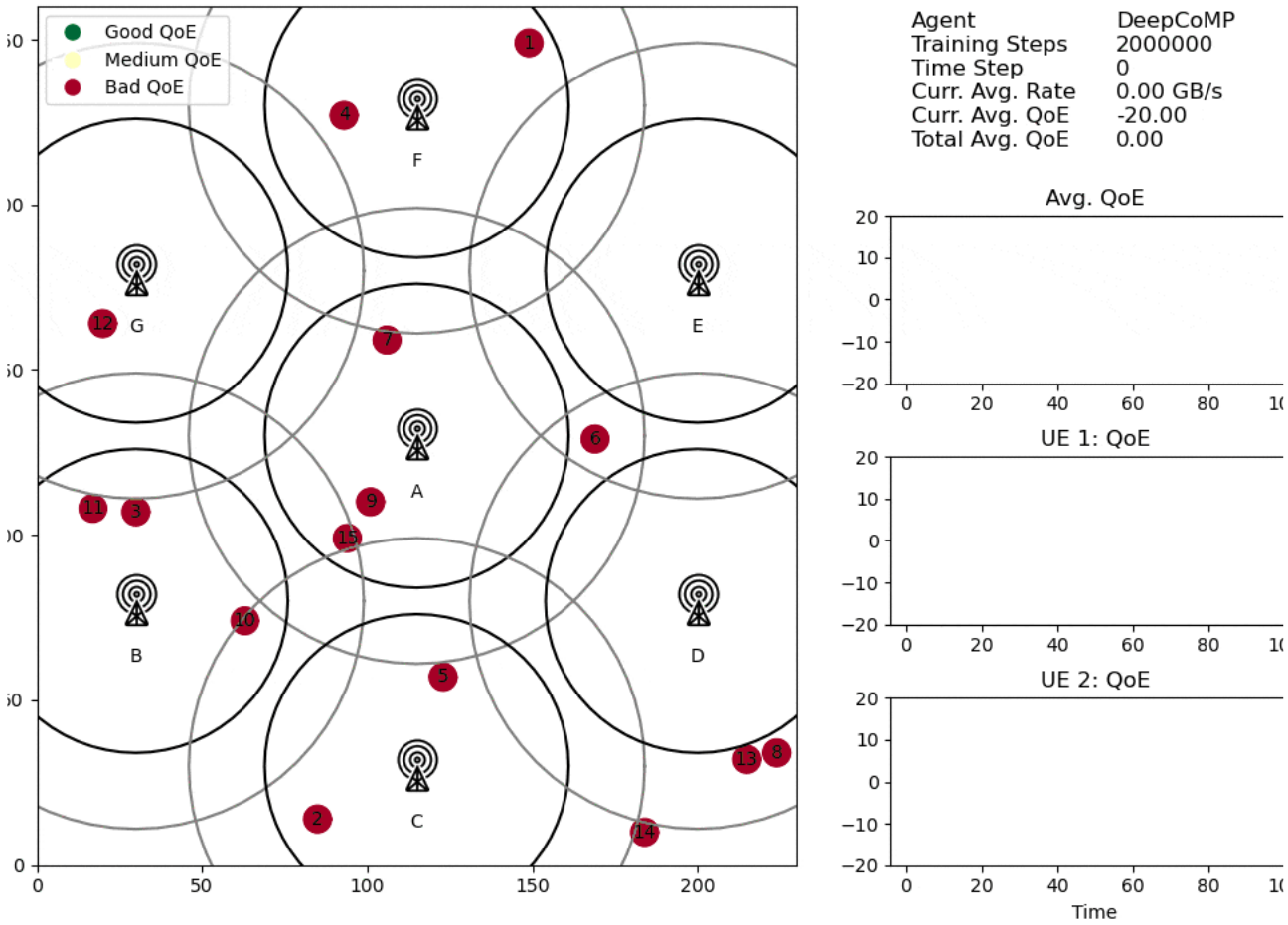
Mobility Management

Owned by David Cherney DISH •••
Last updated: Jul 30, 2023 • Add Workflow

Contents

Mobility management is a term that refers to how a network deals with a UE moving from one cell to another. To maintain connection, data flows need to be handed off between the relevant equipment; antenna to antenna, DU to DU, CU to CU, gNB to gNB and even between instances of network functions.

In the GIF below UE1 starts in the top center with a connection to cell F. It later has a connection with cell E and then with cell A. That is an example of handoff between cells.



Note that while 1G mobile telecommunications technology was introduced in 1973, the technology for handoff was nat introduced until 1982.

iPhone Field mode

In going through this section, you might want to make connection between what you are reading and what your phone does. iPhones have Field Mode, in which you can view some of your connection and mobility information. To enter field mode dial `*3001#12345#*`.

Let's take a step back and understand how a UE comes to be connected to a cellular network.

Network Discovery and Registration

Radio towers constantly send out signals broadcasting the identity of their cells, including which network operator owns the cell. The subscriber identity module (SIM) placed in UE tells the UE to look for a particular network through a parameter called public land mobile network (PLMN). That has two parts;

1. mobile country code MCC (The US has MCC codes 310 through 316, DISH Wireless is on 313.)
2. mobile network code MNC (DISH Wireless's MNC is 340).

Thus, DISH Wireless's PLMN is 313340.

