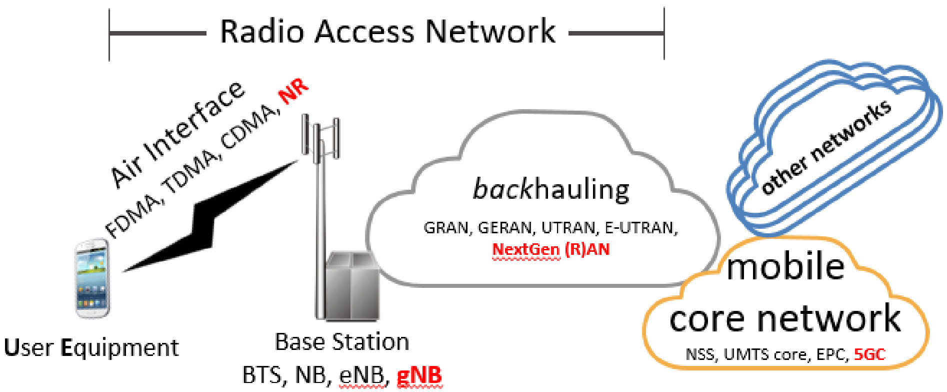
Air Interface



Cells

Area to be covered is tessellated with *cells,* each served by a BS
ideally circular/hexagonal, but precise pattern determined by *planning*

- macrocell < 35 km, typically 5-10 km (terrain dependent), about 25w
- small cells
  - microcell < 2 km about 5w (often used to supplement coverage)
  - picocell < 200 m   < 1w (inside buildings, underground parking, etc.)
  - femtocell ≈ 10 m 100 mW (home, small businesses)

Each cell may use multiple frequencies to support multiple users in same cell
but adjacent cells do not usually use the same frequencies

Cell boundaries may be *dynamic*
e.g., cell breathing is a load balancing mechanism
where overloaded cells reduce transmit power (changing service area)
thus offloading subscribers neighboring cells

Cells may *overlap*

- in CO-MIMO BSs jointly transmit/receive data to/from users
- in CoMP BSs coordinate with each other to reduce interference at edges

Handoff (handover)

True mobility is implemented by *handoff* (called *handover* in Europe)

which is when an active call or data session

is transferred from one cell to another

e.g., when user moves out of serving cell's coverage area

or when serving cell is overloaded

- in *hard* handoff (break before make)

  the existing session breaks before new connection is established

  if performed quickly enough the discontinuity made not be noticed

- in *soft* (smooth) handoff (make before break)

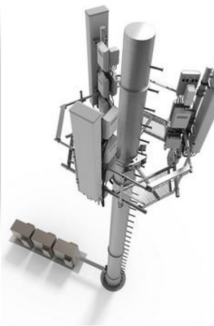
  two sessions exist during handoff

  no discontinuity, but more costly (consumes double resources)

To enable handoff (exact details depend on the protocol)

- UE continually monitors signal strength of serving BS and neighboring BSs
- UE sends these measurements to network
- the network knows channel availability in each cell
- the network decides if/when/how to perform handoff

Antennas



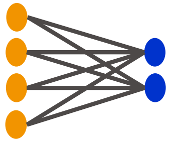
Early cell-phones had retractable antennas
but modern phones have sophisticated active internal antenna arrays

These can be very small due to high RF frequencies
e.g., 1.8 GHz has wavelength 17 cm so ¼ wavelength is about 4 cm

BS mast antennas are large in order since gain and directivity
each sector has many individual antenna elements stacked vertically
to obtain directional gain of 5-9 dBi
and antenna pattern with lobes 20º high and 120º wide

4G and WiFi have started deploying **M**ultiple-**I**nput **M**ultiple-**O**utput (MIMO)
where multiple transmit and receive antennas are used simultaneously
5G pushes MIMO to new directions

4*2 MIMO



X*R MIMO can be exploited for several purposes

- spatial diversity – exploits multiple paths between BS and user
- multi-user MIMO – allows one BS to communicate with multiple users
- beamforming – allows array of BS antennas to focus on user

RF spectral bands

Cellular communications takes place in licensed spectral bands

most common at 400,450,700,750,800,900,1800,1900,2100,2600 MHz

and 5G looking at higher "mm-wave" bands

This differs from unlicensed spectral bands, such as used for WiFi

such as the **I**ndustrial, **S**cientific and **M**edical bands e.g., 2.4, 5, 24 GHz

The spectral bands available for licensing differ by:

- different ITU regions (Americas, Europe, Africa, Asia)
- cellular generation (2G, 3G, 4G)
- modulation details (EVDO/GSM, FDD/TDD, for TDD up/down-link, etc.)

Spectrum is most often allocated to mobile SPs via **auctions**

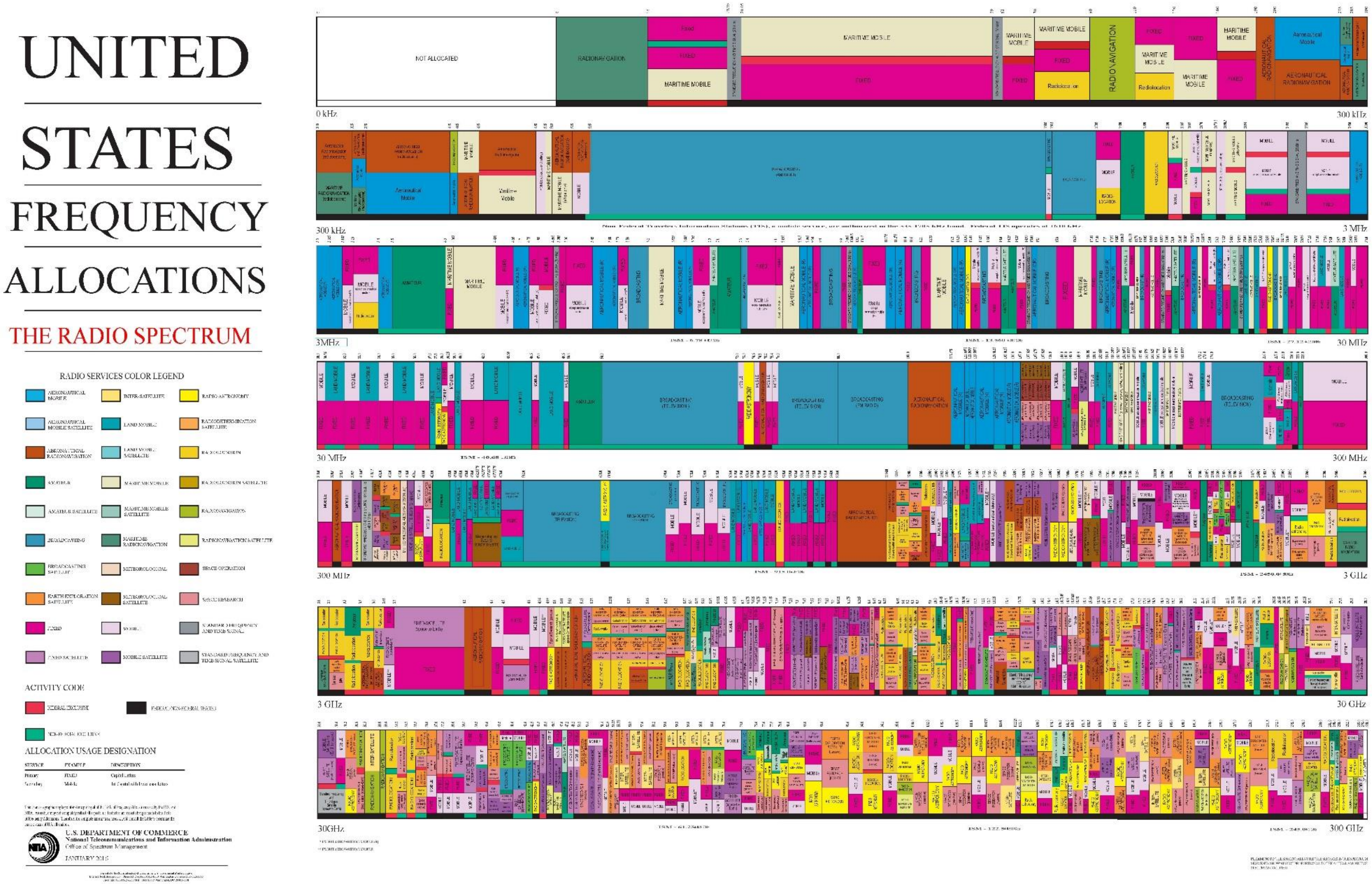
- initiated by the FCC in 1994
- 87 US auctions since 1994 assigned 1000s of licenses and raised >$60B
- 5G auctions have commenced worldwide

although alternatives (hearings, lotteries) exist

Advantages of auctions

- transparent/non-political - allocate spectrum to those who most need it
- raise money provided there is enough competition

Spectrum is tight!



