1

NG-RAN Network – Functional Architecture

COPYRIGHTED MATERIAL

1.1. Functional architecture NSA/SA

Unlike previous generations of mobile networks, the deployment of 5G does not require the simultaneous implementation of the 5G core network (5GC) and the NG-RAN (Next-Generation Radio Access Network).

NSA (Non-Standalone Access) and SA (Standalone Access) are two 5G network models:

– SA is a completely new core service-based architecture: each radio node is autonomously controlled by the 5G core network. A service-based architecture delivers services as a set of NFs (Network Functions). NFs in the 5G core network are cloud native;

– NSA relies either on the 4G core network or on the 5G core network. NSA anchors the control signaling to the core network through a radio MN (Master Node). The MN is either a 4G radio node or a 5G radio node. The MN controls an SN (Secondary Node) (4G radio node or 5G radio node) according to the DC (Dual Connectivity) mechanism.

The 5G-SA architecture requires the deployment of a 5G core network connected to the NG-RAN.

The 5G-NSA architecture and the 5G-SA architecture were introduced in Release 15 of the 3GPP standard.

For a color version of all the figures in this chapter, see www.iste.co.uk/launay/5g.zip.

The 5G-NSA configuration implements the MR-DC (*Multi Radio Dual Connectivity*) architecture.

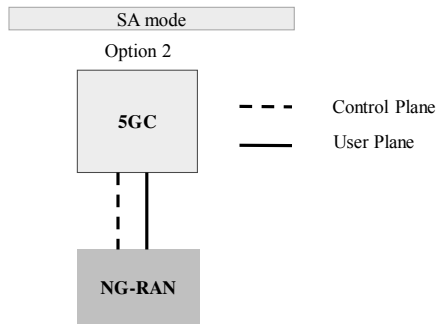


Figure 1.1. *Deployment in the SA mode*

Dual connectivity involves two RAN nodes, i.e. master node (MN) and secondary nodes (SN) which has the following features:

– the MN is connected to the core network for the control plan (signaling) and for the user plane;

– the SN is controlled by the MN. It is connected to the MN for the control plane (C-plane). The user plane (U-plane) is either connected to the MN or connected to the core network;

– the master radio access node controls the secondary radio access node and establishes a bearer, if necessary, for the exchange of data between the two radio nodes.

Dual connectivity defines the "Master Cell Group (MCG) bearer" and the "Secondary Cell Group (SCG) bearer".

The MCG carries data that will be transmitted on the radio resources allocated by the MN. In the case of carrier aggregation, the MN supports data on the PCell (Primary Cell) and SCells (Secondary Cells).

The SCG carries data that will be transmitted on the radio resources allocated by the SN. In the case of carrier aggregation, the SN supports data on the PCell and SCell.

The split bearer consists of routing the traffic between the MN and the SN. According to the U-plane termination, the split bearer consists of splitting either the MCG bearer or the SCG bearer.

For E-RABs configured as “MCG bearers”, the U-plane termination point is located at the MN.

For E-RABs configured as “SCG bearers”, the U-plane termination point is located at the SN.

For the core network configuration, each support (MCG, SCG, split bearer) can end on the MN and/or on the SN. The split bearer is transparent for the core network entities.

Several deployment scenarios (Figure 1.2) have been defined for the 5G-NSA:

– option 3: the E-UTRAN access network is connected to the 4G core network. The master node is the 4G radio node (*eNB – evolved Node B*). The secondary node is the 5G radio node (en-gNB). The MR-DC architecture is called *EN-DC* (*E-UTRAN NR Dual Connectivity*);

– option 4: the NG-RAN access network is connected to the 5G core network. The master node is a 5G radio node (*gNB – next generation Node Base Station*). The secondary node is a 4G radio node (ng-eNB). The MR-DC architecture is called *NE-DC* (*NR – E-UTRAN Dual Connectivity*);

– option 7: the NG-RAN access network is connected to the 5G core network. The master node is a 4G radio node (*ng-eNB – next generation eNB*). The secondary node is a 5G radio node. The MR-DC architecture is called *NGEN-DC* (*NG-RAN E-UTRAN NR Dual Connectivity*).

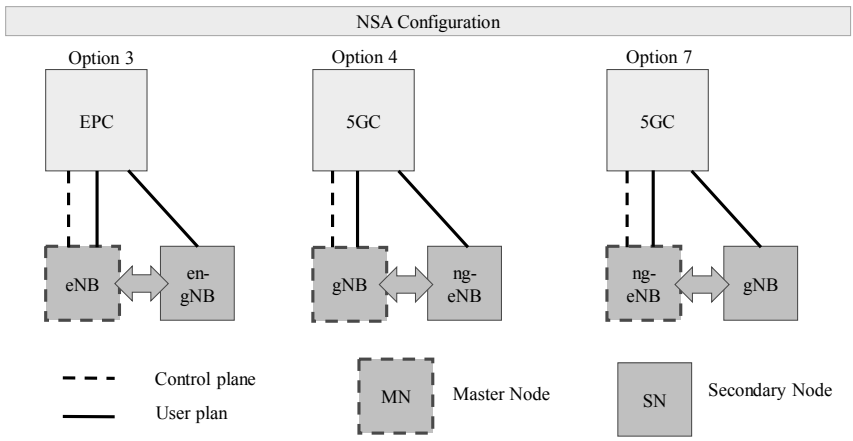


Figure 1.2. *NSA configuration options*

1.1.1. *Option 3*

Option 3 is the non-standalone EN-DC configuration.

Option 3 uses the MN (Master Node) terminated MCG (Master Cell Group) bearer for signaling: the eNB is the master node, and the gNB (gNodeB) acts as the secondary node. The radio access network is connected to EPC.

The 4G base station (eNB) controls the 5G base station (en-gNB) through the X2 interface.

The eNB supports signaling with the MME (*Mobile Management Entity*) through the S1-MME interface and supports the user plane traffic (*MCG bearer*) with the SGW (*Serving Gateway*) entity through the S1-U interface.

The en-gNB base station supports signaling with the eNB. The 5G-NR interface is activated by the eNB over the X2 interface. Once activated, en-gNB controls its own radio resource allocation. The user traffic is either transmitted from the eNB to en-gNB or transmitted from the 4G core network (SGW) over the S1-U interface to en-gNB.