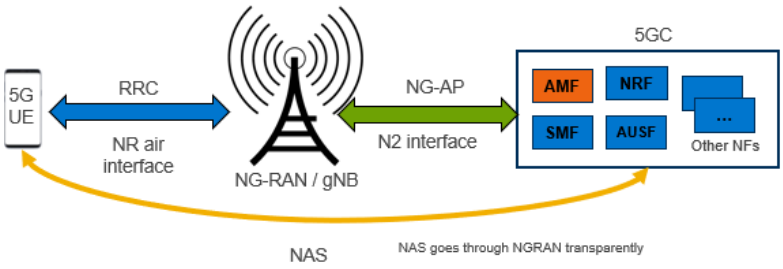
When a UE is on it constantly looks for radio signals that are constantly broadcast from towers. Those radio signals announce the PLMN of the cell. A UE with a SIM is looking for radio signals with a particular PLMN. When a UE wants to make a connection to a network, it selects one of the cells cell from which it detects a signal that is sending the PLMN it is looking for.

RRC Connection

The UE then sends a RRC connection request to that cell. (RRC is for radio resource control.) That message gets sent from the cell's gNB to an AMF instance called the serving AMF. The serving AMF and UE then start a registration procedure to make sure the UE is allowed to access this network; the AUSF verifies that the UE has the same pair of keys that are stored in the UDM for that UE. Since an AMF might get simultaneous requests from two UE claiming to be the same UE (there are nefarious actors out there!), each request is assigned a global unique temporary ID (GUIT) in this procedure. Also, note that this registration event has a feature called `Registration Type` with value `Initial Registration`. In registration of this type, information about the UE's capabilities and settings are sent to the AMF and forwarded to the UDM. An example of such a setting is MICO mode (mobile initiated communication only) in which the UE can send information but is not allowed to receive information that it does not request; this mode is intended for use with IoT devices that need to conserve power and thus should not be asked by the network to do anything.

NAS Connection

After this registration process the UE has a connection to the serving AMF called a non-access stratum (NAS) connection. The reason for the name is that NAS connection does not provide user plane connectivity; a PDU session must be established for that. A NAS connection does, however, provide communication between the UE and the SBI API services of the other network functions in the core; the serving AMF acts as a proxy (e.g. to establish a PDU session) in communication between UE and other NFs.



The NAS connection is a tunnel along the N1 reference point. It uses both the new radio (NR) air interface (NR is the shorthand for 5G radio technology) from UE to gNB and the N2 interface from the gNB to the AMF. Together they form the N1 reference point from UE to AMF. You may recall that the N1 reference point is often visualized as not involving the gNB to emphasize that it is between the UE and AMF.

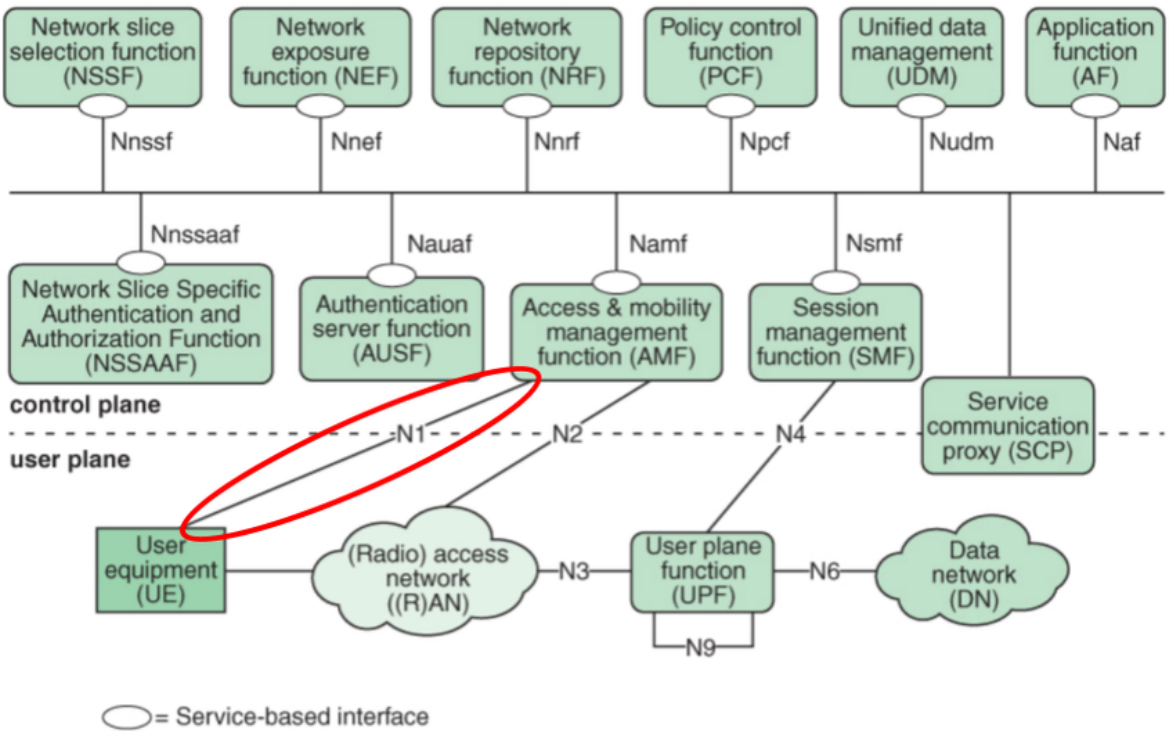
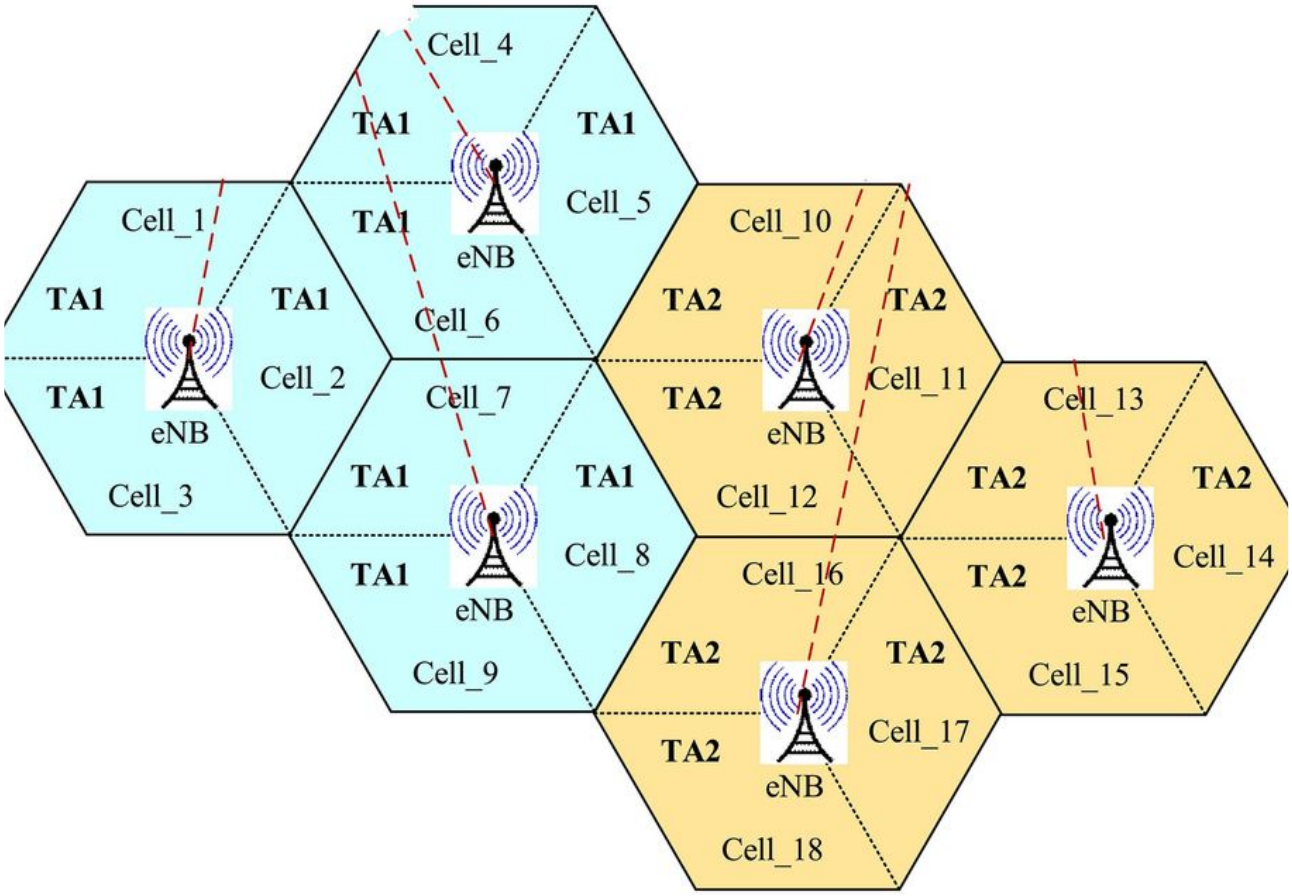


