both coverage and capacity.

24. What are two types of NR-U implementations?

The types of NR-U implementations are

1. Licensed Assisted Access (LAA) NR-U
2. Standalone NR-U

25. What is MIMO?

MIMO stands for Multiple Input and Multiple Output, where several small antennas are grouped together at the transmitter and at the receiver.

26. What is Coordinated Multipoint (CoMP)?

- Coordinated Multipoint (CoMP) is a technique used to improve radio performance, at cell edges by utilising the capability of more than one base station to enable the communication .
- U s i n g C o M P, radio access nodes dynamically coordinate the transmission and reception of signals from the UE to multiple base stations.

27. List the advantages of Coordinated Multipoint (CoMP)

- ❖ Helps the service provider to improve the utilization of network resources. For example, data can be passed through a base station which is not loaded.
- ❖ User notices lesser call drops, as several cell sites come together to provide service to the user.
- ❖ User notices higher bandwidth, as multiple base stations serve the user, at a given time. The impact of the interference gets reduced and interference is used constructively.

28. What is Microcells?

Microcells are designed for providing wireless coverage to a large geographic area and suited for providing network coverage to outdoor applications such as smart communities. Microcells provide coverage for a distance between 500 meters and 2.5 kilometers. Microcells can support nearly 200 users at a time.

29. What are the benefits of 5G to cloud-hosted solutions?

- ✓ Faster Transfer of Data
- ✓ Uninterrupted Storage
- ✓ Reliability with Large Data
- ✓ Enhanced Productivity

30. What is RAN?

The radio access network (RAN) and aggregated backhalls support the capabilities of data transport, radio transmission and reception.

31. What is SCMA?

SCMA (Sparse Code Multiple Access) is introduced as a new multiple access scheme. In SCMA, different incoming data streams are directly mapped to code words of different multi-dimensional cookbooks, where each code word represents a spread transmission layer.

32. What is NOMA?

In non-orthogonal multiple access (NOMA) with advanced receiver, multiple users can use the same time and frequency resource.

33. What are the key baseline functions for the EPC?

The key baseline functions for the EPC based system include support of multiple 3GPP RATs (i.e., GERAN, UTRAN, and E-UTRAN), support for non-3GPP accesses such as W-LAN, and support of fixed wireline access.

34. What are the functionality of the HSS (Home Subscription Server)?

- User security support
- Mobility management
- User identification handling
- Access authorization
- Service authorization support
- Service provision support

35. What is policy control charging (PCC)?

Policy control is a very generic term and in a network there are many different policies that can be implemented, for example, policies related to security, mobility, use of access technologies, etc.

36. What is (Enhanced) Dedicated Core Networks ((e)DECOR)?

(e)DECOR was inspired by the desire and flexibility for the operators to deploy within an operator's network (designated by PLMN ID(s)) multiple core networks and directing users towards specific core networks and thus allowing partitioning off the full core networks.

37. What is virtualized evolved packet core? (Nov/Dec 23)

DECOR and CUPS are two key enablers for the base core network architecture that enhances EPC for 5G based on EN-DC due to the flexibility and versatility they provide for the operators for deployment of differentiated core networks towards specific targeted users.

4G/5G COMMUNICATION NETWORKS

UNIT – III NETWORK ARCHITECTURE AND THE PROCESSES

5G architecture and core, network slicing, multi access edge computing (MEC), visualization of 5G components, end-to-end system architecture, service continuity, relation to EPC and edge computing, 5G protocols: 5G NAS, NGAP, GTP-U, IPSec and GRE.

1. Discuss about Network slicing/ What is network slicing? Explain its requirements, architecture, benefits and management. Discuss same with 100 users (Nov/Dec 23)

11.1 Introduction

- Traditional networks and their one-size-fits-all approach needs to be adapted so that the expected large number of network deployment use cases, many different subscriber types with diverse and sometimes contradictory requirements, and varying application usage can be supported.
- So, instead of using a single monolithic network serving multiple purposes, technology advancements such as Virtualization and SDN allows us to build logical networks on top of a common and shared infrastructure layer. These logical networks are then called Network Slices.
- The meaning of the term Network Slice vary in the industry, but in general a Network Slice is a logical network serving a defined business purpose or customer, consisting of all required network resources configured together.
- A Network Slice is realizing a complete network for any type of access and is an enabler for providing services. The used physical or virtual infrastructure resources may be dedicated to the Network Slice or shared with other Network Slices.
- As the network slicing concept allows multiple logical networks to be created, they can then be accommodated to realize a wanted network characteristic and provide specific network capabilities to address a specific customer need.
- The customer here is not directly the end-user, but a business entity that has requested specific services from the network operator, e.g., an enterprise, another service provider or the network operator itself.
- The Network Slices are orchestrated and managed by management functions. The concept of network slicing and one definition is summarized in [Fig. 11.1](#).

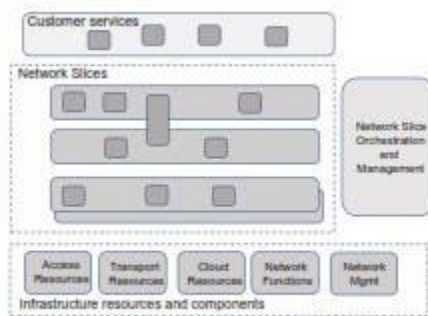


Fig. 11.1 A Network Slice definition.

Network Slice is a **logical network** serving a defined **business purpose** or **customer**, consisting of **all** required network resources **configured** together.

- Complete network within a provider
- Enabler for services
- All access types
- Resources may be physical or virtual, dedicated or shared
- Independent/isolated but may share resources

- What, then, is the benefit with network slicing? The network slicing concept assumes virtualization and automated orchestration and management, and the expectations is that when these are used together they provide:
 - ✓ Better customer experience by per customer adaptations and optimizations
 - ✓ Shorter time-to-market and time-to-customer
 - ✓ Simpler resource management
 - ✓ Increased automation
 - ✓ Flexibility and agility
 - ✓ Reduced risks by separation of concerns.
- Depending on the service type, e.g., eMBB, URLLC, mMTC, and customer expectations, there may be different requirements to be addressed by a Network Slice, for example:
 - Traffic capacity requirements per geographical area
 - Charging requirement
 - Coverage area requirement
 - Degree of isolation requirement
 - End-to-end latency requirement
 - Mobility requirement
 - Overall user density requirement
 - Priority requirement
 - Service availability requirement
 - Service reliability requirement
 - Security requirement
 - UE speed requirement.

- To address the various and possibly diverse requirements when designing Network Slices, the various resources, and logical functions may be placed in different parts of the network. Fig. 11.2 provides example realizations for some type of Network Slices.
- To separate and “Slice” parts of the network is not a new concept. It is supported already by different mechanisms and for different purposes, e.g., the ability to share a radio network between different operators is supported by each operator using separate PLMN identities, or separating PS data can be done by establishing separate data paths (i.e., PDP Contexts in 3G, PDN Connections in 4G and PDU Sessions in 5G),
- These techniques are also available in 5GS, but they have limitations. For example, even if it is possible for an operator to get more than one Mobile Network Code (i.e., MNC part of the PLMN identity) the number of MNC values available are not enough for each Network Slice to be given a separate MNC.
- Also, the separation of PS data paths using separate DNNs would only enable separation of part of the network and would not meet the expectation of complete logical networks dedicated for the customer needs.
- Therefore, it was not possible to re-use any of the existing mechanisms for addressing the complete network slicing concept. However, of course, the existing means can also be used within a Network Slice, to achieve a limited separation between resources.

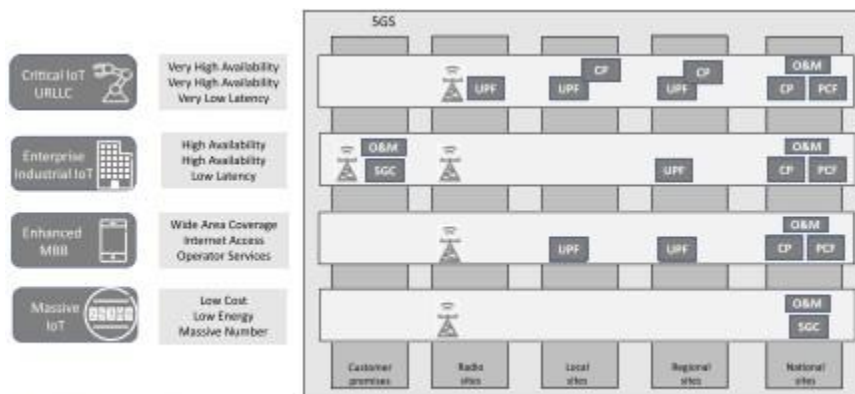


Fig. 11.2 Network Slice examples.

11.2 Management and orchestration

- During the preparation and whole Lifecycle management process, the customer is able to provide its requirements using APIs from which the customer gets information of how the Network Slices perform, and is able to modify its requirements as to adapt to the needs of the customer. Fig. 11.3 provides a high-level view of the process in the preparation and the Lifecycle management of a network Slice Instance (NSI).

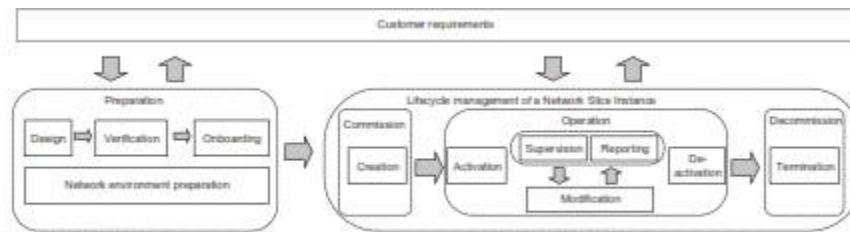


Fig. 11.3 Preparation and Lifecycle management of a Network Slice Instance.

- In each of the steps the nature of isolated Network Slices aids to increase the speed in the process as there are less dependencies to consider.

11.2.1 Preparation

- Network Slice “blueprints” or “templates” are used to simplify the process. If a Network Slice template exists that meets the customer requirements, then the preparation process can be shortened, as either the customer may be able to use an existing NSI, i.e., an existing NSI is then scaled to also meet the requirements from the new customer, or a new NSI is to be created using an existing Network Slice template.
- If that is the case, then the preparation phase can be excluded, and the ordering of the creation and activation can begin.
- If there is no suitable Network Slice template, then a new one is designed using the customer requirements.
- Once a new Network Slice template is designed it is normally added to a catalogue of services and Network Slice templates that allow the preparation phase to be skipped or shortened for the next customer with the same or similar requirements.
- The verification is simplified when done for dedicated Network Slices as there are fewer dependencies to consider compared to when using one network for a large range of customers, applications and services.

- The on boarding includes uploading required information, e.g., the designed templates into the production system, validation of, e.g., templates and virtual machines (VM) images, and everything that is needed by the orchestration system in the next step.
- During the preparation phase the network environment is prepared and other necessary preparations are done as required for the creation of an NSI. Then the ordering of the creation and activation can be done.
- If there is no suitable Network Slice template, then a new one is designed using the customer requirements. Once a new Network Slice template is designed it is normally added to a catalogue of services and Network Slice templates that allow the preparation phase to be skipped or shortened for the next customer with the same or similar requirements.
- The verification is simplified when done for dedicated Network Slices as there are fewer dependencies to consider compared to when using one network for a large range of customers, applications and services.
- The on boarding includes uploading required information, e.g., the designed templates into the production system, validation of, e.g., templates and virtual machines (VM) images, and everything that is needed by the orchestration system in the next step.
- During the preparation phase the network environment is prepared and other necessary preparations are done as required for the creation of an NSI. Then the ordering of the creation and activation can be done.

11.2.2 Commissioning

- NSI provisioning in the commissioning phase includes creation of the NSI. During NSI creation all needed resources are allocated and configured to satisfy the Network Slice requirements.

11.2.3 Operation

- The Operation phase includes the activation, supervision, performance reporting (e.g., for KPI monitoring), resource capacity planning, modification, and de-activation of an NSI.
- Activation makes the NSI ready to support communication services.
- Resource capacity planning includes any actions that calculates resource usage based on an NSI provisioning, and performance monitoring and generates modification policies as a result of the calculation.
- The supervision and performance reporting include, e.g., monitoring, assurance and reporting of the performance according to the KPIs agreed as part of the Service Level Agreements (SLAs) for the NSI.

- NSI modification could include, e.g., capacity or topology changes. The modification can include creation or modification of NSI resources. NSI modification can be triggered by receiving new Network Slice requirements or as the result of supervision/ reporting.
- The deactivation includes actions that make the NSI inactive and stops the communication services.

11.2.4 Decommissioning

- NSI provisioning in the decommissioning phase includes decommissioning of non-shared resources if required and removing the NSI specific configuration from the shared resources. After the decommissioning phase, the NSI is terminated and does not exist anymore.

2. Explain Multi access edge computing (MEC) (Nov/Dec 23)

Multi-access Edge Computing (MEC)

- Multi-Access Edge Computing (MEC) is an integral part of the 5G ecosystem. MEC helps service providers to bring application oriented capabilities closer to the users and support several latency sensitive use cases from the edge.
- The MEC system brings networking and computing capabilities at the edge of the network to optimize the performance for ultra-low latency and high bandwidth services.
- The initial use cases of MEC were very specific to mobile networks and hence it was called as Mobile Edge Computing (MEC).
- However, later on, the industry acknowledged the general applicability of MEC for both wireless and wired networks and hence renamed it to Multi-Access Edge Computing.

Need for MEC

- Compute infrastructure for application services were existent in some form, even in 4G and 3G networks.
- For example, video transcoding, WAN optimization, Content Delivery Network (CDN) and transparent caching services were running previously in the service provider's core network in purpose built network equipment.
- However, with the growth in the number of mobile devices connecting to the network and the explosion of data consumption, it is impossible to offer such application services from a centralized location, without impacting the user experience. Hence, a mobile edge computing infrastructure was conceptualized.

Some of the key drivers for MEC in 5G Network are:

- ✓ Growth in the number of mobile devices connecting to the network (with IOT, it is expected to explode even further)
- ✓ Growth in the volume of data generated by the Over the Top (OTT) applications such as social media, video streaming and online gaming.
- ✓ Need for distributing the infrastructure where the application services are hosted in a service provider network, to improve the application performance and the user experience.
- ✓ Need for running application services in multiple locations to increase the reliability of the services.
- ✓ Need for virtualizing the application services and eliminate the dependencies with purpose-built hardware to simplify the management and orchestration of multi-vendor functions .
- ✓ Dramatically reduce the network latency to support new use cases such as Autonomous Cars, Virtual Reality, Augmented Reality and Robotic Surgeries.

MEC Architecture

MEC architecture resembles NFV architecture. The MEC architecture consists of the following functions:

- ❖ MEC Orchestrator
- ❖ MEC Platform
- ❖ MEC Platform Manager
- ❖ Virtualization Infrastructure
- ❖ MEC Application Services

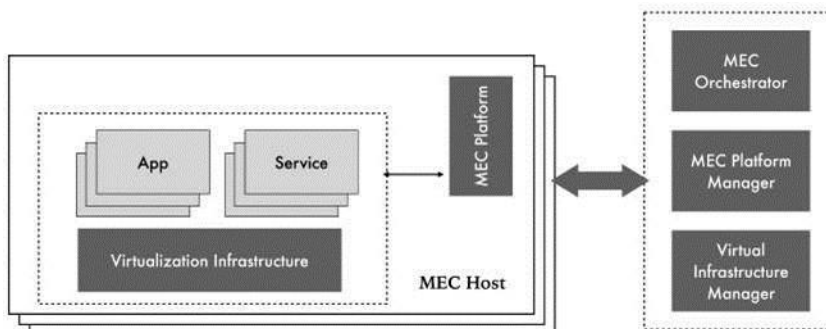


FIGURE 6.1 - MEC SYSTEM ARCHITECTURE

MEC Orchestrator

- MEC Orchestrator is a centralized function and has the complete view of the multi-access edge systems including the topology, available resources in the virtualized infrastructure, available applications and services running on the virtualized infrastructure.
- MEC Orchestrator triggers the life cycle management of the applications and services running on the virtualized infrastructure, including service instantiation, service termination and service relocation.
- It also selects the right set of resources for running the applications and services, to meet the latency requirements.

MEC Platform

- MEC platform provides an environment where applications can discover, advertise, consume and offer mobile edge services.
- It receives regular updates from the MEC platform manager and the various applications or services running in the virtualized infrastructure.
- Some of the updates received by the MEC Platform include activation and deactivation of traffic rules and DNS records.
- For example, MEC Platform would work with the data plane to establish the traffic path for the various applications.
- MEC Platform uses the DNS record updates to configure the DNS proxy or server in the network.
- Thus, DNS records can be used to redirect traffic to a specific application running on the MEC host.

MEC Platform Manager

- MEC Platform manager provides the Fault, Configuration, accounting, Performance and security (F C A P S) management services.
- It periodically receives fault and performance related reports from the Virtual Infrastructure manager and notifies the MEC Orchestrator about the application and service specific events.
- MEC Platform manager also manages the application and service specific rules and policies for managing the traffic.

Virtualization Infrastructure

- Virtualized infrastructure provides shared compute, storage and networking resources for hosting MEC related applications or Virtual Network Functions (VNFs).

- This infrastructure can also be shared with other non-MEC VNFs.

Virtualized Infrastructure Manager

Virtualized Infrastructure Manager manages the infrastructure resources required for the various applications and services hosted on the MEC host. It partitions the physical resources and make them available as multiple tenant spaces for hosting the MEC applications and services.

MEC Applications & Services

The service provider can run their own network applications or services in the MEC. The service provider can also run partner or customer applications on the MEC. A MEC application can belong to one or more network slices that have been configured in the 5G core network.

MEC Deployment modes

MEC can be deployed in one of the four deployment modes, as given below:

Breakout mode - The session connection is redirected to a MEC application which is either hosted locally on the MEC platform or on a remote server. Some examples of breakout applications are local Content Deliver Network (CDN) caches (for ex., Akamai caches), gaming services and media delivery services (for ex., Netflix streaming). Normally, you achieve this by defining forwarding policies

In-line mode - MEC is deployed transparently, in an in-line mode. The session connection is maintained with the original server, while all the traffic traverses and goes through the application running in the MEC. Examples of in-line MEC applications are transparent content caching and security applications.

Tap mode - In Tap mode, data exchanged in a session is selectively duplicated and forwarded to the tap MEC application. Some examples of tap mode applications are virtual network probes, monitoring and security applications.

Independent mode - MEC application and services run independently, but still the MEC application is registered in the MEC platform and will receive other MEC services, such as DNS and radio network information (for ex., radio bearer statistics). Steering the traffic to the MEC is achieved by configuring local DNS or MEC host's data plane.

MEC Deployment scenarios in 5G Network

- MEC can flexibly be deployed in different locations of the 5G network, from near the Base Station to the central Data Network.
- Irrespective of where the MEC is deployed, the User Plane Function (UPF) has to steer the traffic towards the MEC application and back to the network.

- UPF is responsible for traffic forwarding in a 5G network. 5G architecture provides the flexibility to deploy UPF instances at the network edge, as well as at the network core, for improving performance and reducing latency.
 - There are 4 possible deployment scenarios for the MEC system in a 5G network.
 - The location in which the MEC is deployed is dependent on a number of factors such as the infrastructure availability (power, space and cooling), type of applications/services hosted in the MEC, network latency and bandwidth requirements.
1. MEC and the User Plane Function (UPF) can be co-located with the Base Station.
 2. MEC co-located with a transmission node and possibly with a UPF
 3. MEC and the UPF co-located with a network aggregation point
 4. MEC co-located with the Core Network functions, in the same data centre

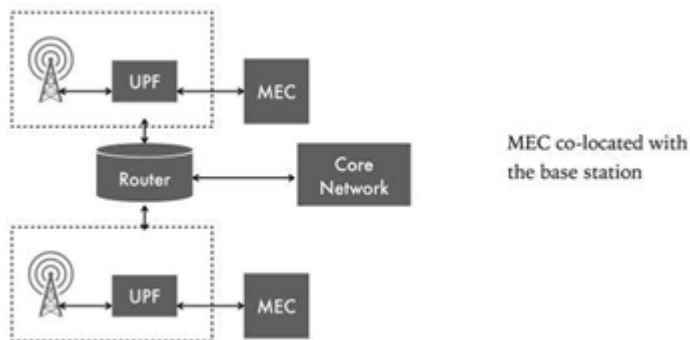


FIGURE 6.2 - MEC CO-LOCATED WITH BASE STATION

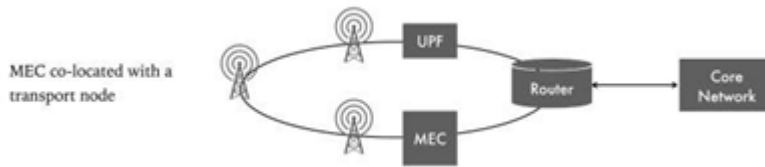


FIGURE 6.3 - MEC CO-LOCATED WITH TRANSPORT NODE

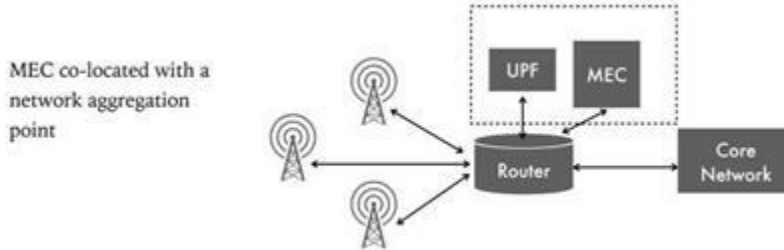


FIGURE 6.4 - MEC CO-LOCATED WITH NETWORK AGGREGATION POINT

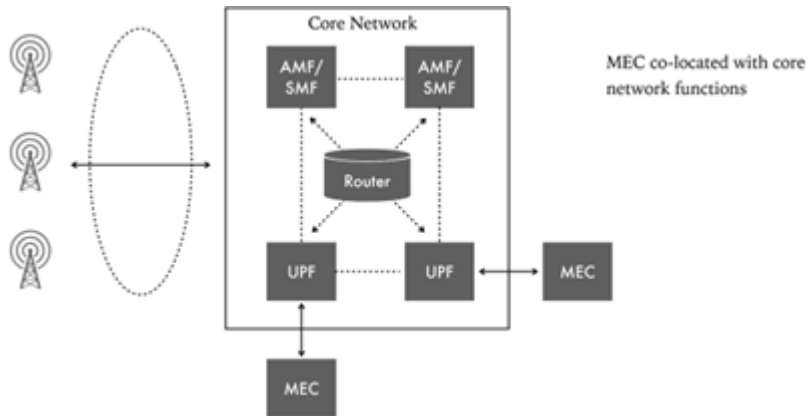


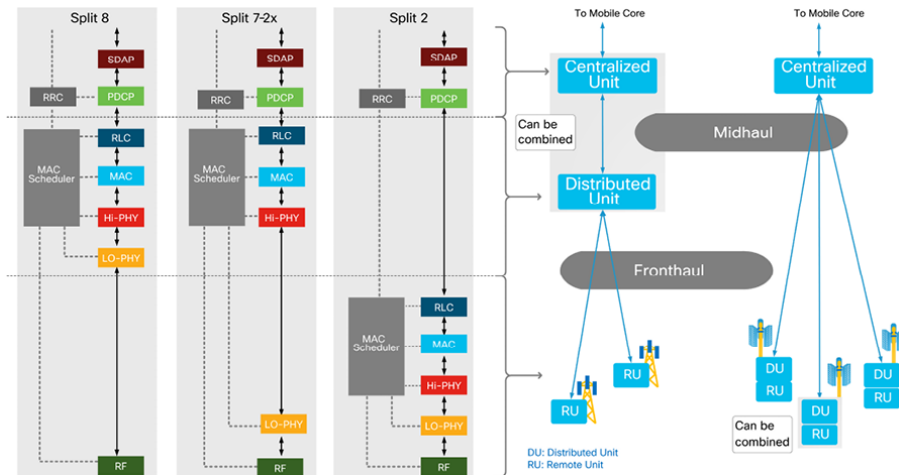
FIGURE 6.5 - MEC CO-LOCATED WITH CORE NETWORK FUNCTIONS

3. Explain 5G end-to-end system architecture

A new end-to-end mobile architecture

- Any new end-to-end mobile network deployed today must adopt basic 5G system architecture principles even if it's using LTE access. Operators should embrace the key principles that underpin the foundation of the 5G system architecture.
- A monolithic functional implementation such as the base transceiver station (BTS) is an example of why the current mobile supply chain is outdated.
- With this implementation, operators must pick one vendor per market and harmonize the macro vendor markets to a "lowest common denominator" set of features.
- The result is a limited set of applications that operators can provide to their customers.
- Sometimes, the vendor dependencies and lock-in can propagate into other domains when proprietary features are implemented.
- For example, mobility management entities (MMEs) are often implemented with the RAN vendor, and small cells are often coupled to the macro RAN using features such as HetNet.
- The new cloud RAN architecture addresses, among other things, the challenges of building multivendor networks and harmonizing to a common feature set.
- One fundamental characteristic is the decomposition of the radio signal processing stack using standardized splits (see Figure 1).