The master node eNB exchanges data in both directions, uplink and downlink, with the mobile.

The secondary node en-gNB allows us to increase both uplink and downlink data rates.

With a DC mechanism, data is transmitted to the mobile according to one of the following variations (Figure 1.3):

– option 3: in plain option 3, all uplink and downlink data flows to and from the LTE part (*MCG split bearer*) of the LTE/NR base station, i.e. to and from the eNB. The eNB then decides which part of the data it wants to forward to the 5G gNB part of the base station over the Xx interface;

– option 3a: both LTE eNG (*MCG bearer*) and 5G-NR en-gNB (*SCG bearer*) exchange traffic to the 4G core network directly. This means that a data bearer allocated to a node cannot share its load over the second node. This option does not suit the case of mobile use;

– option 3x: user data traffic will directly flow to the 5G gNB part of the base station (*SCG split bearer*). The traffic is delivered over the 5G-NR interface to the device, and part of the data can be forwarded over the X2 interface to the 4G eNB.

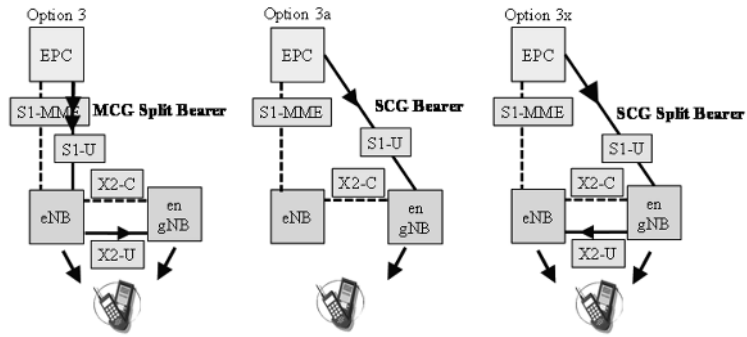


Figure 1.3. *Secondary node addition – option 3*

Option 3 uses the MN terminated MCG bearer for user traffic. The eNB entity splits the S1 bearer into:

– LTE radio support;

– NR support.

Option 3x uses the SN terminated SCG bearer for user traffic. The gNB entity splits the S1 bearer into:

– LTE radio support;

– NR support.

1.1.2. *Option 4*

Option 4 relies on the 5G core (5GC).

The gNB acts as an MN; it supports signaling exchange (*MCG signaling bearer*) with the 5G core network's transport plane through the NG-C interface. The LTE user plane connections go via the 5G-NR through the NG-U interface.

The ng-eNB base station acts as a secondary node. It is controlled by the gNB base station through the Xn-C interface. The ng-eNB is a new generation of the 4G base station.

The gNB controls the ng-eNB through the Xn interface.

The data is transmitted to the ng-eNB entity via one of the following options (Figure 1.4):

– from the master node gNB, which performs the split bearer (option 4, *MN terminated split bearer*);

– from the 5GC network (option 4a, *SCG bearer*).

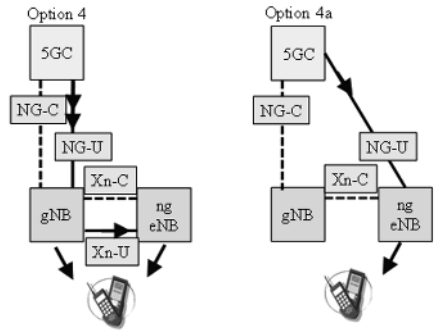


Figure 1.4. *NE-DC architecture – option 4*

1.1.3. *Option 7*

Option 7 relies on the 5G core (5GC).

The ng-eNB acts as an MN; it supports signaling (*MCG signaling bearer*) with the 5GC core network's transport plane through the NG-C interface and exchanges data to the 5G core network's user plane through the NG-U interface.

The gNB base station acts as an SN. It is controlled by the ng-eNB base station via the Xn-C interface.

The ng-eNB (4G base station) controls the gNB through the Xn interface.

The data is transmitted to the gNB entity via one of the following options (Figure 1.5):

– from the ng-eNB base station, which performs a split bearer. This is option 7 (*MN terminated split bearer*);

– from the 5G core network. This is option 7a (*SCG bearer*).

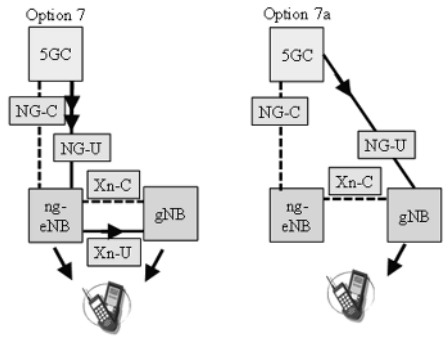


Figure 1.5. *NE-DC architecture – option 7*

1.2. Description of the NG-RAN network

The NG-RAN provides both NR and LTE radio access.

An NG-RAN node is either a gNB (5G base station), providing NR user plane and control plane services, or an ng-eNB (new generation 4G base station) providing the LTE/E-UTRAN services towards the UE (control plane and user plane).

The NG-RAN ensures the connection of mobiles and the reservation of radio resources between:

– the mobile and the ng-eNB base station on a single 4G carrier (LTE) or on several 4G frequency carriers (LTE-Advanced);

– the mobile and the gNB base station on one or more 5G frequency bands (5G-NR).

The gNBs and ng-eNBs are interconnected through the Xn interface. The gNBs and ng-eNBs are also connected, via NG interfaces, to the 5G core (5GC).

The NG interface is the point of reference between the NG-RAN and the 5G core network:

– the NG-C interface is the interface between the radio node and the AMF (Access and Mobility Management Function). It supports signaling via NG-AP (Next Generation Application Protocol);

– the NG-U interface is the interface between the radio node and the UPF (User Plane Function) for tunneling traffic (the IP packet) via GTP-U (GPRS Tunneling Protocol).

The UPF is configured by the SMF (Session Management Function) under the control of the AMF.