FIGURE 9.4 Non-Roaming 5G System Architecture

Tracking Areas and Registration Areas

A UE knows what cell it is in by listening to radio signals constantly sent by cells. However, the network does not get constant updates about the UE's position. This is to save on use of radio resources.

Spectrum is a scarce resource; the more radio communication is happening, the fewer the spectrum bands available and the more interference between signals. To conserve radio resources the core is not notified of UE's cell location constantly. At least not when the UE is "inactive". In particular, when a UE is in the connection management (CM) state called `idle` the UE drops its NAS connection until the UE detects that the UE has moved out of a certain region. That region needs some description.



Cells are grouped together into tracking areas. Further, each UE gets a different grouping of tracking areas called registration areas. That is, different UEs have different boundaries of their registration areas (RAs). UEs in CM state `idle` run their registration process to inform the AMF of their position only when they detect they have change registration areas (via the signals constantly sent out by cells). Since each UE has different registration area borders, when a buss or train full of people with UEs moves across one UE's registration area border, the network will not be overwhelmed with simultaneous registration operations for all of those UEs, but rather just the registration process for the one (or maybe a few) UE that are changing RA.

In addition to running the registration process when a RA border is crossed, UE in cm-idle state run a registration process with some periodicity called the TAU timer (for tracking area update). Both registration events for crossing RA borders and TAU timers have `Registration Type` value `Mobility Registration Update` (MRU). (Compare to the alternative value `Initial Registration`.)

The NAS connection is released after the UE registers, and re-established with each registration event.

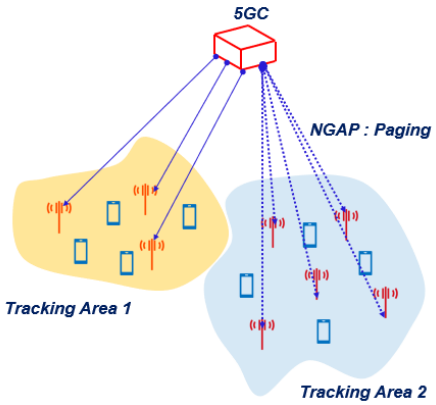
When a UE moves far from the AMF that has been serving its registration requests, it might me assigned a new AMF that is closer to the UE. This allows AMFs to be geographically distributed to distribute load on AMFs. How best to geographically distribute AMFs is a problem to be solved. The UDM keeps a record of which UE is being served by which AMF.