Figure 1. Radio signal processing stack.



- The radio signal processing stack in both 5G NR and LTE is a “service chain” of functions which are processed sequentially.

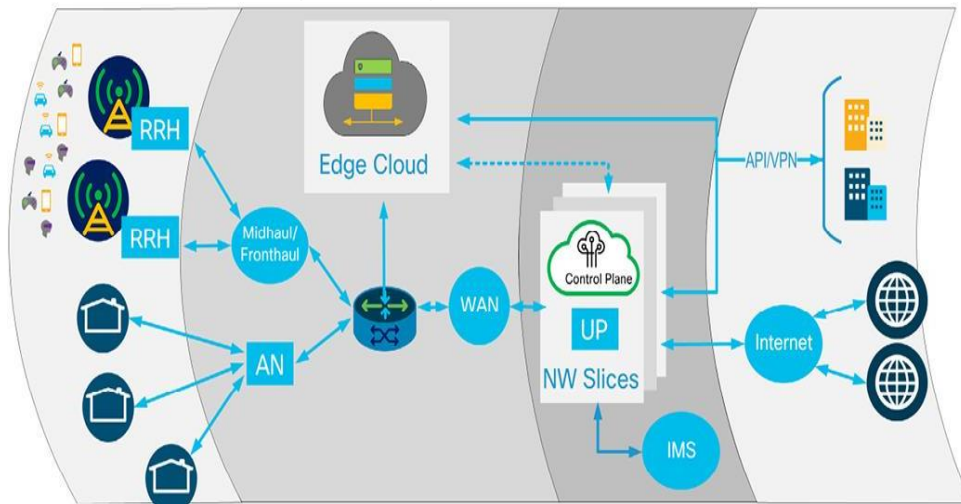
These functions are controlled using signalling derived from the packet core, specifically the MME in the evolved packet core (EPC), and the access and mobility management function (AMF) in 5GC.

The decomposition maps function in the stack as follows:

1. Service delivery adaptation protocol or SDAP (in NR only), packet data convergence protocol (PDCP), and radio resource control (the control functions) to a centralized unit (CU).
 - These stack functions are packet-level manipulations (header compression, over-the-air ciphering) that aren't timing sensitive and can be easily implemented in a virtualized environment.
 - The midhaul links connects the CU to the DU defined below. The CU location is excellent for deploying user plane function (UPF) in the decomposed packet core architectures. The decomposed packet core is known as CUPS in 3GPP

2. The radio link control (RLC), medium access control (MAC); and physical (PHY) layers map to a distributed unit (DU).
 - The DU performs significant preparation for the RF layer including rate adaptation, channel coding, modulation, and scheduling. For the MAC layer, the functions of the DU are time sensitive because a transport block of duration of the Transmit Time Interval (1 ms in LTE) is produced for consumption by the PHY layer.
 - The DU link to the downstream entity is called front and transports digitized RF samples in either the time domain or frequency domain.
3. The baseband radio functions, up and down conversion, and amplification along with any analog beamforming map into a remote unit (RU) deployed at the cell-site also known as a remote radio head (RRH).
 - This decomposition and isolation of functions, along with the well-defined interfaces between them, allow operators to disaggregate software from hardware.
 - Operators can procure software capabilities that will execute in hardware supplied by a different vendor. This disaggregation is natural and relatively simple for the centralized unit (CU) because those functions can be easily virtualized. Disaggregation of the Distributed Unit (DU), is more challenging because of the real-time processing requirements of the MAC and PHY layers.
 - Many in the industry questioned whether it was even possible to achieve the benefits of virtualization for the DU.
 - The decomposition and disaggregation of 5G enables a major architectural shift to an edge infrastructure that combines decomposed subscriber management with access functionality. This shift will also apply to wireline networks with increasing adoption of the 5GC wireline networks (see Figure 2).

Figure 2. Edge infrastructure.



- As the RAN and core become decomposed, it's logical to create an intermediate location. This location would consolidate the CU workloads and a user plane function (UPF) from the mobile core (see the Edge Cloud in Figure 2).
- These distributed UPFs facilitate offload, enabling local virtualized services or more efficient peering at a metro level.
- This unified platform approach supports infrastructure workloads and a variety of service-oriented workloads.
- These workloads can support business-to-business (B2B) services such as tenancies offered to other businesses, and business-to-consumer (B2C) services in support of the operator consumer business.
- Splitting the architecture as described above creates two additional transport domains known as fronthaul and midhaul.
- These domains are implemented when remote radio head (RRH)/DU split or CU/DU splits are exposed.

- In some cases, it's possible to implement fronthaul over dark fiber, which is basically, transport of time domain or frequency domain baseband samples between the DU and the RU/RRH.
- In other cases, transport techniques that rely on Ethernet, IP, or WDM must be employed. Midhaul is basically the transport of GTP-u packets and associated control plane between the CU and the DU. This type of transport is easily implemented over IP.

4. Discuss about 5G service continuity

SSC Modes – Session and Service Continuity in 5G

- In any telecom technology one of the key feature requirement is data session and service continuity to ensure the uninterrupted service experience to the users irrespective of any change of UE IP address or change in Core network anchor point.
- In 4G, EPC provides the continuity of IP session by maintaining the P-GW and UEs PDU session IP regardless of UE's mobility.
- In 5G system architecture, Session and Service Continuity (SSC) support enables to address the various continuity requirements for different applications and services for the UE.
- However, not all applications require guaranteed IP session continuity even if service continuity is needed but 5G being more flexibility and evolved, offers different types of session continuity depending on UE or service type.

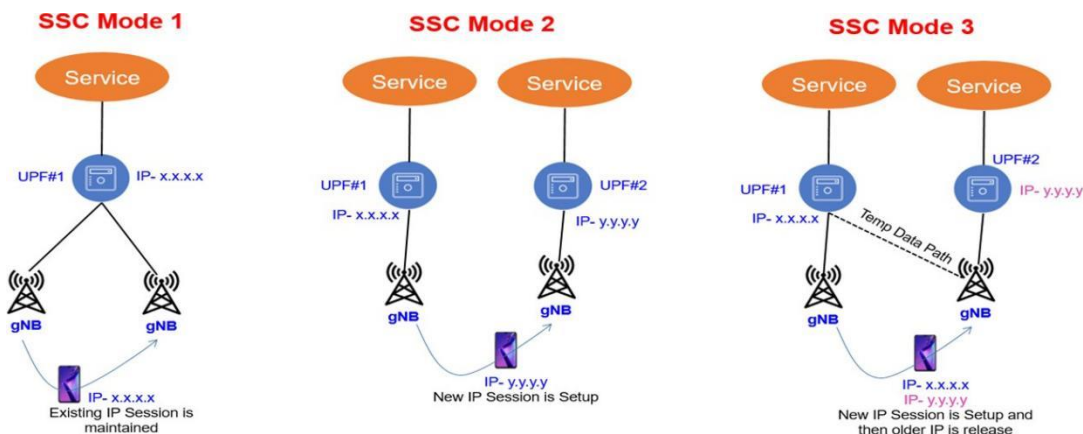
Key Pointer for SSC Modes

- 5G technique allows the service provider to set a specific SSC mode for a given PDU Session.
- SSC Modes determines the flow of the packets across network in mobility scenarios
- A PDU Session is configured to use a specific 'Session and Service Continuity' (SSC) mode
- 3GPP specification provided three different types of SSC modes
- UE can request a specific mode using the 'SSC Mode' field within the NAS: PDU Session Establishment Request
- The SMF specifies the allocated mode using the 'Selected SSC Mode' field within the NAS: PDU Session Establishment Accept

- SSC Mode 1 & 2 can work for PDU Type as IP and Ethernet where as SSC Mode 3 can be only work with PDU session type as IP

Types of SSC Modes

- 5G System architecture provides three types of Session and Service Continuity (SSC) modes to addresses various continuity requirements of different applications/services.
- Once an SSC mode is associated with a PDU Session then it does not change during the lifetime of the PDU Session.
- The 5G architecture allows applications to influence the selection of SSC modes as needed for required data service and shown in following figure.



SSC Mode 1: With SSC mode 1, the 5G network preserves the connectivity service provided to the UE.

- For the PDU session Type as IPv4 or IPv6 or IPv4v6 type, the IP address is preserved.
- In this case the User Plane function (UPF) acting as the PDU session anchor is maintained (remains same) till the point UE release the PDU session.

SSC Mode 2: With SSC mode 2, the 5G network may release the connectivity provided to the UE, i.e. the PDU Session can be released.

- If the PDU Session is being used to transfer IP packets, then the allocated IP address is also released.
- A use case for this is a network may release connectivity if there is a requirement for load balancing at the anchor UPF.
- Here, the PDU Session may be moved onto a different anchor UPF by releasing the existing PDU Session and subsequently establishing a new PDU Session.

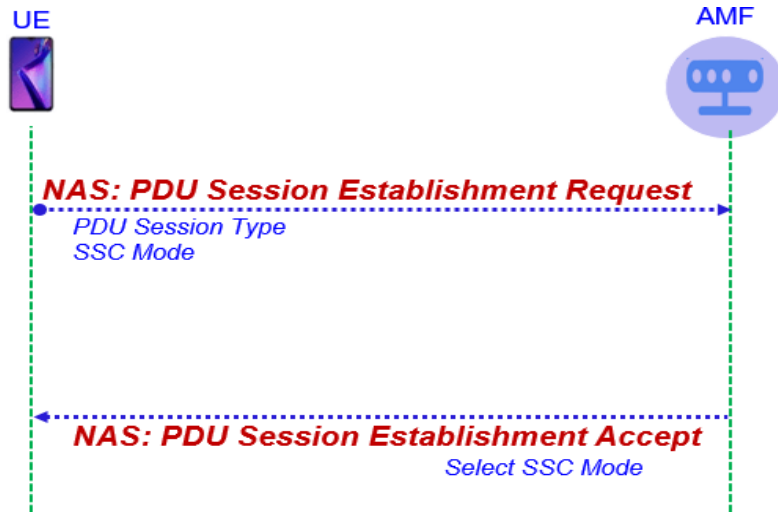
- It works on break and make framework i.e. PDU session will be release from first serving UPF and then a new PDU session is established at new UPF.

SSC Mode 3: With SSC mode 3, the network preserves the connectivity provided to the UE but there may be some impact during certain procedures.

- For example, the IP address allocated to the UE will be updated if the Anchor UPF changes but the change procedure will ensure that connectivity is preserved, i.e. connectivity towards the new Anchor UPF is established before releasing the connection to the old Anchor UPF.
- The 3GPP release 15 only supports Mode 3 for IP based PDU Sessions.
- How does Session and Service Continuity Mode Selection Works? The SSC mode selection policy is used to determine the type of session and service continuity mode associated with an application or group of applications for the UE.
- A MNO may provision the policy rules for UE to determine the type of mode associated with an application or a group of applications. There can be a default policy that matches all applications on UE.
- When a UE initiates a PDU session it determines which SSC mode it has to use for a particular application by checking SSC mode selection policy and included it within PDU session establishment request as “SSC mode” IE.
- In 5G Core, the SMF can receive the list of supported SSC modes and the default SSC mode per DNN per S-NSSAI as part of the subscription information from the UDM. The SMF select the SSC mode by checking against subscriber data and local SMF configuration and allowed SSC mode.
- Based on selection results, the SMF can either accepts or modifies or rejects based on UE subscription or local configuration. If UE does not provide SSC, then SMF selects default SSC depending on the data network in subscription or local configuration.

Call Flow for SSC Mode Information:

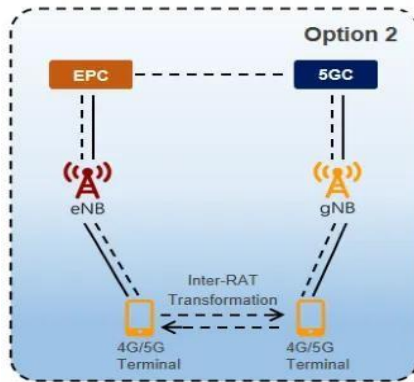
- The **SSC mode** information is exchanged as part of **NAS layer signalling** as shown below.
- The UE sends the PDU session establishment request including the PDU session type and request SSC mode (Optional) toward the Core Network. The Core network response the Selected SSC mode as part of PDU session establishment accept.



Explain the relation to EPC How does core network evolve to 5GC from EPC ?

- EPC (Evolved Packet Core) has ever been the crucial part of 4G LTE. While with the ongoing innovation of the telecommunication industry, 5GC (5G Core Network) defined by 3GPP turns to a rising star in core network field.
- Compared with the earlier generations, 5G deployment can carry both EPC and 5GC in its actual implementation.
- What's more, 5G deployment also introduced 5GC in SA (Standalone) mode and EPC or 5GC in NSA (Non-standalone) mode, which makes it more complicated to explore the best evolution roadmap from EPC to 5GC.
- However, we are always facing both opportunities and challenges. Difficulties may be the chance to develop. Let's see four paths of evolution adapted to specific scenario.

1. 5G SA Network Deployment

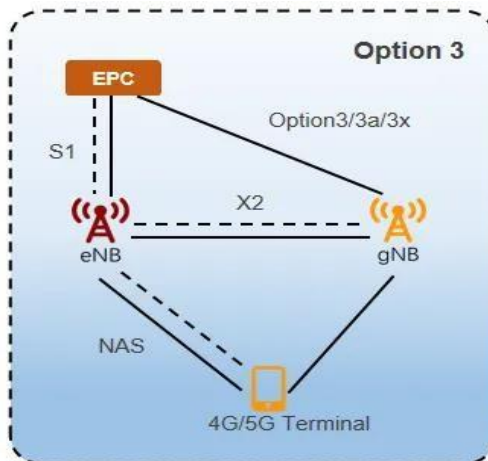


- Option 2 refers to deploy 5G NR (New Radio) with 5GC in 5G SA network deployment.
- When the coverage of 5G is too limited to cover the whole service area, the re-orientation and inter-RAT transformation between LTE/EPC and NR/5GC should be supported under this circumstance.

Advantage:

To make better use of the competence of 5G, we just need to upgrade the existing EPC to complete the interconnection between EPC and 5GC with this solution.

2. 5G NSA Network Deployment



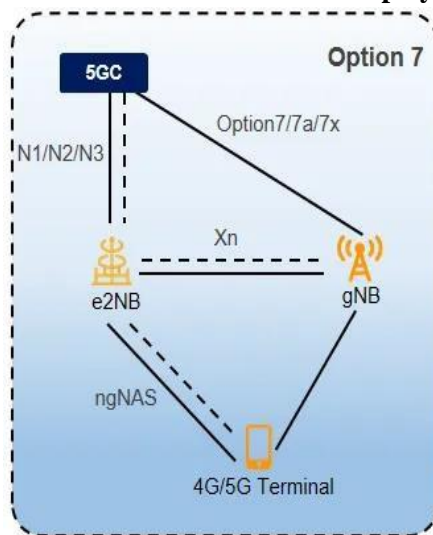
The 5G NSA network architecture with the deployment of gNodeB in existing EPC (Option 3)

- While evolving to 5G NSA mode, Option 3 should always be the best choice. The picture above depicts the dual connectivity of Option 3.
- This option enables UE access simultaneously through LTE or NR, and supports the LTE NSA architecture based on LTE.

Advantage:

There is no need to deploy 5GC in early period of 5G deployment. By implementing this network architecture, the control interface of S1 is only provided to LTE eNodeB, while the user plane interface of S1 can be provided to eNodeB and NR gNodeB at the same time. Therefore, the traffic flows in EPC should be classified.

3. Introduce 5GC in NSA Network Deployment



The 5G NSA network architecture with the deployment of 5GC, gNodeB and eLTE (Option 7)

In essence, Option 7 is the introduction of 5GC in Option 3.

Advantage:

- The eLTE (enhanced LTE) after Release-15 access to the new 5GC with the control and user interfaces, whereas gNodeB only has the user interface connected to 5GC.
- So Option 7 can be smoothly upgraded in 5GC deployment, which could provide better function than SA Option 2.
- However, it's unnecessary to evolve to Option 7 if the deployment of SA Option 2 becomes the mainstream option.

4. Maturity of 5G NR deployment

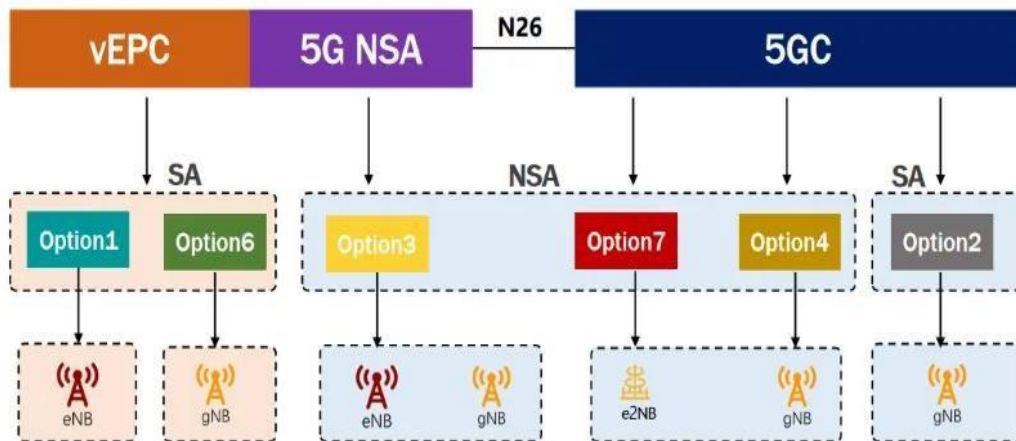
- Once the 5G NR deployment becomes mature, the existing LTE eNodeBs in SA deployment may be replaced by NR gNodeBs.
- In NSA mode, Option 3 and Option 7 can be upgraded to Option 4 which can also adapt to the LTE NSA network architecture applied in dual connectivity.

Advantage:

With the requirement of 5G NR providing continuous coverage, the control plane always set in 5G NR to fully achieve massive and strong function for users. The rapid rising of 5GC has become an irresistible tendency.

- As a prominent mobile core network provider, IPLOOK always keeps pace with the innovation of core network. IPLOOK also constantly develops the path to provide converged core network for our customers. Nowadays, IPLOOK can support seven available options of 5GC deployment with various configuration according to 3GPP.

4. IPLOOK



IPLOOK supports seven available options of 5GC deployment such as UPF, AMF, SMF, UDM, AUSF, PCF, NSSF, NEF, NRF.

5. Explain edge computing

Edge computing

- Edge computing is about bringing the services closer to the location where they are to be delivered.
- Services here includes computing power and memory needed for e.g. running a requested application.

- Edge computing therefore aims to push applications, data and computing power (services) away from centralized points (central data centers) to locations closer to the user (such as distributed data centers).
- The goal is both to achieve a lower latency and to reduce transmission costs.
- Applications that use high data volumes and/or require short response times, e.g. VR gaming, real-time facial recognition, video surveillance etc. are some candidates that could benefit from Edge computing.

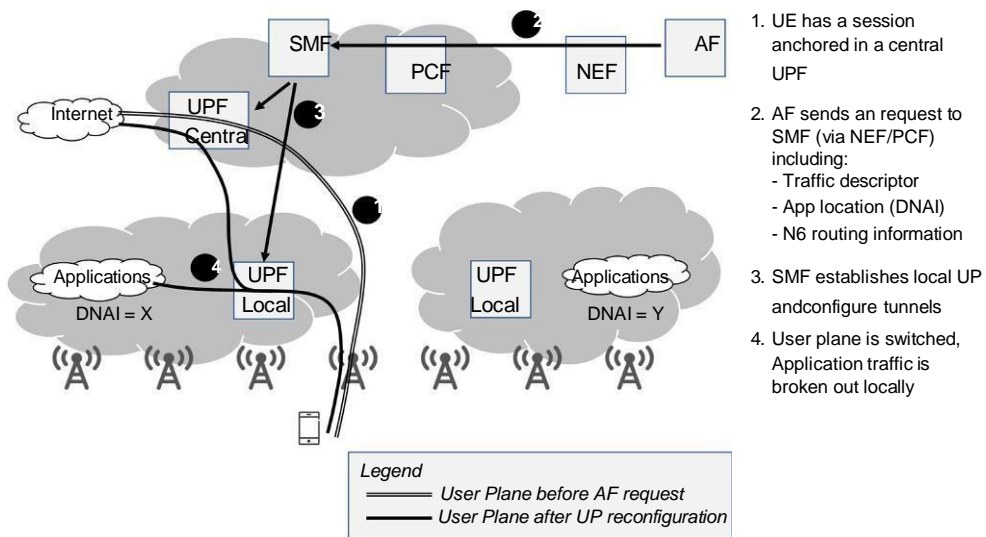


Fig. 6.10 Example use case for AF influence on traffic routing.

- A lot of work in the industry around Edge computing has been done on the application platform for edge applications and related APIs, e.g. by an ETSI Industry Specification Group called MEC (Multi-access Edge Computing).
- In 3GPP however, the focus when it comes to edge computing has so far been concerned with the access and connectivity aspects.
- 3GPP does not specify any special solutions or architecture for Edge computing. Instead 3GPP defines several general tools that can be used to provide an efficient User Plane path.
- These tools, most of which have already been described earlier in this chapter, are not specific to Edge computing but they can be used as enablers in deployments of Edge computing.

The main tools for UP path management are listed below.

- UPF selection
- Selective traffic routing to DN
- Session and Service Continuity (SSC) modes
- AF influence on traffic routing
- Network capability exposure
- LADN

Edge computing can of course also benefit from other general 5GS features such as differentiated QoS and charging.

6. Explain 5G protocols: 5G NAS

14.2 5G non-access stratum (5G NAS)

14.2.1 Introduction

NAS denotes the main Control Plane protocols between the UE and the core network.

The main functions of NAS are:

- Handling of UE registration and mobility, including generic functionality for access control such as connection management, authentication, NAS security handling, UE identification and UE configuration
- Support of Session Management procedures to establish and maintain PDU Session connectivity and QoS for the User Plane between the UE and the DN
- General NAS transport between UE and AMF to carry other types of messages that are not defined as part of the NAS protocol as such. This includes, e.g., transport of SMS, LPP protocol for location services, UDM data such as Steering of Roaming (SOR) messages, as well as UE policies (URSP).
- NAS consists of two basic protocols to support the functionality above; the 5GS Mobility Management (5GMM) protocol and the 5GS Session Management (5GSM) protocol.
- The 5GMM protocol runs between the UE and the AMF and is the basic NAS protocol used for handling UE registrations, mobility, security and also transport of the 5GSM protocol as well as the general NAS transport of other types of messages.
- The 5GSM protocol runs between UE and SMF (via the AMF) and supports management of PDU Session connectivity. It is carried on top of the 5GMM protocol as shown in Fig. 14.1. The 5GMM protocol is also used to transport information between UE and PCF, UE and SMSF, etc. also shown in Fig. 14.1.