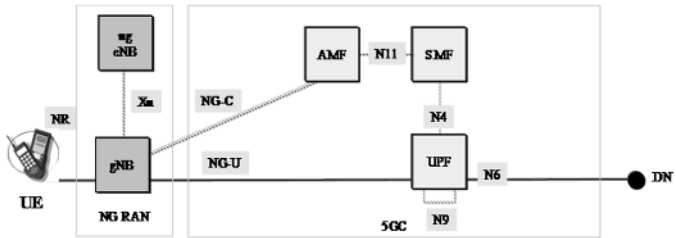


Figure 1.6. *NG-RAN general architecture*

The mobile exchanges data with the DN (Data Network) through logical connections called PDU (Protocol Data Unit) sessions. This logical connection is divided into two parts:

– the NG-RAN ensures the connection of the mobiles with the base station and interconnects the control plane and user plane (traffic) of the mobile UE with the core network;

– the 5G core network interconnects the NG-RAN, provides the interface to the DN, ensures the registration of mobiles, the monitoring of their mobility and the establishment of data sessions with the quality of the corresponding QoS (Quality of Service).

1.2.1. *The NG-RAN*

The NG-RAN provides both an LTE radio interface and a 5G-NR radio interface.

An NG-RAN node is:

– a 5G base station (gNB), which provides the control plane services and the transmission of user plane data through the 5G-NR radio interface;

– an advanced 4G base station (ng-eNB), providing control plane services and data transmission from the user plane to mobiles via the LTE radio interface.

The NG-RAN node is responsible for managing radio resources, controlling the radio bearer establishment of the user plane and managing mobility during the session (handover). The mobile connects to one of the radio nodes.

The NG-RAN node transfers the traffic data from the mobile to the UPF and data from the UPF to the mobile.

When the NG-RAN node receives data from the mobile or from the UPF, it refers to the QFI (QoS Flow Identifier) for the implementation of the data scheduling mechanism.

For outgoing data to the UPF entity, the NG-RAN node performs the marking of the DSCP (DiffServ Code Point) field of the IP (Internet Protocol) header, based on the assigned QFI.

The NG-RAN node performs compression and encryption of traffic data on the radio interface. It can also optionally perform the integrity control of the traffic data exchanged with the mobile.

The NG-RAN node performs the encryption and integrity control of the signaling data exchanged with the mobile on the radio interface.

The NG-RAN node performs the selection of the AMF. The AMF is the function of the core network to which the mobile UE is attached.

The NG-RAN broadcasts the RRC paging received from the AMF.

The NG-RAN node also broadcasts the cell's system information, containing the radio interface characteristics. The devices use these parameters for cell selection and for radio bearer establishment requests.

When a mobile is connected, the NG-RAN uses the measurements made by the mobile to decide on the initiation of a cell change during a session (handover).

In order to manage the services for each connected mobile, the NG-RAN node maintains a UE context information block relating to each mobile. The information saved by the radio node may depend on the mobile usage.

The mobile is either in the RRC connected state (RRC_CONNECTED), the RRC inactive state (RRC_INACTIVE) or the standby RRC state (RRC_IDLE).

When the mobile enters the standby state, the base station is not aware of its presence. Each mobile in the standby state listens to the information broadcasts by the radio node.

There is no UE context at the radio node for the mobile in the RRC_IDLE state.

When the mobile enters the RRC_CONNECTED state or the RRC_INACTIVE state, a mobile radio identifier C-RNTI or I-RNTI, respectively, (Connected/Inactive Radio Network Temporary Identifier) is saved at the radio node. The context of the UE is saved in relation to the RNTI. The context is recorded at the level of the NG-RAN node, which manages the mobile (source node), and it is transmitted to the target node in the event of a handover. The UE context is also created at the level of the MN and at the level of the SN in the event of dual connectivity.

When a mobile is in the connected mode, the NG-RAN node uses measurements made by the mobile to decide whether to trigger a change of node during the session (handover) or to activate or deactivate secondary cells.

1.2.2. *AMF (Access management and Mobility Function)*

The AMF (Access management and Mobility Function) supports:

– the registration of the mobile;

– the access control and the management of mobility on both the NG-RAN and Wi-Fi network access (non-3GPP access);

– network slicing.

The mobile and the AMF exchange data using the NAS (Non-Access Stratum) protocol.

The registration function allows the attachment of the mobile, the detachment of the mobile and the update of its location.

During the attachment, the AMF records the TAI (Tracking Area Identity) location and private identity of the mobile and assigns a 5G-GUTI (5G Globally Unique Temporary Identifier) to the mobile.

5G-GUTI replaces the encrypted private identifier SUCI (Subscription Concealed Identifier) and the private identifier SUPI (Subscription Private Identifier).

Once the attachment procedure is completed, the AMF selects the SMF, according to the DNN (Data Network Name) and the network slice indicator NSSAI (Network Slice Selection Assistance Information).

A load balancing procedure is applied when different SMF can be selected.

The DNN is either communicated by the mobile to the AMF during attachment, or retrieved from the subscriber's profile from the UDR (Unified Data Repository).

The AMF manages a list of TAIs allocated to mobiles, in which the mobile, in the standby state, can move without contacting the AMF to update its location.

The AMF manages the addition and removal of the TNL (Transport Network Layer) association with the entities of the NG-RAN node. In the event of a handover, the source AMF will release the TNL association with the source NG-RAN node and redirect the TNL association to the target NG-RAN node.

1.2.3. *SMF (Session Management Function)*

The SMF (Session Management Function) is responsible for creating, updating and removing PDU (Protocol Data Unit) sessions and managing session context with the UPF (User Plane Function). The SMF injects routing rules to the selected UPFs.

A routing rule corresponds to an entry in the context table of the UPF. This context table contains four fields:

– a correspondence field (PDR (Packet Detection Rule));

– a routing field NH (next hop: IP address, tunnel number TEID (Tunnel End Identifier) or SR (Segment Routing)) to find the next node;

– the quality of service to be applied to the flow (QER (QoS Enhancement Rules));

– the measurement reports to be applied to the flow (URR (Usage Reporting Rules)).

The SMF is responsible for the session management for each DNN and by network slice (S-NSSAI), based on the user profile stored at the UDR.