Paging

PDU sessions can be in a state called `inactive`. This term means more than that data is not flowing across the session; if a PDU session is in state `inactive` then no tunneling endpoint information is stored for data flows in that session. When a PDU session goes from inactive to `active`, the tunneling endpoint information needs to be re-established. When a UE is in CM state `idle` its PDU sessions *do* exist, but are in state `inactive`. That is, for each of the PDU sessions the UE has, address for the UE for that PDU session are not set up.

Paging is the mechanism for getting packets to a UE in cm state `idle`. When downlink packets destined for the UE are sent from the DN to the UPF that anchors one of the UE's inactive PDU sessions, the UPF buffers the packets and sends a message to the AMF requesting that

the UE be paged. This paging is done with radio signals sent in each of the cells in that UE's last known registration area. The amount of data sent is as small as possible; 10 bits.



Despite the small amount of data sent, paging is costly since there can be so many cells in a registration area (which, recall, has multiple tracking areas). Because of this, it is more spectrum efficient to have small registration areas. On the other hand, having registration areas that are too small leads to many UE having the same RA borders. Thus, a compromise needs to be calculated; that is part of the work of RF optimization. In fact, data about a UEs past positions can allow a network operator to use ML to assign UE optimal RAs. This is another problem to be solved.

When the UE receives a page, it enters CM state `connected` and sends a request to the AMF to establish user plane resources to resume data flow across its existing PDUs. That is, it requests tunnel endpoints.

Paging and MICO

A feature new to 5G is MICO mode, in which a UE can not be paged. This mode is intended for use by IoT devices that need long battery life; the mode protects the IoT devices from network triggered battery use.

CM State: Connected

A UE in CM state `connected` can have n PDU sessions for n in {0,1,2,…}.

To determine when to handover connection from one cell to another, the UE periodically measures the strength of signal from each cell it is set to check. The network periodically sends the UE a list of cells to check based on the UE's current cell. The UE can not send or receive user plane or control plane data while it is checking the strength of these signals; time gaps in data transmission are created to allow for this periodic check.

If handoff is from a cell with one gNB to another cell with a different gNB then the nodes are called a source gNB and target gNB.

The UPFs serving as PDU session anchors for the UE must be informed of the address of the target gNB so that the UPFs can apply that new address to downlink packet GTP-U headers.

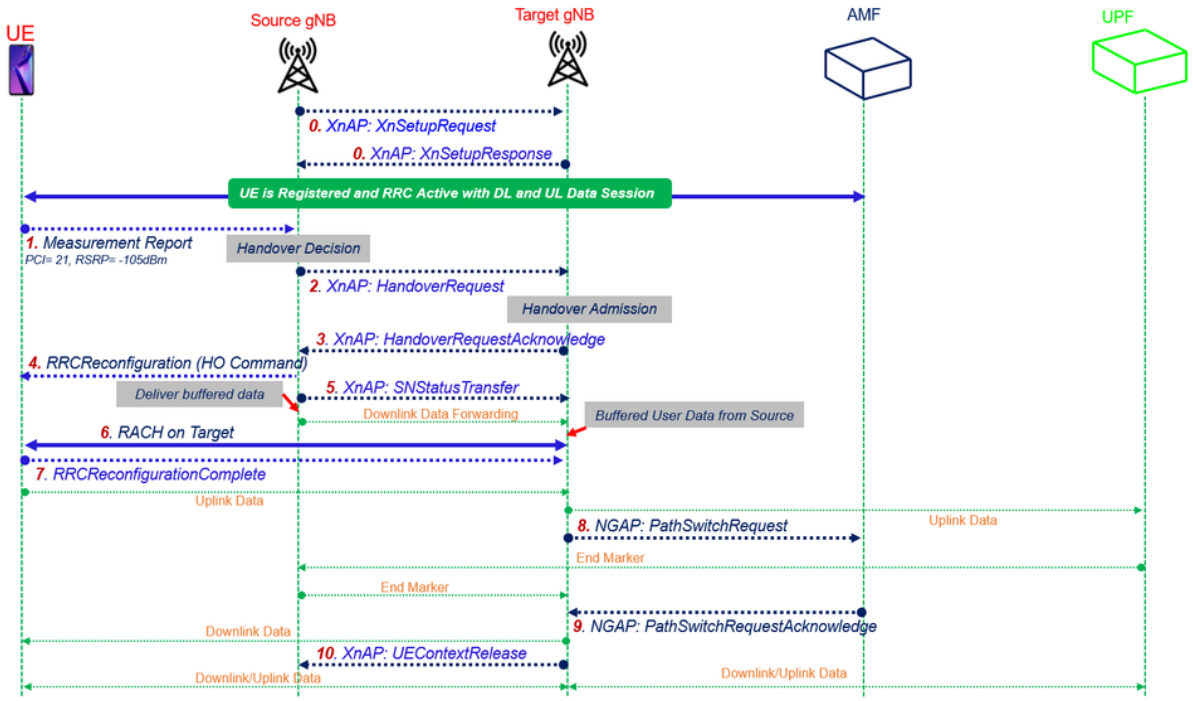
Besides this change of address, one of two handover mechanisms is needed: handover using the Xn connection between two gNBs, or handover using the N2 interface between UE and AMF.