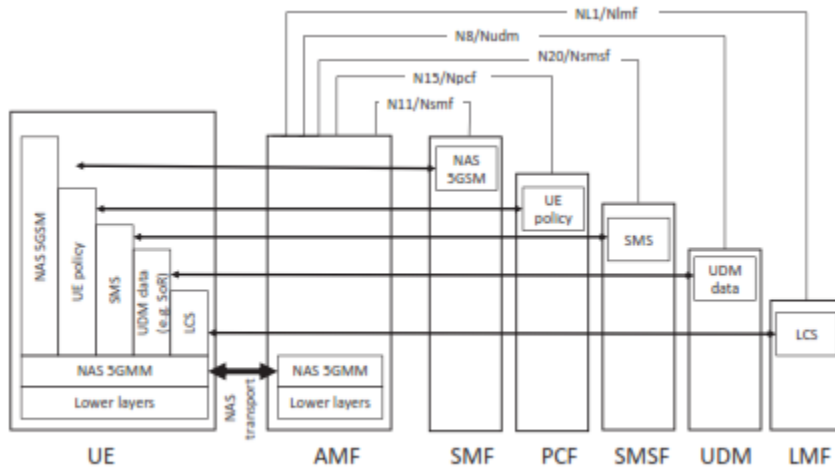


Fig. 14.1 NAS protocol stack with NAS-MM and NAS-MM protocols.

14.2.2 5G mobility management

- 5GMM procedures are used to keep track of the whereabouts of the UE, to authenticate the UE and control integrity protection and ciphering.
- The 5GMM procedures also allow the network to assign new temporary identities to the UE (5G-GUTI) and also request identity information (SUCI and PEI) from the UE.
- In addition, the 5GMM procedures provide the UE's capability information to the network and the network may also inform the UE about information regarding specific services in the network.
- The 5GMM protocol thus operates on a UE level (per Access Type) in contrast to the 5GSM protocol that is on a per PDU Session level. The 5GMM NAS signalling takes place between the UE and the AMF.

The basic 5GMM procedures are:

- Registration
- Deregistration
- Authentication
- Security mode control
- Service request
- Notification

14.2.3 5G session management

- 5GSM procedures are used to manage the PDU Sessions and QoS for the User Plane.
- This includes procedures for establishing and releasing PDU Sessions as well as modification of PDU Sessions to add, remove or modify QoS rules.
- The 5GSM procedures are also used to carry out the secondary authentication for a PDU Session.
- The 5GSM protocol thus operates on a PDU Session level in contrast to the 5GMM protocol that works on a UE level.

The basic 5GSM procedures are:

- PDU Session establishment
- PDU Session release
- PDU Session modification
- PDU Session authentication and authorization
- 5GSM status (to exchange PDU Session status information)

14.2.4 Message structure

- The NAS protocols are implemented as standard 3GPP L3 messages in accordance with [3GPP TS 24.007](#). Standard 3GPP L3 according to 3GPP TS 24.007 and its predecessors have also been used for NAS signalling messages in previous generations (2G, 3G, 4G).
- The encoding rules have been developed to optimize the message size over the air interface and to allow extensibility and backwards compatibility without the need for version negotiation.
- Each NAS message contains a Protocol Discriminator and a Message Type.
- The Protocol Discriminator is a value that indicates the protocol being used, i.e., for 5G NAS messages it is either 5GMM or 5GSM (to be precise, for 5G, an Extended Protocol Discriminator had to be defined as the available spare numbers of the original Protocol Discriminator was running out).
- The organization of a plain 5GMM NAS message is shown in [Fig. 14.2](#) and of a plain 5GSM message is shown in [Fig. 14.3](#).
- When an NAS message is security protected the plain NAS message is encapsulated as shown in [Fig. 14.4](#).
- This format applies to all 5GSM messages since they are always security protected. It also applies to security-protected 5GMM messages.
- In these security protected NAS messages, the first Extended Protocol Discriminator indicates that it is a 5GMM message since NAS security is part of the 5GMM NAS protocol.

- The plain NAS message inside the security protected NAS message has additional Extended Protocol Discriminator(s) that indicate whether it is a 5GMM or a 5GSM message.
- Further encapsulation may be done in the plain NAS message inside the security protected NAS message.
- The plain NAS message could, e.g., be a UL NAS transport (5GMM) message that contains a PDU Session Establishment request (5GSM) message.

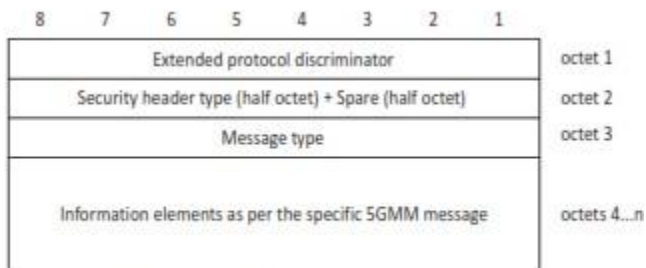


Fig. 14.2 Frame structure of plain 5GMM NAS message.

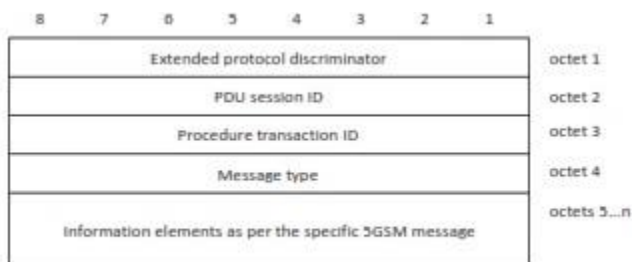


Fig. 14.3 Frame structure of plain 5GSM NAS message.

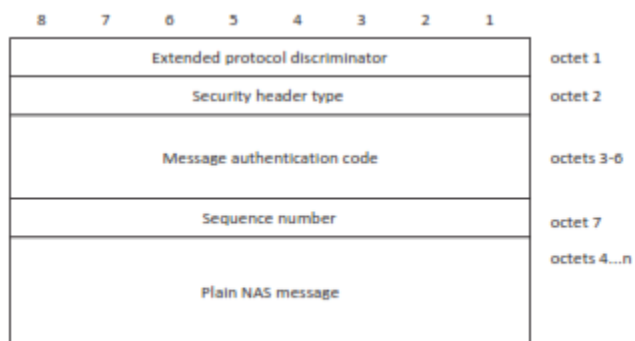


Fig. 14.4 Security protected NAS message.

7. Explain NG application protocol (NGAP)

14.3.1 Introduction

- The NGAP protocol is designed for use on the N2 interface between the (R)AN and AMF.
- It can be noted that the 3GPP RAN groups have given the name NG to the RAN-AMF interface that in the overall system architecture is called N2.
- The protocol name NGAP is thus derived from the interface name NG with the addition of AP (Application Protocol), which is a term that has been used many times by 3GPP to denote a signalling protocol between two network functions.

14.3.2 Basic principles

- NGAP supports all mechanisms necessary to handle the procedures between AMF and (R)AN, and it also supports transparent transport for procedures that are executed between the UE and the AMF or other core network functions.
- NGAP is applicable both to 3GPP access and non-3GPP accesses integrated with 5GC. This is a key difference to EPC where S1AP was designed for use only with 3GPP access (E-UTRAN) and not non-3GPP accesses.
- However, even though NGAP is applicable to any access, the design has been primarily targeted at 3GPP accesses (NG-RAN), which can also be noticed in the protocol specification defined in [3GPP TS 38.413](#). Support for specific parameters related to non-3GPP accesses have been added to the protocol when needed.

The NGAP interactions between AMF and (R)AN are divided into two groups:

- **Non UE-associated services:** These NGAP services are related to the whole NG interface instance between the (R)AN node and AMF. They are, e.g., used to establish the NGAP signalling connection between AMF and (R)AN, handle some overload situations and to exchange RAN and AMF configuration data.
- **UE-associated services:** These NGAP services are related to one UE. This NGAP signalling is thus related to procedures where a UE is involved, e.g., at Registration, PDU Session Establishment, etc.

The NGAP protocol supports the following functions:

- NG (i.e., N2) interface management functions, for example initial NG interface setup as well as Reset, Error Indication, Overload Indication and Load Balancing.
- Initial UE Context Setup functionality for establishment of an initial UE context in the (R)AN node.
- Provision of the UE capability information to the AMF (when received from the UE).

- Mobility functions for UEs in order to enable handover in NG-RAN, e.g., Path Switch request.
- Setup, modification, and release of PDU Session resources (User Plane resources)
- Paging, providing the functionality for 5GC to page the UE.
- NAS signalling transport functionality between the UE and the AMF
- Management of the binding between a NGAP UE association and a specific transport network layer association for a given UE
- Status transfer functionality (transfers PDCP Sequence Number status information from source NG-RAN node to target NG-RAN node (via AMF) in support of in-sequence delivery and duplication avoidance for handover).
- Trace of active UEs.
- UE location reporting and positioning protocol support.
- Warning message transmission.

14.3.3 NGAP elementary procedures

- NGAP consists of Elementary Procedures. An Elementary Procedure is a unit of interaction between the (R)AN (e.g., NG-RAN node) and AMF.
- These Elementary Procedures are defined separately and are intended to be used to build up complete sequences in a flexible manner.
- The Elementary Procedures may be invoked independently of each other as standalone procedures, which can be active in parallel.
- Some elementary procedures are specifically related to only Non UE-associated services (e.g., the NG Setup procedure) while others are related to only UE-associated services (e.g., the PDU Session Resource Modify procedure).
- Some elementary procedures may be using either Non UE-associated or UE-associated signalling depending on the scope and the context, e.g., the Error Indication procedure that uses UE-associated signalling if the error was related to a reception of a UE-associated signalling message, while it uses Non UE-associated signalling otherwise.
- In some cases, the independence between some Elementary Procedures is restricted; in this case the particular restriction is specified in the NGAP protocol specification.
- [Tables 14.3 and 14.4](#) list the elementary procedures in NGAP. Some of the procedures are request-response type of procedures, where the initiator gets a response from the receiver of the request, indicating whether the request was successfully handled or not.

Table 14.3 NGAP elementary procedures with a response to indicate success or failure (table based on Table 8.1-1 in 3GPP TS 38.413).

Elementary procedure	Initiating NGAP message	Successful outcome NGAP response message	Unsuccessful outcome NGAP response message
AMF configuration update	AMF configuration update	AMF configuration update acknowledge	AMF configuration update failure
RAN configuration update	RAN configuration update	RAN configuration update acknowledge	RAN configuration update failure
Handover cancellation	Handover cancel	Handover cancel acknowledge	
Handover preparation	Handover required	Handover command	Handover preparation failure
Handover resource allocation	Handover request	Handover request acknowledge	Handover failure
Initial context setup	Initial context setup request	Initial context setup response	Initial context setup failure
NG reset	NG reset	NG reset acknowledge	
NG setup	NG setup request	NG setup response	NG setup failure
Path switch request	Path switch request	Path switch request acknowledge	Path switch request failure
PDU session resource modify	PDU session resource modify request	PDU session resource modify response	
PDU session resource modify indication	PDU session resource modify indication	PDU session resource modify confirm	
PDU session resource release	PDU session resource release command	PDU session resource release response	
PDU session resource setup	PDU session resource setup request	PDU session resource setup response	
UE context modification	UE context modification request	UE context modification response	UE context modification failure
UE context release	UE context release command	UE context release complete	
Write-replace warning	Write-replace warning request	Write-replace warning response	
PWS cancel	PWS cancel request	PWS cancel response	
UE radio capability check	UE radio capability check request	UE radio capability check response	

Table 14.4 Elementary NGAP procedures without response (table based on Table 8.1-2 in 3GPP TS 38.413).

Elementary procedure	NGAP message
Downlink RAN configuration transfer	Downlink RAN configuration transfer
Downlink RAN status transfer	Downlink RAN status transfer
Downlink NAS transport	Downlink NAS transport
Error indication	Error indication
Uplink RAN configuration transfer	Uplink RAN configuration transfer
Uplink RAN status transfer	Uplink RAN status transfer
Handover notification	Handover notify
Initial UE message	Initial UE message
NAS non delivery indication	NAS non delivery indication
Paging	Paging
PDU session resource notify	PDU session resource notify
Reroute NAS request	Reroute NAS request
UE context release request	UE context release request
Uplink NAS transport	Uplink NAS transport
AMF status indication	AMF status indication
PWS restart indication	PWS restart indication
PWS failure indication	PWS failure indication
Downlink UE associated NRPPa transport	Downlink UE associated NRPPa transport
Uplink UE associated NRPPa transport	Uplink UE associated NRPPa transport
Downlink Non UE associated NRPPa transport	Downlink non UE associated NRPPa transport
Uplink non UE associated NRPPa transport	Uplink non UE associated NRPPa transport
Trace start	Trace start
Trace failure indication	Trace failure indication
Deactivate trace	Deactivate trace
Cell traffic trace	Cell traffic trace
Location reporting control	Location reporting control
Location reporting failure indication	Location reporting failure indication

Table 14.4 Elementary NGAP procedures without response (table based on Table 81-2 in).—cont'd

Elementary procedure	NGAP message
Location report	Location report
UE TNLA binding release	UE TNLA binding release request
UE radio capability info indication	UE radio capability info indication
RRC inactive transition report	RRC inactive transition report
Overload start	Overload start
Overload stop	Overload stop

8. Explain GPRS tunnelling protocol for the User Plane (GTP-U)

.14.7 GPRS tunnelling protocol for the User Plane (GTP-U)

- The two main components of GTP are the Control Plane part of GTP (GTP-C) and the User Plane part of GTP (GTP-U). GTP-C is the control protocol used in 3G/GPRS and 4G/EPS to control and manage PDN Connections and the User Plane tunnels that build up the User Plane path.
- The GTP-U uses a tunnel mechanism to carry the user data traffic and runs over UDP transport. In 5GS, GTP-U has been re-used to carry User Plane data over N3 and N9 (and N4) but the control protocol to manage the tunnel identities, etc.
- GTP-U tunnels are used between two corresponding GTP-U nodes to separate traffic into different communication flows.
- A local Tunnel Endpoint Identifier (TEID), the IP address, and the UDP port uniquely identify a tunnel endpoint in each node, where the TEID assigned by the receiving entity must be used for the communication.
- In 5GC, GTP-U tunnels are established by providing GTP-U TEIDs and IP addresses between (R)AN and SMF.
- This signalling is carried by HTTP/2 between SMF and AMF and by NGAP between AMF and (R)AN.
- There is thus no use of GTP-C in 5GC to manage GTP-U tunnels. The user plane protocol stack for a PDU Session is shown in [Fig. 14.15](#).
- A GTP path is identified in each node with an IP address and a UDP port number. A path may be used to multiplex GTP tunnels and there may be multiple paths between two entities supporting GTP.

- The TEID that is present in the GTP-U header indicates which tunnel a particular payload belongs to. Thus, packets are multiplexed and DE multiplexed by GTP-U between a given pair of Tunnel Endpoints. The GTP-U header is shown in Fig. 14.16.

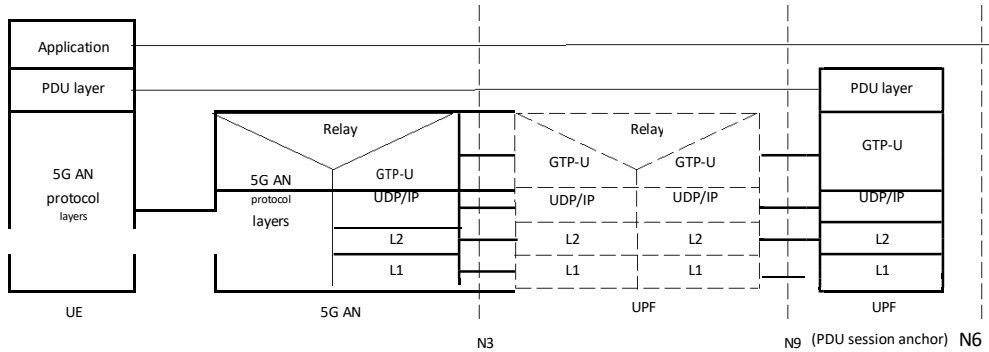


Fig. 14.15 User plane protocol stack for a PDU session.

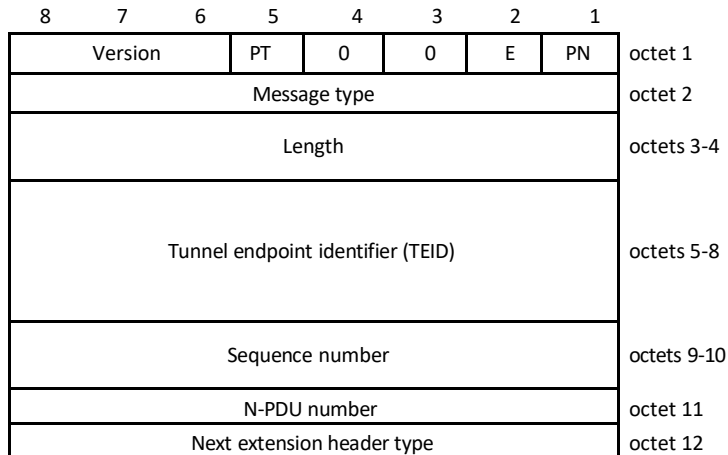


Fig. 14.16 GTP-U header.

9. Explain IP security (IPSec) Protocol (Nov/Dec 23)

IP security (IPSec)

Introduction

- IPsec is a very wide topic and many books have been written on this subject.
- It is not the intention or ambition of this chapter to provide a complete overview and tutorial on IPsec.
- Instead, we will give a high-level introduction to the basic concepts of IPsec focusing on the parts of IPsec that are used in 5GS.
- IPsec provides security services for both IPv4 and IPv6.
- It operates at the IP layer, offers protection of traffic running above the IP layer, and it can also be used to protect the IP header information on the IP layer.
- 5GS uses IPsec to secure communication on several interfaces, in some cases between nodes in the core network and in other cases between the UE and the core network.
- For example, IPsec is used to protect traffic in the core network as part of the NDS/IP framework. IPsec is also used between the UE and the N3IWF to protect NAS signalling and User Plane traffic.

IPsec overview

The set of security services provided by IPsec include:

- Access control
 - Data origin authentication
 - Connection-less integrity
 - Detection and rejection of replays
 - Confidentiality
 - Limited traffic flow confidentiality.
- By access control we mean the service to prevent unauthorized use of a resource such as a particular server or a particular network.
 - The data origin authentication service allows the receiver of the data to verify the identity of the claimed sender of the data.
 - Connectionless integrity is the service that ensures that a receiver can detect if the received data has been modified on the path from the sender.
 - However, it does not detect if the packets have been duplicated (replayed) or reordered.
 - Data origin authentication and connectionless integrity are typically used together.
 - Detection and rejection of replays is a form of partial sequence integrity, where the receiver can detect if a packet has been duplicated. Confidentiality is the service that protects the traffic from being read by unauthorized parties.

- The mechanism to achieve confidentiality with IPsec is encryption, where the content of the IP packets is transformed using an encryption algorithm so that it becomes unintelligible.
- Limited traffic flow confidentiality is a service whereby IPsec can be used to protect some information about the characteristics of the traffic flow, e.g., source and destination addresses, message length, or frequency of packet lengths.
- In order to use the IPsec services between two nodes, the nodes use certain security parameters that define the communication, such as keys, encryption algorithms, and so on.
- In order to manage these parameters, IPsec uses Security Associations (SAs). A SA is the relation between the two entities, defining how they are going to communicate using IPsec.
- A SA is unidirectional, so to provide IPsec protection of bidirectional traffic a pair of SAs is needed, one in each direction.
- Each IPsec SA is uniquely identified by a Security Parameter Index (SPI), together with the destination IP address and security protocol (AH or ESP; see below).
- The SPI can be seen as an index to a Security Associations data-base maintained by the IPsec nodes and containing all SAs.
- As will be seen below, the IKE protocol can be used to establish and maintain IPsec SAs.
- IPsec also defines a nominal Security Policy Database (SPD), which contains the policy for what kind of IPsec service is provided to IP traffic entering and leaving the node.
- The SPD contains entries that define a subset of IP traffic, for example using packet filters, and points to a SA (if any) for that traffic.

Encapsulated Security Payload and Authentication Header

- IPsec defines two protocols to protect data, the Encapsulated Security Payload (ESP) and the Authentication Header (AH).
- ESP can provide integrity and confidentiality while AH only provides integrity.
- Another difference is that ESP only protects the content of the IP packet (including the ESP header and part of the ESP trailer), while AH protects the complete IP packet, including the IP header and AH header. See [Figs. 14.18 and 14.19](#) for illustrations of ESP- and AH-protected packets.
- The fields in the ESP and AH headers are briefly described below. ESP and AH are typically used separately but it is possible, although not common, to use them together. If used together, ESP is typically used for confidentiality and AH for integrity protection.

- The SPI is present in both ESP and AH headers, and is a number that, together with the destination IP address and the security protocol type (ESP or AH), allows the receiver to identify the SA to which the incoming packet is bound.
- The Sequence number contains a counter that increases for each packet sent. It is used to assist in replay protection.
- The Integrity Check Value (ICV) in the AH header and ESP trailer contains the cryptographically computed integrity check value.
- The receiver computes the integrity check value for the received packet and compares it with the one received in the ESP or AH packet.
- ESP and AH can be used in two modes: transport mode and tunnel mode. In transport mode ESP is used to protect the payload of an IP packet.
- The Data field as depicted in Fig. 14.17 would then contain, for example, a UDP or TCP header as well as the application data carried by UDP or TCP.
- See Fig. 14.20 for an illustration of a UDP packet that is protected using ESP in transport mode. In tunnel mode, on the other hand, ESP and AH are used to protect a complete IP packet.
- The Data part of the ESP packet in Fig. 14.17 now corresponds to a complete IP packet, including the IP header. See Fig. 14.21 for an illustration of a UDP packet that is protected using ESP in tunnel mode.
- Transport mode is often used between two endpoints to protect the traffic corresponding to a certain application.
- Tunnel mode is typically used to protect all IP traffic between security gateways or in VPN connections where a UE connects to a secure network via an unsecure access.

Internet key exchange

- In order to communicate using IPsec, the two parties need to establish the required IPsec SAs.
- This can be done manually by simply configuring both parties with the required parameters. However, in many scenarios a dynamic mechanism for authentication, key generation, and IPsec SA generation is needed.
- This is where Internet Key Exchange (IKE) comes into the picture. IKE is used for authenticating the two parties and for dynamically negotiating, establishing, and maintaining SAs. (One could view IKE as the creator of SAs and IPsec as the user of SAs.) There are in fact two versions of IKE: IKE version 1 (IKEv1) and IKE version 2 (IKEv2).

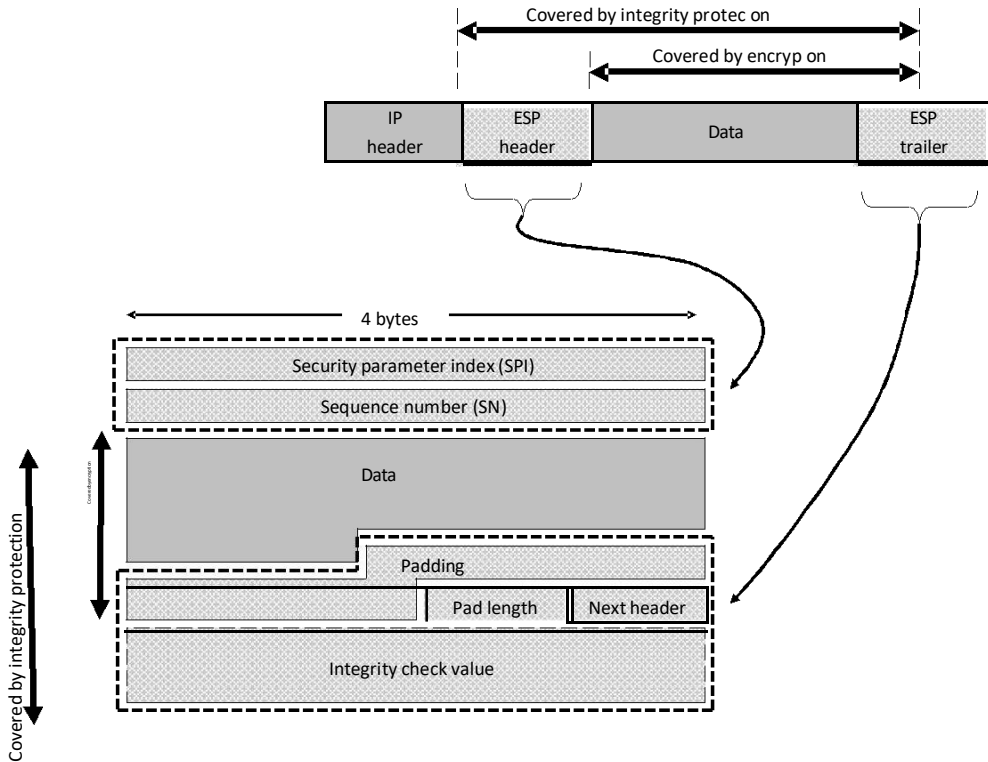


Fig. 14.18 IP packet (data) protected by ESP.