When requesting a session to be established, the SMF selects a UPF or queries the NRF (Network Repository Function) to obtain the address of the UPF.

The SMF grants an IPv4 or IPv6 address to the mobile. An IP address is provided for each PDU session, based on the address range of the PSA (PDU Session Anchor) selected to join the IP data network. The address range is obtained by either directly querying the selected UPF or by querying the NRF. If the assigned

IPv4 address is a private address, the UPF entity performs NAPT (Network Address and Port Translation) in order to translate the IP address and TCP (Transmission Control Protocol) or UDP (User Datagram Protocol) port numbers.

At the end of the IP session, when the mobile enters the standby state, the SMF releases the session by removing the context at the UPF.

In the event of incoming packets, if the mobile is in the idle state, the SMF sends a notification to the AMF (Downlink Data Notification).

1.2.4. *UPF (User Plane Function)*

The UPF (User Plane Function) manages the routing of user traffic and implements traffic filtering functions.

The PSA UPF is the traffic gateway connecting the 5GC network to the DN (Data Network). The PSA constitutes the anchor point for inter-UPF mobility.

The UPF is the anchor point for traffic when the mobile is moving from one NG-RAN node to another.

The UPF measures the quantity of data consumed for each UE.

The UPF can also implement traffic optimization functions and NAT (Network Address Translation), either from a private IPv4 address to a public IPv4 address, or from an IPv4 address to an IPv6 address and vice versa.

When the UPF receives data from the DN:

– in the absence of a routing context concerning the incoming flow, the UPF informs the SMF. The UPF either stores the data or transmits it to the SMF;

– in the presence of a routing context stored at the UPF level concerning the incoming flow, the flow is either transmitted to an NG-RAN node or to another UPF.

The UPF implements traffic routing rules by configuring the DSCP field of the IP based on the QFI. The QFI is defined by QoS rules which are injected by the SMF to the UPF when establishment of a session is requested. For each incoming piece of data, the UPF performs a traffic inspection (DPI (Deep Packet Inspection))

and classifies the packets into IP flow groups according to the SDF (Service Data Flow) service templates.

The UPF is a branch point supporting the multi-homing function.

The UPF performs replications of the mobile traffic data within the framework of lawful interception.

1.3. Functional separation between the NG-RAN radio interface and the 5G core network

The AMF is in charge of managing the 5G core network and services. It authenticates and registers each mobile and manages their mobility. Once registered, the AMF authorizes services according to the user's profile.

Figure 1.7 summarizes the functions managed on the NG-RAN and on the 5GC.

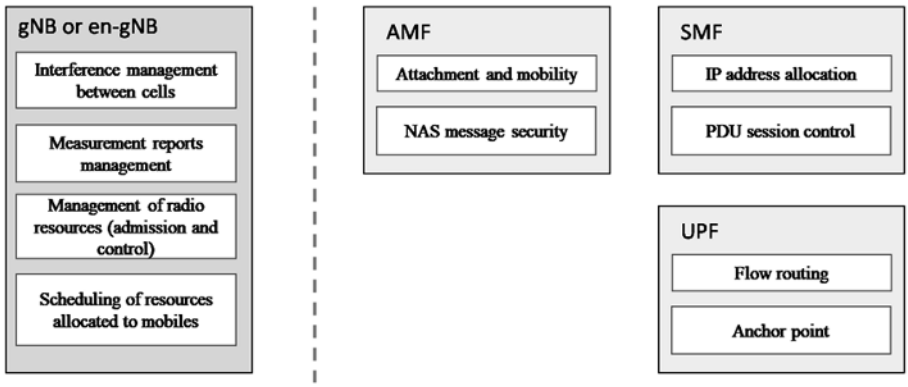


Figure 1.7. *The functional separation between NG-RAN and 5GC*

1.3.1. *Mobile identities*

1.3.1.1. *The identity of the mobile at the level of the AMF*

Registration procedure occurs when the mobile switches on. If authentication succeeds, the state of the mobile changes from the RM-DEREGISTERED state to

the RM-REGISTERED state and a user context (UE Context) is created on the AMF.

During the registration procedure, the AMF registers the IMSI (International Mobile Subscriber Identity) from its private identifier SUPI or private and hidden identifier SUCI.

The SUPI and SUCI identifiers allow the core network to identify the subscription associated with the mobile. The identifier format matches with the description of the NAI (Network Access Identifier) in order to be compatible with the DNS (Domain Name Server) servers by respecting the RFC7542 specification.

The SUPI identifier consists of two fields: the type of SUPI identifier (IMSI Identifier (International Mobile Subscriber Identity) or an identifier specific to the operator network) and the IMSI identity value or specific NAI value.

The SUCI identifier is made up of six fields defined as follows (Figure 1.8):

type <supi type>.hni <home network identifier>.rid <routing indicator>.schid <protection scheme id>.hnkey <home network public key id>.out <HPLMN defined scheme out>.



Figure 1.8. *The fields of the SUCI identifier*

The SUPI type value is used to indicate whether the SUPI identifier corresponds to the IMSI identity (type = 0) or a network-specific identifier (type = 1).

The hni field corresponds to the country code (MCC (Mobile Network Country)) and the operator code (MNC (Mobile Network Code)).

The ri field is defined over four digits. The default value is 0, but a specific value is used to identify on which partition of the UDR database the mobile subscription is stored (in the case of UDR, composed of different memory stacks).

The hnkey field identifies the key used for the encryption of the SUCI identifier.

The outfield is the result of the encryption of the mobile's IMSI identity (MSIN (Mobile Subscriber Identification Number)) including the message authentication value (MAC (Message Authentication Code)).

After authentication, the AMF provides a 5G-GUTI. Each 5G-GUTI is unique. This 5G-TMSI identifier is concatenated with the AMF identifier, in order to constitute the 5G-GUTI identifier (Figure 1.9).

The GUAMI identity corresponds to the AMF address. The GUAMI identity is the concatenation of the identity of the PLMN and the identity of the AMF:

<GUAMI> = <MCC><MNC><AMF Identifier>.

The 5G-S-TMSI identifier combines the AMF identity (AMF pool ID and pointer) as well as the 5G-TMSI identifier:

<5G-S-TMSI> = <AMF Set ID><AMF Pointer><5G-TMSI>.