Spectral efficiency

Given a spectral bandwidth the obtainable data rate
depends on the *spectral efficiency* specified in bits/sec per Hz (b/s/Hz)

Achievable spectral efficiency for a given BW depends on *path loss* and *noise*
since Shannon's capacity law states $C = BW \log_2(SNR + 1)$

Noise sources (see ITU-R P.372-13)

- thermal noise (including that of the receiver front end)
- atmospheric noise (about 40 lightning flashes per second worldwide)
- cosmic background noise
- man-made noise

Spectral efficiency has improved from cellular generation to generation

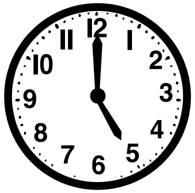
- GSM channel spacing is 200 kHz and the digital rate is 16 kBit/sec
  so the spectral efficiency is 0.08 b/s/Hz
- 3G HSDPA can provide about 14 Mbps in 5 MHz bandwidth
  so the spectral efficiency is 2.8 b/s/Hz
- LTE R8 DL max channel bandwidth is 20 MHz and max bit rate is 300 Mbps
  so the channel efficiency is 15 b/s/Hz
- 5G NR may achieve > 50 b/s/Hz !

Synchronization

BSs of neighboring cells use different frequencies
but how are these frequency generated?

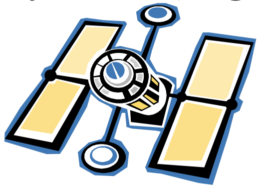
Any error in frequency will lead to spectral overlap and hence interference
but leaving *guard* frequencies would lead to inefficient use of spectrum

Some modulation/duplexing/multiple access techniques
additionally need accurate phase (time)
in particular to enable smooth handoff



Two synchronization methods are commonly used:

- deriving frequency/time from **G**lobal **N**avigation **S**atellite **S**ystem (e.g., GPS)
- deriving frequency/phase from the backhaul network
  - frequency from physical layer - SDH or SyncE
  - time from time distribution protocols
    e.g., IEEE 1588v2 and NTP



Duplexing

Almost all cellular networks use **F**requency **D**omain **D**uplexing

- (D)AMPS 824-849 MHz uplink (UE→BS), 869-894 MHz downlink (BS→UE)
- 2G GSM, e.g., 1710.2-1784.8 uplink, 1805.2-1879.8 DL
- 3G UMTS and 4G LTE FDD, e.g., 1920-1980 UL 2110-2170 DL

The alternative is **T**ime **D**omain **D**uplexing, used in:

- 802.16 WiMAX
- 3G TD-SCDMA (Chinese version)
- 4G LTE TDD (2.3 GHz and 2.5 GHz bands)

TDD allocates UL and DL time slots in the same frequency band

TDD has advantages

- easy to adapt to asymmetric data rates
- efficient for bursty data
- can attain higher spectral efficiency than FDD (less guard waste)
- channel reciprocity
- facilitates beamforming

but suffers from the need for highly accurate time synchronization

Multiple Access

The challenge of **M**ultiple **A**ccess in a single cell
is that the sources to be muxed are (at least initially) uncoordinated

Different multiple access methods have been used in cellular standards

- **F**requency **D**omain **MA** (AMPS)
- **O**rthogonal **FDMA** (LTE BS→UE DL, UL uses similar SC-FDMA to improve PAR)
- **T**ime **D**omain **MA** (D-AMPS)
- combined FDMA and TDMA (GSM)
- Code Domain **MA** (2G IS-95, 3G EV-DO)

In FDMA users are allocated frequencies in the frequency band

OFDMA optimizes spectral efficiency by allowing minimal channel spacing

In TDMA users are allocated time slots in a periodic frame

In GSM there are 125 different UL/DL frequency pairs
each supporting 8 time slots

In CDMA users are allocated (nearly) orthogonal spreading codes

Resource allocation is performed by the RAN

Modulation types

Hundreds of modulation techniques have been developed
and any have been applied to cellular traffic

1G AMPS used analog **F**requency **M**odulation (30 kHz channel BW)

2G GSM and GPRS (8 kbps) use **G**aussian **M**inimum **S**hift **K**eying (PSK variant)

2.75G EDGE data (384kbps) uses 8PSK and evolved to (1.5 Mbps) 16/32-QAM

IS-136 CDMA (8 kbps) uses DQPSK, IS-95 cdmaONE (8 kbps) QPSK/OQPSK

3G UMTS WCDMA (2Mbps) BPSK,QPSK

3G CDMA2000 EV-DO (2 Mbps) uses BPSK,QPSK,16-QAM

3.5G HSPA DL (14Mbps) BPSK,QPSK,16-QAM

3.75G HSPA+ DL (168Mbps) dual carrier, 2*2 MIMO, up to 64-QAM

4G LTE DL (100 Mbps) OFDM + PSK/QAM up to 64-QAM

4.5G LTE-A (300 Mbps) improves LTE using up to 256-QAM (+ MIMO)

5G **N**ew **R**adio is also based on LTE's OFDMA, but with improvements

History of cellular networks in Israel

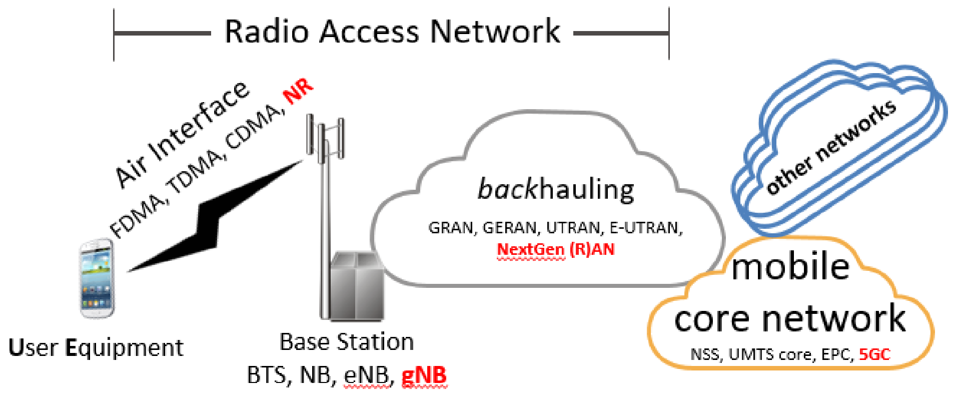
Israel was the only jurisdiction to host all three 2G technologies

- Pelephone founded in 1985 joint venture Motorola/Tadiran (now Bezeq)
  - started with 1G AMPS
  - 1995 2G IS-95 CDMA, closed 2017
  - 2003 3G CDMA2000/EV-DO
  - 2009 3G UMTS, 2010 upgraded to 3.75G HSPA+
  - 2014 LTE
- Cellcom founded in 1994
  - 2G TDMA, closed 2011
  - 2002 launched a 2G GSM network
  - 2006 launched 3G UMTS-2100, upgraded to 3.5G HSDPA
  - 2014 LTE
- Partner(formerly Orange Israel) founded in 1999
  - first 2G GSM
  - 3G UMTS, upgraded to HSPA+
  - LTE network