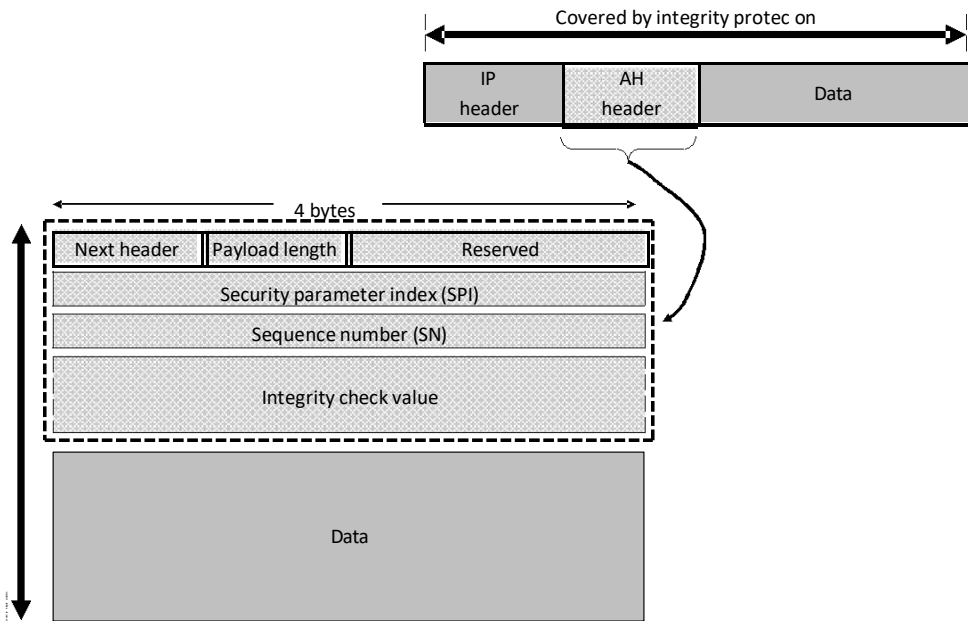


Fig. 14.19 IP packet (data) protected by AH.

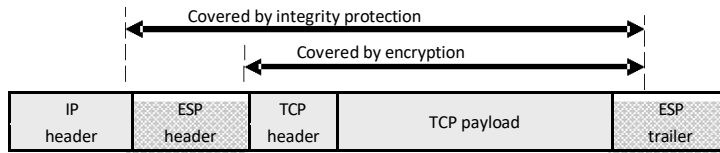


Fig. 14.20 Example of IP packet protected using ESP in transport mode

10. Explain Generic routing encapsulation (GRE)

Generic routing encapsulation (GRE)

Introduction

- The GRE is a protocol designed for performing tunnelling of a network layer protocol over another network layer protocol.
- It is generic in the sense that it provides encapsulation of one arbitrary network layer protocol (e.g., IP or MPLS) over another arbitrary network layer protocol.
- This is different from many other tunnelling mechanisms, where one or both of the protocols are specific, such as IPv4-in-IPv4 2003 or Generic Packet Tunnelling over IPv6.
- GRE is also used for many different applications and in many different network deployments outside the telecommunications area.
- It is not the intention of this book to discuss aspects for all those scenarios. Instead, we focus on the properties of GRE that are most relevant to 5GS.

Basic protocol aspects

- The basic operation of a tunnelling protocol is that one network protocol, which we call the payload protocol, is encapsulated in another delivery protocol.
- It should be noted that encapsulation is a key component of any protocol stack where an upper layer protocol is encapsulated in a lower layer protocol.
- This aspect of encapsulation, however, should not be considered as tunnelling. When tunnelling is used, it is often the case that a layer-3 protocol such as IP is encapsulated in a different layer-3 protocol or another instance of the same protocol. The resulting protocol stack may look like that shown in Fig. 14.26.

We use the following terminology:

- **Payload packet and payload protocol:** The packet and protocol that needs to be encapsulated (the three topmost boxes in the protocol stack in Fig. 14.26).
- **Encapsulation (or tunnel) protocol:** The protocol used to encapsulate the payload packet, i.e., GRE (the third box from the bottom in Fig. 14.26).

- **Delivery protocol:** The protocol used to deliver the encapsulated packet to the tunnel endpoint (the second box from the bottom in Fig. 14.26).

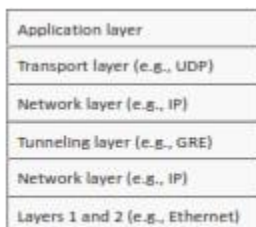


Fig. 14.26 Example of protocol stack when GRE tunneling is used.

- The basic operation of GRE is that a packet of protocol A (the payload protocol) that is to be tunnelled to a destination is first encapsulated in a GRE packet (the tunnelling protocol).
- The GRE packet is then encapsulated in another protocol B (the delivery protocol) and sent to the destination over a transport network of the delivery protocol.
- The receiver then decapsulates the packet and restores the original payload packet of protocol type.
- In 5GS, GRE is primarily used to carry the packets (PDUs) between UE and N3IWF.
- GRE here allows the QFI value and the RQI indicator for reflective QoS to be carried in the GRE header together with the encapsulated PDU. The QFI and RQI are included in the GRE key field (see below). Fig. 14.27 shows an example of an PDU carried in a GRE tunnel between UE and N3IWF over an IP delivery protocol.

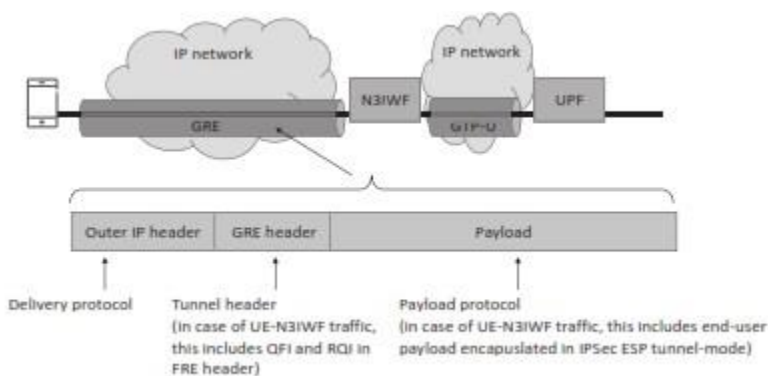


Fig. 14.27 Example of GRE tunnel between two network nodes with IPv4 delivery protocol.

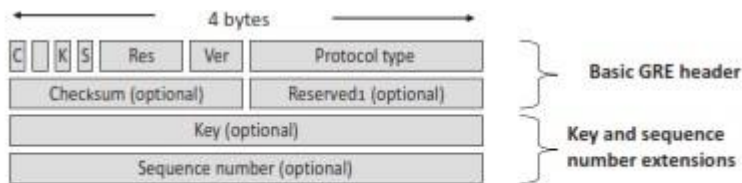


Fig. 14.28 GRE header format including the basic header as well as the key and sequence number extensions.

14.11.3 GRE packet format

The GRE header format is illustrated in Fig. 14.28.

- The C flag indicates whether the Checksum and Reserved1 fields are present. If the C flag is set, the Checksum and Reserved1 fields are present.
- In this case the Checksum contains a checksum of the GRE header as well as the payload packet.
- The Reserved1 field, if present, is set to all zeros. If the C flag is not set, the Checksum and Reserved1 fields are not present in the header.
- The K and S flags respectively indicate whether or not the Key and/or Sequence number is present.
- The Protocol Type field contains the protocol type of the payload packet. This allows the receiving endpoint to identify the protocol type of the decapsulated packet.
- The intention of the Key field is to identify an individual traffic flow within a GRE tunnel.
- GRE in itself does not specify how the two endpoints establish which Key field(s) to use.
- This is left to implementations or is specified by other standards making use of GRE. The Key field could, for example, be statically configured in the two endpoints, or be dynamically established using some signalling protocol between the endpoints.
- In 5GS the key field is used between UE and N3IWF to carry the QFI value and the RQI. The QFI takes 6 bits and RQI a single bit out of the available 32 bits in the key field.
- The Sequence number field is used to maintain the sequence of packets within the GRE tunnel.
- The node that performs the encapsulation inserts the Sequence number and the receiver uses it to determine the order in which the packets were sent.

Part – A Two marks

1. What is network slicing?

A Network Slice is realizing a complete network for any type of access and is an enabler for providing services. The used physical or virtual infrastructure resources may be dedicated to the Network Slice or shared with other Network Slices.

2. List the benefits of network slicing.

- ✓ Better customer experience by per customer adaptations and optimizations
- ✓ Shorter time-to-market and time-to-customer
- ✓ Simpler resource management
- ✓ Increased automation
- ✓ Flexibility and agility
- ✓ Reduced risks by separation of concerns.

3. What are the different requirements to be addressed by a Network Slice?

- ✓ Traffic capacity requirements per geographical area
- ✓ Charging requirement
- ✓ Coverage area requirement
- ✓ Degree of isolation requirement
- ✓ End-to-end latency requirement
- ✓ Mobility requirement
- ✓ Overall user density requirement
- ✓ Priority requirement and etc.

4. What is Management and orchestration?

During the preparation and whole Lifecycle management process, the customer is able to provide its requirements using APIs from which the customer gets information of how the Network Slices perform, and is able to modify its requirements as to adapt to the needs of the customer. It provides a high-level view of the process in the preparation and the Lifecycle management of a network Slice Instance (NSI).

5. What is network slice preparation?

Network Slice “blueprints” or “templates” are used to simplify the process. If a Network Slice template exists that meets the customer requirements, then the preparation process can be shortened, as either the customer may be able to use an existing NSI.

6. What is network slice commissioning?

NSI provisioning in the commissioning phase includes creation of the NSI. During NSI creation all needed resources are allocated and configured to satisfy the Network Slice requirements.

7. What is network slice decommissioning?

NSI provisioning in the decommissioning phase includes decommissioning of non-shared resources if required and removing the NSI specific configuration from the shared resources. After the decommissioning phase, the NSI is terminated and does not exist anymore.

8. What is Multi-access Edge Computing (MEC)?

Multi-Access Edge Computing (MEC) is an integral part of the 5G ecosystem. MEC helps service providers to bring application oriented capabilities closer to the users and support several latency sensitive use cases from the edge.

9. What are the needs for MEC?

- ✓ Growth in the number of mobile devices connecting to the network (with IOT, it is expected to explode even further)
- ✓ Growth in the volume of data generated by the Over the Top (OTT) applications such as social media, video streaming and online gaming.
- ✓ Need for distributing the infrastructure where the application services are hosted in a service provider network, to improve the application performance and the user experience etc.

10. What are the MEC architecture functions?

- ❖ MEC Orchestrator
- ❖ MEC Platform
- ❖ MEC Platform Manager
- ❖ Virtualization Infrastructure
- ❖ MEC Application Services

11. What is the use of Virtualization Infrastructure?

Virtualized infrastructure provides shared compute, storage and networking resources for hosting MEC related applications or Virtual Network Functions (VNFs).

12. What MEC Breakout mode?

The session connection is redirected to a MEC application which is either hosted locally on the MEC platform or on a remote server. Some examples of breakout applications are local Content Deliver Network (CDN) caches (for ex., Akamai caches), gaming services and media delivery services (for ex., Netflix streaming). Normally, you achieve this by defining forwarding policies

13. What is MEC In-line mode ?

MEC is deployed transparently, in an in-line mode. The session connection is maintained with the original server, while all the traffic traverses and goes through the application running in the MEC. Examples of in-line MEC applications are transparent content caching and security applications.

14. What is MEC Tap mode?

In Tap mode, data exchanged in a session is selectively duplicated and forwarded to the tap MEC application. Some examples of tap mode applications are virtual network probes, monitoring and security applications.

15. **What is MEC Independent mode** - MEC application and services run independently, but still the MEC application is registered in the MEC platform and will receive other MEC services, such as DNS and radio network information (for ex., radio bearer statistics). Steering the traffic to the MEC is achieved by configuring local DNS or MEC host's data plane.

16. What is meant by service continuity?

In any telecom technology one of the key feature requirement is data session and service continuity to ensure the uninterrupted service experience to the users irrespective of any change of UE IP address or change in Core network anchor point.

17. Give the types of SSC modes?

- SSC mode 1
- SSC mode 2
- SSC mode 3

18. List the advantages of 5G SA network deployment.

To make better use of the competence of 5G, we just need to upgrade the existing EPC to complete the interconnection between EPC and 5GC with this solution.

19. List the advantages of 5G NSA network deployment.

There is no need to deploy 5GC in early period of 5G deployment. By implementing this network architecture, the control interface of S1 is only provided to LTE eNodeB, while the user plane interface of S1 can be provided to eNodeB and NR gNodeB at the same time.

20. List the advantages of 5G NR network deployment.

With the requirement of 5G NR providing continuous coverage, the control plane always set in 5G NR to fully achieve massive and strong function for users. The rapid rising of 5GC has become an irresistible tendency.

21. What edge computing?

Edge computing is about bringing the services closer to the location where they are to be delivered. Services here includes computing power and memory needed for e.g. running a requested application.

22. Mention the 5G protocols.

- 5G NAS
- NGAP
- GTP-U
- IPSec
- GRE

23. What is 5G mobility management?

5GMM procedures are used to keep track of the whereabouts of the UE, to authenticate the UE and control integrity protection and ciphering. The 5GMM procedures also allow the network to assign new temporary identities to the UE (5G-GUTI) and also request identity information (SUCI and PEI) from the UE.

24. What are the basic 5GMM procedures?

- Registration
- Deregistration
- Authentication
- Security mode control
- Service request
- Notification

25. What are the set of security services provided by IPsec?

- Access control
- Data origin authentication
- Connection-less integrity
- Detection and rejection of replays
- Confidentiality
- Limited traffic flow confidentiality.

26. Give an overview of GRE Protocol aspects. (Nov/Dec 23)

The GRE is a protocol designed for performing tunnelling of a network layer protocol over another network layer protocol.

It is generic in the sense that it provides encapsulation of one arbitrary network layer protocol (e.g., IP or MPLS) over another arbitrary network layer protocol.

27. Write the main functions of the 5G non-access stratum (NAS)? (Nov/Dec 23)

Handling of UE registration and mobility, including generic functionality for access control such as connection management, authentication, NAS security handling, UE identification and UE configuration

4G/5G COMMUNICATION NETWORKS

UNIT – IV DYNAMIC SPECTRUM MANAGEMENT AND MM-WAVES

Mobility management, command and control, spectrum sharing and spectrum trading, cognitive radio based on 5G, millimetre waves.

1. Explain Mobility management. (Nov/Dec 23)

The mobility is a core feature of 5GS. Mobility Management is required to ensure the following:

- That the network can “reach” the user, for example to notify the user about incoming messages and calls,
- That a user can initiate communication toward other users or services such as Internet access, and that connectivity and ongoing sessions can be maintained as the user moves, within or between access technologies.

As a result, there are several optional 5GS mobility management related functions that differ from previous 3GPP systems:

- Service Area Restriction: mobility with session continuity is controlled at UE level at certain areas
- Local Area Data Network (LADN): mobility with session continuity is controlled at PDU Session level making communication available at certain areas
- Mobile Initiated Connection Only (MICO): paging capability (as part of the mobility service) is optional.

5G Mobility Management (5GMM) related procedures are divided into three categories depending on the purpose of the procedure and on how they can be initiated:

1. Common procedures; can always be initiated when the UE is in CM-CONNECTED state.
2. Specific procedures; only one UE initiated specific procedure can be running for each of the Access Types.

3. Connection management procedures; used to establish a secure signalling connection between the UE and the network, or to request the resource reservation for sending data, or both.

7.2 Establishing connectivity

7.2.1 Network discovery and selection

- Network discovery and selection procedures for 5GS do not differ much from EPS and the principles used when a 3GPP access type is selected, have been maintained.
- Before the UE can receive and use the services and capabilities from the 5GS, e.g. Session Management services from SMF, the UE needs to establish a connection to the 5GS.
- To achieve this, the UE first selects a network/PLMN and a 5G-AN. For 3GPP access i.e. NG-RAN, the UE selects a cell, then the UE establishes a RRC connection to the NG-RAN.
- Based on the content (e.g. selected PLMN, Network Slice information) provided by the UE in establishing the RRC connection the NG-RAN selects an AMF and forwards the UE NAS MM message to the AMF in the 5GC using the N2 reference point.
- Using the AN connection (i.e. RRC connection) and the N2, the UE and the 5GS complete a Registration procedure. Once the Registration procedure is completed, the UE is registered in the 5GC i.e. the UE is known and the UE has a NAS MM connection to the AMF, the UE's entry point to the 5GC, which is used as the NAS connection to the 5GC.
- Further communication between the UE and other entities in the 5GC uses the established NAS connection as NAS transport from that point forward.
- To save resources the NAS connection is released while the UE is still registered and known in the 5GC, i.e. to re-establish the NAS connection the UE or 5GC initiates a Service Request procedure.

7.2.2 Registration and Mobility

- Idle-mode Mobility Management for 5GS using NR and E-UTRA is built on similar concepts to LTE/E-UTRAN (EPS), GSM/WCDMA, and CDMA.
- Radio networks are built by cells that range in size from tens and hundreds of meters to tens of kilometers and the UE updates the network about its location on a regular basis.

- It is not practical to keep track of a UE in idle mode every time it moves between different cells due to the amount of signalling it would cause, nor to search for a UE across the entire network for every terminating event (e.g. an incoming call).
- In order to create efficiencies, therefore, cells are grouped together into Tracking Areas (TA), and one or more Tracking Areas may be assigned to the UE as a Registration Area (RA).
- RA is used as a base for the network to search for the UE and for the UE to report its location.

A summary of the idle mobility procedure in 5GS is:

- A TA consists of a set of cells,
- The Registration Area in 5GS is a list of one or more Tracking Areas (TA list),
- The UE performs Registration update due to mobility when moving outside its Registration Area i.e. TA list,
- The UE in idle state also performs periodic Registration update when the periodic Registration update timer expires.

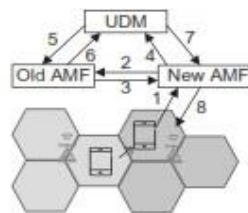


Fig. 7.1 Mobility Registration Update procedure.

7.2.3 Cellular connected mode mobility

- Great effort has been put into optimized connected mode mobility for cellular systems.
- The basic concept is somewhat similar across different technologies with some variations in the functional distribution between UE and networks.

- While in connected mode, the UE has a connected signalling connection and zero, one or more connected user plane resources, and data transmission may be ongoing.
- To limit interference and provide the UE with good data communication, the UE changes cells through handover when there is a cell that can provide better service than the cell that the UE is currently using.
- To save on complexity in the UE design and power, the systems are designed to ensure that the UE only needs to listen to a single gNB/ng-eNB at a time.
- In addition, for inter-RAT handover (e.g. NR to E-UTRAN HO) the UE only needs to have a single radio technology connected at a time.

7.3 Reachability

7.3.1 Paging

- Paging is used to search for Idle UEs and establish a signalling connection. Paging is, for example, triggered by downlink packets arriving to the UPF.
- When the UPF receives a downlink packet destined for an Idle UE, it does not have an NG-RAN User Plane tunnel address to which it can send the packet.
- The UPF instead buffers the packet and informs the SMF that a downlink packet has arrived.
- The SMF asks the AMF to setup User Plane resources for the PDU Session, and the AMF which knows in which RA the UE is located and sends a paging request to the NG-RAN within the RA.
- The NG-RAN calculates at which occasion the UE is to be paged using parts of the UE's 5G-S-TMSI (10 bits) as input, and then the NG-RAN pages the UE.

7.3.2 Mobile Initiated Connection Only (MICO) mode

- Mobile Initiated Connection Only (MICO) mode was introduced to allow paging resources to be saved for UEs that don't need to be available for Mobile Terminating communication.
- When the UE is in MICO mode, the AMF considers the UE as unreachable when the UE is in CM-IDLE state.

- MICO mode is negotiated (and re-negotiated) during Registration procedures, i.e. the UE may indicate its preference for MICO mode and the AMF decides whether MICO mode can be enabled taking into account the UE's preference as well as other information such as the user's subscription and network policies.

7.3.3 UE's reachability and location

- 5GS also supports location services in a similar way to EPS, but 5GS also provides the possibility for any authorized NF (e.g. SMF, PCF or NEF) in the 5GC to subscribe to UE mobility related event reporting.
- The NF subscribing to a UE mobility related event can do so by providing the following information to the AMF:
 - Whether UE location or the UE mobility in relation to an area of interest is to be reported
 - In case an area of interest is requested, then the NF specifies the area as:
List of Tracking Areas, list of cells or list of NG-RAN nodes.
If the NF wants to get an LADN area, the NF (e.g. SMF) provides the LADN DNN to refer the LADN service area as the area of interest.
If a Presence Reporting Area is requested as area of interest, then the NF (e.g. SMF or PCF) may provide an identifier to refer to a predefined area configured in the AMF.
 - Event Reporting Information: event reporting mode (e.g., periodic reporting), number of reports, maximum duration of reporting, event reporting condition (e.g. when the target UE moved into a specified area of interest).
 - The notification address i.e. address of NF that the AMF is to provide the notifications which can be another NF than the NF subscribing to the event
 - The target of event reporting that indicates a specific UE, a group of UE(s) or any UE (i.e. all UEs).

7.4 Additional MM related concepts

7.4.1 RRC Inactive

- It includes support for efficient communication with minimal signalling by using a concept called RRC Inactive which affects the UE, NG-RAN and 5GC.
- RRC Inactive is a state where a UE remains in CM-CONNECTED state (i.e. at NAS level) and can move within an area configured by NG-RAN (the RAN Notification Area – RNA) without notifying the network.
- The RNA is a subset within the RA allocated by the AMF. When the UE is in RRC Inactive state the following applies:
 - UE reachability is managed by the NG-RAN, with assistance information from 5GC;
 - UE paging is managed by the NG-RAN;
 - UE monitors for paging with part of the UE's 5GC (5G S-TMSI) and NG-RAN identifier.
- In RRC Inactive, the last serving NG-RAN node keeps the UE context and the UE-associated NG (N2 and N3) connections with the serving AMF and UPF. Therefore, there is no need for the UE to signal toward the 5GC before sending User Plane data.

7.5 N2 management

- In EPS, when a UE attaches to EPC and is assigned a 4G-GUTI, the 4G-GUTI is associated to a specific MME and if there is a need to move the UE to another MME the UE needs to be updated with a new 4G-GUTI.

- This may be a drawback e.g. if the UE is using some power saving mechanism or if a large amount of UEs are to be updated at the same time. With 5GS and N2 there is support for moving one or multiple UEs to another AMF without immediately requiring updating the UE with a new 5G-GUTI.
- The 5G-AN and the AMF are connected via a Transport Network Layer that is used to transport the signalling of the NGAP messages between them. The transport protocol used is SCTP.
- The SCTP endpoints in the 5G-AN and the AMF sets up SCTP associations between them that are identified by the used transport addresses.
- An SCTP association is generically called a Transport Network Layer Association (TNLA).

The N2 (also called NG in RAN3 specifications) reference point between the 5G-AN and the 5GC (AMF) supports different deployments of the AMFs e.g. either

(1) An AMF NF instance which is using virtualization techniques such that it can provide the services toward the 5G-AN in a distributed, redundant, stateless, and scalable manner and that it can provide the services from several locations, or

(2) An AMF Set which uses multiple AMF NF instances within the AMF Set and the multiple AMF Network Functions are used to enable the distributed, redundant, stateless, and scalable characteristics.

- The N2 reference point supports a form of self-automated configuration. During this type of configuration the 5G-AN nodes and the AMFs exchange NGAP information of what each side supports e.g. the 5G-AN indicates supported TAs, while the AMF indicates supported PLMN IDs and served GUAMIs.
- The exchange is performed by the NG SETUP procedure and, if updates are required, the RAN or AMF CONFIGURATION UPDATE procedure. The AMF CONFIGURATION UPDATE procedure can also be used to manage the TNL associations used by the 5G-AN.