The 5G-S-TMSI identifier is used as a radio identification for paging notifications.

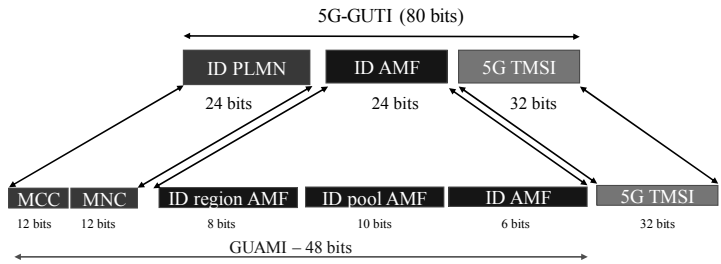


Figure 1.9. *The fields of the 5G-GUTI identifier*

Once registered, the connection state of the mobile at the AMF level is either the connected state (CM-CONNECTED) or the standby state (CM-IDLE).

In the CM-CONNECTED state, a NAS connection is established between the mobile and the AMF function. Because the NAS connection is encapsulated in the RRC message, the mobile is both in the CM-CONNECTED state with the AMF and in the RRC_CONNECTED state with the NG-RAN node. The NG-RAN node creates a UE context with RNTI (Radio Network Temporary Identifier).

After a period of inactivity, the NG-RAN node can:

– suspend the radio connection. The mobile goes to the RRC_INACTIVE state, the context is kept by the NG-RAN node and the UE context is kept at the AMF level (the mobile is still in the CM-CONNECTED state);

– release the radio bearer and remove the UE context. The mobile goes to the RRC_IDLE state and the UE context goes to the CM_IDLE state at the AMF level.

In the RRC_IDLE state, the mobile listens to the information sent by the NG-RAN node. In the RRC_IDLE state, cell reselection is managed by the mobile.

1.3.1.2. *The identity of the mobile at the level of NG-RAN*

Each NG-RAN node manages a set of mobiles. The NG-RAN node assigns a specific RNTI radio identifier to each mobile. A broadcast identifier is also used to broadcast information like the common control channel and information system.

Consequently, for any mobile in the standby mode, the NG-RAN node uses the following radio identities:

– P-RNTI (Paging) to send a paging request;

– SI-RNTI (System Information) to broadcast SIB messages;

– RA-RNTI (Random) to identify a mobile when requesting radio access (random access procedure; see Chapter 3);

– TC-RNTI (Temporary Cell) which helps to retrieve the information exchanged during the connection procedure, following the random procedure.

If the mobile is in the RRC_INACTIVE state (in the standby mode) on the NG-RAN, its identifier is called I-RNTI.

If the mobile is in the RRC_CONNECTED state, connected with the NG-RAN node, then several radio identities are used per mobile:

– C-RNTI (Cell): unique identification of the mobile for the RRC connection and for scheduling;

– CS-RNTI (Configured Scheduling): radio identifier of the mobile for the RRC connection and for semi-persistent scheduling;

– SP-CSI-RNTI (Semi-Persistent): radio identifier of the mobile used for the transmission of radio channel information (CSI: Channel State Information) on the uplink traffic channel with semi-persistent scheduling;

– MCS-RNTI (Modulation Coding Scheme): radio identifier of the mobile to indicate the modulation and coding scheme (MCS) of the mobile for the transmission of data on the downlink and uplink traffic links;

– TPC-RNTI (Transmit Power Control RNTI) is used to retrieve the information corresponding to the power control for both the traffic and uplink control channels.

1.3.2. *Mobile mobility*

1.3.2.1. *Mobility in the idle mode*

When the mobile is in the idle state, it is located by the AMF from the TAI (Tracking Area Identity) value.

The mobile listens to the broadcast information and to the paging information sent by the NG-RAN node.

If the mobile is in the discontinuous reception mode (DRX), the frame that carries the paging notification is calculated from the 5G-identity S-TMSI of the UE modulo 1024.

The mobile becomes aware of the paging when the content of the paging message carries the 5G-S-TMSI identifier or the I-RNTI identifier of the mobile.

The mobile informs the AMF of a change of tracking zone when it falls under the coverage of a radio node whose TAI is different from the previous TAI coverage zone.

The selection or reselection of cells is based on the transmitted synchronization signal CD-SSB (Cell Defining Synchronization Signal and PBCH Block) specific to each cell.

1.3.2.2. *Mobility in RRC_Inactive*

The mobile is in the RRC_INACTIVE state for the NG-RAN and in the CM_CONNECTED state for the 5G core network. The mobile context is always stored on the node where the mobile was connected.

The mobile can move in the RAN Notification Area (RNA) without notifying the NG-RAN nodes. An RNA zone comprises one or several cells located in the same TAI zone.

In the case of mobile terminating data, the UPF transfers the packet to the last NG-RAN node on which the mobile had established a radio connection. The NG-RAN node broadcasts a paging request over the air interface and transmits (through the Xn interface) a paging order to neighboring NG-RAN nodes that are configured on the same RNA area. If the paging request fails (the mobile is not reachable), then the NG-RAN node informs the AMF.

If the mobile establishes a radio connection with an NG-RAN node other than the last serving node, then the new NG-RAN node initiates a procedure to retrieve the context of the mobile UE from the old serving node. In the case where the request fails, the NG-RAN node triggers a new RRC connection.

1.3.2.3. *Mobility in the RRC_Connected state*

When the mobile is in the RRC_CONNECTED state and CM_CONNECTED state, its mobility is controlled by the NG-RAN node via the handover mechanism between a source radio node and a target radio node, or between two beams of the same radio node.

In the case of a handover, the NG-RAN node exchanges RRC signaling with the mobile. There are several types of handover:

– on the Xn interface between two neighboring connected nodes;

– on the NG interface with the UPF.

In the case of beam-based coverage, the beam selection is carried out at the MAC (Medium Access Control) layer of the mobile from mobile measurements. No RRC signaling occurs when the beam is changing since the mobile was configured at the beginning of the radio connection in terms of the measurements to be carried out (Measurement Objects).