Air interfaces in Israel today

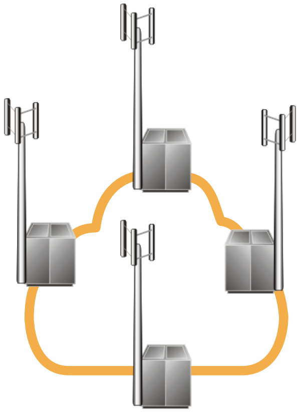
	2G GSM	3G	4G
850 MHz		Pelephone, Cellcom	
900 MHz	Partner	Partner	
1800 MHz	Cellcom, Partner		All
2100 MHz		All	

Radio Access Network



The need for backhaul

We can imagine connecting cellular base-stations directly to a core network
and modern cellular networks are indeed approaching this ideal
logically if not *physically*



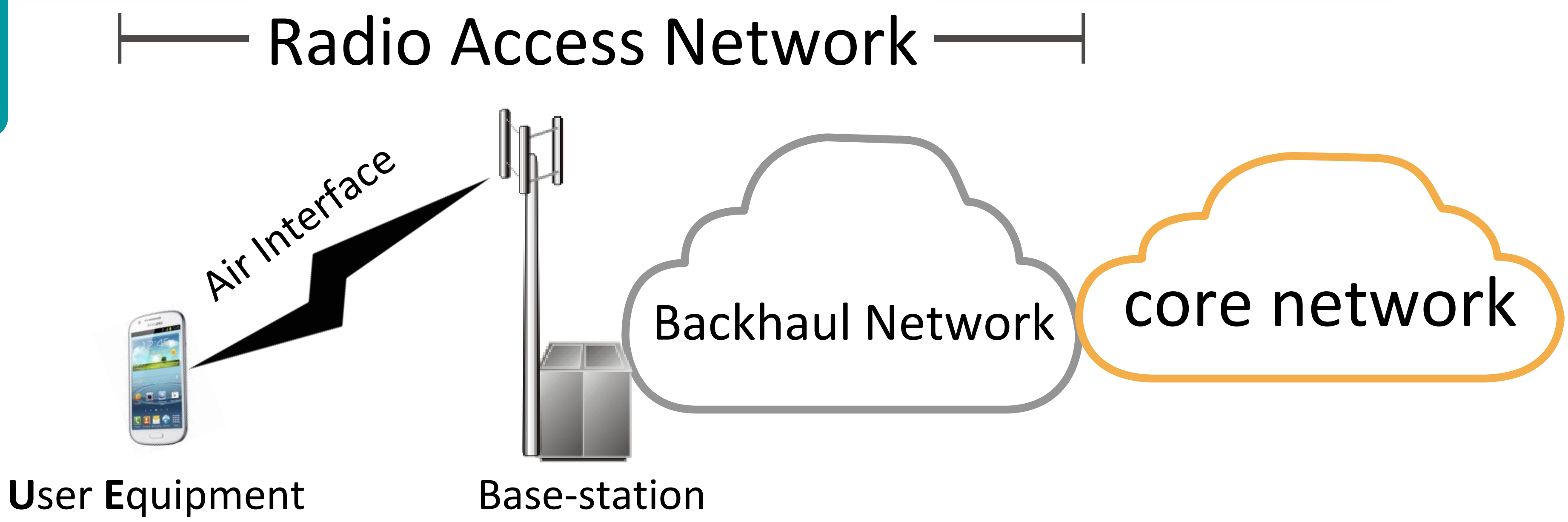
However, such a simplistic approach suffers from drawbacks
due to the large number of base-stations

- base-stations would require many physical ports (expense)
- physical ports would need to support high rates
- many fiber runs would be needed to interconnect
- base-stations would need to be routers
- base-stations would need to support redundancy for other base-stations
- core network would extend to edge with security implications

Instead, base-station connectivity is supported by a *backhaul* network
that connects the base-stations to the core network

The air interface and the backhaul network together form
the **R**adio **A**ccess **N**etwork (RAN)

Backhauling



The backhaul segment (AKA transport segment) is the part of the RAN between the base-station and the core network (AKA backbone network)

Backhaul transport is conventionally provided by

- transport division of the mobile service provider or
- 3rd party transport network provider

Backhaul topologies and protocols

Access networks are never richly connected meshes
they are conventionally limited to

- star (for small backhaul networks)
- daisy chain (e.g., GSM Abis chaining)
- tree (implemented actively or using PON)
- rings

Backhaul networks rely on various transport technologies

- PDH and/or SONET/SDH and/or ATM (2G and 3G)
- Point-to-point microwave
- Carrier Ethernet networks
- IP (IPv4 and IPv6)
- MPLS
- DSL
- satellite

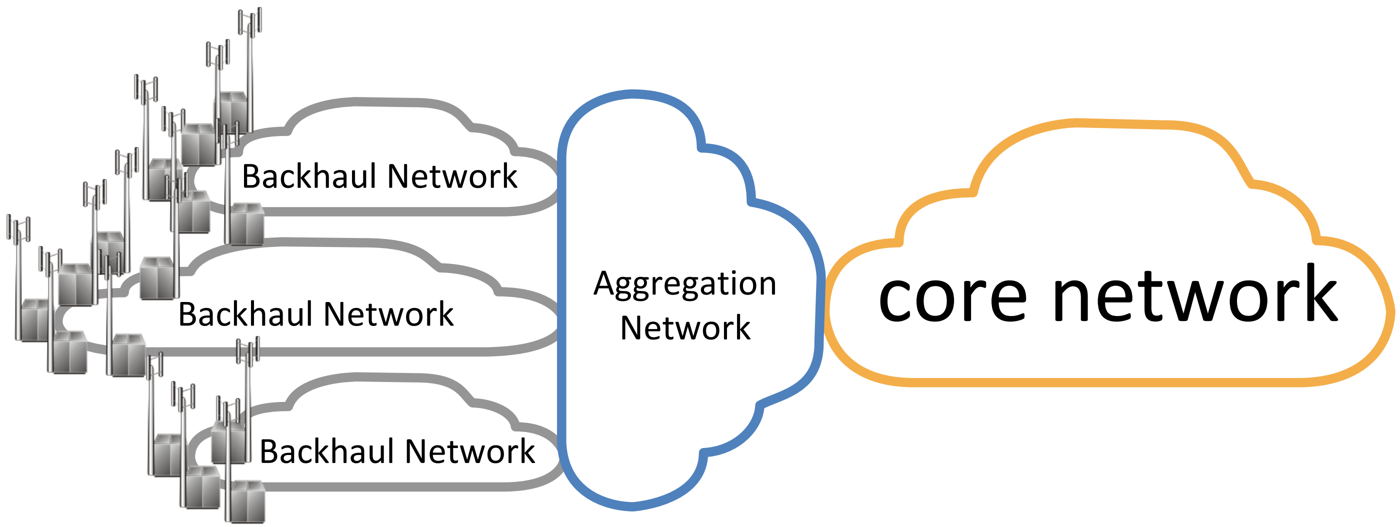
Aggregation segment

Backhaul networks are designed for

- relatively low data rates from base-stations
- very high scale
- constrained cost (CAPEX and OPEX)

Furthermore, often backhaul network protocols are not fully standardized necessitating vendor-homogenous backhaul networks

In order to support these conflicting requirements *aggregation* network***s*** are often employed



Fronthaul

Fronthaul was initially a solution to transport RF
from the antenna at the top of a tower
and the BS processing at the bottom of the tower

When transporting analog RF in coax there is signal loss and noise
to eliminate, we can perform the A/D conversion close to the antenna
and transport a digital bit-stream down the tower

For this purpose 2 protocols were developed:

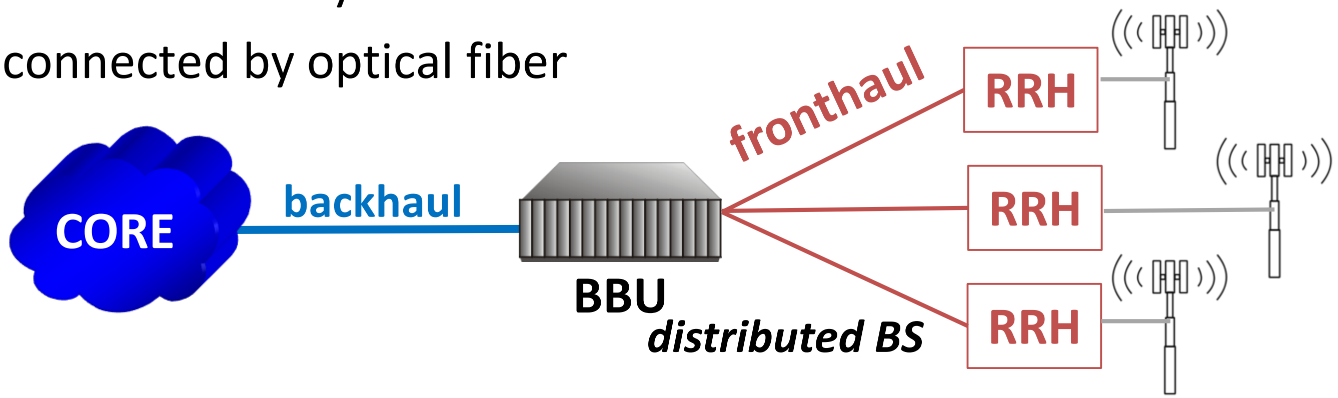
- **C**ommon **P**ublic **R**adio **I**nterface (CPRI)
  - specified by forum consisting of (Nortel), Ericsson, Huawei, NEC, Nokia
  - uses complex (I/Q) sampling
  - carried over dark fiber or OTN
- **O**pen **B**ase **S**tation **A**rchitecture **I**nitiative (OBSAI)
  - specified by forum consisting of Hyundai, LG, Nokia, Samsung, ZTE
  - uses real (Nyquist) sampling

Fronthaul architecture (before 5G)

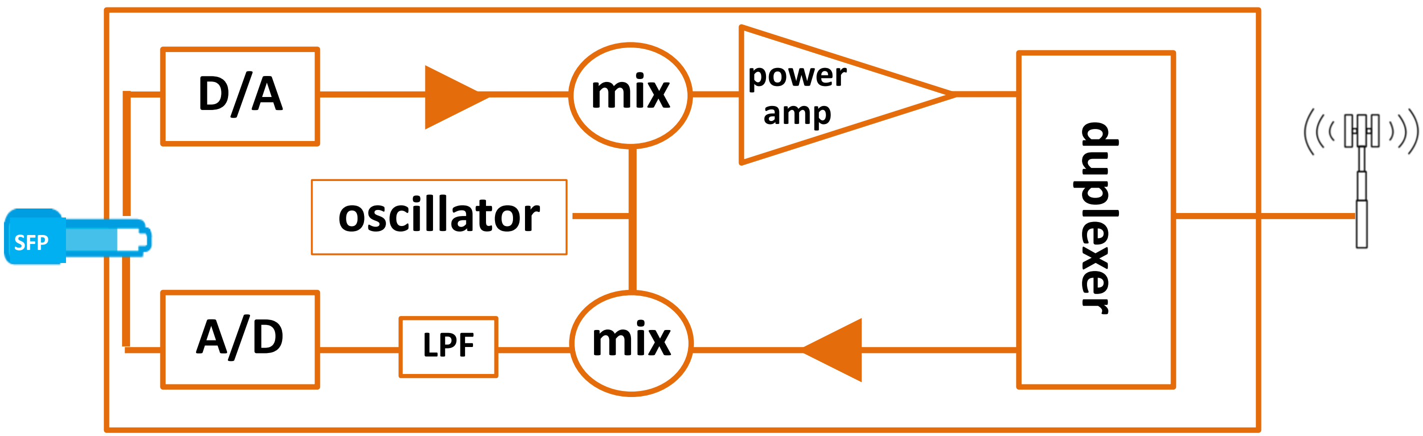
Fronthaul decomposes the BS into

- **R**emote **R**adio **H**ead at antenna (general 3 per installation – 1 per sector)
- **B**ase**B**and **U**nit remotely located

which are connected by optical fiber



The RRH is simple and inexpensive



CPRI and OBSAI rates

CPRI option	Line Bit Rate	Line Coding
1	0.6144 Gbps	8B/10B
2	1.2288 Gbps	8B/10B
3	2.4576 Gbps	8B/10B
4	3.0720 Gbps	8B/10B
5	4.9152 Gbps	8B/10B
6	6.1440 Gbps	8B/10B
7A	8.1100 Gbps	64B/66B
7	9.8304 Gbps	8B/10B
8	10.1376 Gbps	64B/66B
9	12.1651 Gbps	64B/66B
10	24.3302 Gbps	64B/66B

OBSAI Line Bit Rate	Line Coding
768 Mbps	8B/10B
1.536 Gbps	8B/10B
3.072 Gbps	8B/10B
6.144 Gbps	8B/10B

CRAN, CRAN, and vRAN (before 5G)

Fronthaul decomposes the BS into RRHs and a BBU
 but we needn't locate the BBU at the foot of the tower

It may make sense to have one BBU serve several base-stations
 using fronthaul links, that now need to be much longer

Baseband processing places strict limits on the round-trip delay
 and thus on the fronthauling distance (e.g., 20 km for LTE)

This architecture is called **C**entralized **RAN** (CRAN)
 and was first promoted by China Mobile in 2010
 in order to improve scaling of the world's largest mobile network

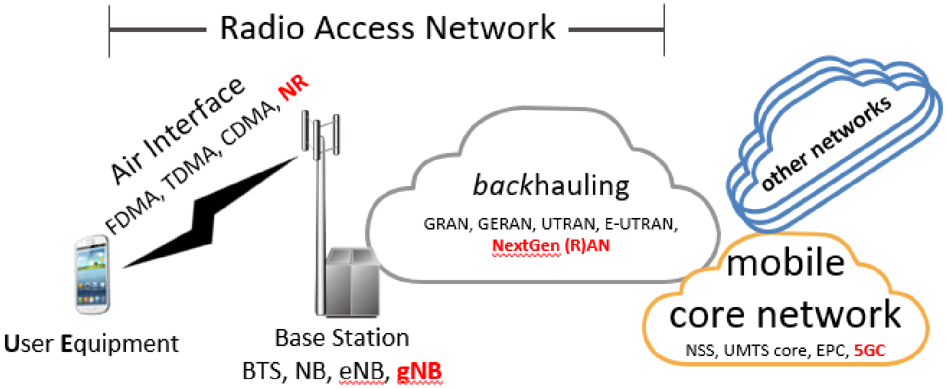
Using load-balancing/resilience techniques developed for cloud computing
 may allow a central site to service 100s of RRHs

This is called BBU Hotelling or **C**loud **RAN** (CRAN)

If we are already using cloud techniques
 we may implement at least some of the BBU functionality in software
 (e.g., packet processing, control functionality)
 running as virtual machines inside a Data Center

This architecture is called **V**irtualized **RAN**

Mobile Core



Radio Access Network
Air Interface
FDMA, TDMA, CDMA, NR
User Equipment
Base Station
BTS, NB, eNB, gNB
backhauling
GRAN, GERAN, UTRAN, E-UTRAN,
NextGen (R)AN
other networks
mobile
core network
NSS, UMTS core, EPC, 5GC

Mobile core networks

Mobile core networks were originally circuit switched (SDH)
but have migrated to packet switching using IP technologies
However, unlike the connectionless (CL) Internet
the PSTN and mobile networks are connection oriented (CO)
and mobile cores need to maintain sessions despite mobility

Like all core networks, the mobile core handles transport of user data, with
- very high data rates
- relatively small number of network elements and links
- relatively stable environment

But PSTN and mobile core networks
are very different from other IP core networks (e.g., Internet ASes)
- the PSTN assumes dumb terminals and intelligence in the network
- the Internet assumes smart terminals and dumb pipes in the network
(the *end-to-end principle*)

The PSTN and mobile core networks have a rich set of functionalities
not present in the Internet model

Some mobile core functionalities

Some of the functionalities of a mobile core network:

- end-to-end transport of voice traffic (at least until 4G)
- connection oriented data transport (see session management)
- maintenance of sessions despite mobility (see mobility management)
- *user* (not necessarily true end-user) management
  - authentication and registration
  - mobility management (tracking where users are)
  - user profile, home location, roaming
  - billing (AKA charging)
- session management (call establishment, management and termination)
- lawful interception (CALEA) and metadata collection
- QoS enforcement (network neutrality is usually not relevant for mobile)

Elements of the core network (before 5G)

Home **S**ubscriber **S**erver / **H**ome **L**ocation **R**egister
- database containing user-related and subscriber-related information
- support mobility management, call/session setup, user authentication

Serving **G**ate**W**ay (SGSN)
- transport traffic between UE and network
- mobility anchor point

Packet Data **G**ate**W**ay (GGSN)
- interconnection between mobile core and external packet networks

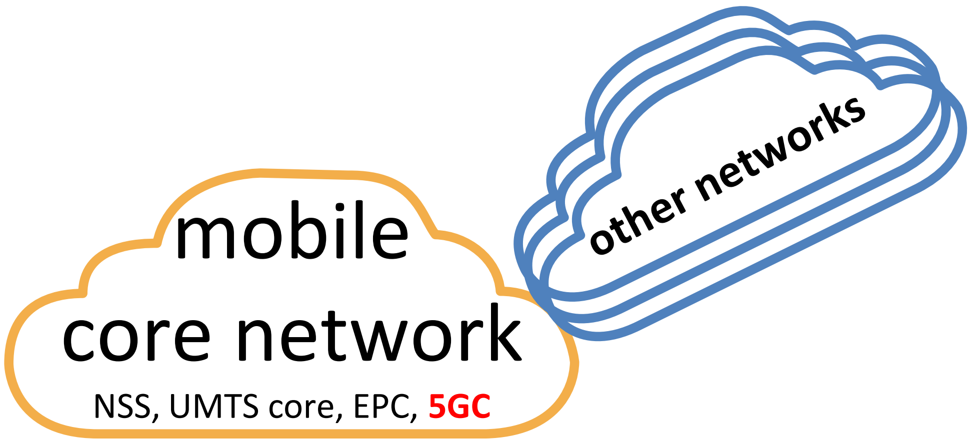
Policy and **C**harging **R**ules **F**unction
- software component that determines and disseminates policy rules
- separated rom PCEF enforcement function

Mobility **M**anagement **E**ntity
- control plane entity handling signalling related to mobility and security
- tracking and paging UEs in idle-mode

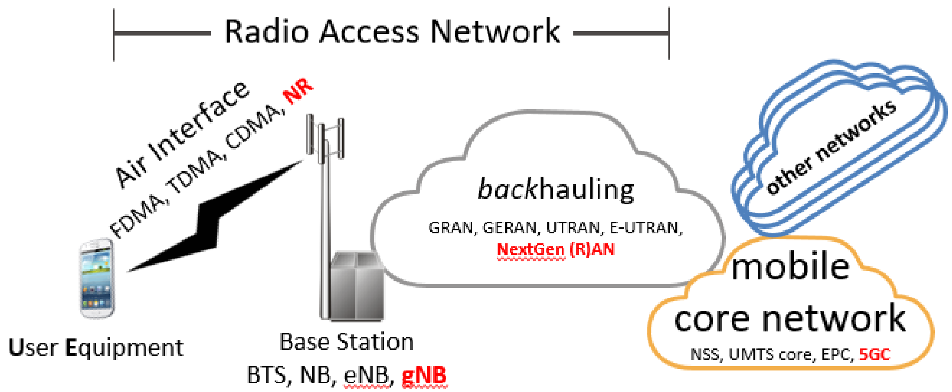
Other Networks

The mobile core network needs to interconnect with other networks, e.g.,

- **P**ublic **S**witched **T**elephone **N**etwork – for voice calls to non-cellular users
- Core networks of other mobile service providers
- Internet, however there may be connectivity in the RAN
  - local breakout (facilitated by **M**obile **E**dge **C**omputing) is more efficient
  - **C**ontent **D**elivery **N**etworks change the architecture
- Data Centers (belonging to the SP or having connectivity agreements)
- Corporate Networks



Architecture



Cellular system architecture

Telecommunications systems architectures are composed of

- **N**etwork **E**lements or **N**etwork **F**unctions
  in general NEs may be composed of many NFs
  in which case the NFs may be *microservices*
- Interfaces or reference points between them

Architectures differ in

- types of NEs/NFs
- granularity of NEs/NFs
- flat networks vs. hierarchical/heterogeneous
- separation of user, control, and management planes

Proper (top-down) design starts with architecture
and only afterwards designs protocols

2G GSM interfaces and NEs

In the beginning, there was the **A interface** – the standard PSTN voice trunk

2G mobile utilized framed compressed voice, leading to the **Abis interface**

The user mobile (air) interface was called the **Um interface**

The BS is called the **B**ase **T**ransceiver **S**tation

A **B**ase **S**tation **C**ontroller controls a group of BTSs (often co-located with one)
- manages radio resources, allocates channels, controls handoff

The **M**obile Services **S**witching **C**entre acts like a PSTN switch, but provides
- interconnection to the PSTN and other mobile networks
- authentication, registration, call location/routing, inter-MSC handoff

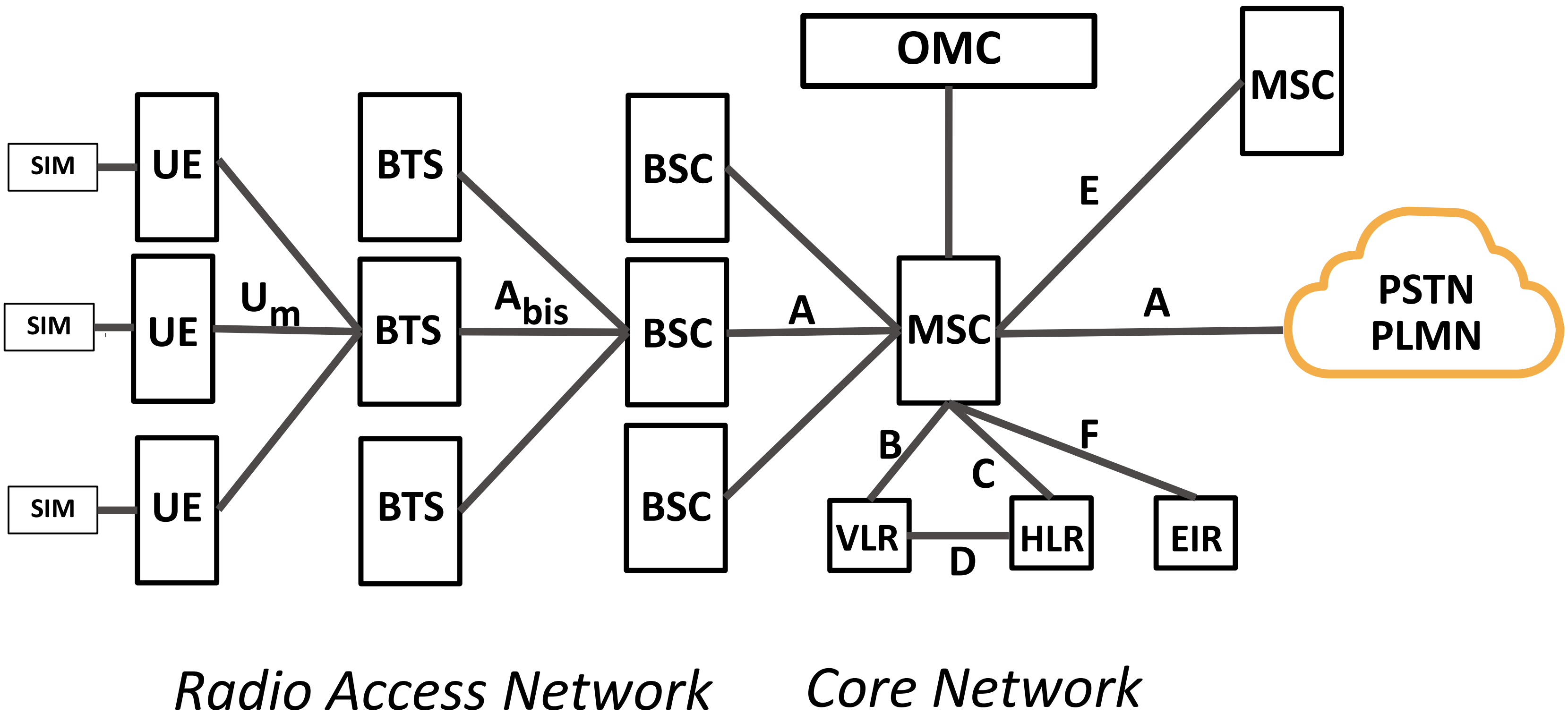
The **H**ome **L**ocation **R**egister database contains all user information
- including last known location (even if the UE is idle)

