Software-i-zation and AI at the wireless edge

AI-Edge summer REU program

Jim Kurose
University of Massachusetts, Amherst

EDGE Institute

Overview of today's REU seminar

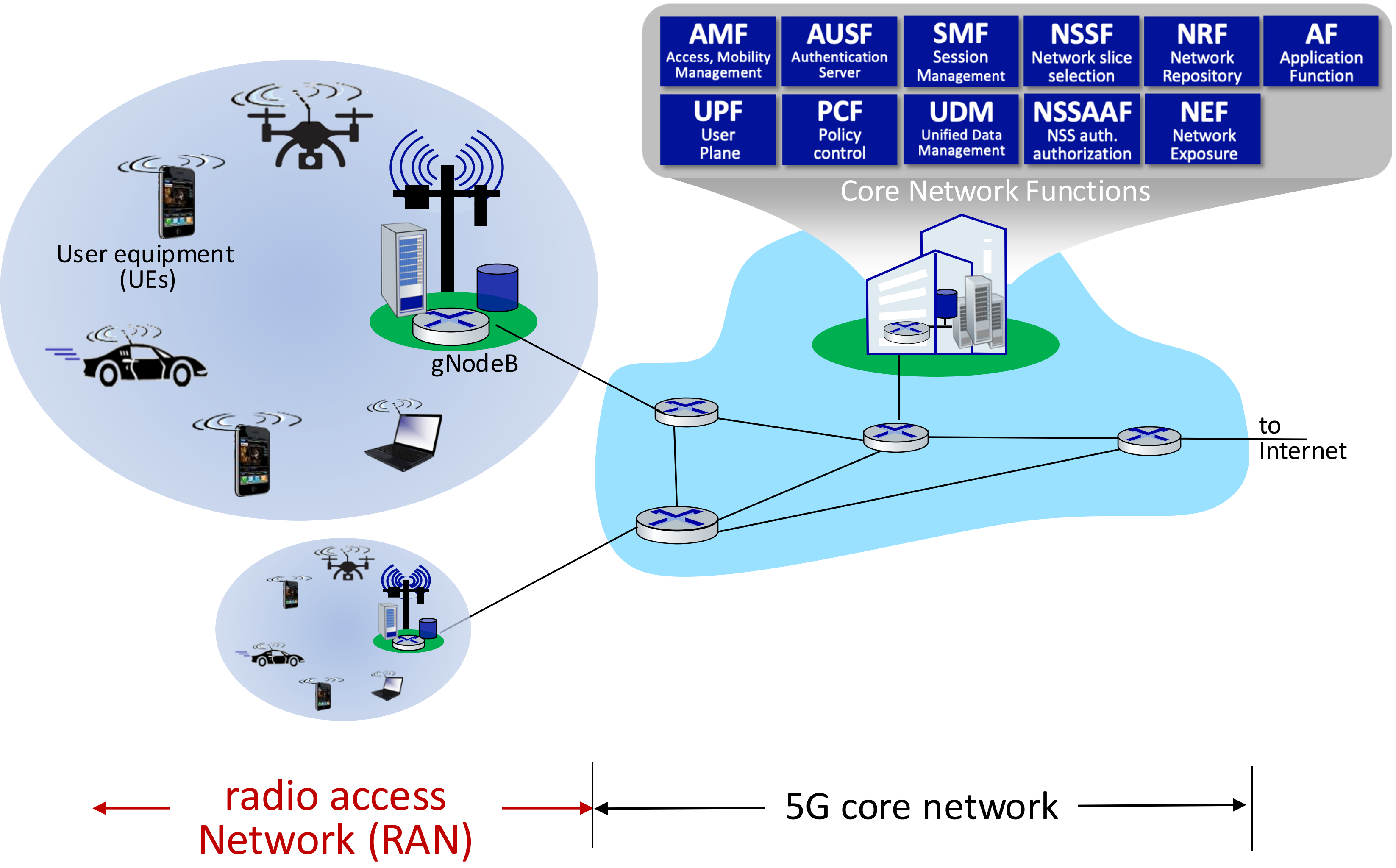
- wireless (5G, nG) networks: review
- network software-i-zation
- three 5G/nG network control/management challenges

