5G Core
and
Service Based Architecture

Strata

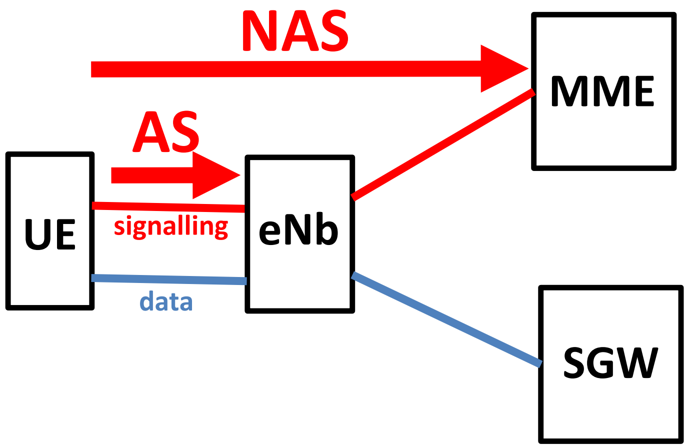
This time we will discuss the mobile *core* functionalities

In all networks we differentiate between user (Data, forwarding) plane
and the control (and/or management) plane

In mobile networks we further split the control plane into *strata*
Access **S**tratum and **N**on**A**ccess **S**tratum

The AS is the signaling between the UE and the base station (NB, eNB, gNB)
and deals with all the aspects of the air interface

The NAS is between the UE and the core (in 4G with the MME)
and handles establishing sessions and maintaining continuity as UE moves



Bearers

In reading standards you will come across the term *bearers*
although we won't need it here

In the physical layer we talked about *channels*
in higher layers we talk about *bearers*

A bearer is a transparent connection between UE and DN

We differentiate between

- data (user plane) bearers
- signaling (control plane) bearers, which can be AS or NAS bearers

For example, on the air interface, we distinguish 3 types of signaling bearers:

- Signaling Radio Bearer 0 (SRB0)
  - AS messages over **C**ommon **C**ontrol logical **CH**annel
- Signaling Radio Bearer 1 (SRB1)
  - NAS messages over **D**edicated **C**ontrol logical **CH**annel
- Signaling Radio Bearer 2 (SRB2)
  - high priority AS messages over DCCH logical channel

AS - RRC

The **A**ccess **S**tratum only controls the air interface
and thus only the connection between the UE and one base station

The highest layer of the AS is called **R**adio **R**esource **C**ontrol

RRC messages include:

- system information broadcast (MIB, SIBs)
- information for idle UEs (cell selection parameters, neighboring cell info)
- emergency broadcast messages (Earthquake and Tsunami Warning System)
- paging
- connection establishment/modification/release
- UE state (idle/connected) handling
- handoff management (including security handling)
- radio configuration (ARQ configuration, HARQ configuration, etc.)
- assignment/release of user RBs
- QoS control
- recovery from radio link failure
- measurement configuration and reporting

NAS messages

The **N**on**A**ccess **S**tratum controls the connection between the UE and the core independent of the serving base station

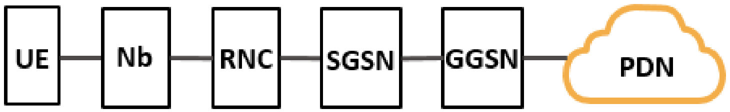
NAS messages include:

- identity management
  - identity request and response
  - authentication request and response
- session management
  - session (PDN connection) request and response
  - session detach request and response
- mobility management messages
  - tracking area update
  - mobility attach request and response
  - mobility detach request and response

Cores from 3G to 4G-CUPS

3G data the Nb+RNC connect to the SGSN and GGSN

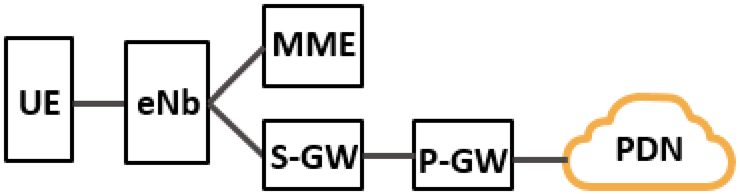
SGSN and GGSN handle both data and control