7.5.1 AMF management

- The 5GC, including N2, supports the possibility to add and remove AMFs from AMF Sets.
- Within 5GC, the NRF is updated (and DNS system for interworking with EPS) with new NFs when they are added, and the AMF's NF profile includes which GUAMI(s) the AMF handles.
- For a GUAMI there may also be one or more backup AMF registered in the NRF (e.g. to be used in case of failure or planned removal of an AMF).
- A planned removal of an AMF can be done either through the AMF storing the registered UEs' contexts in a UDSF (Unstructured Data Storage Function), or with the AMF deregistering itself from the NRF, in which case the AMF notifies the 5G-AN that the AMF will be unavailable for processing transactions for the GUAMI(s) configured on this AMF.
- Additionally, the AMF can initially decrease the load by changing the weight factor for the AMF toward the 5G-AN, e.g. setting it to zero, causing the 5G-AN to select other AMFs within the AMF Set for new UEs entering the area.

7.5.2 5GC assistance for RAN optimizations

- As the UE context information is not kept in the NG-RAN when the UE transition to RRC-IDLE, it may be hard for the NG-RAN to optimize the logic related to the UE as UE specific behavior is unknown unless the UE has been in RRC-CONNECTED state for some time.
- There are NG-RAN specific means to retrieve such UE information e.g. UE history information can be transferred between NG-RAN nodes.
- To further facilitate an optimized decision in NG-RAN e.g. for UE RRC state transition, CM state transition decision and optimized NG-RAN strategy for RRC-INACTIVE state, the AMF may provide 5GC assistance information to NG-RAN.
- 5GC has a better method to store UE related information for a longer time and a means to retrieve information from external entities through external interfaces. When calculated by the 5GC (AMF) the algorithms used and related criteria, and the decision when it is considered suitable and stable to send to the NG-RAN are vendor specific.
- Therefore, along with the assistance information sent to NG-RAN, it often is accompanied with the information whether it is derived by statistics or retrieved via subscription information (e.g. set by agreements or via an API).

5GC assistance information is divided into 3 parts:

- Core Network assisted RAN parameters tuning;
 - Core Network assisted RAN paging information;
 - RRC Inactive Assistance Information.
- Core Network assisted RAN parameters tuning provides the NG-RAN with a way to understand UE behavior so as to optimize NG-RAN logic e.g. how long to keep the UE in specific states.
- Besides the content listed in Table 7.4, the 5GC also provides the source of the information e.g. if it is subscription information or derived based on statistics.

7.5.3 Service Area and Mobility Restrictions

Mobility Restrictions enables the network, mainly via subscriptions, to control the Mobility Management of the UE as well as how the UE accesses the network. Similar logic as used in EPS is applied in 5GS, but with some new functionality added as well.

The 5GS supports the following:

- RAT restriction:

Defines the 3GPP Radio Access Technologies a UE is not allowed to access in a PLMN and may be provided by the 5GC to the NG-RAN as part of the Mobility Restrictions. The RAT restriction is enforced by the NG-RAN at connected mode mobility.

- Forbidden Area:

A Forbidden Area is an area in which the UE is not permitted to initiate any communication with the network for the PLMN.

- Core Network type restriction:

Defines whether UE is allowed to access to 5GC, EPC or both for the PLMN.

- **Service Area Restriction:**

Defines areas controlling whether the UE is allowed to initiate communication for services as follows:

- **Allowed Area:** In an Allowed Area, the UE is permitted to initiate communication with the network as allowed by the subscription.
- **Non-Allowed Area:** In a Non-Allowed Area a UE is “service area restricted” meaning that neither the UE nor the network is allowed to initiate signalling to obtain user services (both in CM-IDLE and in CM-CONNECTED states).

The UE performs mobility related signalling as usual, e.g. mobility Registration updates when moving out of the RA. The UE in a Non-Allowed Area replies to 5GC initiated messages, which makes it possible to inform the UE that e.g. the area is now allowed.

7.6 Control of overload

- 5GS supports the ability to control the amount of load UEs produce toward the 5GS through different mechanisms.
- Mechanisms for 5GC to balance load across NFs and also to scale the amount of resources consumed for the NFs are often enough to handle normal fluctuations of load impacting the 5GC.
- In order to protect itself from overload situations, the 5GC supports a number of mechanisms including instructing UEs to back-off through NAS back-off timers (for Mobility Management as well as Session Management messages) such that the UE does not re-attempt to connect while the back-off timer is running.
- 5GS also supports the possibility to indicate to the NG-RAN that the load toward the AMF needs to be reduced using different criteria in an NGAP Overload Start message sent to the NG-RAN.
- Different methods can be used to handle possible bottlenecks in the NG-RAN Control Plane which also protects the 5GC. The mechanism used often depends on the load in the system. These are summarized below.

- Congestion in control channel resources: 5QI-based scheduling controls cases when e.g. the number of users awaiting scheduling exceeds the number of users that can be admitted such that the random access procedure fails.
- The random access procedure is a lower layer procedure used when the UE wants to initiate communication e.g. the UE gets synchronized with the network from a timing perspective.
- Congestion in random access channel (RACH) resources: random access back-off. This pushes some UEs into a longer back-off. This is when there are so many access attempts on the RACH that the UE provided preambles cannot be detected anymore.
- Release/reject UE RRC connection: If there are not enough resources to process RRC connection requests, Releasing RRC connection or rejecting RRC connection attempts can be used.

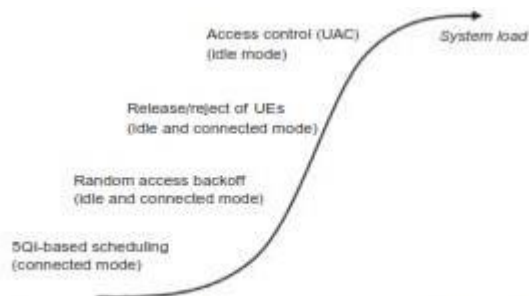


Fig. 7.6 Access and congestion control mechanisms as a function of the system load.

7.6.1 Unified Access Control

- EPS supports multiple variants of access barring mechanisms as they were developed in different releases to address different needs for congestion control.
- The 5GS supports one mechanism called Unified Access Control (UAC) which is extensible, flexible (e.g. each operator can define their own category when to apply access control) and supports a variety of scenarios.
- The UAC affects UEs in all RRC states i.e. RRC_IDLE, RRC_INACTIVE and RRC_CONNECTED state.
- In the case of multiple 5GC sharing the same NG-RAN, the NG-RAN provides UAC for each PLMN individually.

2. Explain Interworking with EPC

7.8 Interworking with EPC

7.8.1 General

- To ensure successful interworking with the appropriate EPS functionality, only one PGW-C + SMF is allocated per APN for a given UE, and that is enforced e.g. by the HSS + UDM providing one PGW-C + SMF FQDN per APN to the MME.
- Interworking with EPC while using non-3GPP access in 5GS is also applicable and in such cases, NR would be replaced with N3IWF and access specific entities underneath e.g. Wi-Fi Access Point.
- Furthermore, it is also possible to interwork between EPC connected to non-3GPP while using 3GPP access toward the 5GC, and in such case the MME and SGW would be replaced with an ePDG and the HSS with a 3GPP AAA server.
- For interworking to be possible it is required that the UE supports both EPC NAS procedures as well as 5GC NAS procedures.

- If this is not the case, then the UE will be directed toward the Core Network that the UE supports, and no interworking will be applicable.

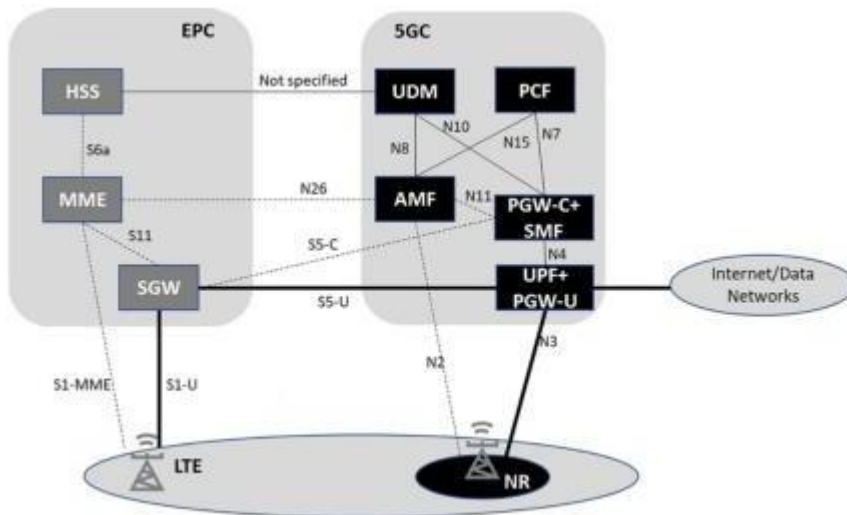


Fig. 7.7 Detailed architecture for interworking between EPC and 5GC.

3. Explain Interworking with EPC using 3GPP access

7.8.2 Interworking with EPC using 3GPP access

7.8.2.1 General

- When a UE is selecting networks – or PLMNs – or camping on a cell that is connected to both EPC and 5GC, the UE needs to select which Core Network to register with.
- That decision can be operator controlled or user controlled.

- The operator can control the decision e.g. by influencing the network selection using an operator controlled prioritized list in the USIM by which the operator is able to steer the network selection including which Access Technology e.g. NG-RAN or E-UTRAN to prioritize, or the operator can set the subscription to only allow either EPC, 5GC or both, or the operator can control RRM procedures per UE as to prioritize certain radio access to be used.
- Interworking with EPC is specified both with usage of N26 and without N26, and the UE may operate in single-registration mode or dual-registration mode for 3GPP access.

In single-registration mode;

- The UE has one active Mobility Management state for 3GPP access toward the Core Network and is either in 5GC NAS mode or in EPC NAS mode dependent on which Core Network the UE is connected to;
- The UE context information is transferred between the two systems when the UE moves back and forth, which either is done via N26 or by the UE moving each PDN Connection or PDU Session to the other system when interworking without an N26 interface.
- To enable the RAN in the target system to select the same Core Network entity which the UE was registered to in the source system (if it is available) and to enable the retrieval of UE context over N26, the UE maps the 4G-GUTI to 5G GUTI during mobility between EPC and 5GC and vice versa as described in Fig. 7.8.
- For handling of security contexts, Chapter 8 describes how to enable an efficient re-use of a previously established 5G security context when returning to 5GC.

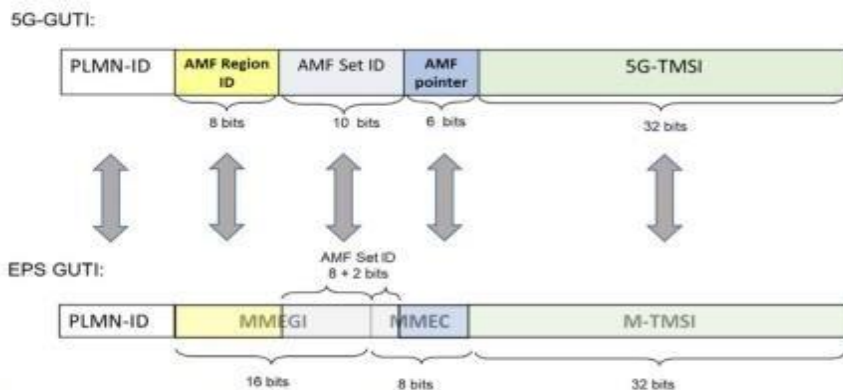


Fig. 7.8 Mapping between 5G-GUTI and EPS GUTI.

In dual-registration mode;

- The UE maintains independent Mobility Management states for 3GPP access toward the 5GC and EPC using separate RRC connections.
- In this mode, UE maintains 5G-GUTI and 4G-GUTI independently, and the UE may be registered to 5GC only, EPC only, or to both 5GC and EPC.
- It should be noted that N26 is used only for 3GPP access.
- Mobility of PDU Sessions between 3GPP access and non-3GPP access in the EPC and 5GC systems are driven by the UE and is supported without N26.
- When the UE moves from one system to the other, the UE provides its UE temporary identity in the format of the target system.
- If the UE has previously been registered/ attached to another system or has not registered/attached at all in the target system and does not hold any UE temporary identity of the target system, the UE provides a mapped UE temporary identity as described in Fig. 7.9.

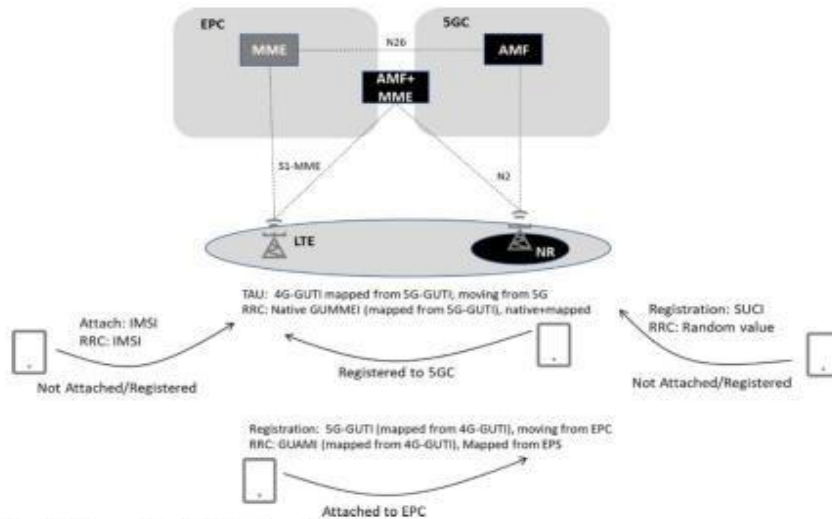


Fig. 7.9 UE provided UE identity at NAS and RRC.

- When the UE initially attaches to EPS the UE uses its IMSI as UE identity toward both E-UTRAN (in RRC) and EPC (in NAS).
- However, in 5GS, the UE uses a SUCI toward 5GC (in NAS) which conceals the UE's Identity. In both cases, there is no stored UE context in the network i.e. the network creates the UE context.

4. Explain interworking using the N26 interface

7.8.2.2 Interworking using the N26 interface

- When N26 interface is used for interworking procedures, the UE operates in single-registration mode, and the UE context information is exchanged over N26 between AMF and MME.
- The AMF and MME keeps one MM state (for 3GPP access) for the UE, i.e. either in the AMF or MME (and the MME or AMF is registering in the HSS + UDM when it holds the UE context).
- The interworking procedures provide IP address continuity at inter-system mobility between 5GS and EPS and are required to enable seamless session continuity (e.g. for voice services).
- The PGW-C + SMF keeps a mapping between PDN Connection and PDU Session related parameters e.g. PDN Type/PDU Session Type, DNN/APN, APN-AMBR/Session AMBR and QoS parameter mapping. Fig. 7.10 shows at a high level the interactions.

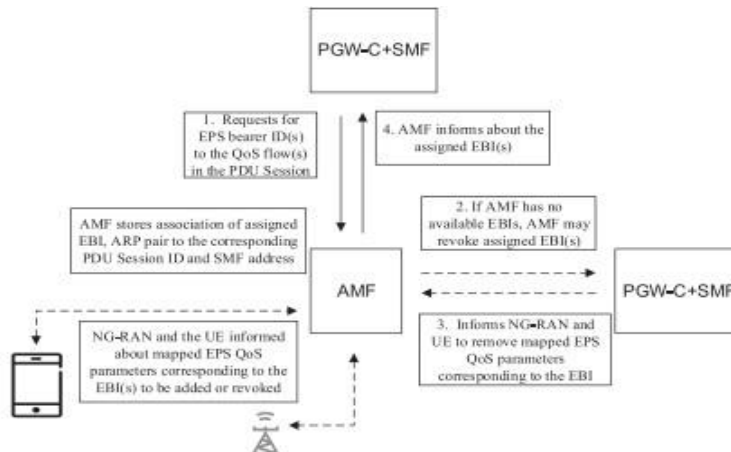


Fig. 7.10 EBI allocation and revocation.

5. Explain command and control.

4.1 Command and Control Method

- The conventional method for allocating spectrum is known as “Command and Control Method” shown in the Figure 4.1.
- There are some countries following this technique of spectrum allocation.
- In this method radio spectrum is divided into different spectrum bands that are allowed to specific radio communication services such as satellite services, mobile, broadcast on an exclusive basis.
- This method guarantees that the radio frequency spectrum will be exclusively licensed to an authorized user and can use spectrum without any interference.
- This method of spectrum allocation is not efficient because :
 - Spectrum assigned to a particular radio communication service cannot be replaced by other services even though it is witnessed that spectrum is underutilized.
 - There is no possibility of questioning the user once the spectrum is allocated to him (during the licensing period) as per the norms, provided he fulfills the terms and conditions.

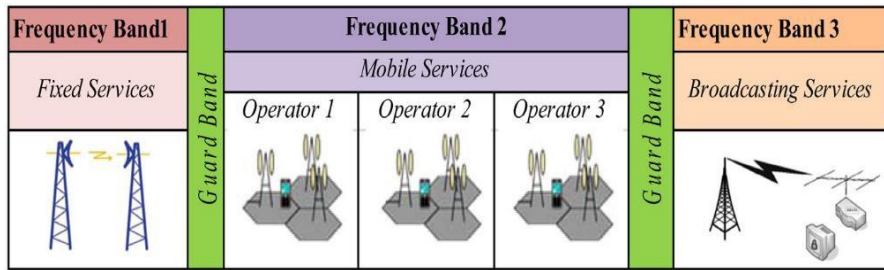


Figure 4.1 Command and Control Method.

- This method does not make spectrum to be utilized efficiently in rural areas as the spectrum utilization is heavy in the urban regions and underutilized in the rural areas.
- It is sad to see the spectrum underutilized and not accessible to all. It is difficult for some nations to provide 4G services itself.
- There is a need for taking serious steps in handling spectrum issues by implementing sophisticated technologies for the development of nation.
- In these cases the techniques like spectrum trading would be a successful solution. This will only lead to the development of 5G communications in these types of countries.

6. Explain spectrum sharing and trading/ Dynamic spectrum sharing and its advantages (Nov/Dec 23)

4.2 Spectrum Sharing

- The demand for multimedia content and information processing, services such as e-education and e-health, mobile broadcasting, enormous increase in the electronic gadgets necessitate efficient use of all available and usable frequency spectrum.
- The new generation of mobile broadband networks will require supporting higher data throughput rates.
- Many sophisticated technologies have been implemented for making the efficient use of available spectrum.

- For example, line-of-sight (LOS) systems are usable up to 100 GHz now. Running down the size of electronic components and systems introduces the multiple frequency bands in single equipment leading to the efficient use of available spectrum by the enhanced dynamic sharing of frequency bands.
- Spectrum management should be in such a way that there should be always optimum spectral sharing.
- Greater sharing of frequencies and bands allows more data to be sent by different users in the same amount of available spectrum.
- Spectrum sharing has basically three dimensions: frequency, time and location.
- The Collective Use of Spectrum (CUS) allows spectrum to be used by more than one user simultaneously without requiring a license.
- Some of the examples that come under spectrum sharing are frequency reuse concept in the existing telecom networks, FDMA, TDMA.
- Another important challenge is the sharing of spectrum among the heterogeneous networks.
- While it is easier to achieve efficient and successful spectrum sharing among the homogeneous or similar networks or applications, there arises complexity in heterogeneous networks.
- The spectrum sharing methods are classified into three categories based on based on the priority level of accessing the radio spectrum as follows :
 - a. Horizontal spectrum sharing: all the devices have equal rights to access the spectrum.
 - b. Vertical spectrum handover only: the primary users are allotted priorities to access the spectrum.
 - c. Hierarchical spectrum sharing: it is an enhanced variant of the vertical spectrum sharing.

4.2.1 Spectrum Using SDR and Cognitive Radio – Dynamic Sharing

- Evolution of software defined radio (SDR) and cognitive radio (CR) are the two major milestones in the mobile communications.
- Dynamic sharing of spectrum improves the spectrum efficiency and the above mentioned technologies play a vital role in this aspect.
- Conventionally, transmitters were tuned to specific frequencies, and facilities for multiple frequencies would cost high.
- But after the development of these technologies, tuning the transmitters to the multiple frequencies has become easier, i.e., switching to the different frequencies in a dynamic way would be possible at a reasonable cost.
- Cognitive radio first detects the occupation of the channel, and if it is occupied, it helps the users to switch to the other vacant channels.
- Also the carrier signals are sensed regularly for usage in other. There is always a need of large amount of spectrum in case of emergency or public safety conditions compared to that of normal conditions.
- In these emergency cases, dynamic sharing of spectrum would be a promising solution.
- In some countries spectrum regulators are used for the encouraging dynamic sharing spectrum with public safety requirements.
- It is to be noted that CR is a combination of administrative (regulatory), technical, and market based techniques to enhance the efficiency of spectrum utilization.
- Another area of utility for dynamic sharing is White Spaces (TV Band).
- Normally, the TV broadcasters repeat the same channel/carrier at relatively longer distances, to avoid any interference especially at the border/edger of the coverage areas that are on the border of two adjacent broadcast transmissions on same channel.
- However, there are very few receivers in this area, and the spectrum utility is not effective and could be utilized for other purposes.
- The broadcasters are generally quite protective for their signal Transmissions, even in areas beyond the theoretical coverage areas.
- Hence, only low power systems that cause minimal interference can be considered for shared usage with the TV spectrum.

- However, gradually with time building collective confidence amongst the users that includes the broadcasters, higher power based systems could be considered.

7. Explain Spectrum Trading

4.3 Spectrum Trading

- Spectrum trading is a case of spectrum sharing with the involvement of commercial activities.
- Spectrum trading is found to be a more economical way of efficient use of spectrum. It is an option through which flexibility can be increased and spectrum assigned to a particular service, and can be easily transferred for other usage.
- To explain it in brief, spectrum trading is a market based mechanism where buyers and the sellers determine the assignments of spectrum and its uses in which seller transfers the right of spectrum usage, in full or part, to buyer while retaining the ownership.
- In many countries spectrum trading is already running and the trading procedure is confined to specific bands, which are in demand for commercial use with specified conditions.
- Spectrum trading improves the efficiency and facilitates new services to enter in the market by making slight modification in the regulatory provisions.

The difference between spectrum sharing and spectrum trading can be explained as follows:

- In spectrum trading the usage rights are transferred completely from the seller for a specified period.

- However, in spectrum sharing buyer gets a temporary right of spectrum usage with the exclusive rights resting with the seller.
- Trading becomes effective only when it is clubbed with liberalization. Spectrum trading can be implemented if there is solid base in understanding advanced technologies and operating systems as the spectrum flexibility demands new approaches and practical methods for monitoring compliance, enforcement and conflict resolution.

4.3.1 Spectrum Trading Merits

The merits of spectrum trading are as follows:

- Improves efficient spectrum usage
- Facilitates the evaluation of spectrum licenses, and gaining knowledge of market value of spectrum
- Quicker process, with better and faster decision-making by those with information
- Removes barriers to entry by allowing small operators and start-ups to acquire spectrum rights of use more readily, thereby facilitating the development of market competition
- There is an opportunity for more rapid redeployment and faster access for spectrum
- Encourages new technologies to gain access to spectrum more quickly
- Existing operators gain an opportunity to sell unused or under-used spectrum and make more flexible use of spectrum
- Reduction in the transactions costs of acquiring rights to use spectrum
- Allows operators increased flexibility to accommodate shifting demand driven by market changes.

8. Explain Cognitive radio based on 5G.

4.4 Cognitive Radio

- IEEE approved definition of cognitive radio (CR) is a radio in which communication systems are aware of their environment and internal state, and can make decisions about their radio operation based on that information and predefined objectives.
- The environmental information may not include location information related to communication systems.
- Cognitive radio is a very good solution for increasing the spectrum utilization.
- Cognitive radios should be able to self-organize their communication based on sensing and reconfigurable functions as stated below :
 - ***Spectrum resource management:*** this scheme is necessary to manage and organize efficiently spectrum holes information among cognitive radios.
 - ***Security management:*** cognitive radio networks (CRNs) are heterogeneous networks in essence and this heterogeneous property introduces a lot of security issues. So this scheme helps in providing security functions in dynamic environment.
 - ***Mobility and connection management:*** this scheme can help neighbour- hood discovery, detect available Internet access, and support vertical handoffs, which help cognitive radios to select route and networks.