Xn Based Handover

There are fiber connections between some gNBs. These connections are called Xn connections. If two gNBs are served by the same AMF then handoff of a UE's data flows from source to target gNB is eligible for Xn handover. The call flow for this handover is shown below with the following acronyms;

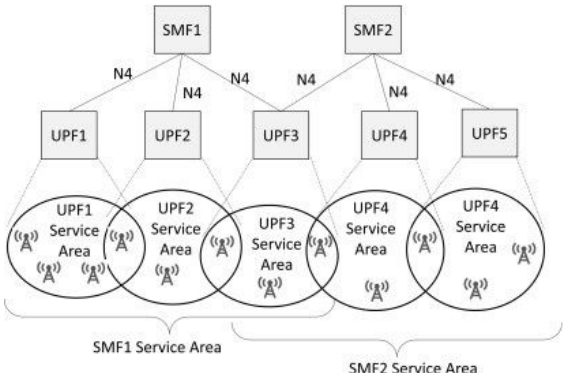
- XnAP is Xn application protocol, the protocol for Xn transmission. It is built on SCTP on top of IP.

- PCI is for physical cell identifier. In this case the source gNB hosts a cell with PCI 22 and the target gNB hosts a cell with PCI 21, and the latter cell will be taking over the flow of traffic.
- RSRP is for reference signal received power. This is the measurement of the strength of the radio signal from the cell with PCI 21 to the UE in units of milli-decibels (dBm). The number is negative because decibels are logarithms base 10 of intensity, and the power *P* is less than 10^{-3} watts;
  - RSRP := $10\log_{10}$ ($P \div 10^{-3}$W).
- RACH is for random access channel. This is the channel through which UE can make their initial communication with a cell before any spectrum channel is assigned to the UE.
- RRC is radio resource control, the protocol for radio communication built on top of MAC protocol.



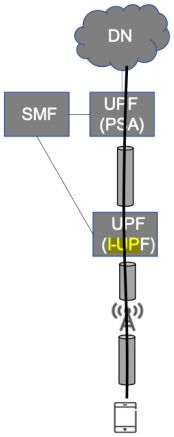
UPF and SMF and Mobility Management: I-UPFs and I-SMFs

If a UE has a PDU session anchored at a UPF instance, and the UE leaves the UPF service area of that instance by moving from a source gNB to a target gNB, then there is no physical connection from the target gNB to the anchor UPF.



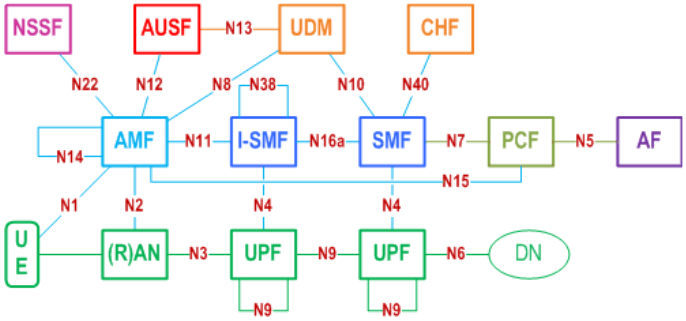
The PDU session anchor (PSA) does not change in this case; instead, another UPF from the target gNB's UPF service area is inserted into the data plane. The inserted UPF instance is called an intermediate UPF (I-UPF). A new packet forwarding control protocol (PFCP) session along N4 is established between the SMF and the I-UPF. The packet detection rules from the GPRS session between the PSA and SMF

are handed over to the I-UPF. For the sake of example, lets say that in the diagram above the PSA is UPF1, the UE moves to a gNB that is in the UPF2 service are but not in the UPF1 service area. Then a UPF2 becomes the I-UPF.



If, from that state, the UE moves to a new UPF serving area that is in the same SMF serving area, then the I-UPF changes. Continuing the example, if the UE moves to a gNB in UPF3 services area that is not in UPF2 service area, then UPF2 stops being the I-UPF and UPF3 becomes the I-UPF. Note that this means a configuration with a chain of I-UPFs does not happen.

If the UE moves to a UPF serving area that is outside the SMF serving area, say to UPF5 serving area, then new I-UPF will not have a connection to the SMF serving the PSA. In this case an intermediate SMF also needs to be inserted.