4G Nb+RNC were unified into the eNB

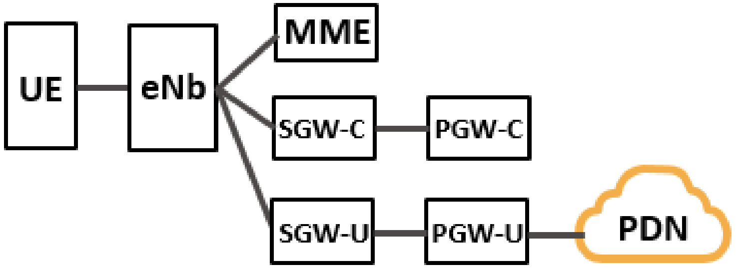
eNB connects to S-GW and P-GW

Mobility management was separated

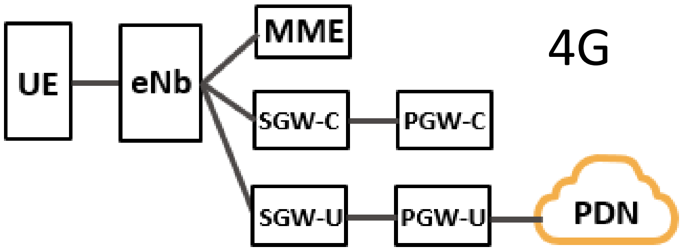


4G CUPS (R14) separates into UPF and CPFs

S-GW-C and S-GW-U,   P-GW-C, P-GW-U

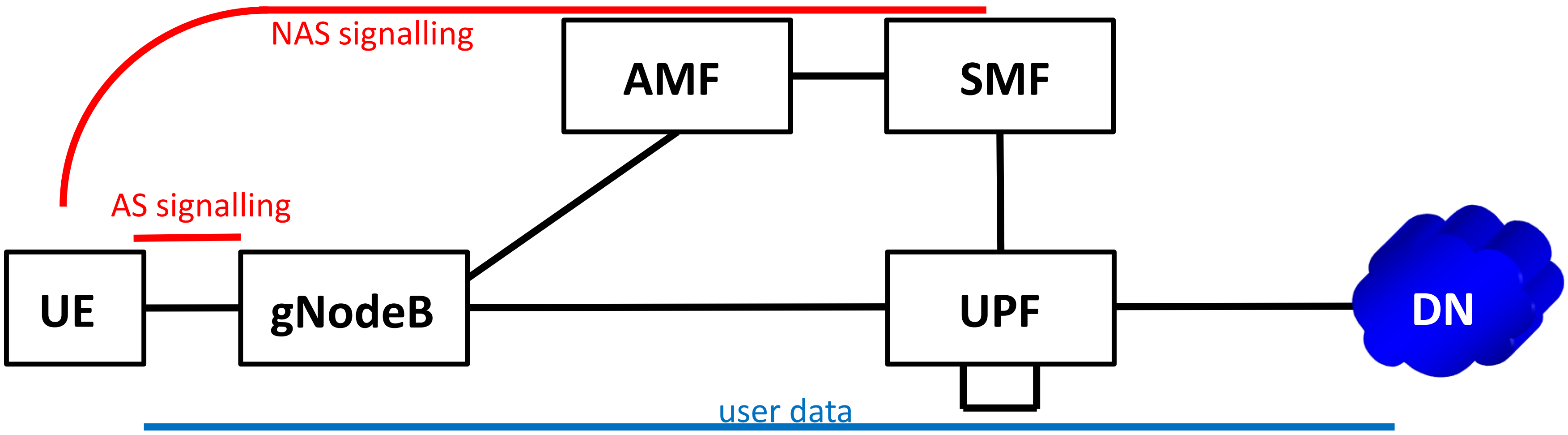


Cores from 4G-CUPS to 5G

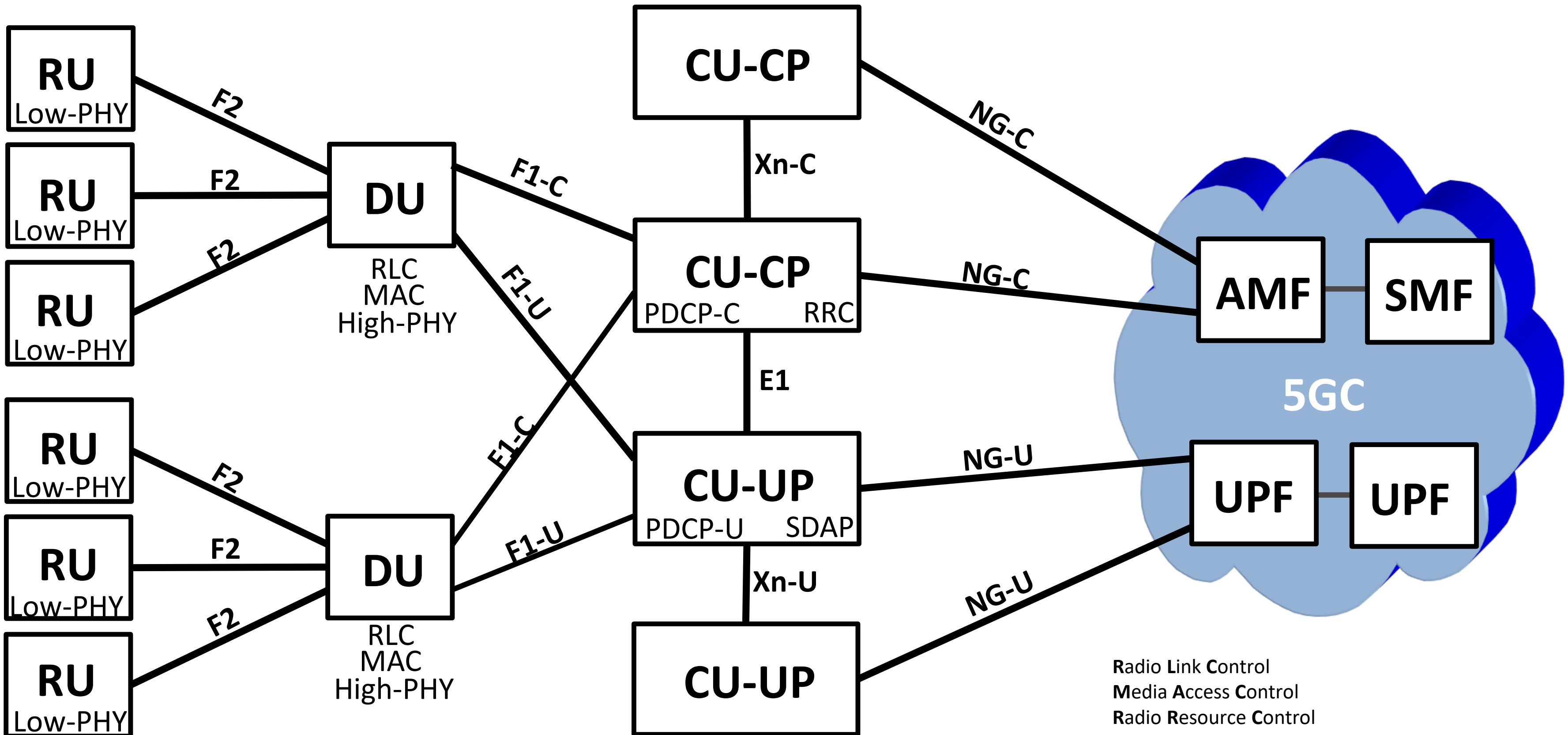


5G

- decomposes the MME into AMF and SMF
- unifies S-GW-U and P-GW-U (and TDF) into UPF(s)
- unifies S-GW-C, P-GW-C and MME session management into SMF



5G RAN architecture with CUPS



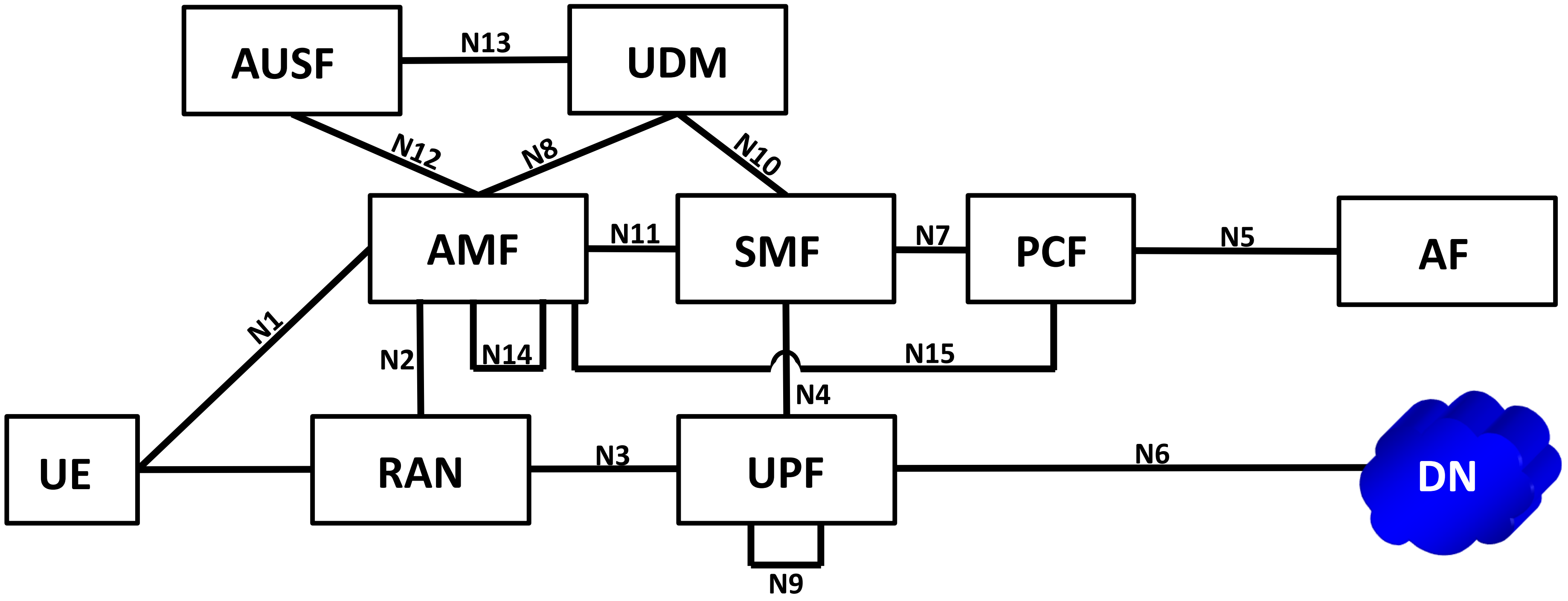
Radio **L**ink **C**ontrol
Media **A**ccess **C**ontrol
Radio **R**esource **C**ontrol
Packet **D**ata **C**onvergence **P**rotocol
Service **D**ata **A**daptation **P**rotocol

5G core (simplified)

AUthentication **S**erver **F**unction
Access & **M**obility Management **F**unction

Unified **D**ata **M**anagement
Session **M**anagement **F**unction
Policy **C**ontrol **F**unction
User **P**lane **F**unction

Application **F**unction
Data **N**etwork



5G “N” reference points (for *reference*)

The 5GC architecture currently defines the following reference points:

- **NG**: RAN - core
- **N1**: UE - AMF
- **N2**: RAN - AMF
- **N3**: RAN - UPF
- **N4**: SMF - UPF
- **N5**: PCF - AF
- **N6**: UPF - DN
- **N7**: SMF - PCF
- **N8**: UDM - AMF
- **N9**: two UPFs
- **N10**: UDM - SMF
- **N11**: AMF - SMF
- **N12**: AMF - AUSF
- **N13**: UDM - AUSF
- **N14**: two AMFs
- **N15**: PCF - AMF
- **N16**: two SMFs
- **N17**: AMF - 5G-EIR
- **N18**: any NF – UDSF
- **N19**: two PSAs
- **N20**: AMF - SMSF
- **N21**: UDM - SMSF
- **N22**: AMF - NSSF
- **N23**: PCF - NWDAF
- **N24**: visited PCF – home PCF
- **N25**: PCF - UDR
- **N26**: AMF – 4G MME
- **N27**: visited NRF - home NRF
- **N28**: PCF - CHF
- **N29**: SMF – NEF
- **N30**: PCF – NEF
- **N31**: visited NSSF – home NSSF
- **N32**: visited SEPP – home SEPP
- **N33**: AF – NEF
- **N34**: NSSF – NWDAF
- **N40**: SMF - CHF

UPF

The 5GC's **U**ser **P**lane **F**unction performs all the user plane functions handled in 4G by S-GW, P-GW, and TDF, including:

- anchor for mobility
- connection to external data networks (e.g., Internet)
- optionally Firewall and Network Address Translation (NAT) functions
- packet queuing
- packet routing and forwarding
- packet inspection (optionally DPI), classification, QoS handling
- policy enforcement
- packet marking
- lawful intercept
- traffic usage statistics collection and reporting
- IPv4 ARP and IPv6 neighbor solicitation

Why decouple AMF and SMF ?

The 4G MME has 2 distinguishable functions

1. access/mobility management
   - contacting the HSS, handling UE authorization and key distribution
   - allocating **T**emporary **M**obile **S**ubscriber **I**dentity
   - managing handoff
   - lawful interception
2. session management
   - creating/updating/removing data sessions
   - allocating IP addresses
   - managing context for the UPF

A single RRC message often performs access and session attaches!

But a single UE can simultaneously participate in multiple sessions

Access/mobility and session management
can be separated into micro-services
to increase flexibility and scalability

AMF — Access and Mobility Function

The AMF performs the access and mobility functions
that were handled by the 4G MME, S-GW-C and P-GW-C

- NAS signaling for access and mobility management
- UE authentication
- allocation of **G**lobally **U**nique **T**emporary **I**dentity and **T**emporary **M**obile **S**ubscriber **I**dentity
- UE security context management
- registration management
- connection management
- reachability management
- mobility management
- apply mobility related policies from PCF (e.g., mobility restrictions)

SMF — Session Management Function

The SMF performs the session management functions
that were handled by the 4G MME, SGW-C, and PGW-C

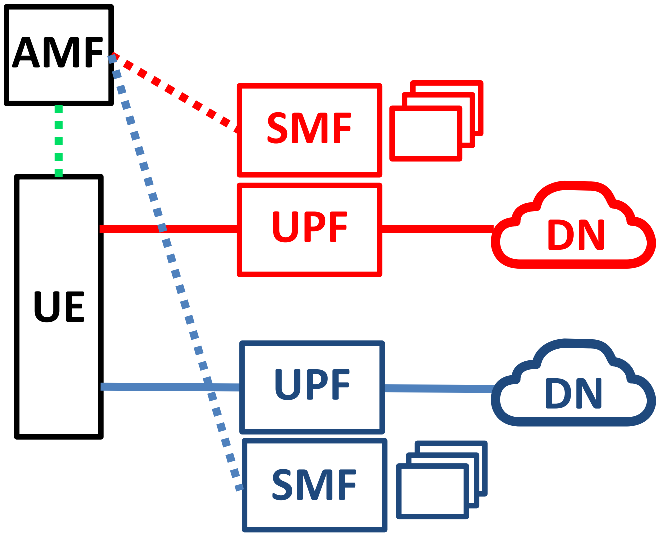
- NAS signaling for session management
- managing the PDU sessions
- allocates IP addresses to UEs (DHCP server)
- selection and control of UPF
- sends QoS and policy information to RAN via the AMF
- downlink data notification
- supports MEC by selecting a UPF close to the edge
- applies policy and charging for services
- control plane for lawful interception

Slicing

A single UE can participate in more than one slice

Each UE is served by a single AMF

but each slice has its own SMF and UPF



Capability exposure

In order to enable new service types and integrate with vertical industries
5G core functionalities will be made available to 3rd parties, including
- application service providers
- end-users (vehicles, factories, smart cities, etc.)

5G learned from MEC the importance of capability exposure
and defined the **N**etwork **E**xposure **F**unction