4.4.1 CR Device Concept

This section explains the features of CR whose implementation in a single device offers a very smart and high performance user terminal – CR terminal. The Figure 4.2 shows the CR properties.

A. Spectrum sensing

Spectrum sensing operation can be divided into the three step functions:

- *Signal Detection:* In this step of operation existence of the signal is sensed. There is no need to know the type of signal in this step.
- *Signal Classification:* In this step of operation the type of signal is detected, which is done by extracting the features of the signal.
- *Channel availability decision:* In this channel availability is detected. Once the free channels are detected, the step next to it is, sharing the spectrum holes which can be achieved by the spectrum allocation scheme.

The CR technology also brings new *security and pricing challenges* which are shown in the Figure 4.2.

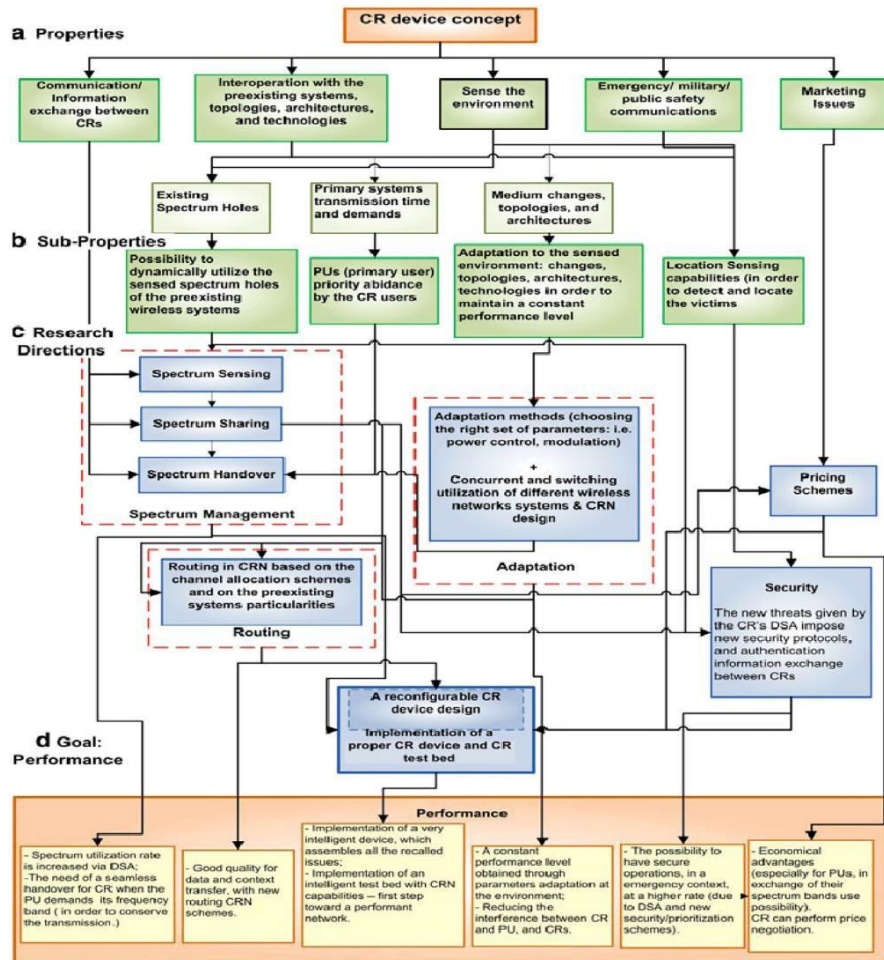


Figure 4.2 Concept of CR Device .

- New security threats appear with the dynamic spectrum access concept, as well as the CR's authentication needs.
- The pricing is very much influenced by the used channel allocation scheme. Additionally, CRs must be designed with strong capabilities to negotiate the available channels' price.

B. Spectrum handover

The phenomenon of frequency changing dynamically is said to be spectrum handover. A secondary user changes its frequency on appearance of a primary user or due to transmission degradation. This necessitates designing a handover scheme.

C. Environment adaptation

Different changes like topological changes, noise or interference power may occur while sensing the information. In order to adapt to these changes and to maintain the constant performance new adaptation techniques have to be implemented which is an important point of concern.

D. CR routing

CR routing is based on the requirement for CR device to interoperate with different systems, and is influenced by the spectrum sharing techniques.

CRNs inherit the PSs (Primary Systems) network characteristics: infrastructure - based, mesh, ad-hoc, sensor networks, etc. and these architecture types impose a specific routing algorithm, which must also include the CR devices and the possibility for a CR to be a relay node for another CR.

4.4.2 CR based on 5G

The various functionalities for 5G that could be met with CR usage are as follows:

- Advanced PHY and MAC technologies.
- Implementation of novel and flexible protocols.
- Capacity to support homogenous and heterogeneous systems.
- Adaptation to different changes like environment changes, dynamic frequency changes, etc.

Correlation between WISDOM and CR in reference to 5G could be given as:

- “5G brings the convergence concept through WISDOM and CR represents the technological tool to implement it.”
- The 5G technology eliminates the radio terminals that are specific to particular wireless technologies and proposes a universal terminal which must include all of the predecessor features in a single device.
- This terminal convergence is supported by the users’ needs and demands and is strongly found in CR terminal .

9. Discuss about Millimetre Waves.

4.5 Millimetre Waves

- Most of the radio communications including TV, satellite communications, GPS, Bluetooth are utilizing frequency band ranging from 300 MHz to 3 GHz.
- But this band is getting crowded and the focus is on releasing and utilizing the additional spectrum. In mm-waves are the promising solution for this problem.
- The spectrum bands identified under the IMT umbrella do not have the capacity to carry such enormous data required for 5G services.
- Therefore, mm-waves could be the candidate bands for 5G mobile communications due to high data carrying capacity. The mm-waves have the following advantages :
 - (a) Not much operation at mm-waves so more spectrum is available at mm- waves
 - (b) Very large blocks of contiguous spectrum to support future applications.
 - (c) Due to high attenuation in free space, frequency reuse is possible at shorter distance
 - (d) Spatial resolution is better at mm-waves hardware with CMOS technology
 - (e) Advancement in semiconductor technology allows low cost equipment

(f) Small wavelength makes possible use of large antenna arrays for adaptive beam forming

(g) Small size of antenna at mm-waves facilitates easy integration on chip and installation at suitable locations.

- In mm-waves allow larger bandwidth and offer high data transfer and low latency rate that are suitable for high speed reliable Internet services.
- The small wavelength facilitates small size antenna and other part of radio hardware, which reduces costs and also easy to install.
- The transmitter's antenna would be like a lamppost, which could be installed on building, street lamppost, etc.
- High directionality attained in this band can be used to increase spatial multiplexing.
- The size of antenna required for an mm-waves radio can be one-tenth or less of an equivalent lower frequency radio which is an advantage to the manufactures to build smaller and lighter systems.
- Beam width is the measure of how a transmitted beam spreads out as it gets farther from its point of origin.
- But due to limited availability of radio frequency (RF) bands the fifth generation wireless communication systems will move to ultra-high capacity mm-wave bands.
- High frequency makes mm-wave band more attractive for wireless communication system and these frequencies are used in terrestrial and satellite communications. Wireless products that use millimeter waves already exist for fixed, LOS transmissions
- But the absorption rate of the mm-wave electromagnetic signal poses great challenges for their utilization in the non-LOS and mobile connections.
- On the other hand, high directionality achieved in this band can be used to increase spatial multiplexing.

Wireless backhaul will be another key enabler of 5G-mm-wave small cells.

- Within the mm frequencies, the frequency band of 60 GHz has attracted the researchers to work with, as the large amounts of bandwidth are unallocated in this band, bandwidths that are required for communication systems at the intended data rates of 100 Mbps and above.

- Also, another advantage of 60-GHz band is due to a physical property of the propagation channel at this frequency that provides a natural way for reduction of frequency reuse factor, which tends to compact cell size.
- It is a general property of the mm-wave propagation that the behaviour of the propagation rays is well characterized by the geometric optics.
- That is, the waves do not penetrate the walls or other obstacles and wave reflection is the main mechanism leading to a multipath.
- In mm-waves have the potential to support broad-band service access which is especially relevant because of the advent of Broadband Integrated Service Digital Network (B-ISDN).
- With the development of personal wireless communication systems, two things are appearing to be significant:
 - Exploiting high frequency bands, such as mm-waves to provide broad-band for high rate data transmission.
 - To integrate multi-tasks in one system which greatly extend the application of wireless device.

The utility of mm-waves for the micro cells that form the WISDOM based GIMCV are well positioned to be served by these mm-waves. It has been elaborated in these following points:

- It is relatively easy to get licenses for big blocks of mm-wave spectrum, which would allow carriers to deploy large backhaul pipes over 1 Gbps in size. While a single small cell may not need that much capacity, the complexity of heterogeneous networks will require daisy-chaining many small cells together, each cell passing its load down the line.
- Small cell backhaul makes the best use high frequency characteristics of mm-waves. The higher the frequency the shorter distance a wave propagates unless it gets a serious power boost. But the heterogeneous network by definition will be composed of densely packed cells in urban environments, meaning no mm-waves will have to travel far between hops.

- The traditional uses of the mm-waves include radio navigation, space research, radio astronomy, earth exploration satellite, radar, military weapons and other applications.
- The backbone/backhaul networks (point to point network) for existing telecom network to connect base station to main switching centre (MSC), Local Multipoint Distribution System (LMDS), indoor WLAN, high capacity dense networks are also present in the mm-waves. The typical microwave backhaul bands are at 6.0 GHz, 11.0 GHz, 18.0 GHz, 23.0 GHz, and 38.0 GHz frequency bands.
- The light use of mm-waves could be attributed to high attenuation and low penetration. At such high frequency, waves are more prone to rain and other atmospheric attenuation.
- The wavelength is in the order of millimeters, and rain drops are also of the same size. Rains absorb high frequency waves and make it difficult for propagation.
- However, the experimental results show that in heavy rain condition, attenuation is 1.4 dB and 1.6 dB for 200 meters distance at 28 GHz and 38 GHz, respectively. The rain attenuations at 60 GHz for a rainfall rate of 50 mm/h, is approximately 18 dB/km. A proper link design with slightly high transmit power may take care of rain attenuation.
- Slight change in the position would affect the signal strength at the receiving end, due to which mm-waves are deeply affected by scattering, reflection and refraction. The root mean square (RMS) delay spread for mm-waves is of the order of few nano seconds, and it is high for non-LOS (NLOS) links than (LOS) links.
- Similarly, path loss exponent for NLOS links is higher than LOS links.
- Due to higher path loss and RMS delay spread, it is assumed that mm-waves are not suitable for (NLOS) links.
- However, these difficulties could be managed by using carrier aggregation, high order MIMO, steerable antenna, beam-forming techniques.
- The frequency bands around 60 GHz is best suited for pico and femto cell due to high data carrying capacity and small reuse distance due to strong oxygen absorption at the rate of 15 dB/Km.

- The usage in frequency bands around 60 GHz is highly sparse, which provides freedom to allocate a large bandwidth to every channel. Moreover, equipment can be made very compact due to the very small antenna size. The figure for mm wave measurement is shown below.

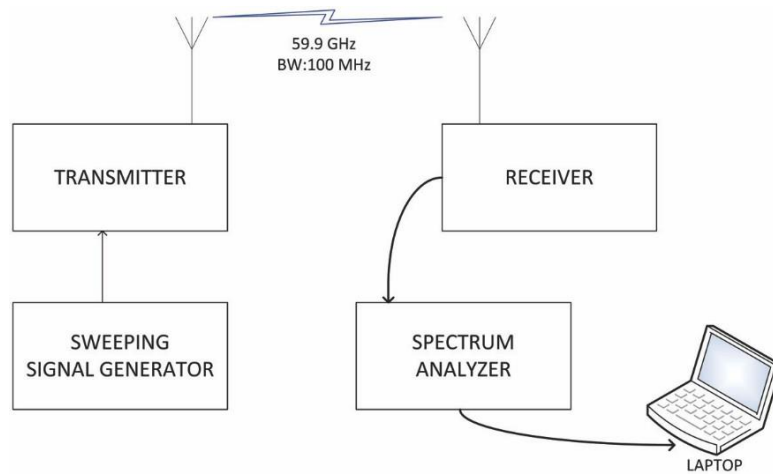


Figure 4.3 Measurement setup.

Part – A Two marks

1. What is mobility management?

The mobility is a core feature of 5GS. Mobility Management is required to ensure the following:

- That the network can “reach” the user, for example to notify the user about incoming messages and calls,
- That a user can initiate communication toward other users or services such as Internet access, and that connectivity and ongoing sessions can be maintained as the user moves, within or between access technologies.

2. What are the mobility management functions?

- Service Area Restriction
- Local Area Data Network (LADN)
 - Mobile Initiated Connection Only (MICO)

3. What is Service Area Restriction?

Mobility with session continuity is controlled at UE level at certain areas

4. Local Area Data Network (LADN)?

Mobility with session continuity is controlled at PDU Session level making communication available at certain areas

5. Mobile Initiated Connection Only (MICO)?

Paging capability (as part of the mobility service) is optional.

6. What are the 5G Mobility Management (5GMM) related procedures?

- ✓ Common procedures
- ✓ Specific procedures
- ✓ Connection management procedures

7. What is Registration and Mobility?

Radio networks are built by cells that range in size from tens and hundreds of meters to tens of kilometers and the UE updates the network about its location on a regular basis.

8. What is paging?

Paging is sending message and it is used to search for Idle UEs and establish a signalling connection. Paging is, for example, triggered by downlink packets arriving to the UPF.

9. What is Mobile Initiated Connection Only (MICO) mode?

Mobile Initiated Connection Only (MICO) mode was introduced to allow paging resources to be saved for UEs that don't need to be available for Mobile Terminating communication. When the UE is in MICO mode, the AMF considers the UE as unreachable when the UE is in CM-IDLE state.

10. What is RRC Inactive?

It includes support for efficient communication with minimal signalling by using a concept called RRC Inactive which affects the UE, NG-RAN and 5GC. RRC Inactive is a state where a UE remains in CM-CONNECTED state (i.e. at NAS level) and can move within an area configured by NG-RAN (the RAN Notification Area – RNA) without notifying the network.

11. How are the 5GC assistance information divided?

- ✓ Core Network assisted RAN parameters tuning.
- ✓ Core Network assisted RAN paging information.
- ✓ RRC Inactive Assistance Information.

12. What is Service Area and Mobility Restrictions?

Mobility Restrictions enables the network, mainly via subscriptions, to control the Mobility Management of the UE as well as how the UE accesses the network. Similar logic as used in EPS is applied in 5GS, but with some new functionality added as well.

13. What is RAT restriction?

Defines the 3GPP Radio Access Technologies a UE is not allowed to access in a PLMN and may be provided by the 5GC to the NG-RAN as part of the Mobility Restrictions. The RAT restriction is enforced by the NG-RAN at connected mode mobility.

14. What is Forbidden Area?

A Forbidden Area is an area in which the UE is not permitted to initiate any communication with the network for the PLMN.

15. What is Core Network type restriction?

Defines whether UE is allowed to access to 5GC, EPC or both for the PLMN.

16. What is Service Area Restriction?

Defines areas controlling whether the UE is allowed to initiate communication for services.

17. What is Allowed Area?

In an Allowed Area, the UE is permitted to initiate communication with the network as allowed by the subscription.

18. What is Non-Allowed Area?

In a Non-Allowed Area a UE is “service area restricted” meaning that neither the UE nor the network is allowed to initiate signalling to obtain user services (both in CM-IDLE and in CM-CONNECTED states).

19. What is control overhead?

5GS supports the ability to control the amount of load UEs produce toward the 5GS through different mechanisms. Mechanisms for 5GC to balance load across NFs and also to scale the amount of resources consumed for the NFs are often enough to handle normal fluctuations of load impacting the 5GC.

20. What are the mechanisms used to control overload in 5G?

- ✓ Congestion in control channel resources
- ✓ Congestion in random access channel (RACH) resources
- ✓ Release/reject UE RRC connection

21. What is Congestion in control channel resources?

5QI-based scheduling controls cases when e.g. the number of users awaiting scheduling exceeds the number of users that can be admitted such that the random access procedure fails.

22. What is Congestion in random access channel (RACH) resources?

Random access back-off. This pushes some UEs into a longer back-off. This is when there are so many access attempts on the RACH that the UE provided preambles cannot be detected anymore.

23. What is Release/reject UE RRC connection?

If there are not enough resources to process RRC connection requests, Releasing RRC connection or rejecting RRC connection attempts can be used.

24. What is Unified Access Control?

EPS supports multiple variants of access barring mechanisms as they were developed in different releases to address different needs for congestion control. The 5GS supports one mechanism called Unified Access Control (UAC) which is extensible, flexible (e.g. each operator can define their own category when to apply access control) and supports a variety of scenarios.

25. What are the Interworking with EPC using 3GPP access options?

- ✓ In single-registration mode
- ✓ In dual-registration mode

26. What is Command and Control Method?

The conventional method for allocating spectrum is known as “Command and Control Method. There are some countries following this technique of spectrum allocation. In this method radio spectrum is divided into different spectrum bands that are allowed to specific radio communication services such as satellite services, mobile, broadcast on an exclusive basis.

27. What is Spectrum Sharing?

It is sharing of available spectrum. The demand for multimedia content and information processing, services such as e-education and e-health, mobile broadcasting, enormous increase in the electronic gadgets necessitate efficient use of all available and usable frequency spectrum.

28. What is Spectrum Trading?

Spectrum trading is a case of spectrum sharing with the involvement of commercial activities. Spectrum trading is found to be a more economical way of efficient use of spectrum. It is an option through which flexibility can be increased and spectrum assigned to a particular service, and can be easily transferred for other usage.

29. List the merits of spectrum trading?

- Improves efficient spectrum usage
- Facilitates the evaluation of spectrum licenses, and gaining knowledge of market value of spectrum
- Quicker process, with better and faster decision-making by those with information
- Removes barriers to entry by allowing small operators and start-ups to acquire spectrum rights of use more readily, thereby facilitating the development of market competition
- There is an opportunity for more rapid redeployment and faster access for spectrum

30. Define Cognitive Radio.(Nov/Dec 23)

IEEE approved definition of cognitive radio (CR) is a radio in which communication systems are aware of their environment and internal state, and can make decisions about their radio operation based on that information and predefined objectives.

31. What are the self-organizing features of Cognitive radios?

- ✓ Spectrum resource management
- ✓ Security management
- ✓ Mobility and connection management

32. What is Spectrum Handover?

The phenomenon of frequency changing dynamically is said to be spectrum handover. A secondary user changes its frequency on appearance of a primary user or due to transmission degradation. This necessitates designing a handover scheme.

33. Define Millimeter (MM) Wave (Nov/Dec 23)

Millimeter wave (MM wave), also known as millimeter band, is the band of spectrum with wavelengths between 10 millimeters (30 GHz) and 1 millimeter (300 GHz). It is also known as the extremely high frequency (EHF) band by the International Telecommunication Union (ITU).

34. List the Advantages of millimeter (MM) wave

Millimeter wave is a band of electromagnetic spectrum that can be used in a broad range of products and services, such as high-speed, point-to-point wireless local area networks (WLANs) and broadband access. In telecommunications, millimeter wave is used for a variety of services on mobile and wireless networks, as it enables higher data rates than at lower frequencies, such as those used for Wi-Fi and current cellular networks.

4G/5G COMMUNICATION NETWORKS

UNIT – IV DYNAMIC SPECTRUM MANAGEMENT AND MM-WAVES

Mobility management, command and control, spectrum sharing and spectrum trading, cognitive radio based on 5G, millimetre waves.

1. Explain Mobility management. (Nov/Dec 23)

The mobility is a core feature of 5GS. Mobility Management is required to ensure the following:

- That the network can “reach” the user, for example to notify the user about incoming messages and calls,
- That a user can initiate communication toward other users or services such as Internet access, and that connectivity and ongoing sessions can be maintained as the user moves, within or between access technologies.

As a result, there are several optional 5GS mobility management related functions that differ from previous 3GPP systems:

- Service Area Restriction: mobility with session continuity is controlled at UE level at certain areas
- Local Area Data Network (LADN): mobility with session continuity is controlled at PDU Session level making communication available at certain areas
- Mobile Initiated Connection Only (MICO): paging capability (as part of the mobility service) is optional.

5G Mobility Management (5GMM) related procedures are divided into three categories depending on the purpose of the procedure and on how they can be initiated:

1. Common procedures; can always be initiated when the UE is in CM-CONNECTED state.
2. Specific procedures; only one UE initiated specific procedure can be running for each of the Access Types.

3. Connection management procedures; used to establish a secure signalling connection between the UE and the network, or to request the resource reservation for sending data, or both.

7.2 Establishing connectivity

7.2.1 Network discovery and selection

- Network discovery and selection procedures for 5GS do not differ much from EPS and the principles used when a 3GPP access type is selected, have been maintained.
- Before the UE can receive and use the services and capabilities from the 5GS, e.g. Session Management services from SMF, the UE needs to establish a connection to the 5GS.
- To achieve this, the UE first selects a network/PLMN and a 5G-AN. For 3GPP access i.e. NG-RAN, the UE selects a cell, then the UE establishes a RRC connection to the NG-RAN.
- Based on the content (e.g. selected PLMN, Network Slice information) provided by the UE in establishing the RRC connection the NG-RAN selects an AMF and forwards the UE NAS MM message to the AMF in the 5GC using the N2 reference point.
- Using the AN connection (i.e. RRC connection) and the N2, the UE and the 5GS complete a Registration procedure. Once the Registration procedure is completed, the UE is registered in the 5GC i.e. the UE is known and the UE has a NAS MM connection to the AMF, the UE's entry point to the 5GC, which is used as the NAS connection to the 5GC.
- Further communication between the UE and other entities in the 5GC uses the established NAS connection as NAS transport from that point forward.
- To save resources the NAS connection is released while the UE is still registered and known in the 5GC, i.e. to re-establish the NAS connection the UE or 5GC initiates a Service Request procedure.

7.2.2 Registration and Mobility

- Idle-mode Mobility Management for 5GS using NR and E-UTRA is built on similar concepts to LTE/E-UTRAN (EPS), GSM/WCDMA, and CDMA.
- Radio networks are built by cells that range in size from tens and hundreds of meters to tens of kilometers and the UE updates the network about its location on a regular basis.

- It is not practical to keep track of a UE in idle mode every time it moves between different cells due to the amount of signalling it would cause, nor to search for a UE across the entire network for every terminating event (e.g. an incoming call).
- In order to create efficiencies, therefore, cells are grouped together into Tracking Areas (TA), and one or more Tracking Areas may be assigned to the UE as a Registration Area (RA).
- RA is used as a base for the network to search for the UE and for the UE to report its location.

A summary of the idle mobility procedure in 5GS is:

- A TA consists of a set of cells,
- The Registration Area in 5GS is a list of one or more Tracking Areas (TA list),
- The UE performs Registration update due to mobility when moving outside its Registration Area i.e. TA list,
- The UE in idle state also performs periodic Registration update when the periodic Registration update timer expires.

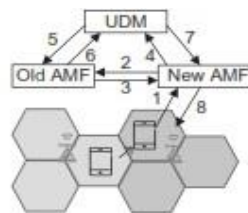


Fig. 7.1 Mobility Registration Update procedure.

7.2.3 Cellular connected mode mobility

- Great effort has been put into optimized connected mode mobility for cellular systems.
- The basic concept is somewhat similar across different technologies with some variations in the functional distribution between UE and networks.

- While in connected mode, the UE has a connected signalling connection and zero, one or more connected user plane resources, and data transmission may be ongoing.
- To limit interference and provide the UE with good data communication, the UE changes cells through handover when there is a cell that can provide better service than the cell that the UE is currently using.
- To save on complexity in the UE design and power, the systems are designed to ensure that the UE only needs to listen to a single gNB/ng-eNB at a time.
- In addition, for inter-RAT handover (e.g. NR to E-UTRAN HO) the UE only needs to have a single radio technology connected at a time.