The **E**quipment **I**dentity **R**egister that lists block phones
- connects to the MSC over the F interface

The **V**isitor **L**ocation **R**egister receives information from the HLR

The **O**perations and **M**aintenance **C**enter manages the entire system

2G GSM architecture (simplified)



2.5/3G GPRS NEs

The **G**eneral **P**acket **R**adio **S**ervice extends GSM for best effort packet service

In 3G the BTS is replaced by the **N**ode**B**

The BSC is replaced by the **R**adio **N**etwork **C**ontroller, which

- is responsible for controlling the NBs
- performs radio resource and some mobility management functions
- performs encryption of data sent to and from UE

The MSC is the same as in the GSM architecture (replaced in R5 by MGW)

- but supplemented by a **G**ateway **MSC** to connect to the PSTN

Serving **G**PRS **S**upport **N**ode for packet traffic is similar to MSC+VLR for voice

- performs access control, security functions and tracks UE location

Gateway **G**PRS **S**upport **N**ode for packet traffic is similar to the GMSC

- acts as an IP router to connect to external IP networks
- also handles billing, filtering and firewall functions

2.5/3G GPRS interfaces

The U_m air interface is replaced with the $\mathbf{U_u}$ interface

Iu-CS interface is the interface between the RNC and MSC (MGW in R5)

Iu-PS interface is the interface between the RNC and the SGSN
originally over ATM but migrated to IP (SIGTRAN SS7 over IP)

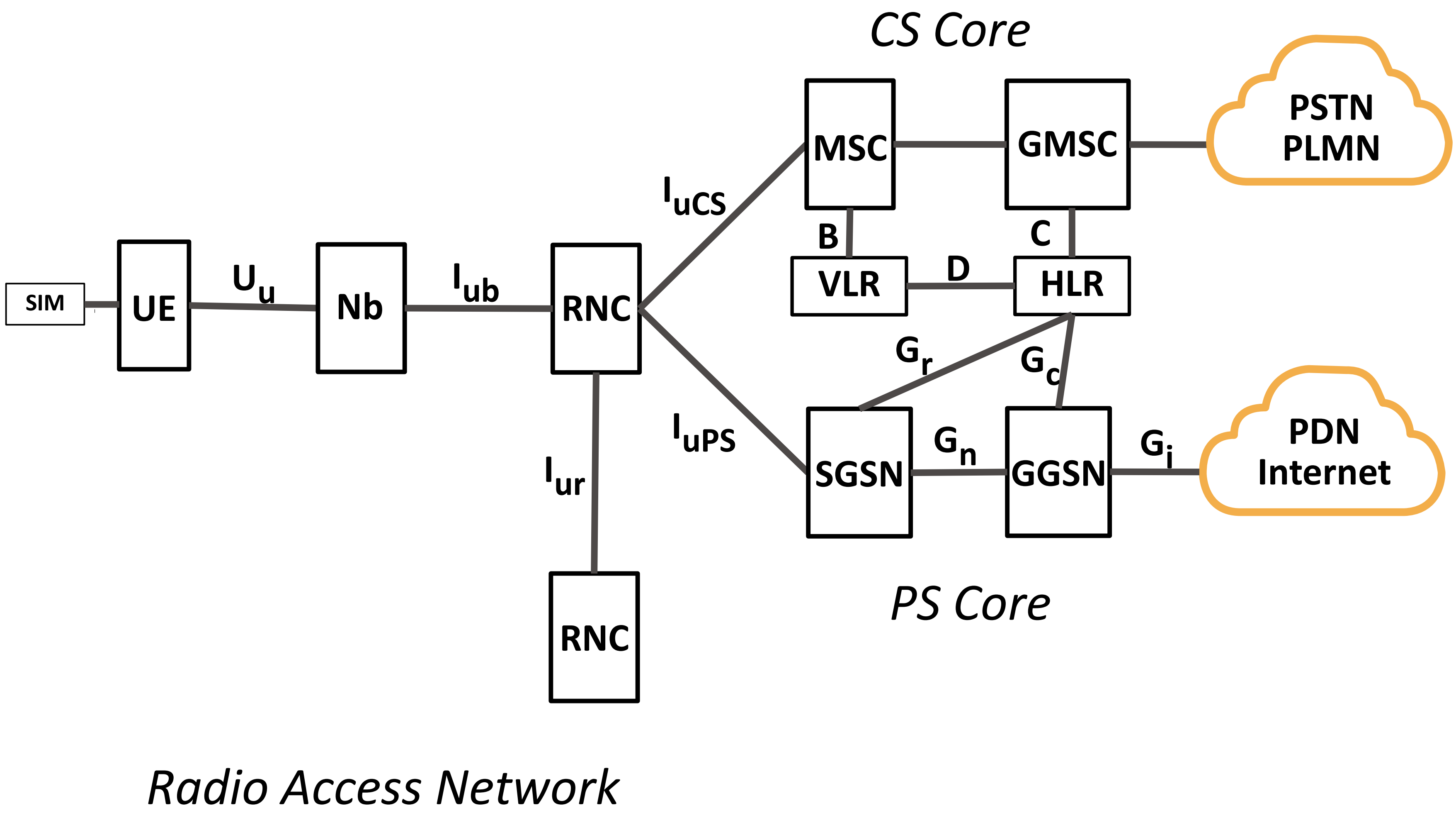
Iur is an optional interface between RNCs for smooth handoff

The **Gn** interface transports user data and signaling between SGSN and GGSN
it uses **G**PRS **T**unnelling **P**rotocol to tunnel through IP networks

The **Gp** interface is the same as Gn interface but instead to foreign networks

Other **G** interfaces are for mobility management (HLR, authentication), etc.

R99 GPRS architecture (simplified)



4G LTE NEs and interfaces

Long **T**erm **E**volution was developed as 3GPP **S**ystem **A**rchitecture **E**volution

- **E**volved **U**MTS **T**errestrial **R**adio **A**ccess **N**etwork (E-UTRAN)
- **E**volved **P**acket **C**ore

The architecture is simpler in that

- the BS is a single NE called the **e**volved **N**ode **B** (replacing the NB and RNC)
- the EPC has more user/control separation
- the EPC is a single flat IP network, consisting of
  - **S**erving **G**ate**W**ay (S-GW) connecting to the E-UTRAN
  - **M**obility **M**anagement **E**ntity (most of the gateway signaling)
  - **P**acket Data Network **G**ate**W**ay (P-GW) to connect to other PDNs
  - **H**ome **S**ubscriber **S**erver (HSS) database (instead of HLR)
  - R11 introduced the **T**raffic **D**etection **F**unction (**D**eep **Packet** **I**nspection)