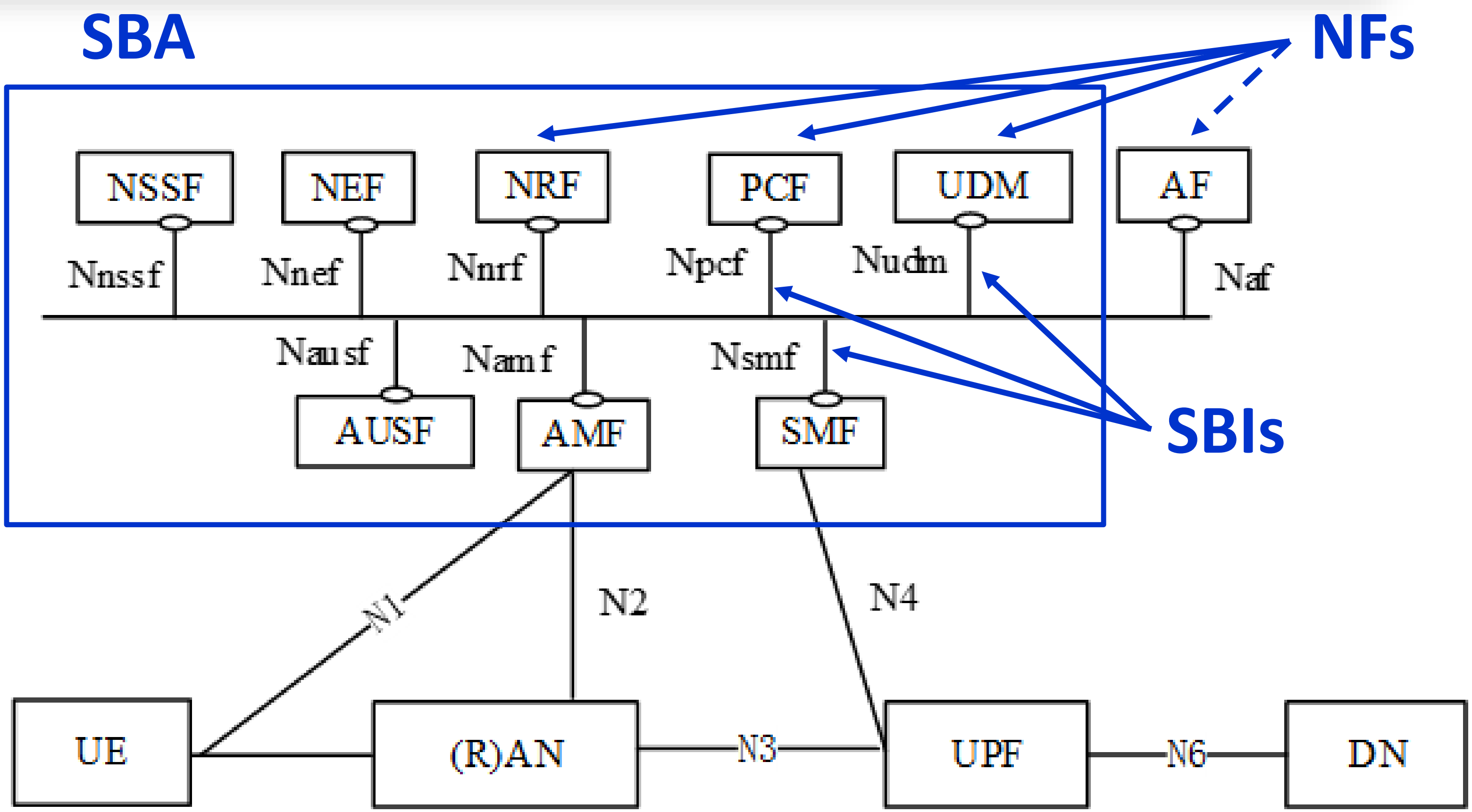
The NEF, like MEC's **M**obile **E**dge **P**latform, can be queried via an API
to discover available services

Capability exposure is a very common feature of web-based services
and the modern way of providing such services is via RESTful APIs

3GPP CT4 decided to completely re-architect the core to be RESTful
resulting in the **S**ervice **B**ased **A**rchitecture

In SBA, all the core network functions are defined as RESTful servers
with APIs called **S**ervice **B**ased **I**nterfaces

Simplified 5G core – SBA



The NFs are interconnected via a *logical* bus

i.e., every NF can communicate with every other NF

The software of one NF may or may not be on the same server as another NF

REST

Representational **S**tate **T**ransfer, defined by Roy Fielding (in his PhD thesis) is a software architectural style for services, not a precise protocol

REST breaks down transactions into component interactions

In order to guarantee performance, scalability, simplicity, and reliability REST architecture imposes 6 specific properties, including (3 / 6)

- client-server (consumer-producer) architecture
- stateless (servers do not maintain information on clients)
- uniform interfaces
  - CRUD operations
    - **C**reate (POST)
    - **R**ead (GET)
    - **U**pdate (PUT)
    - **D**elete (DELETE)
  - usually using **U**niform **R**esource **I**dentifiers and HTTP+JSON/XML

An API that conforms to REST principles is called a RESTful API while an API that violates any of the principles is not RESTful

Using RESTful APIs

Let's see how a RESTful API could be used for a fictitious social network

GET https://api.friendnet.com/members
    will return a list of all members of *friendnet* (in JSON or XML format)

GET https://api.friendnet.com/members/yjstein
    will return profile information of a member named *yjstein*

GET https://api.friendnet.com/members/yjstein/job
    will return only the member's job information

PUT https://api.friendnet.com/members/yjstein/job {new info}
    will update the member's job information

POST https://api.friendnet.com/members/yjstein/blog {content}
    will create a new blog entry in yjstein's profile

POST https://api.friendnet.com/members {new member information}
    will create a new profile

DELETE https://api.friendnet.com/members/yjstein
    will delete the member's profile

JSON and XML

An HTTP server responds with *status codes* and a body in JSON or XML

- 1xx Informational
- 2xx Successful (e.g., 200 OK, 201 created)
- 3xx Redirection
- 4xx Client Error (e.g., 400 bad request, 401 unauthorized, 402 payment required, 404 not found)
- 5xx Server Error (e.g., 500 Internal Server Error, 501 Not Implemented, 503 Service Unavailable)

For example, to GET https://api.friendnet.com/people/yjstein
the *friendnet* server may respond with 200 (OK) and one of :