7.3 Reachability

7.3.1 Paging

- Paging is used to search for Idle UEs and establish a signalling connection. Paging is, for example, triggered by downlink packets arriving to the UPF.
- When the UPF receives a downlink packet destined for an Idle UE, it does not have an NG-RAN User Plane tunnel address to which it can send the packet.
- The UPF instead buffers the packet and informs the SMF that a downlink packet has arrived.
- The SMF asks the AMF to setup User Plane resources for the PDU Session, and the AMF which knows in which RA the UE is located and sends a paging request to the NG-RAN within the RA.
- The NG-RAN calculates at which occasion the UE is to be paged using parts of the UE's 5G-S-TMSI (10 bits) as input, and then the NG-RAN pages the UE.

7.3.2 Mobile Initiated Connection Only (MICO) mode

- Mobile Initiated Connection Only (MICO) mode was introduced to allow paging resources to be saved for UEs that don't need to be available for Mobile Terminating communication.
- When the UE is in MICO mode, the AMF considers the UE as unreachable when the UE is in CM-IDLE state.

- MICO mode is negotiated (and re-negotiated) during Registration procedures, i.e. the UE may indicate its preference for MICO mode and the AMF decides whether MICO mode can be enabled taking into account the UE's preference as well as other information such as the user's subscription and network policies.

7.3.3 UE's reachability and location

- 5GS also supports location services in a similar way to EPS, but 5GS also provides the possibility for any authorized NF (e.g. SMF, PCF or NEF) in the 5GC to subscribe to UE mobility related event reporting.
- The NF subscribing to a UE mobility related event can do so by providing the following information to the AMF:
 - Whether UE location or the UE mobility in relation to an area of interest is to be reported
 - In case an area of interest is requested, then the NF specifies the area as:
List of Tracking Areas, list of cells or list of NG-RAN nodes.
If the NF wants to get an LADN area, the NF (e.g. SMF) provides the LADN DNN to refer the LADN service area as the area of interest.
If a Presence Reporting Area is requested as area of interest, then the NF (e.g. SMF or PCF) may provide an identifier to refer to a predefined area configured in the AMF.
 - Event Reporting Information: event reporting mode (e.g., periodic reporting), number of reports, maximum duration of reporting, event reporting condition (e.g. when the target UE moved into a specified area of interest).
 - The notification address i.e. address of NF that the AMF is to provide the notifications which can be another NF than the NF subscribing to the event
 - The target of event reporting that indicates a specific UE, a group of UE(s) or any UE (i.e. all UEs).

7.4 Additional MM related concepts

7.4.1 RRC Inactive

- It includes support for efficient communication with minimal signalling by using a concept called RRC Inactive which affects the UE, NG-RAN and 5GC.
- RRC Inactive is a state where a UE remains in CM-CONNECTED state (i.e. at NAS level) and can move within an area configured by NG-RAN (the RAN Notification Area – RNA) without notifying the network.
- The RNA is a subset within the RA allocated by the AMF. When the UE is in RRC Inactive state the following applies:
 - UE reachability is managed by the NG-RAN, with assistance information from 5GC;
 - UE paging is managed by the NG-RAN;
 - UE monitors for paging with part of the UE's 5GC (5G S-TMSI) and NG-RAN identifier.
- In RRC Inactive, the last serving NG-RAN node keeps the UE context and the UE-associated NG (N2 and N3) connections with the serving AMF and UPF. Therefore, there is no need for the UE to signal toward the 5GC before sending User Plane data.

7.5 N2 management

- In EPS, when a UE attaches to EPC and is assigned a 4G-GUTI, the 4G-GUTI is associated to a specific MME and if there is a need to move the UE to another MME the UE needs to be updated with a new 4G-GUTI.

- This may be a drawback e.g. if the UE is using some power saving mechanism or if a large amount of UEs are to be updated at the same time. With 5GS and N2 there is support for moving one or multiple UEs to another AMF without immediately requiring updating the UE with a new 5G-GUTI.
- The 5G-AN and the AMF are connected via a Transport Network Layer that is used to transport the signalling of the NGAP messages between them. The transport protocol used is SCTP.
- The SCTP endpoints in the 5G-AN and the AMF sets up SCTP associations between them that are identified by the used transport addresses.
- An SCTP association is generically called a Transport Network Layer Association (TNLA).

The N2 (also called NG in RAN3 specifications) reference point between the 5G-AN and the 5GC (AMF) supports different deployments of the AMFs e.g. either

(1) An AMF NF instance which is using virtualization techniques such that it can provide the services toward the 5G-AN in a distributed, redundant, stateless, and scalable manner and that it can provide the services from several locations, or

(2) An AMF Set which uses multiple AMF NF instances within the AMF Set and the multiple AMF Network Functions are used to enable the distributed, redundant, stateless, and scalable characteristics.

- The N2 reference point supports a form of self-automated configuration. During this type of configuration the 5G-AN nodes and the AMFs exchange NGAP information of what each side supports e.g. the 5G-AN indicates supported TAs, while the AMF indicates supported PLMN IDs and served GUAMIs.
- The exchange is performed by the NG SETUP procedure and, if updates are required, the RAN or AMF CONFIGURATION UPDATE procedure. The AMF CONFIGURATION UPDATE procedure can also be used to manage the TNL associations used by the 5G-AN.

7.5.1 AMF management

- The 5GC, including N2, supports the possibility to add and remove AMFs from AMF Sets.
- Within 5GC, the NRF is updated (and DNS system for interworking with EPS) with new NFs when they are added, and the AMF's NF profile includes which GUAMI(s) the AMF handles.
- For a GUAMI there may also be one or more backup AMF registered in the NRF (e.g. to be used in case of failure or planned removal of an AMF).
- A planned removal of an AMF can be done either through the AMF storing the registered UEs' contexts in a UDSF (Unstructured Data Storage Function), or with the AMF deregistering itself from the NRF, in which case the AMF notifies the 5G-AN that the AMF will be unavailable for processing transactions for the GUAMI(s) configured on this AMF.
- Additionally, the AMF can initially decrease the load by changing the weight factor for the AMF toward the 5G-AN, e.g. setting it to zero, causing the 5G-AN to select other AMFs within the AMF Set for new UEs entering the area.

7.5.2 5GC assistance for RAN optimizations

- As the UE context information is not kept in the NG-RAN when the UE transition to RRC-IDLE, it may be hard for the NG-RAN to optimize the logic related to the UE as UE specific behavior is unknown unless the UE has been in RRC-CONNECTED state for some time.
- There are NG-RAN specific means to retrieve such UE information e.g. UE history information can be transferred between NG-RAN nodes.
- To further facilitate an optimized decision in NG-RAN e.g. for UE RRC state transition, CM state transition decision and optimized NG-RAN strategy for RRC-INACTIVE state, the AMF may provide 5GC assistance information to NG-RAN.
- 5GC has a better method to store UE related information for a longer time and a means to retrieve information from external entities through external interfaces. When calculated by the 5GC (AMF) the algorithms used and related criteria, and the decision when it is considered suitable and stable to send to the NG-RAN are vendor specific.
- Therefore, along with the assistance information sent to NG-RAN, it often is accompanied with the information whether it is derived by statistics or retrieved via subscription information (e.g. set by agreements or via an API).

5GC assistance information is divided into 3 parts:

- Core Network assisted RAN parameters tuning;
 - Core Network assisted RAN paging information;
 - RRC Inactive Assistance Information.
- Core Network assisted RAN parameters tuning provides the NG-RAN with a way to understand UE behavior so as to optimize NG-RAN logic e.g. how long to keep the UE in specific states.
- Besides the content listed in Table 7.4, the 5GC also provides the source of the information e.g. if it is subscription information or derived based on statistics.

7.5.3 Service Area and Mobility Restrictions

Mobility Restrictions enables the network, mainly via subscriptions, to control the Mobility Management of the UE as well as how the UE accesses the network. Similar logic as used in EPS is applied in 5GS, but with some new functionality added as well.

The 5GS supports the following:

- RAT restriction:

Defines the 3GPP Radio Access Technologies a UE is not allowed to access in a PLMN and may be provided by the 5GC to the NG-RAN as part of the Mobility Restrictions. The RAT restriction is enforced by the NG-RAN at connected mode mobility.

- Forbidden Area:

A Forbidden Area is an area in which the UE is not permitted to initiate any communication with the network for the PLMN.

- Core Network type restriction:

Defines whether UE is allowed to access to 5GC, EPC or both for the PLMN.

- **Service Area Restriction:**

Defines areas controlling whether the UE is allowed to initiate communication for services as follows:

- **Allowed Area:** In an Allowed Area, the UE is permitted to initiate communication with the network as allowed by the subscription.
- **Non-Allowed Area:** In a Non-Allowed Area a UE is “service area restricted” meaning that neither the UE nor the network is allowed to initiate signalling to obtain user services (both in CM-IDLE and in CM-CONNECTED states).

The UE performs mobility related signalling as usual, e.g. mobility Registration updates when moving out of the RA. The UE in a Non-Allowed Area replies to 5GC initiated messages, which makes it possible to inform the UE that e.g. the area is now allowed.

7.6 Control of overload

- 5GS supports the ability to control the amount of load UEs produce toward the 5GS through different mechanisms.
- Mechanisms for 5GC to balance load across NFs and also to scale the amount of resources consumed for the NFs are often enough to handle normal fluctuations of load impacting the 5GC.
- In order to protect itself from overload situations, the 5GC supports a number of mechanisms including instructing UEs to back-off through NAS back-off timers (for Mobility Management as well as Session Management messages) such that the UE does not re-attempt to connect while the back-off timer is running.
- 5GS also supports the possibility to indicate to the NG-RAN that the load toward the AMF needs to be reduced using different criteria in an NGAP Overload Start message sent to the NG-RAN.
- Different methods can be used to handle possible bottlenecks in the NG-RAN Control Plane which also protects the 5GC. The mechanism used often depends on the load in the system. These are summarized below.

- Congestion in control channel resources: 5QI-based scheduling controls cases when e.g. the number of users awaiting scheduling exceeds the number of users that can be admitted such that the random access procedure fails.
- The random access procedure is a lower layer procedure used when the UE wants to initiate communication e.g. the UE gets synchronized with the network from a timing perspective.
- Congestion in random access channel (RACH) resources: random access back-off. This pushes some UEs into a longer back-off. This is when there are so many access attempts on the RACH that the UE provided preambles cannot be detected anymore.
- Release/reject UE RRC connection: If there are not enough resources to process RRC connection requests, Releasing RRC connection or rejecting RRC connection attempts can be used.

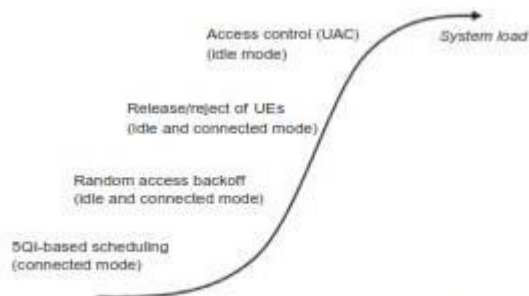


Fig. 7.6 Access and congestion control mechanisms as a function of the system load.

7.6.1 Unified Access Control

- EPS supports multiple variants of access barring mechanisms as they were developed in different releases to address different needs for congestion control.
- The 5GS supports one mechanism called Unified Access Control (UAC) which is extensible, flexible (e.g. each operator can define their own category when to apply access control) and supports a variety of scenarios.
- The UAC affects UEs in all RRC states i.e. RRC_IDLE, RRC_INACTIVE and RRC_CONNECTED state.
- In the case of multiple 5GC sharing the same NG-RAN, the NG-RAN provides UAC for each PLMN individually.

2. Explain Interworking with EPC

7.8 Interworking with EPC

7.8.1 General

- To ensure successful interworking with the appropriate EPS functionality, only one PGW-C + SMF is allocated per APN for a given UE, and that is enforced e.g. by the HSS + UDM providing one PGW-C + SMF FQDN per APN to the MME.
- Interworking with EPC while using non-3GPP access in 5GS is also applicable and in such cases, NR would be replaced with N3IWF and access specific entities underneath e.g. Wi-Fi Access Point.
- Furthermore, it is also possible to interwork between EPC connected to non-3GPP while using 3GPP access toward the 5GC, and in such case the MME and SGW would be replaced with an ePDG and the HSS with a 3GPP AAA server.
- For interworking to be possible it is required that the UE supports both EPC NAS procedures as well as 5GC NAS procedures.

- If this is not the case, then the UE will be directed toward the Core Network that the UE supports, and no interworking will be applicable.

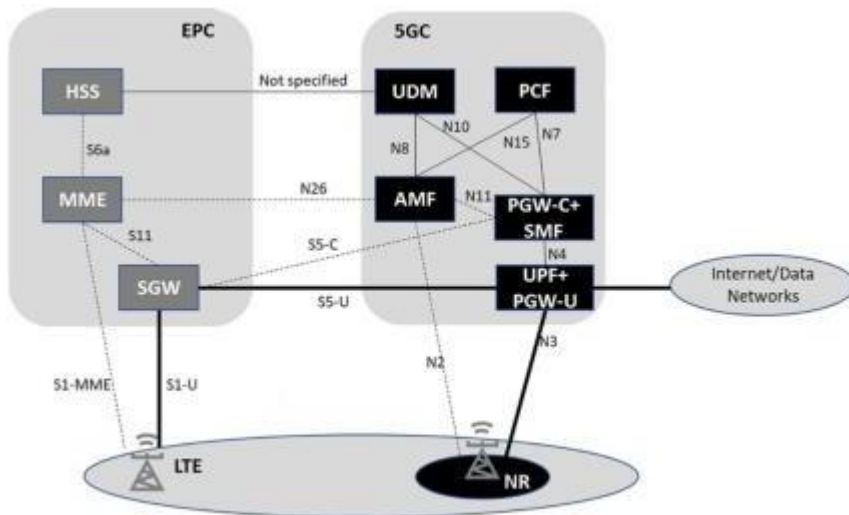


Fig. 7.7 Detailed architecture for interworking between EPC and 5GC.

3. Explain Interworking with EPC using 3GPP access

7.8.2 Interworking with EPC using 3GPP access

7.8.2.1 General

- When a UE is selecting networks – or PLMNs – or camping on a cell that is connected to both EPC and 5GC, the UE needs to select which Core Network to register with.
- That decision can be operator controlled or user controlled.

- The operator can control the decision e.g. by influencing the network selection using an operator controlled prioritized list in the USIM by which the operator is able to steer the network selection including which Access Technology e.g. NG-RAN or E-UTRAN to prioritize, or the operator can set the subscription to only allow either EPC, 5GC or both, or the operator can control RRM procedures per UE as to prioritize certain radio access to be used.
- Interworking with EPC is specified both with usage of N26 and without N26, and the UE may operate in single-registration mode or dual-registration mode for 3GPP access.

In single-registration mode;

- The UE has one active Mobility Management state for 3GPP access toward the Core Network and is either in 5GC NAS mode or in EPC NAS mode dependent on which Core Network the UE is connected to;
- The UE context information is transferred between the two systems when the UE moves back and forth, which either is done via N26 or by the UE moving each PDN Connection or PDU Session to the other system when interworking without an N26 interface.
- To enable the RAN in the target system to select the same Core Network entity which the UE was registered to in the source system (if it is available) and to enable the retrieval of UE context over N26, the UE maps the 4G-GUTI to 5G GUTI during mobility between EPC and 5GC and vice versa as described in Fig. 7.8.
- For handling of security contexts, Chapter 8 describes how to enable an efficient re-use of a previously established 5G security context when returning to 5GC.

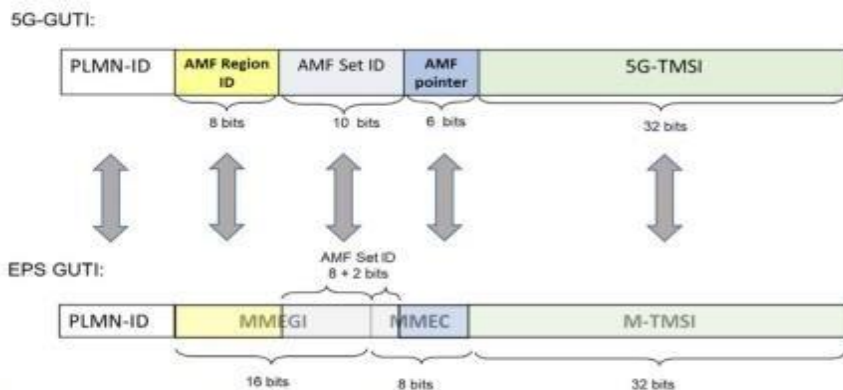


Fig. 7.8 Mapping between 5G-GUTI and EPS GUTI.

In dual-registration mode;

- The UE maintains independent Mobility Management states for 3GPP access toward the 5GC and EPC using separate RRC connections.
- In this mode, UE maintains 5G-GUTI and 4G-GUTI independently, and the UE may be registered to 5GC only, EPC only, or to both 5GC and EPC.
- It should be noted that N26 is used only for 3GPP access.
- Mobility of PDU Sessions between 3GPP access and non-3GPP access in the EPC and 5GC systems are driven by the UE and is supported without N26.
- When the UE moves from one system to the other, the UE provides its UE temporary identity in the format of the target system.
- If the UE has previously been registered/ attached to another system or has not registered/attached at all in the target system and does not hold any UE temporary identity of the target system, the UE provides a mapped UE temporary identity as described in Fig. 7.9.

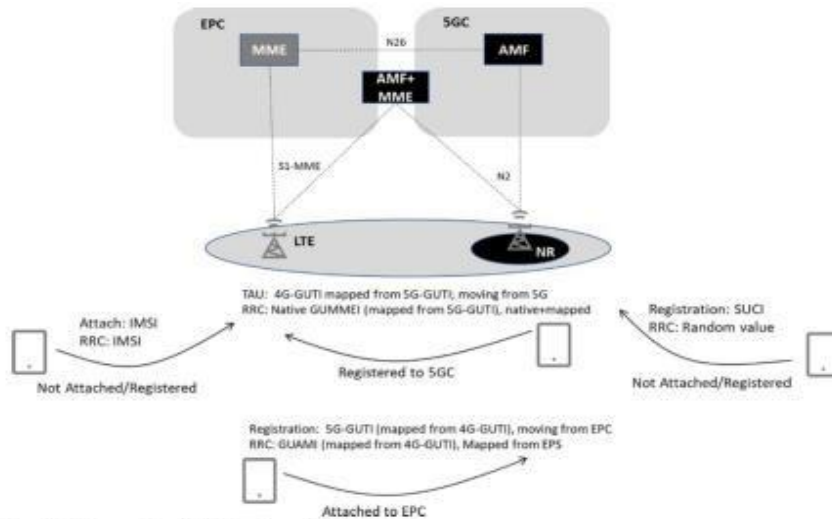


Fig. 7.9 UE provided UE identity at NAS and RRC.

- When the UE initially attaches to EPS the UE uses its IMSI as UE identity toward both E-UTRAN (in RRC) and EPC (in NAS).
- However, in 5GS, the UE uses a SUCI toward 5GC (in NAS) which conceals the UE's Identity. In both cases, there is no stored UE context in the network i.e. the network creates the UE context.

4. Explain interworking using the N26 interface

7.8.2.2 Interworking using the N26 interface

- When N26 interface is used for interworking procedures, the UE operates in single-registration mode, and the UE context information is exchanged over N26 between AMF and MME.
- The AMF and MME keeps one MM state (for 3GPP access) for the UE, i.e. either in the AMF or MME (and the MME or AMF is registering in the HSS + UDM when it holds the UE context).
- The interworking procedures provide IP address continuity at inter-system mobility between 5GS and EPS and are required to enable seamless session continuity (e.g. for voice services).
- The PGW-C + SMF keeps a mapping between PDN Connection and PDU Session related parameters e.g. PDN Type/PDU Session Type, DNN/APN, APN-AMBR/Session AMBR and QoS parameter mapping. Fig. 7.10 shows at a high level the interactions.

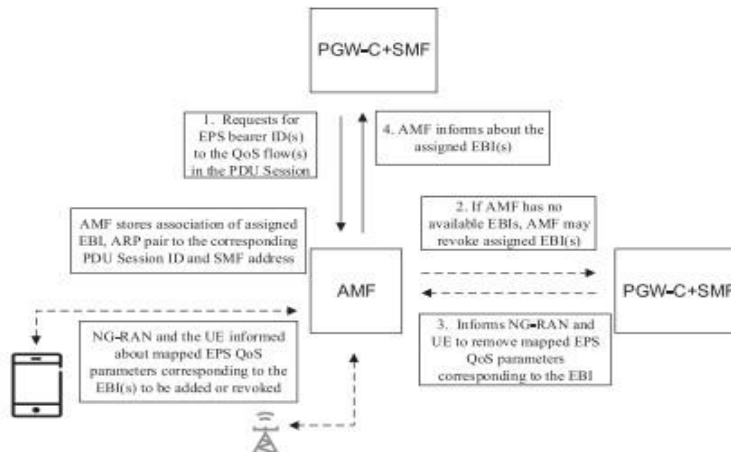


Fig. 7.10 EBI allocation and revocation.

5. Explain command and control.

4.1 Command and Control Method

- The conventional method for allocating spectrum is known as “Command and Control Method” shown in the Figure 4.1.
- There are some countries following this technique of spectrum allocation.
- In this method radio spectrum is divided into different spectrum bands that are allowed to specific radio communication services such as satellite services, mobile, broadcast on an exclusive basis.
- This method guarantees that the radio frequency spectrum will be exclusively licensed to an authorized user and can use spectrum without any interference.
- This method of spectrum allocation is not efficient because :
 - Spectrum assigned to a particular radio communication service cannot be replaced by other services even though it is witnessed that spectrum is underutilized.
 - There is no possibility of questioning the user once the spectrum is allocated to him (during the licensing period) as per the norms, provided he fulfills the terms and conditions.

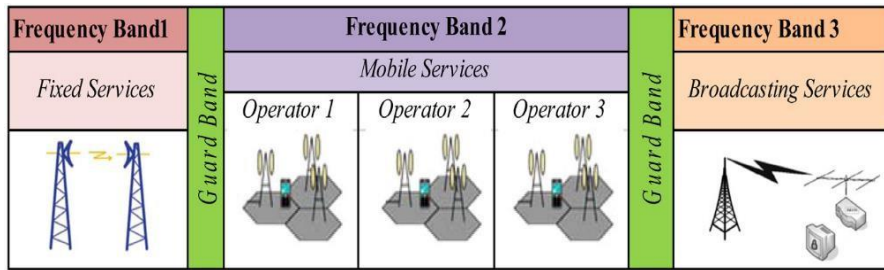


Figure 4.1 Command and Control Method.

- This method does not make spectrum to be utilized efficiently in rural areas as the spectrum utilization is heavy in the urban regions and underutilized in the rural areas.
- It is sad to see the spectrum underutilized and not accessible to all. It is difficult for some nations to provide 4G services itself.
- There is a need for taking serious steps in handling spectrum issues by implementing sophisticated technologies for the development of nation.
- In these cases the techniques like spectrum trading would be a successful solution. This will only lead to the development of 5G communications in these types of countries.

6. Explain spectrum sharing and trading/ Dynamic spectrum sharing and its advantages (Nov/Dec 23)

4.2 Spectrum Sharing

- The demand for multimedia content and information processing, services such as e-education and e-health, mobile broadcasting, enormous increase in the electronic gadgets necessitate efficient use of all available and usable frequency spectrum.
- The new generation of mobile broadband networks will require supporting higher data throughput rates.
- Many sophisticated technologies have been implemented for making the efficient use of available spectrum.

- For example, line-of-sight (LOS) systems are usable up to 100 GHz now. Running down the size of electronic components and systems introduces the multiple frequency bands in single equipment leading to the efficient use of available spectrum by the enhanced dynamic sharing of frequency bands.
- Spectrum management should be in such a way that there should be always optimum spectral sharing.
- Greater sharing of frequencies and bands allows more data to be sent by different users in the same amount of available spectrum.
- Spectrum sharing has basically three dimensions: frequency, time and location.
- The Collective Use of Spectrum (CUS) allows spectrum to be used by more than one user simultaneously without requiring a license.
- Some of the examples that come under spectrum sharing are frequency reuse concept in the existing telecom networks, FDMA, TDMA.
- Another important challenge is the sharing of spectrum among the heterogeneous networks.
- While it is easier to achieve efficient and successful spectrum sharing among the homogeneous or similar networks or applications, there arises complexity in heterogeneous networks.
- The spectrum sharing methods are classified into three categories based on based on the priority level of accessing the radio spectrum as follows :
 - a. Horizontal spectrum sharing: all the devices have equal rights to access the spectrum.
 - b. Vertical spectrum handover only: the primary users are allotted priorities to access the spectrum.
 - c. Hierarchical spectrum sharing: it is an enhanced variant of the vertical spectrum sharing.