The mobile makes radio measurements on one or more beams of a cell and determines:

– the radio quality of the beam by filtering the measurements on the L1 layer;

– the quality of the radio link of the cell by averaging the measurement of the different beams at the level of the L3-RRC layer.

The mobile also carries out measurements on the quality of the SSB block (Synchronization Signal and Broadcast) of the beams under the coverage of the cell, and also the quality of the neighboring intra- and inter-band cells. Each SSB measurement is seen by the mobile as a different cell.

Table 1.1 summarizes the different cases.

<table>
<tr><td>RRC_IDLE</td><td>RRC_INACTIVE</td><td>RRC_CONNECTED</td></tr>
<tr><td colspan="2">Cell selection is controlled by the UE in the function of radio access network parameters (AS: Access Stratum).</td><td rowspan="3">UE mobility is known by the core network.
The core network knows the identity of the radio node with which the terminal is connected.</td></tr>
<tr><td colspan="2">The UE listens to broadcast signal.</td></tr>
<tr><td>Paging notifications are initiated by the core network.</td><td>Paging notifications are initiated by the NG-RAN node.</td></tr>
<tr><td rowspan="2">The mobile has a temporary identity created by the core network.
No UE context is stored on the NG-RAN node.</td><td>The NG-RAN node knows the RNA location area on which the mobile is camped</td><td rowspan="2"></td></tr>
<tr><td colspan="2">The mobile radio context is stored at the mobile side and the NG-RAN node side. The core network and the NG-RAN node exchange information through NG-C and NG-U interfaces.</td></tr>
</table>

Table 1.1. *RRC mobile states*

1.4. Scheduling and QoS

1.4.1. *Scheduling*

Scheduling enables us to share radio resources among all connected users. A scheduling task is performed in real time to share radio resources among all mobiles. Scheduling attribution is calculated based on:

– the quality of the radio link for each mobile;

– requirements in terms of the quality of service expected by each mobile;

– the state of the mobile buffer.

The scheduler is performed at the MAC layer of the NG-RAN node and defines the scheduling for downlink and uplink transmissions.

The quality of the radio link is used to define the modulation and coding scheme (MCS) for a given mobile, as well as the transmission power. The MCS is dynamically adjusted according to the HARQ (Hybrid Automatic Repeat Request) retransmission rate.

To meet mobile requirements, several strategies can be defined at the scheduler level:

– fairness, a strategy for which each mobile receives the same resource allocation regardless of the modulation scheme;

– proportional fairness which allocates more frequency resources for mobiles with less efficient MCS;

– round-robin, a strategy which consists of allocating equal resources to all mobiles;

– max-CQI, a strategy that aims to maximize cell capacity by prioritizing the allocation of radio resources to mobiles which have the best radio quality.

The scheduler is based on measurement reports:

– the state of the buffer BSR (Buffer Status Report);

– the quality of the radio link (CSI-RS);

– the rising power margin PHR (Power Headroom Reports);

– Inter-Cell Interference Coordination (ICIC) between NG-RAN nodes.

From this information, the scheduler defined:

– the frequency radio resources to be allocated for each mobile;

– the number of spatial layers that can be used, depending on the category of the mobile;

– the transmission TTI (Time Transmission Interval) instants.

The mobile listens to the information transmitted over the PDCCH (Physical Downlink Control Channel) logical control channel and decodes the information channel when it detects its radio identifier C-RNTI.

The scheduler makes its decisions at each slot. The duration of the TTI slot depends on the spacing between SCS (SubCarrier Spacing). While the scheduler decision is 1 ms in 4G, it is variable from 1 ms, 500 µs or 250 µs for 5G on the FR1 (Frequency Range 1) band, and can go down to 125 µs or 62.5 µs for the FR2 band.

Transmission on the downlink direction can be pre-empted for critical communication (low latency). The NG-RAN node informs the mobiles by transmitting the INT-RNTI identifier over the PDCCH control channel.

Semi-persistent scheduling allows us to periodically allocate radio resources for a mobile. The periodicity of the messages is transmitted over the RRC layer and the resource allocation is transmitted to the mobile via the CS-RNTI identifier.

1.4.2. *Support for quality of service on radio link*

The QoS (Quality Of Service) control consists of implementing the maximum quality of service applicable to a data flow.

Like the 4G mobile network, only the core network is aware of the service requirements: QoS management is under the control of the core network (AMF). The NG-RAN has no knowledge of the service to be managed. Thus, when establishing a PDU session, the AMF entity establishes QoS rules between the radio node and one or more UPF entities.

The PDU session carries IP flows with one or more different qualities of service for all flows. Each flow is associated with a QFI flow profile identifier. The flow profile corresponds to a QoS indicator (5QI: 5G QoS Identifier) and an allocation and retention priority (ARP). The QFI flag is unique within a PDU session. The flag is either configured during the PDU session establishment procedure or during the PDU session modification procedure.

The value of a QFI is configured by the AMF during the procedure of session establishment; the AMF querying the unified UDR database to know the user's authorized QoS. For the establishment of dedicated services, the SMF chooses the QoS characteristics (5QI/ARP) according to the values stored at the SMF or by querying the PCF entity. The 5QI/ARP combination defined by the PCF link in a PDU session is a QoS flow binding.

The 5QI indicator is a parameter standardized by the 3GPP standard, allowing us to define:

– the type of resource (Guaranteed Bit Rate or not): GBR or non-GBR;

– priority;

– the maximum transmission delay within the 5GS mobile network;

– the residual error rate.

The standardization of the 5QI value makes it possible to indicate how the flow is processed on each element of the user's plan, in order that processing is consistent between the entities of the 5G core network and of the NG-RAN access.

The 5QI indicator is identical to the QCI (QoS Class Identifier) indicator for the 4G network for non-critical services (indicator from 1 to 80). New QCI values (81–85) have been defined in the case of URLLC services to guarantee speed and critical delay (delay critical GBR).