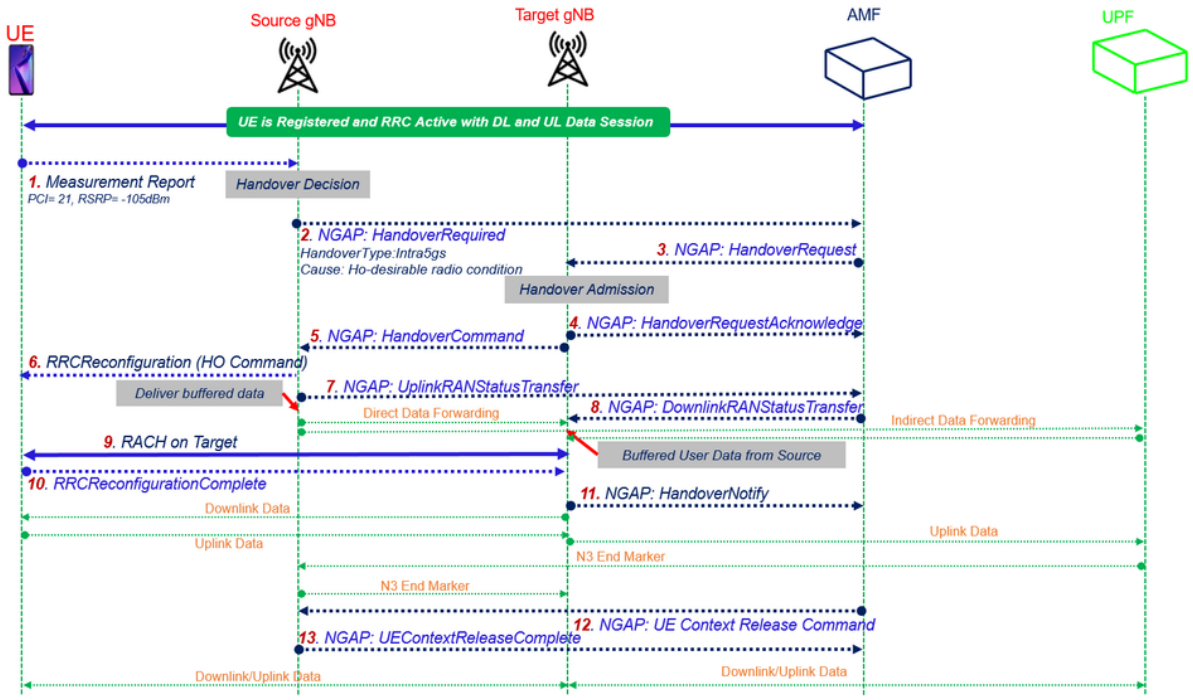


None of this mobility management is considered roaming because it all takes place within the same PLMN.

N2 Based Handover

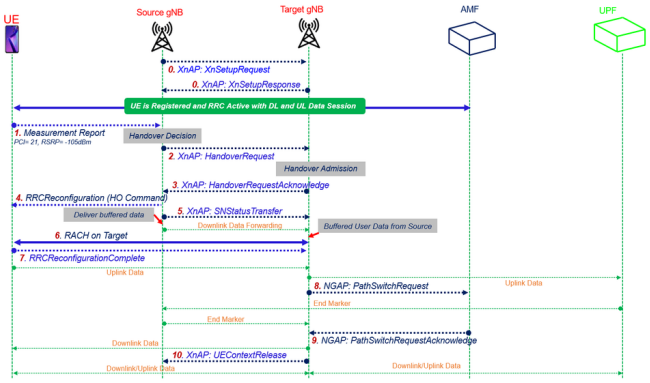
If the target gNB is served by a different AMF than the source, or if otherwise the target and source gNB do not have an Xn connection, then the Xn based handover is not possible. In this case, communication within the 5G core can manage the handover, albeit more slowly than Xn based handover. The call flow for this handover is shown below with many of the same acronyms in the diagram for Xn based handover above with one more;

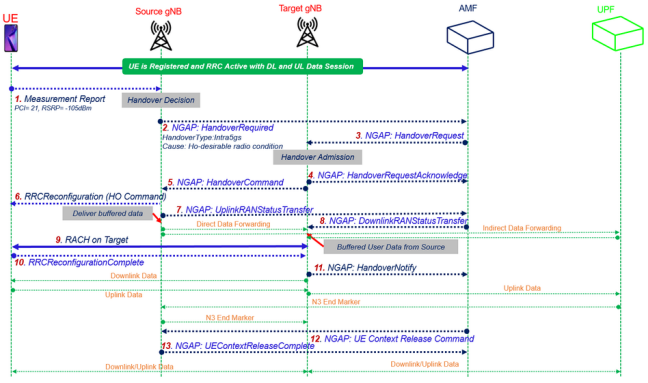
- NGAP is for next generation application protocol. This is the protocol used on the N2 reference point between gNB and AMF. There is a mistake in the diagram in step 5; the handover command is sent from the AMF to the source gNB; there is no connection from target to source gNB other than with AMF as a proxy.



The diagram presents the simplest case N2 handover; there are no I-SMF or I-UPF insertions of changes. There is also a more complicated version of N2 handover where serving AMF is changed.

The diagrams for Xn based handover and N2 based handover are presented side by side for comparison.





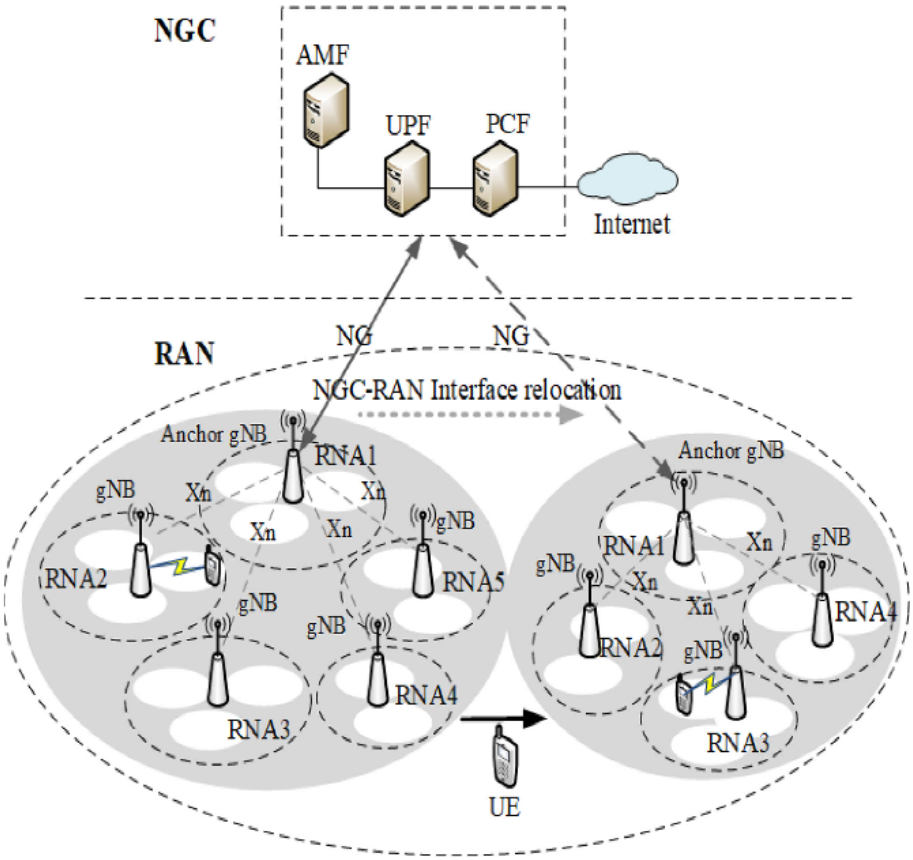
Location Services

A 5G specific feature is the ability of network functions (including AFs) to get UE location data from the AMF. This data is only available if the UE is in both CM state `connected` and RRC state `connected`.