The air interface is still called $\mathbf{U_u}$

The main backhaul from eNodeB to core is called the **S1** interface(s)

but there are other S interfaces, such as S2 for non-3GPP access

eNodeBs are interconnected (for handoff and other reasons) by **X2** interfaces

Functions of some EPC NEs

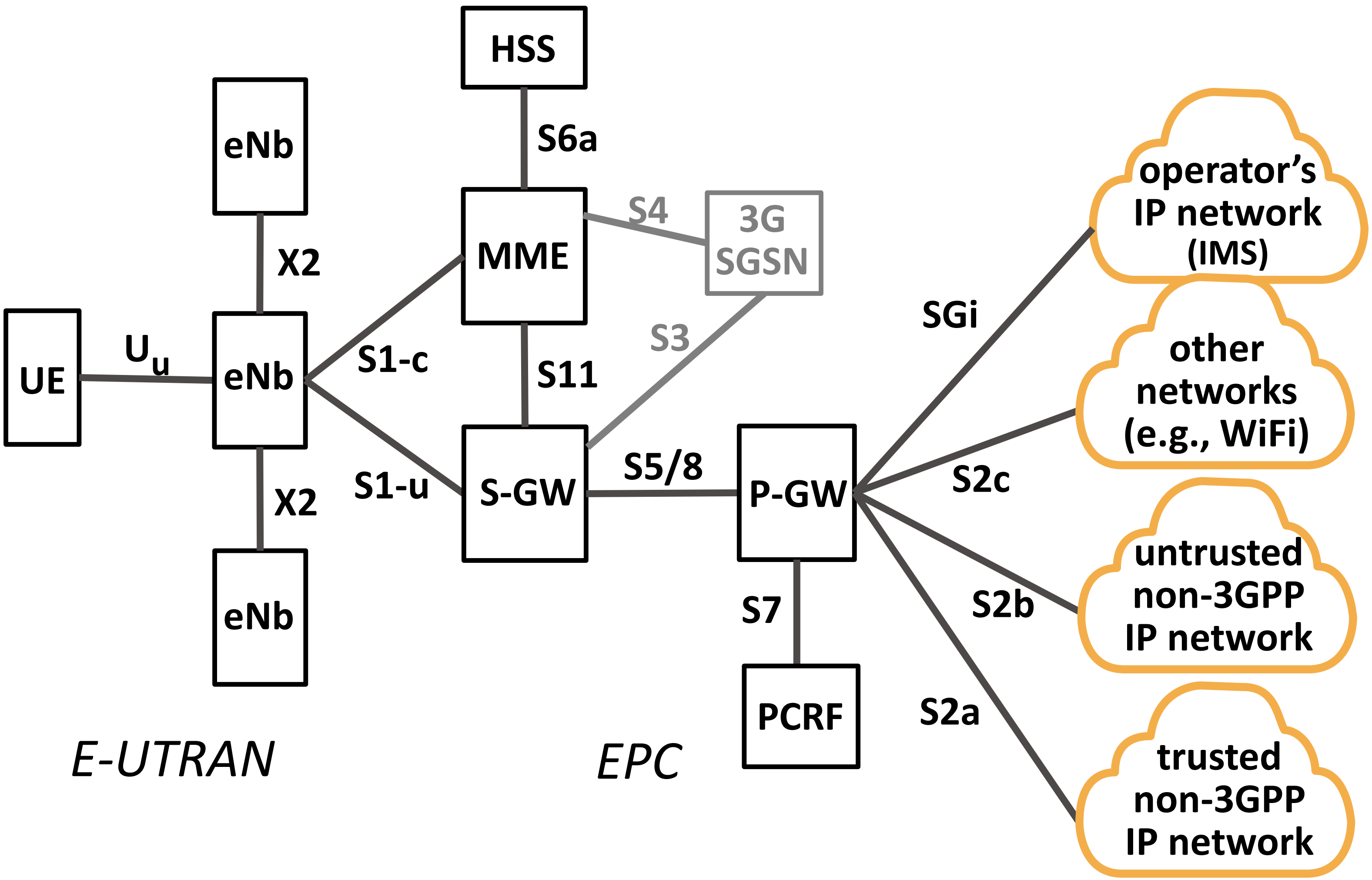
Mobility Management Entity (MME)
- handling of security keys
- transitions between idle to active states
- sending paging messages to the eNodeBs
- bearer setup procedures
- contacting the HSS for subscriber information

Serving Gateway (S-GW)
- connecting RAN to EPC
- mobility anchor when UEs move between eNodeBs
- terminates data tunnels between eNodeB and PDN-GW

Packet Data Network Gateway (PDN-GW)
- connecting EPC to the internet (and other PDNs)
- assigning IP addresses for terminals

Home Subscriber Server (HSS)
- subscriber database
- allow seamless roaming between networks
- connects to MME via S6 interface

4G LTE architecture



4G R14 and CUPS

In 3G the SGSN and GGSN handle both user and control plane functions

In LTE there is better separation GWs vs. MME

in order to enable different scaling – traffic volume vs UEs/signaling load
(e.g., 2009 iPhone AT&T 3G network meltdown)

Release 14 introduced EPC **C**ontrol and **U**ser **P**lane **S**eparation, in order to

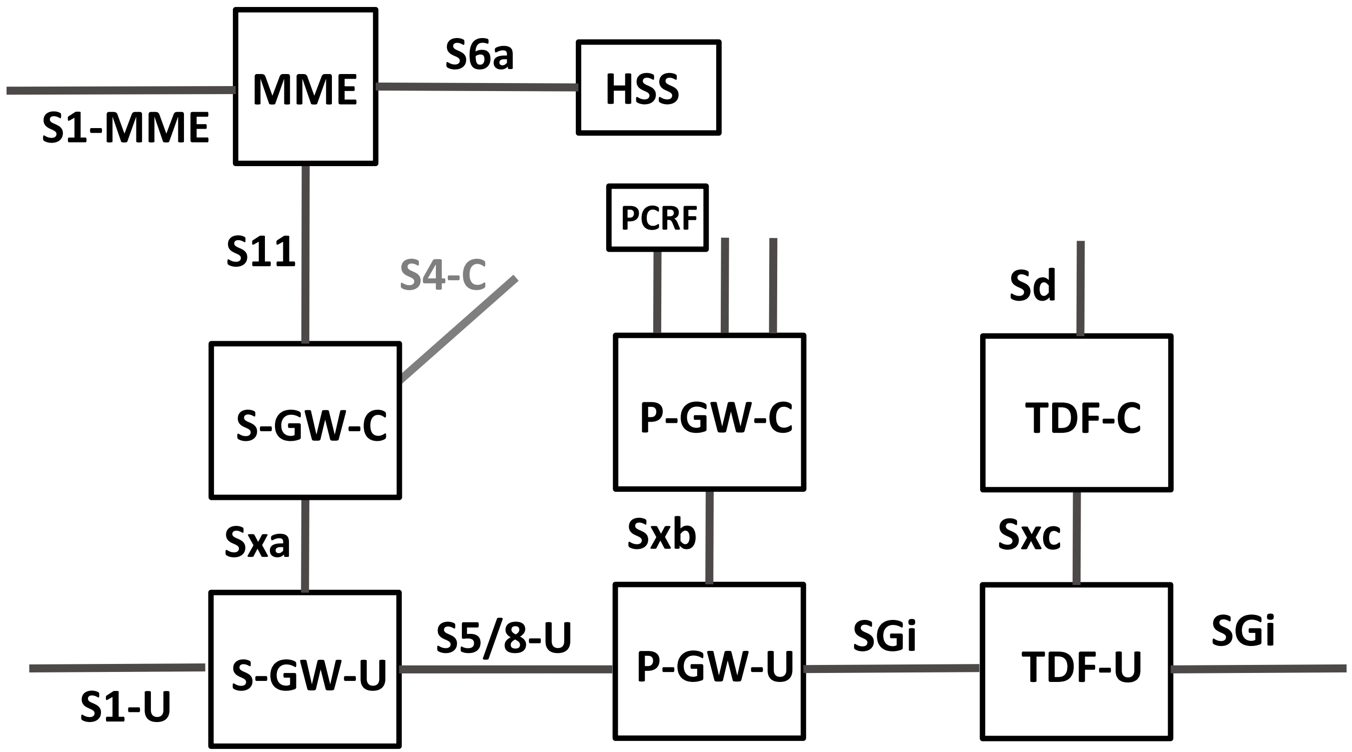
- support increased traffic volume and reducing latency
  by adding user plane NEs w/o adding control nodes
- independently locating UP and CP EPC nodes
- independent evolving UP and CP functions
- enabling distributed or centralized control
- enabling **S**oftware **D**efined **N**etworking