JSON


```
{
 "first-name": "Yaakov",
 "last-name":"Stein",
 "job":"CTO - RAD",
 "education":"PhD Theoretical Physics, HUJI",
 "web-site":"www.dspcsp.com"
}
```


XML


```
<member>
 <first-name>Yaakov</first-name>
 <last-name>Stein</last-name>
 <job>CTO - RAD</job>
 <education>PhD Theoretical Physics, HUJI</education>
 <web-site>www.dspcsp.com</web-site>
</member>
```

REST in SBA using NRF

To see 5G SBA REST principles, start with the **N**etwork **R**epository **F**unction which allows every NF to discover the services offered by other NFs

- registering services (network function instances)
- maintaining profile of available NF instances
- exposing services

Before service instance NF0 can be used, it registers with the NRF

- NF0 is the *client*, NRF is the *server*
- NF0 sends to the NRF an HTTP PUT with its profile in the body
- NRF responds with a 201 message "created success" acknowledgement

Instance NF1 desiring to consume service provided by NF0 queries the NRF

- NF1 is the *client*, NRF is the *server*
- NF1 sends to the NRF an HTTP POST with desired query in the body
- NRF responds with a 200 "OK" message with a list of NFs containing NF0

NF1 can now consume service from NF0

- NF1 is the *client*, NF0 is the *server*
- NF1 sends to NF0 an HTTP POST with request for service/session in body
- NF0 responds with a 200 or 201 message (depending if 1-time read or opening session)

Simplified example – UE service request

Let's assume that a UE has already registered with a gNB (via RRC)
and the gNB has selected an AMF for it
and it has connected (N1 messaging)

The UE now wants to consume some service (with a type and attributes)

1 the SMF registers the services it provides with the NRF
2 the gNB forwards a registration request to the selected AMF
3 the AMF queries the NRF for an appropriate SMF
and receives the address of a registered SMF
4 the AMF now sends a post to the selected SMF (N11 messaging)
5 the SMF accesses the UDM (N10 messaging) to check authorization
6 the SMF selects an appropriate UPF, initializes it (N4 messaging)
and returns 200 with IP address, tunnel identifiers, etc. in body
7 the SMF communicates with PCF (N7 messaging) to configure rules
8 the SMF returns "created" to AMF
9 the AMF informs the UE that it can start consuming the service

Network Exposure Function

In 5G end-user and service provider application functions (AFs)
also need access to the mobile network's resources (mostly NFs)

Likewise, the 5G network wants information from external AFs
such as expected traffic patterns and mobility behavior

Allowing external AFs full access via the NRF would present security issues
so 5G defined a secure, intelligent, service-aware *gateway* function

5G adopted from **M**obile **E**dge **C**omputing the idea of an exposure function

- in MEC it is called Mobile Edge Platform service discovery function
- in the 5G core it is called the Network Exposure Function (NEF)

The NEF provides a RESTful API for external users to discover services

The basic idea of an NEF actually started with 4G
which defined a **S**ervice **C**apability **E**xposure **F**unction
for transferring small amounts of IoT data in signaling messages
without need to set up a user plane connection

Statelessness and the UDM

5G core servers are stateless, but often need access to state information

For this purpose there is a **U**nified **D**ata **M**anagement function
that offers data storage (currently to AMF, SMF, SMSF, NEF and AUSF)

The UDM access 2 other functions

- **U**nstructured **D**ata **S**torage **F**unction
- **S**tructured **D**ata **S**torage **F**unction

Using a *unified* data store simplifies its management, resilience, security, etc.

Modern data stores are fast, handle huge amounts of data, and *cloud native*

UDM is used by AMF and SMF
to retrieve the UE's subscription data (like 3G HLR and 4G HSS)

UDM is used by AFs to subscribe or un-subscribe to data change notifications

The AUSF retrieves information from the UDM to authenticate
and informs the UDM about successful or unsuccessful authentications

Future revisions will expand the use of the UDM

AUSF, PCF, AF

The 5G **Au**thentication **S**erver **F**unction
implements the part of the 4G HSS not in the UDM

The AMF (using the NRF) selects an AUSF to authenticate the UE to the core

The AUSF employs EAP authentication

The 5G **P**olicy **C**ontrol **F**unction replaces the PCRF in 4G networks

- provides policy rules for control plane functions
  including slicing, roaming and mobility management
- accesses subscription information for policy decisions taken by the UDM
- supports new 5G QoS policy and charging control functions

The **A**pplication **F**unction interfaces MEC functions with the SBA

- identifies relevant traffic (e.g., by 5-tuple)
- specifies locations of processing functions
- influences selection of UPF
- supplies application related information to the PCF

CHF

The **Ch**arging **F**unction (which is part of

assists mobile operators and application service providers

to monetize their services

The CHF is part of the **C**onverged **C**harging **S**ystem in the BSS/OSS

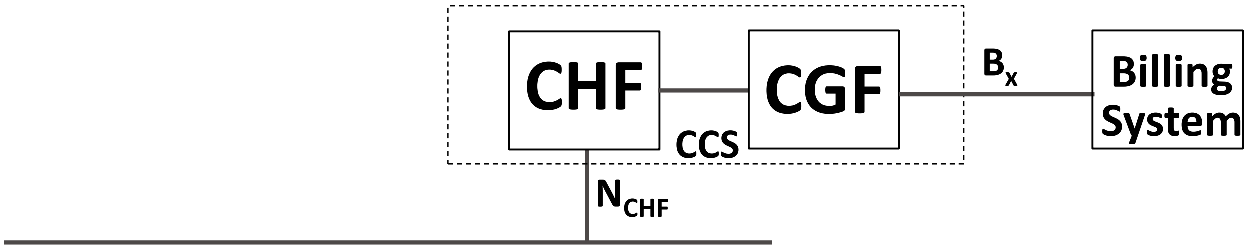
The CHF can differentiate billing rates according to