5QI	Type	Priority	Packet Delay Budget	PLER	Services
1	GBR	2	100 ms	10^{-2}	Conversational Voice
2		4	150 ms	10^{-3}	Conversational Video (Live Streaming)
3		3	50 ms	10^{-3}	Real Time Gaming, V2X messages
4		5	300 ms	10^{-6}	Non-Conversational Video (Buffered Streaming)
65		0.7	75 ms	10^{-2}	Mission Critical user plane Push To Talk voice (e.g. MC-PTT)
66		2	100 ms	10^{-2}	Non-Mission-Critical user plane Push To Talk voice
75		2.5	50 ms	10^{-2}	V2X messages
5	Non-GBR	1	100 ms	10^{-6}	IMS Signaling
6		6	300 ms	10^{-6}	Video (Buffered Streaming) TCP-Based (e.g. www, email, chat, ftp, p2p, etc.)
7		7	100 ms	10^{-3}	Voice, Video (Live Streaming), Interactive Gaming
8		8	300 ms	10^{-6}	Video (Buffered Streaming) TCP-Based (e.g. www, email, chat, ftp, p2p, etc.)
9		9	300 ms	10^{-6}	Video (Buffered Streaming) TCP-Based (e.g. www, email, chat, ftp, p2p, etc.). Typically used as default bearer
69		0.5	60 ms	10^{-6}	Mission Critical delay sensitive signaling (e.g. MC-PTT signaling)
70		5.5	200 ms	10^{-6}	Mission Critical Data (e.g. example services are the same as QCI 6/8/9)
79		6.5	50 ms	10^{-2}	V2X messages
80		6.8	10 ms	10^{-6}	Low latency eMBB applications (TCP/UDP-based); Augmented Reality
82	Delay Critical GBR	1.9	10 ms	10^{-4}	Discrete Automation (small packets)
83		2.2	10 ms	10^{-4}	Discrete Automation (big packets)
84		2.4	30 ms	10^{-5}	Intelligent Transport Systems
85		2.1	5 ms	10^{-5}	Electricity Distribution-high voltage

Table 1.2. *5G QoS characteristics*

The ARP parameter allows the NG-RAN node to choose whether the bearer establishment request should be made or rejected in the event of congestion.

The QFI value is coded on 6 bits. The 5QI value is set between 1 and 85. For any 5QI value less than or equal to 64, the QFI indicator and 5QI can be the same.

When the mobile is in the RRC_CONNECTED state, the management of QoS rules is delegated to the 5G-NR radio interface.

A user's plane traffic in a PDU session with the same QFI flag is handled with the same traffic routing rules (e.g. sequencing rules, admission level).

The role of the radio node is to configure one or more radio data bearers (RAB: Radio Access Bearer) and to perform a mapping between the QFI and the bearer(s) from a TFT flow filtering template (Traffic Flow Template).

For uplink, there are two ways to control the mapping between the radio bearers and the QoS of IP flows:

– reflective QoS for which the mobile replicates QoS rules received in downlink (configuration of the TFT flow policy rules);

– explicit configuration for which the uplink QoS configuration is defined by configuring the radio bearer.

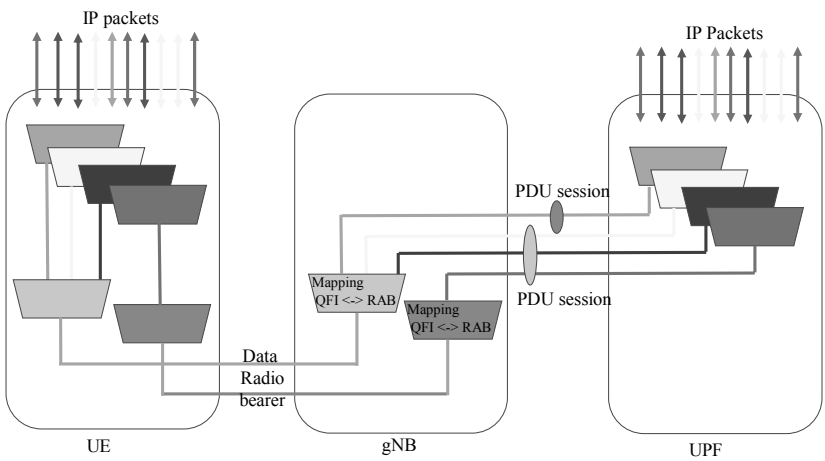


Figure 1.10. *QFI management in the user's plane*

1.5. Security architecture

The security architecture implemented on the 5G mobile is based on:

– mutual authentication between the 5GC core network and mobile (UICC);

– ciphering and integrity of NAS signaling messages exchanged between the mobile and the AMF;

– AS security through the 5G-NR radio interface between the mobile and the NG-RAN node. Security concerns the integrity control and encryption of RRC messages and IP packets. Integrity on IP packets is optional.

Data integrity:

– ensures that the data have not been altered by a third party between transmission and reception;

– verifies the transmitting source;

– ensures that a message already received is not reused.

Encryption ensures the confidentiality of data exchanged between two entities.

The security of the NAS and AS messages consists of deriving different keys at the level of the mobile and at the level of the following entities (Figure 1.11):

– The AMF:

– K_{AMF} key;

– K_{NASint} key from the K_{AMF} key for the integrity check of the NAS signaling;

– K_{NASenc} key from the K_{AMF} key for the encryption of the NAS signaling.

– The radio node:

– K_{gNB} key from the K_{AMF} key;

– K_{RRCenc} key derived from the K_{gNB} key for the encryption of RRC signaling on the 5G-NR interface;

– K_{RRCint} key derived from the K_{gNB} key for the integrity check of RRC signaling on the 5G-NR interface;

– K_{UPenc} key derived from the K_{gNB} key for encrypting IP traffic on the 5G-NR interface;

– optionally, a K_{UPint} key derived from the K_{gNB} key for the integrity check of IP traffic on the 5G-NR interface.

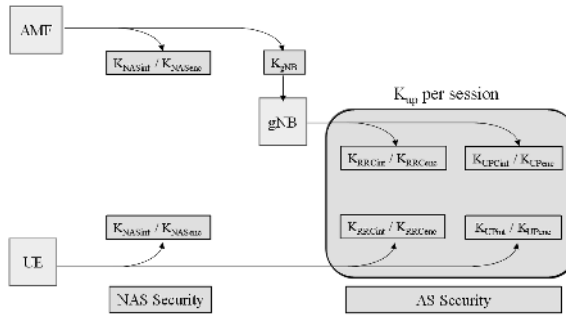


Figure 1.11. *Security architecture*

The mobile must support the NAS security based on information transmitted by the 5G core network and AS security, according to the indications sent by the NG-RAN access node.