CUPS separates S-GW, P-GW, and TDF

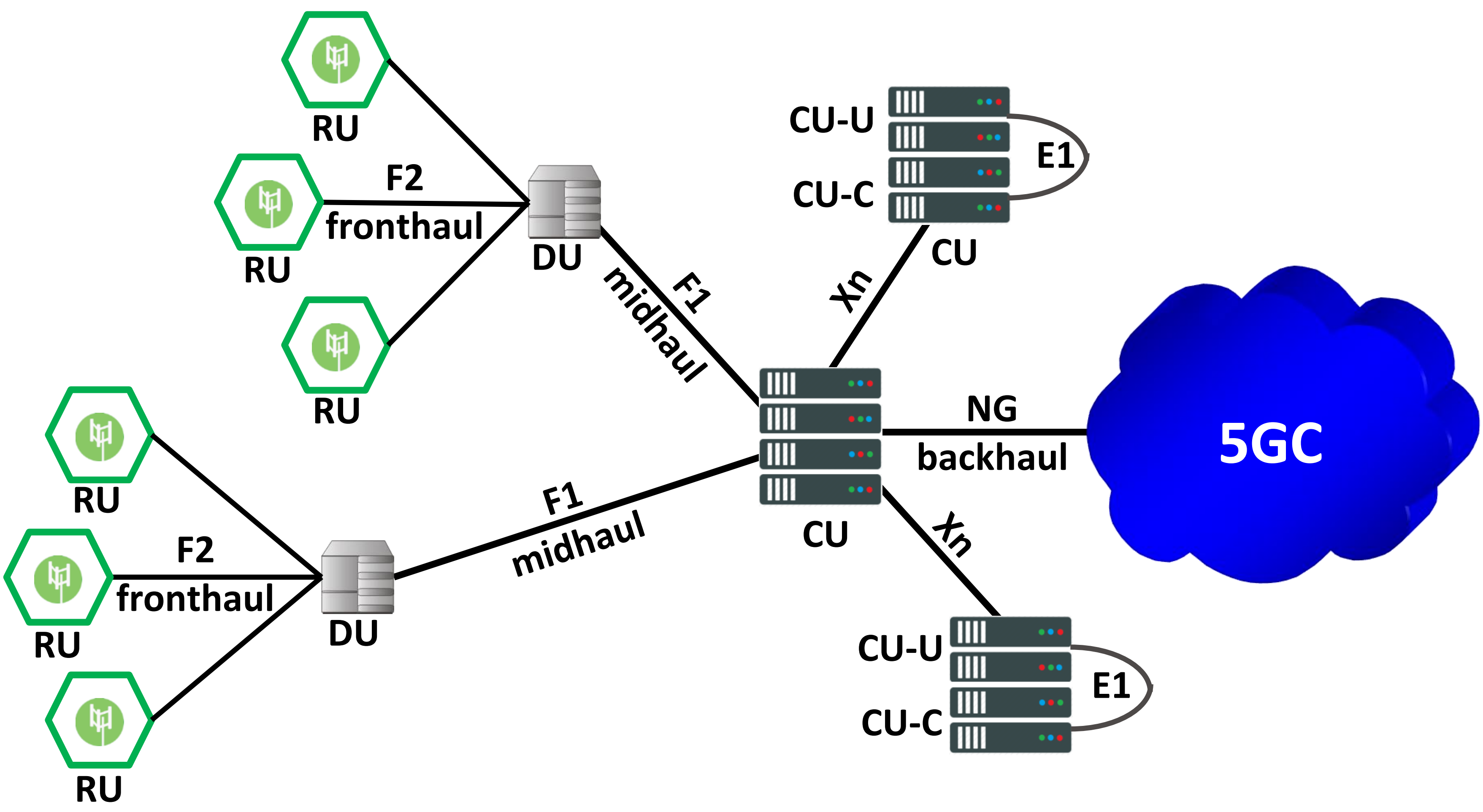
into **U**ser **P**lane **F**unctions and **C**ontrol **P**lane **F**unctions
S-GW-C and S-GW-U,   P-GW-C, P-GW-U,   TDF-C and TDF-U
and introduces new Sx interfaces

The CUPS architecture leads to the 5G **S**ervice **B**ased **A**rchitecture

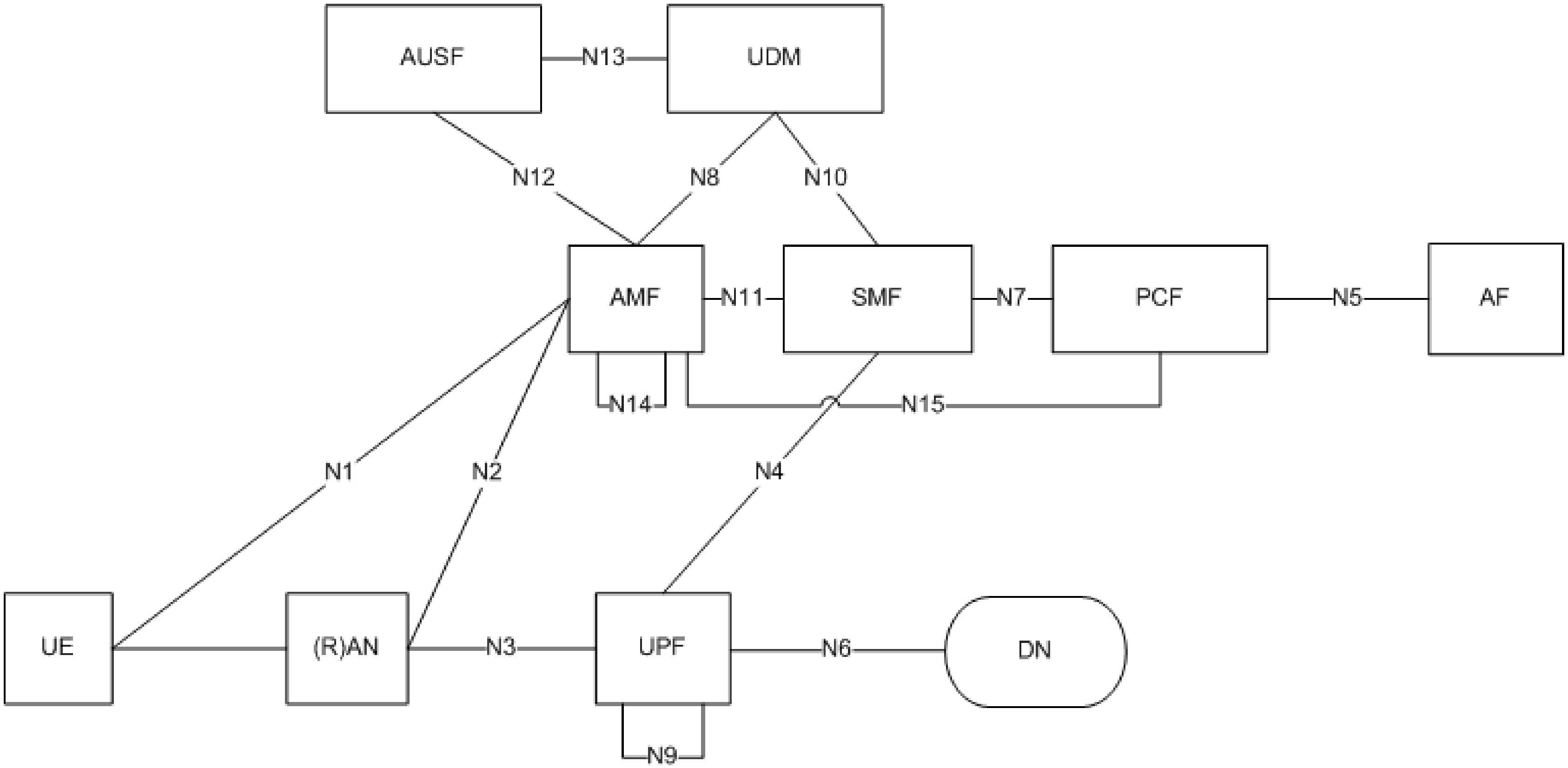
CUPS EPC architecture detail



5G architecture - RAN



5G core (simplified) without SBA



5G core (simplified) with SBA