- network slice
- QoS parameters
- application functions consumed

CHF also supports spending limiting by interaction with the PCF (N28)

The CHF also enables

- unified charging for multi-operator cases
- charging for non-3GPP access



NWDAF

The **N**et**W**ork **D**ata **A**nalytics **F**unction
analyzes network performance data to other core functions

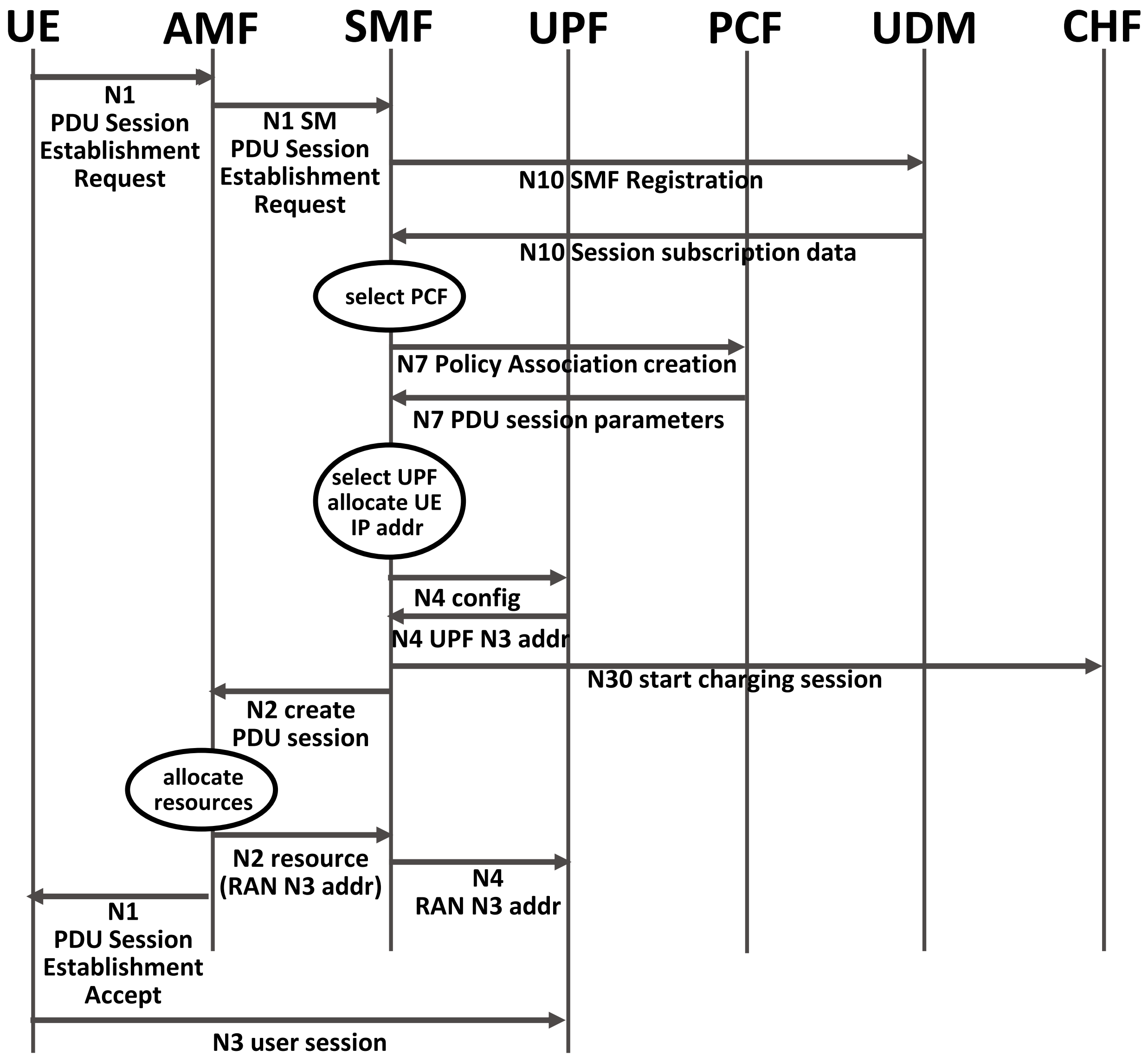
At present only the load level of a particular slice instance is reported

For example:

- the NSSF needs load level for intelligent slice selection (N34 interface)
  which it can access over the N34 interface
- the PCF requires load level to steer traffic or assign additional resources
  which it can access over the N23 interface

A core function may *request* analytics data when needed
or may *subscribe* to be notified by the NWDAF
when a slice's load level changes or passes a threshold