4.2.1 Spectrum Using SDR and Cognitive Radio – Dynamic Sharing

- Evolution of software defined radio (SDR) and cognitive radio (CR) are the two major milestones in the mobile communications.
- Dynamic sharing of spectrum improves the spectrum efficiency and the above mentioned technologies play a vital role in this aspect.
- Conventionally, transmitters were tuned to specific frequencies, and facilities for multiple frequencies would cost high.
- But after the development of these technologies, tuning the transmitters to the multiple frequencies has become easier, i.e., switching to the different frequencies in a dynamic way would be possible at a reasonable cost.
- Cognitive radio first detects the occupation of the channel, and if it is occupied, it helps the users to switch to the other vacant channels.
- Also the carrier signals are sensed regularly for usage in other. There is always a need of large amount of spectrum in case of emergency or public safety conditions compared to that of normal conditions.
- In these emergency cases, dynamic sharing of spectrum would be a promising solution.
- In some countries spectrum regulators are used for the encouraging dynamic sharing spectrum with public safety requirements.
- It is to be noted that CR is a combination of administrative (regulatory), technical, and market based techniques to enhance the efficiency of spectrum utilization.
- Another area of utility for dynamic sharing is White Spaces (TV Band).
- Normally, the TV broadcasters repeat the same channel/carrier at relatively longer distances, to avoid any interference especially at the border/edger of the coverage areas that are on the border of two adjacent broadcast transmissions on same channel.
- However, there are very few receivers in this area, and the spectrum utility is not effective and could be utilized for other purposes.
- The broadcasters are generally quite protective for their signal Transmissions, even in areas beyond the theoretical coverage areas.
- Hence, only low power systems that cause minimal interference can be considered for shared usage with the TV spectrum.

- However, gradually with time building collective confidence amongst the users that includes the broadcasters, higher power based systems could be considered.

7. Explain Spectrum Trading

4.3 Spectrum Trading

- Spectrum trading is a case of spectrum sharing with the involvement of commercial activities.
- Spectrum trading is found to be a more economical way of efficient use of spectrum. It is an option through which flexibility can be increased and spectrum assigned to a particular service, and can be easily transferred for other usage.
- To explain it in brief, spectrum trading is a market based mechanism where buyers and the sellers determine the assignments of spectrum and its uses in which seller transfers the right of spectrum usage, in full or part, to buyer while retaining the ownership.
- In many countries spectrum trading is already running and the trading procedure is confined to specific bands, which are in demand for commercial use with specified conditions.
- Spectrum trading improves the efficiency and facilitates new services to enter in the market by making slight modification in the regulatory provisions.

The difference between spectrum sharing and spectrum trading can be explained as follows:

- In spectrum trading the usage rights are transferred completely from the seller for a specified period.

- However, in spectrum sharing buyer gets a temporary right of spectrum usage with the exclusive rights resting with the seller.
- Trading becomes effective only when it is clubbed with liberalization. Spectrum trading can be implemented if there is solid base in understanding advanced technologies and operating systems as the spectrum flexibility demands new approaches and practical methods for monitoring compliance, enforcement and conflict resolution.

4.3.1 Spectrum Trading Merits

The merits of spectrum trading are as follows:

- Improves efficient spectrum usage
- Facilitates the evaluation of spectrum licenses, and gaining knowledge of market value of spectrum
- Quicker process, with better and faster decision-making by those with information
- Removes barriers to entry by allowing small operators and start-ups to acquire spectrum rights of use more readily, thereby facilitating the development of market competition
- There is an opportunity for more rapid redeployment and faster access for spectrum
- Encourages new technologies to gain access to spectrum more quickly
- Existing operators gain an opportunity to sell unused or under-used spectrum and make more flexible use of spectrum
- Reduction in the transactions costs of acquiring rights to use spectrum
- Allows operators increased flexibility to accommodate shifting demand driven by market changes.

8. Explain Cognitive radio based on 5G.

4.4 Cognitive Radio

- IEEE approved definition of cognitive radio (CR) is a radio in which communication systems are aware of their environment and internal state, and can make decisions about their radio operation based on that information and predefined objectives.
- The environmental information may not include location information related to communication systems.
- Cognitive radio is a very good solution for increasing the spectrum utilization.
- Cognitive radios should be able to self-organize their communication based on sensing and reconfigurable functions as stated below :
 - ***Spectrum resource management:*** this scheme is necessary to manage and organize efficiently spectrum holes information among cognitive radios.
 - ***Security management:*** cognitive radio networks (CRNs) are heterogeneous networks in essence and this heterogeneous property introduces a lot of security issues. So this scheme helps in providing security functions in dynamic environment.
 - ***Mobility and connection management:*** this scheme can help neighbour- hood discovery, detect available Internet access, and support vertical handoffs, which help cognitive radios to select route and networks.

4.4.1 CR Device Concept

This section explains the features of CR whose implementation in a single device offers a very smart and high performance user terminal – CR terminal. The Figure 4.2 shows the CR properties.

A. Spectrum sensing

Spectrum sensing operation can be divided into the three step functions:

- *Signal Detection:* In this step of operation existence of the signal is sensed. There is no need to know the type of signal in this step.
- *Signal Classification:* In this step of operation the type of signal is detected, which is done by extracting the features of the signal.
- *Channel availability decision:* In this channel availability is detected. Once the free channels are detected, the step next to it is, sharing the spectrum holes which can be achieved by the spectrum allocation scheme.

The CR technology also brings new *security and pricing challenges* which are shown in the Figure 4.2.

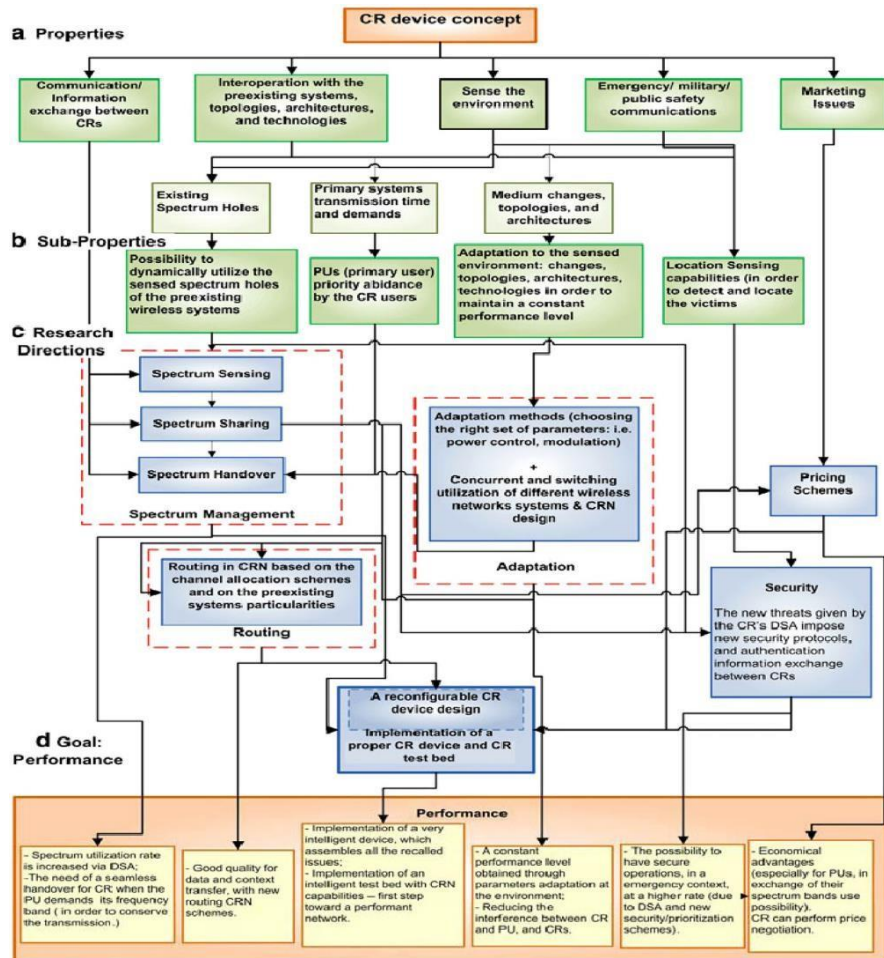


Figure 4.2 Concept of CR Device .

- New security threats appear with the dynamic spectrum access concept, as well as the CR's authentication needs.
- The pricing is very much influenced by the used channel allocation scheme. Additionally, CRs must be designed with strong capabilities to negotiate the available channels' price.

B. Spectrum handover

The phenomenon of frequency changing dynamically is said to be spectrum handover. A secondary user changes its frequency on appearance of a primary user or due to transmission degradation. This necessitates designing a handover scheme.

C. Environment adaptation

Different changes like topological changes, noise or interference power may occur while sensing the information. In order to adapt to these changes and to maintain the constant performance new adaptation techniques have to be implemented which is an important point of concern.

D. CR routing

CR routing is based on the requirement for CR device to interoperate with different systems, and is influenced by the spectrum sharing techniques.

CRNs inherit the PSs (Primary Systems) network characteristics: infrastructure - based, mesh, ad-hoc, sensor networks, etc. and these architecture types impose a specific routing algorithm, which must also include the CR devices and the possibility for a CR to be a relay node for another CR.

4.4.2 CR based on 5G

The various functionalities for 5G that could be met with CR usage are as follows:

- Advanced PHY and MAC technologies.
- Implementation of novel and flexible protocols.
- Capacity to support homogenous and heterogeneous systems.
- Adaptation to different changes like environment changes, dynamic frequency changes, etc.

Correlation between WISDOM and CR in reference to 5G could be given as:

- “5G brings the convergence concept through WISDOM and CR represents the technological tool to implement it.”
- The 5G technology eliminates the radio terminals that are specific to particular wireless technologies and proposes a universal terminal which must include all of the predecessor features in a single device.
- This terminal convergence is supported by the users’ needs and demands and is strongly found in CR terminal .

9. Discuss about Millimetre Waves.

4.5 Millimetre Waves

- Most of the radio communications including TV, satellite communications, GPS, Bluetooth are utilizing frequency band ranging from 300 MHz to 3 GHz.
- But this band is getting crowded and the focus is on releasing and utilizing the additional spectrum. In mm-waves are the promising solution for this problem.
- The spectrum bands identified under the IMT umbrella do not have the capacity to carry such enormous data required for 5G services.
- Therefore, mm-waves could be the candidate bands for 5G mobile communications due to high data carrying capacity. The mm-waves have the following advantages :
 - (a) Not much operation at mm-waves so more spectrum is available at mm- waves
 - (b) Very large blocks of contiguous spectrum to support future applications.
 - (c) Due to high attenuation in free space, frequency reuse is possible at shorter distance
 - (d) Spatial resolution is better at mm-waves hardware with CMOS technology
 - (e) Advancement in semiconductor technology allows low cost equipment

(f) Small wavelength makes possible use of large antenna arrays for adaptive beam forming

(g) Small size of antenna at mm-waves facilitates easy integration on chip and installation at suitable locations.

- In mm-waves allow larger bandwidth and offer high data transfer and low latency rate that are suitable for high speed reliable Internet services.
- The small wavelength facilitates small size antenna and other part of radio hardware, which reduces costs and also easy to install.
- The transmitter's antenna would be like a lamppost, which could be installed on building, street lamppost, etc.
- High directionality attained in this band can be used to increase spatial multiplexing.
- The size of antenna required for an mm-waves radio can be one-tenth or less of an equivalent lower frequency radio which is an advantage to the manufactures to build smaller and lighter systems.
- Beam width is the measure of how a transmitted beam spreads out as it gets farther from its point of origin.
- But due to limited availability of radio frequency (RF) bands the fifth generation wireless communication systems will move to ultra-high capacity mm-wave bands.
- High frequency makes mm-wave band more attractive for wireless communication system and these frequencies are used in terrestrial and satellite communications. Wireless products that use millimeter waves already exist for fixed, LOS transmissions
- But the absorption rate of the mm-wave electromagnetic signal poses great challenges for their utilization in the non-LOS and mobile connections.
- On the other hand, high directionality achieved in this band can be used to increase spatial multiplexing.

Wireless backhaul will be another key enabler of 5G-mm-wave small cells.

- Within the mm frequencies, the frequency band of 60 GHz has attracted the researchers to work with, as the large amounts of bandwidth are unallocated in this band, bandwidths that are required for communication systems at the intended data rates of 100 Mbps and above.

- Also, another advantage of 60-GHz band is due to a physical property of the propagation channel at this frequency that provides a natural way for reduction of frequency reuse factor, which tends to compact cell size.
- It is a general property of the mm-wave propagation that the behaviour of the propagation rays is well characterized by the geometric optics.
- That is, the waves do not penetrate the walls or other obstacles and wave reflection is the main mechanism leading to a multipath.
- In mm-waves have the potential to support broad-band service access which is especially relevant because of the advent of Broadband Integrated Service Digital Network (B-ISDN).
- With the development of personal wireless communication systems, two things are appearing to be significant:
 - Exploiting high frequency bands, such as mm-waves to provide broad-band for high rate data transmission.
 - To integrate multi-tasks in one system which greatly extend the application of wireless device.

The utility of mm-waves for the micro cells that form the WISDOM based GIMCV are well positioned to be served by these mm-waves. It has been elaborated in these following points:

- It is relatively easy to get licenses for big blocks of mm-wave spectrum, which would allow carriers to deploy large backhaul pipes over 1 Gbps in size. While a single small cell may not need that much capacity, the complexity of heterogeneous networks will require daisy-chaining many small cells together, each cell passing its load down the line.
- Small cell backhaul makes the best use high frequency characteristics of mm-waves. The higher the frequency the shorter distance a wave propagates unless it gets a serious power boost. But the heterogeneous network by definition will be composed of densely packed cells in urban environments, meaning no mm-waves will have to travel far between hops.

- The traditional uses of the mm-waves include radio navigation, space research, radio astronomy, earth exploration satellite, radar, military weapons and other applications.
- The backbone/backhaul networks (point to point network) for existing telecom network to connect base station to main switching centre (MSC), Local Multipoint Distribution System (LMDS), indoor WLAN, high capacity dense networks are also present in the mm-waves. The typical microwave backhaul bands are at 6.0 GHz, 11.0 GHz, 18.0 GHz, 23.0 GHz, and 38.0 GHz frequency bands.
- The light use of mm-waves could be attributed to high attenuation and low penetration. At such high frequency, waves are more prone to rain and other atmospheric attenuation.
- The wavelength is in the order of millimeters, and rain drops are also of the same size. Rains absorb high frequency waves and make it difficult for propagation.
- However, the experimental results show that in heavy rain condition, attenuation is 1.4 dB and 1.6 dB for 200 meters distance at 28 GHz and 38 GHz, respectively. The rain attenuations at 60 GHz for a rainfall rate of 50 mm/h, is approximately 18 dB/km. A proper link design with slightly high transmit power may take care of rain attenuation.
- Slight change in the position would affect the signal strength at the receiving end, due to which mm-waves are deeply affected by scattering, reflection and refraction. The root mean square (RMS) delay spread for mm-waves is of the order of few nano seconds, and it is high for non-LOS (NLOS) links than (LOS) links.
- Similarly, path loss exponent for NLOS links is higher than LOS links.
- Due to higher path loss and RMS delay spread, it is assumed that mm-waves are not suitable for (NLOS) links.
- However, these difficulties could be managed by using carrier aggregation, high order MIMO, steerable antenna, beam-forming techniques.
- The frequency bands around 60 GHz is best suited for pico and femto cell due to high data carrying capacity and small reuse distance due to strong oxygen absorption at the rate of 15 dB/Km.

- The usage in frequency bands around 60 GHz is highly sparse, which provides freedom to allocate a large bandwidth to every channel. Moreover, equipment can be made very compact due to the very small antenna size. The figure for mm wave measurement is shown below.

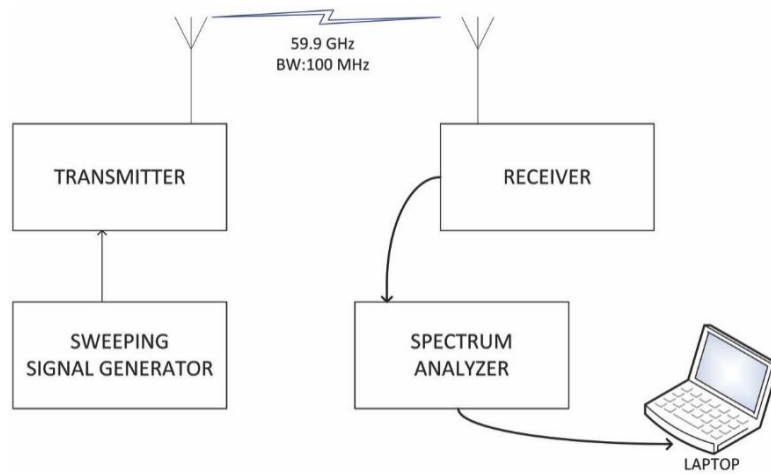


Figure 4.3 Measurement setup.

Part – A Two marks

1. What is mobility management?

The mobility is a core feature of 5GS. Mobility Management is required to ensure the following:

- That the network can “reach” the user, for example to notify the user about incoming messages and calls,
- That a user can initiate communication toward other users or services such as Internet access, and that connectivity and ongoing sessions can be maintained as the user moves, within or between access technologies.

2. What are the mobility management functions?

- Service Area Restriction
- Local Area Data Network (LADN)
 - Mobile Initiated Connection Only (MICO)

3. What is Service Area Restriction?

Mobility with session continuity is controlled at UE level at certain areas

4. Local Area Data Network (LADN)?

Mobility with session continuity is controlled at PDU Session level making communication available at certain areas

5. Mobile Initiated Connection Only (MICO)?

Paging capability (as part of the mobility service) is optional.

6. What are the 5G Mobility Management (5GMM) related procedures?

- ✓ Common procedures
- ✓ Specific procedures
- ✓ Connection management procedures

7. What is Registration and Mobility?

Radio networks are built by cells that range in size from tens and hundreds of meters to tens of kilometers and the UE updates the network about its location on a regular basis.

8. What is paging?

Paging is sending message and it is used to search for Idle UEs and establish a signalling connection. Paging is, for example, triggered by downlink packets arriving to the UPF.

9. What is Mobile Initiated Connection Only (MICO) mode?

Mobile Initiated Connection Only (MICO) mode was introduced to allow paging resources to be saved for UEs that don't need to be available for Mobile Terminating communication. When the UE is in MICO mode, the AMF considers the UE as unreachable when the UE is in CM-IDLE state.

10. What is RRC Inactive?

It includes support for efficient communication with minimal signalling by using a concept called RRC Inactive which affects the UE, NG-RAN and 5GC. RRC Inactive is a state where a UE remains in CM-CONNECTED state (i.e. at NAS level) and can move within an area configured by NG-RAN (the RAN Notification Area – RNA) without notifying the network.

11. How are the 5GC assistance information divided?

- ✓ Core Network assisted RAN parameters tuning.
- ✓ Core Network assisted RAN paging information.
- ✓ RRC Inactive Assistance Information.

12. What is Service Area and Mobility Restrictions?

Mobility Restrictions enables the network, mainly via subscriptions, to control the Mobility Management of the UE as well as how the UE accesses the network. Similar logic as used in EPS is applied in 5GS, but with some new functionality added as well.

13. What is RAT restriction?

Defines the 3GPP Radio Access Technologies a UE is not allowed to access in a PLMN and may be provided by the 5GC to the NG-RAN as part of the Mobility Restrictions. The RAT restriction is enforced by the NG-RAN at connected mode mobility.

14. What is Forbidden Area?

A Forbidden Area is an area in which the UE is not permitted to initiate any communication with the network for the PLMN.

15. What is Core Network type restriction?

Defines whether UE is allowed to access to 5GC, EPC or both for the PLMN.

16. What is Service Area Restriction?

Defines areas controlling whether the UE is allowed to initiate communication for services.

17. What is Allowed Area?

In an Allowed Area, the UE is permitted to initiate communication with the network as allowed by the subscription.

18. What is Non-Allowed Area?

In a Non-Allowed Area a UE is “service area restricted” meaning that neither the UE nor the network is allowed to initiate signalling to obtain user services (both in CM-IDLE and in CM-CONNECTED states).

19. What is control overhead?

5GS supports the ability to control the amount of load UEs produce toward the 5GS through different mechanisms. Mechanisms for 5GC to balance load across NFs and also to scale the amount of resources consumed for the NFs are often enough to handle normal fluctuations of load impacting the 5GC.

20. What are the mechanisms used to control overload in 5G?

- ✓ Congestion in control channel resources
- ✓ Congestion in random access channel (RACH) resources
- ✓ Release/reject UE RRC connection

21. What is Congestion in control channel resources?

5QI-based scheduling controls cases when e.g. the number of users awaiting scheduling exceeds the number of users that can be admitted such that the random access procedure fails.

22. What is Congestion in random access channel (RACH) resources?

Random access back-off. This pushes some UEs into a longer back-off. This is when there are so many access attempts on the RACH that the UE provided preambles cannot be detected anymore.

23. What is Release/reject UE RRC connection?

If there are not enough resources to process RRC connection requests, Releasing RRC connection or rejecting RRC connection attempts can be used.

24. What is Unified Access Control?

EPS supports multiple variants of access barring mechanisms as they were developed in different releases to address different needs for congestion control. The 5GS supports one mechanism called Unified Access Control (UAC) which is extensible, flexible (e.g. each operator can define their own category when to apply access control) and supports a variety of scenarios.

25. What are the Interworking with EPC using 3GPP access options?

- ✓ In single-registration mode
- ✓ In dual-registration mode

26. What is Command and Control Method?

The conventional method for allocating spectrum is known as “Command and Control Method. There are some countries following this technique of spectrum allocation. In this method radio spectrum is divided into different spectrum bands that are allowed to specific radio communication services such as satellite services, mobile, broadcast on an exclusive basis.

27. What is Spectrum Sharing?

It is sharing of available spectrum. The demand for multimedia content and information processing, services such as e-education and e-health, mobile broadcasting, enormous increase in the electronic gadgets necessitate efficient use of all available and usable frequency spectrum.

28. What is Spectrum Trading?

Spectrum trading is a case of spectrum sharing with the involvement of commercial activities. Spectrum trading is found to be a more economical way of efficient use of spectrum. It is an option through which flexibility can be increased and spectrum assigned to a particular service, and can be easily transferred for other usage.

29. List the merits of spectrum trading?

- Improves efficient spectrum usage
- Facilitates the evaluation of spectrum licenses, and gaining knowledge of market value of spectrum
- Quicker process, with better and faster decision-making by those with information
- Removes barriers to entry by allowing small operators and start-ups to acquire spectrum rights of use more readily, thereby facilitating the development of market competition
- There is an opportunity for more rapid redeployment and faster access for spectrum

30. Define Cognitive Radio.(Nov/Dec 23)

IEEE approved definition of cognitive radio (CR) is a radio in which communication systems are aware of their environment and internal state, and can make decisions about their radio operation based on that information and predefined objectives.

31. What are the self-organizing features of Cognitive radios?

- ✓ Spectrum resource management
- ✓ Security management
- ✓ Mobility and connection management

32. What is Spectrum Handover?

The phenomenon of frequency changing dynamically is said to be spectrum handover. A secondary user changes its frequency on appearance of a primary user or due to transmission degradation. This necessitates designing a handover scheme.

33. Define Millimeter (MM) Wave (Nov/Dec 23)

Millimeter wave (MM wave), also known as millimeter band, is the band of spectrum with wavelengths between 10 millimeters (30 GHz) and 1 millimeter (300 GHz). It is also known as the extremely high frequency (EHF) band by the International Telecommunication Union (ITU).

34. List the Advantages of millimeter (MM) wave

Millimeter wave is a band of electromagnetic spectrum that can be used in a broad range of products and services, such as high-speed, point-to-point wireless local area networks (WLANs) and broadband access. In telecommunications, millimeter wave is used for a variety of services on mobile and wireless networks, as it enables higher data rates than at lower frequencies, such as those used for Wi-Fi and current cellular networks.