Such location data is much lower spatial resolution than GPS location services, which most smartphones run with a separate program and separate antenna; this location data from the AMF can be at the level of tracking area, gNB, or cell.

RRC mode `inactive`

Another 5G specific feature addresses the problems caused by having just two CM states, `idle` and `connected`. The CM state `idle` exists to reduce mobility update signaling. But there are certainly times when the UE is in state CM state `connected` where mobility updates are not needed at the cell level. So, 3GPP introduced a new RRC state value: `inactive`. When a UE is in CM state `connected` and in RRC state `inactive` the UE does not need to notify the AMF when it changes cells unless it leaves a region (set of TACs) called a RAN notification area (RNA). In the image below the light regions are TACs, the dotted circles are RNAs, and the grey circles are registration areas.



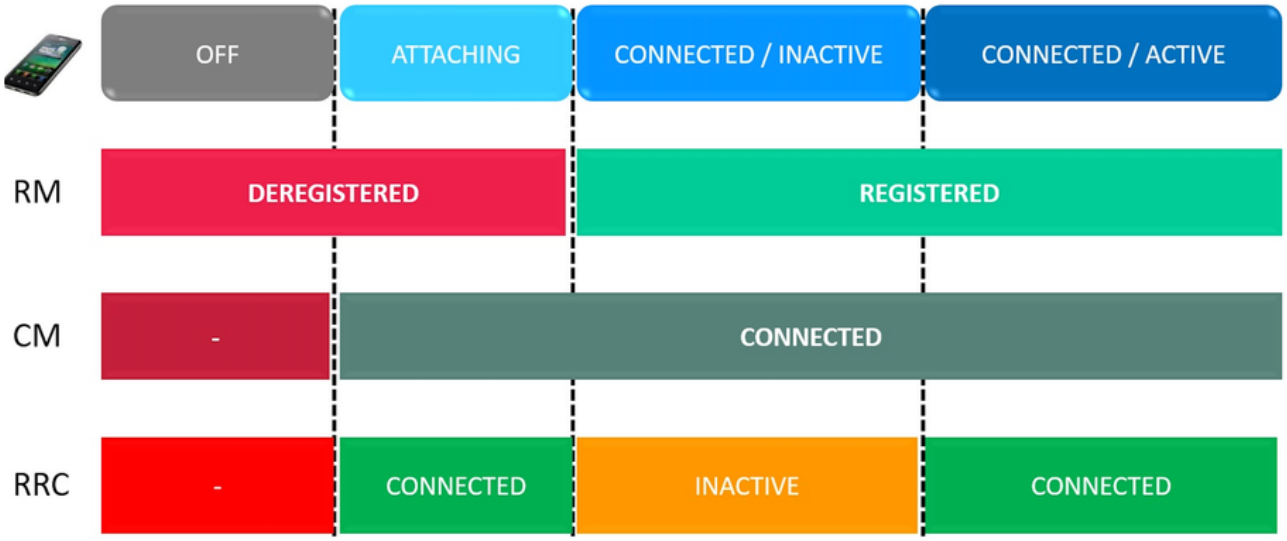
The RAN decides, with input from the core, when the UE should be put in RRC state `inactive`. This is because the resources that are freed by doing so are RAN resources, in particular spectrum resources.

To compare, UE need to register

- when changing RAs if in CM state `idle`
- when changing RNAs if in CM state `connected`

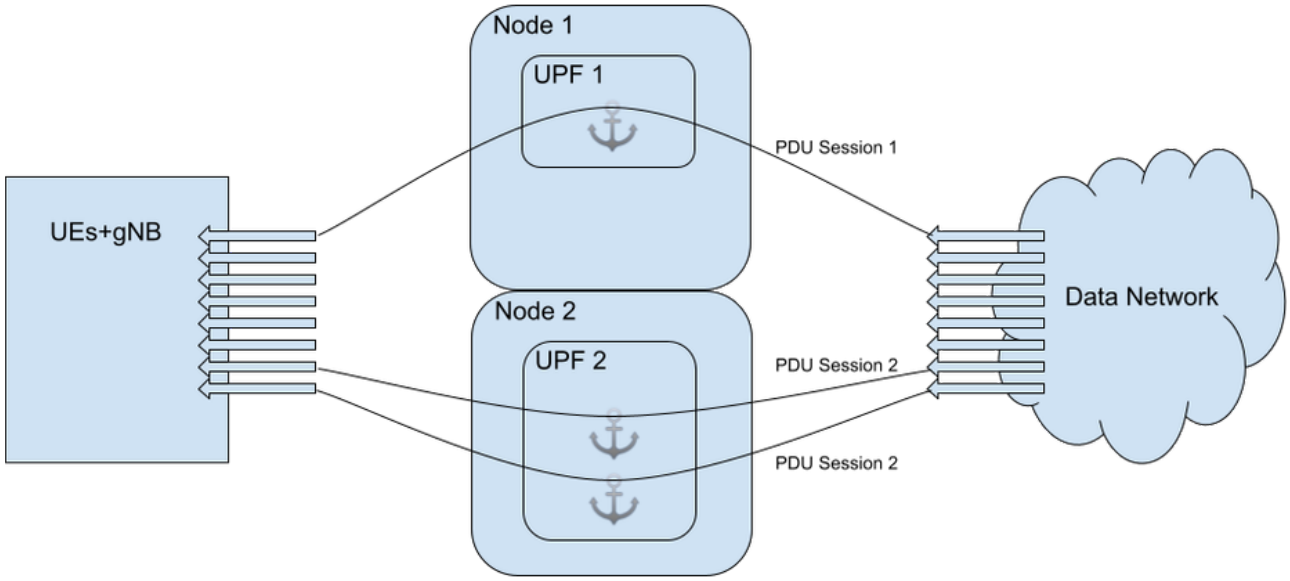
since RAN notification areas (RNAs) are subsets of registration areas (RAs), UE need to re-register less often in the combination RRC state `inactive` plus CM state `connected` than in CM state `idle`.