5G security is based on the use of:

– NEA encryption algorithms (Encryption Algorithm for 5G);

– NIA (Integrity Algorithm for 5G) integrity control algorithms;

– the K_{UPenc}, K_{RRCenc}, K_{NASenc} encryption keys consist of 128 bits.

The encryption and integrity control algorithms are similar to those used on the LTE interface:

– NEA0/NIA0: no ciphering;

– 128-NEA1/128-NIA1: algorithm SNOW 3G (flow ciphering);

– 128-NEA2/128-NIA2: algorithm AES (bloc ciphering);

– 128-NEA3/128-NIA3: algorithm ZUC (flow ciphering).

Encryption and integrity are based on the following parameters:

– a 32-bit counter;

– the identity of the 5-bit bearer;

– the direction of the connection on one bit;

– the length of the message.

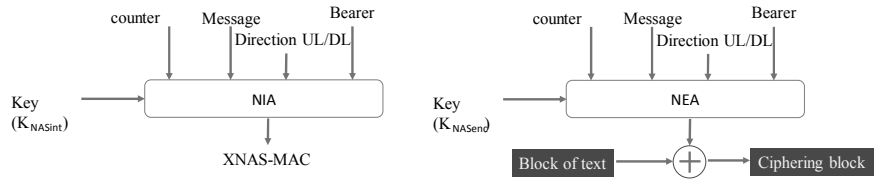


Figure 1.12. *Ciphering and integrity*

1.6. Network slicing

Network slicing is the embodiment of the concept of running multiple logical networks as virtually independent business operations on a common physical infrastructure in an efficient and economical way.

Virtualization is a hardware abstraction to partition network resources into distinct logical segments.

Network partitioning makes it possible to allocate a part of server hardware resources (NFVI: Network Function Virtualization Infrastructure) to network functions (VNF: Virtualized Network Functions).

Hardware capabilities are dynamically managed based on the number of users, on the one hand, and the profile of each user, on the other hand. By default, the 3GPP standard has defined four types of services:

– eMBB: evolved Mobile BroadBand to manage smartphone services such as high speeds, several session establishments, handover management, low latency;

– mMTC: massive Machine Type Communication to manage the sessions of IoT terminals (low speed, little transmission and mainly in the upstream direction, long delay);

– URLLC: Ultra-Reliable Low-Latency communication for critical communications requiring very low latency (less than 1 ms for the user plane) and efficient management of the handover;

– V2X: Vehicle to Everything for autonomous vehicles (between vehicles, with radio infrastructure, etc.).

Network slicing provides all the functionality of the 5G network, including optimization of the radio access network and core network entities to meet the service level agreement (SLA) requirements requested by the user.

Virtualization allows us:

– to allocate a set of material resources (storage capacity, network performance, computing capacity in terms of the number of CPUs);

– to deploy optimized software instances on the hardware resources. The instances correspond to the NFV network functions to be deployed:

- in the radio access network by dividing the radio functions into two entities gNB-CU and gNB-DU,

- in the 5G core network (AMF, SMF, PCF, etc.),

- to deploy optimized network functions (content cache, video optimizer, malware detection, etc.).

The set of hardware and software resources form a Network Slice Instance (NSI). The network instance is split into an RAN Slice Instance (RSI) and a Core Network instance.

From a user point of view, the mobile requests registration on a network instance from the 5GS network. The mobile profile allows the network to define the optimized network instances through the S-NSSAI (Single Network Slice Selection Assistance Information) identifier.

The S-NSSAI indicator is composed of two fields:

– SST: Slice Service Type defined user profile (1: eMBB, 2: URLLC, 3: mMTC, 4: V2X);

– SD: Slice Differentiator to differentiate specific services within an SST service type.

The S-NSSAI indicator is stored at the UDM database for each user profile and stored in the mobile. Each mobile can subscribe to up to eight S-NSSAI. S-NSSAI indicators are integrated into the NSSAI indicator.

When requesting registration, the mobile sends the desired NSSAI flag in the RRC request. The gNB-CU entity selects the AMF entity from the NSSAI indicator if possible; otherwise, it selects a default AMF entity. The AMF entity consults the UDM entity to know the value of the NSSAI indicator that will actually be implemented. The answer depends on the customer’s profile and the radio access network (3GPP, non-3GPP or roaming).

At the 5GC core network, when the mobile wants to establish a logical connection for a data service, it sends the RRC Service Request to the AMF with the S-NSSAI flag. The AMF selects the most suitable SMF.

On NG-RAN, the gNB-CU entity selects the gNB-DU entity based on the S-NSSAI flag. The division of the network on radio access is defined by:

– common RRC functions (management of sharing of radio resources between slices) and specific RRC functions (DRX, eDRX, timers, QoS, etc.);

– PDCP, RLC, common or specific functions (header compressions, acknowledgment, etc.);

– sequencing (MAC function) and prioritization on the physical layer.

The objective of virtualizing the network is to provide network flexibility and dynamic adaptation to the needs of different users, in order to meet performance indicators specific to the services requested (latency, throughput, packet loss, etc.). This flexibility is provided by an orchestrator which supervises the network functions and delegates the traffic management to the radio controllers (SD-RAN – Software Defined RAN) and network controllers (SDN: Software Defined Networking). In addition, virtualization allows isolation of the different slices.

1.7. References

All standards can be downloaded from the ETSI website: https://www.etsi.org/standards.

3GPP TS 37.340
Universal Mobile Telecommunications System (UMTS); LTE; 5G; NR; Multi-connectivity; Overall description; Stage-2
Version 15.3.0 Release 15

3GPP TS 38.401
5G - NG-RAN: Architecture description
Version 15.2.0 Release 15

3GPP TS 24.501
5G - Non-Access-Stratum (NAS) protocol for 5G System (5GS); Stage 3
Version 15.0.0 Release 15

3GPP TS 23.003
Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); Numbering, addressing and identification
Version 15.5.0 Release 15

3GPP TS 33.501
5G - Security architecture and procedures for 5G System
Version 15.2.0 Release 15