Example session establishment



SA and NSA

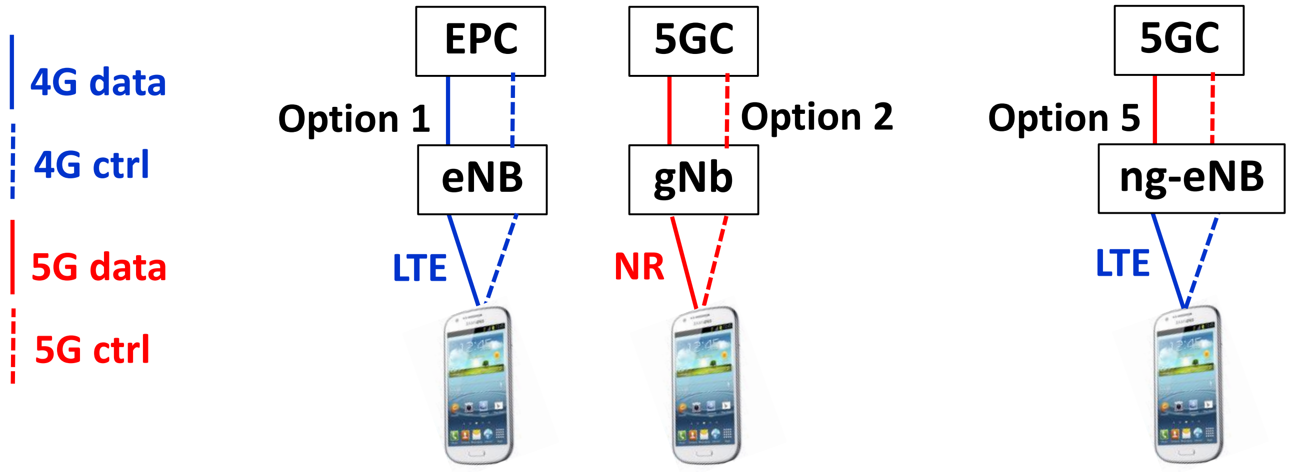
Up to now we have been talking about **S**tand**A**lone access
where a 5G gNB connects to a 5G core

Most of the initial deployments will be **N**on**S**tand**A**lone access
where parts of 4G LTE network are utilized

3GPP has defined a 3 SA options

The obvious two are pure SA 4G (option 1) and 5G (option 2)

For the distant future there is option 5 for supporting legacy phones
with a 4G LTE air interface connecting to a 5G core



Option(s) 3

The most important NSA option in the near term is called option 3

which assumes installation of a gNB for the advantages of NR

but does not (yet) upgrade the core to 5GC

and so provides higher data rates but not full 5G capabilities

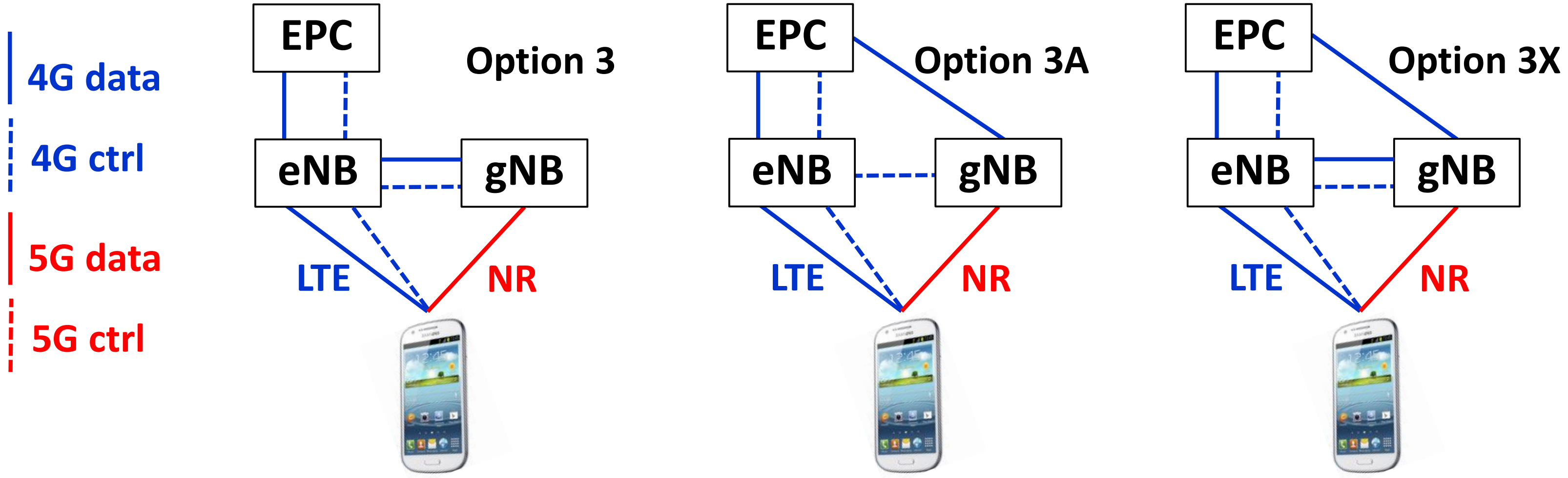
This will enable fast deployment of gNBs for eMBB

In option 3 there is no direct connection between gNB and EPC

user and control data flow through the eNB via X2-U and X2-C interfaces

In option 3A there is an S1-U connection from gNB to EPC (but no X2-U)

Option 3X has both X2 and S1 to enable load balancing



Other options

For the distant future there are options 4 and 7

which only have a 5G core (no EPC)

but support 4G legacy UEs via upgraded ng-eNBs

In option 4 gNB is the master and ng-eNB connects via Xn interface

In option 7 ng-eNB is the master and gNB connects via Xn interface

These too have variations (4, 4A, 7, 7A, and 7X)