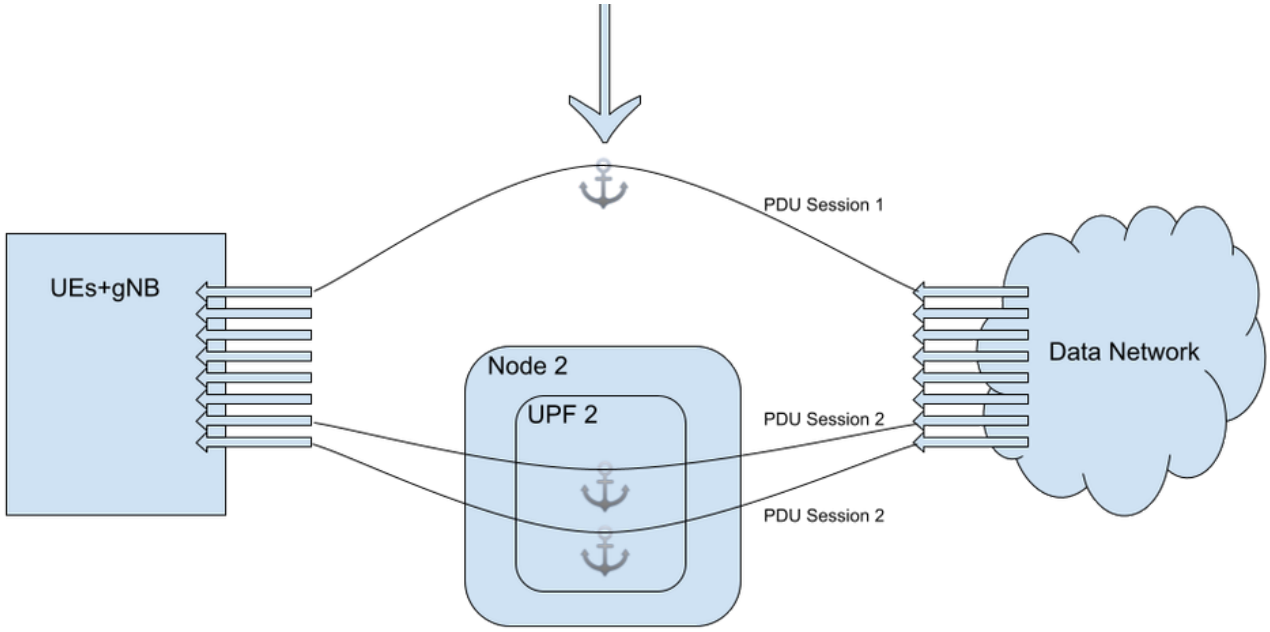
Note that there are 4 combination of registration management (RM), RRC, and CM states.



Mobility and Anchors in the Cloud

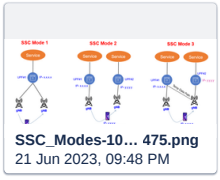
The defining feature of a mobile network is the ability of user equipment to move while the system maintains service continuity. The core is the general idea of a place where connectivity is maintained throughout mobility. The core has evolved to be spatially distributed, especially with the placement of UPFs at edge data centers to facilitate edge computing. This distributed core facilitates more diverse uses of mobile networks including faster user plane connectivity. It does, however, lead to increased complexity of the core. One feature that has remained is anchoring of data sessions in the plane; one point stays constant in PDU sessions, the anchoring UPF. The MSS BOAT named RESPONS found that this anchoring is an obstruction to dynamically scaling user plane resources; if the number of Kubernetes nodes hosting UPFs is decreased, then the PDU sessions anchored in those nodes experience service disruption.





However, 3GPP has acknowledged that this permanency of anchors is not sustainable by defining service and session continuity modes 2 and 3.

- Mode 1: The PDU session is permanent under mobility handoff events.
- Mode 2: Upon a mobility handoff event, the core may request that the UE terminate a PDU session and then create a new PDU session which takes over flows that were handled by the original. This is the break-before-make option.
- Mode 3: Upon a mobility handoff event, the core may request that the UE create a new PDU session, transfer flows from the original PDU session to the new PDU session, and then terminate the original PDU session. This is the make-before-break option.


SSC_Modes-10… 475.png
21 Jun 2023, 09:48 PM

UE with the capability can create PDU sessions with SSC mode 2 or 3 (1 is the default) and these enable new PDU sessions to take over the data flows of an old PDU session. This has the effect of changing the anchoring UPF in the process. Note that the PDU session does not move from one PSA to another (and thus the IP address of the UE for the PDU session is not maintained); rather new PDU sessions are created, including a new IP address for the new PDU session.