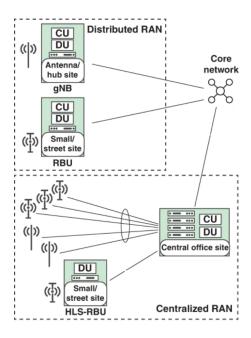
NG-RAN



5G new radio is a combination of radio technologies giving mobile devices access to the 5G core using radio interfaces. It is a new radio access technology. A network of access points that provide 5G new radio access is called a NG-RAN.

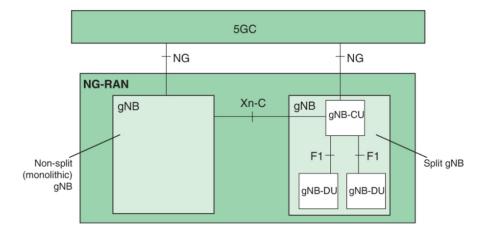
One example of how NG-RAN equipment is novel is the separation of the gNB into DU and CU, allowing multiple antennas (each with its own DU) to be served by the same gNB (with a single CU). This allows several network deployment options. There are two categories of deployment options.

- In a distributed RAN (DRAN) deployment the CU and DU of each gNB are collocated at the cell site.
- In a centralized RAN (C-RAN, pronounced "see ran"... the dash is to prevent you from saying "crayon") deployment the DUs for several cell sites are located at a common point that is not any of the cell sites.
 - Small cell sites like antennas on street light poles can have their DU collocated, and can share a CU with the other cell sites.



gNBs

Radio access network nodes in 5G are called gNBs. A new feature relative to 4G (eNBs) is the option to split the node into distributed and central parts, DU and CU, respectively. When a gNB has this split it is called a split gNB. Shen it does not, it is called a monolithic gNB. The connection between gNBs is called the Xn connection. When one of the connected gNBs is a split gNB the connection is to the CU, and the connections is called a Xn-C.

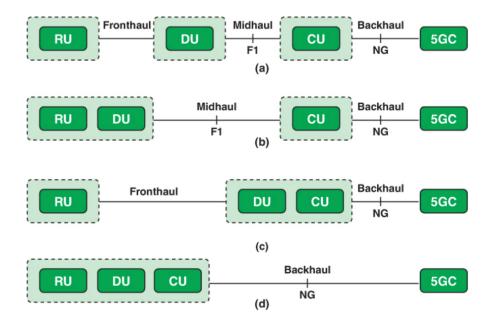


The general nature of the DU-CU split is that the

- CU processes non-real time protocols and services while the
- DU processes real time physical layer services.

One DU supports one cell (no matter if that cell has 3 or 6 sectors).

Note also that there are a variety of ways for (non)-split gNBs to be connected to radio units (RU) and the the 5G core.



In 5G the CU is always virtualized. In newer deployments the DU is also virtualized. In older literature this virtualization of RAN was called cloud-RAN; however vRAN has become the acronym of choice.

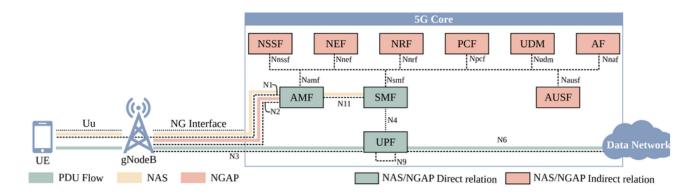
Interfaces

The fronthaul connection is specified by 3GPP and is the specification is called the common public radio interface (CPRI). The midhaul reference point (between DU and CU) is called the F1 reference point,

The reference point between the UE and gNB is called the Uu reference point. A particular connection between a UE and a gNB is called a Uu interface. (This is like an object from a class.) Transmission over such an interface uses a protocol named radio resource control (RRC) . It carries

· user plane data in PDU session flows via indefinitely open connections initiated and terminated by the UE in protocol RRC

• non access stratum (NAS) signaling for messaging between UE and control plane via API calls via short message-acknowledgement exchange in protocol RRC.



An interface between the gNB and core is called a NG interface. It carries

- · user plane traffic in PDU session flows via indefinitely open connections initiated and terminated by the UE in protocol GTP,
- NAS signaling between UE and core via API calls in HTTP/2 protocol,
- control signaling between the gNB and AMF via the protocol NGAP (on top of stream control transmission protocol (SCTP)) in sustained connections.

Layers

The NG-RAN is designed to have two layers, radio network layer (RNL) and transport network layer (TNL). These are structured differently for the control and user plane, but both are built on IP protocol.

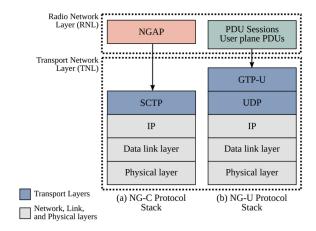
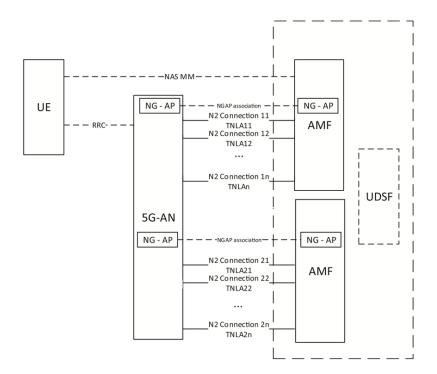


Figure 3. NG-C and NG-U protocol stack.

TNL Associations

The connection between a gNB and an AMF instance, that is a connection over the NG interface, is called a TNL association (TNLA). A single gNB shall have the capability to support multiple TNLAs with a single AMF. Each AMF can support at most 32 TNLAs.



Author's note: I have not been able to find information on the number of UE or PDU sessions per TNLA. My best guess is that each UE is assigned one TNLA to host all of its PDU sessions. However, note that (1) in 5G it is possible for a UE to use dual connectivity, connecting to two gNBs at the same time (2) in 5G a single gNB can have connection to many AMF instances and one can imagine different PDU sessions for the same UE going to different AMFs.

N2 Management

A novel feature of 5G is the ability to move UE connections from one AMF to another. When the UE is in CM state connected AMF instances have the sole authority to do this moving of connection from one AMF to another. The mechanism is rebinding of TNLAs to a new AMF.

This novel feature enables adding and removing AMF instances from AMF sets; that is, it enables dynamic scaling of AMF compute resources. To supplement this new ability, the network repository function (NRF) entry for an AMF instance includes a key served-guami.backup-amf with optional value; putting a value here allows the 5G system to use a planned removal process for an AMF instance. The value is a list, allowing several AMF instances to be potential backups.

There are two ways for an AMF to go through planned removal (I'm not clear on the distinction here);

- 1. through use of UE context data stored in the unstructured data support function (UDSF) to ensure that all TNLAs for all UE contexts are transferred to a new AMF
- 2. through a AMF sending a de-reregistration request to the NRF. After this, when a NF queries the NRF for NFs that can provide a service, the deregistered AMF is not included in the list. That will remove the load on the AMF's SBI and N1 load. But what of the load on the AMF through N2? Optionally, the AMF scheduled for removal can also set its own (load balancing) relative weight factor to 0 so that the AN never again chooses to form TNLAs with this AMF.

Both of these mechanisms for planned removal of AMFs require AMF instances to have the the functionality to rebind TNLAs to new AMFs. (Binding is associating multiple addresses to the same interface, and in this case TNLA endpoints at the AMF under planned removal need to be bound to a new AMF's network interface.)