  … and opportunities for AI/ML

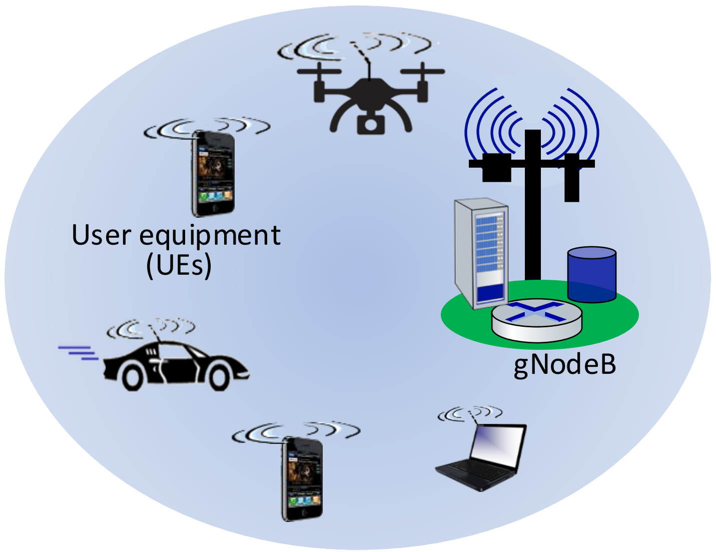
For more information:

- Networking: http://gaia.cs.umass.edu/kurose_ross
- Wireless Networking: https://gaia.cs.umass.edu/wireless_and_mobile_networking

Architectural Elements of 5G



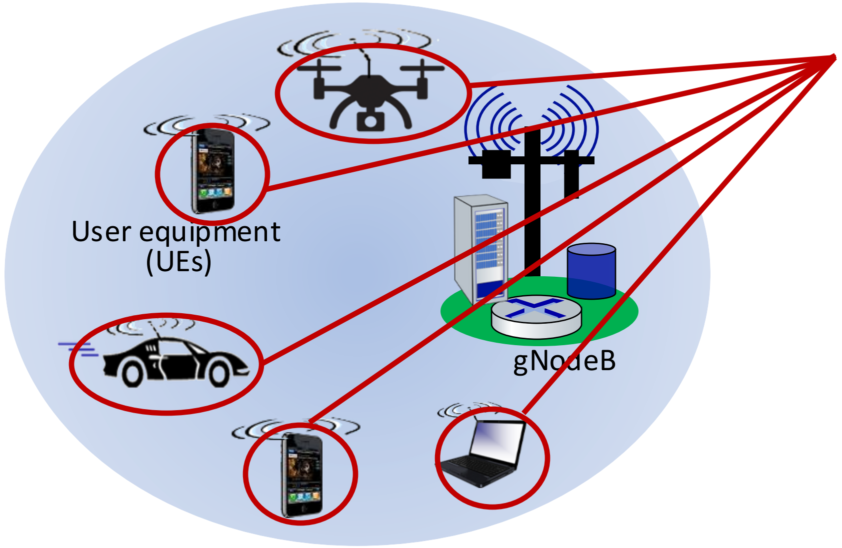
5G Radio Access Network (RAN)



5G RAN: edge network connecting devices (UEs) to base station (gNodeB)

- provides link-layer service, as first hop between devices and larger network
- limited geographic scope
- under control of a single service provider
- somewhat analogous to WiFi LAN
- RAN components:
  - many devices (User Equipment: UE)
  - radio channel (New Radio: NR)
  - one base station (Next Generation Node B: gNodeB, gNB)

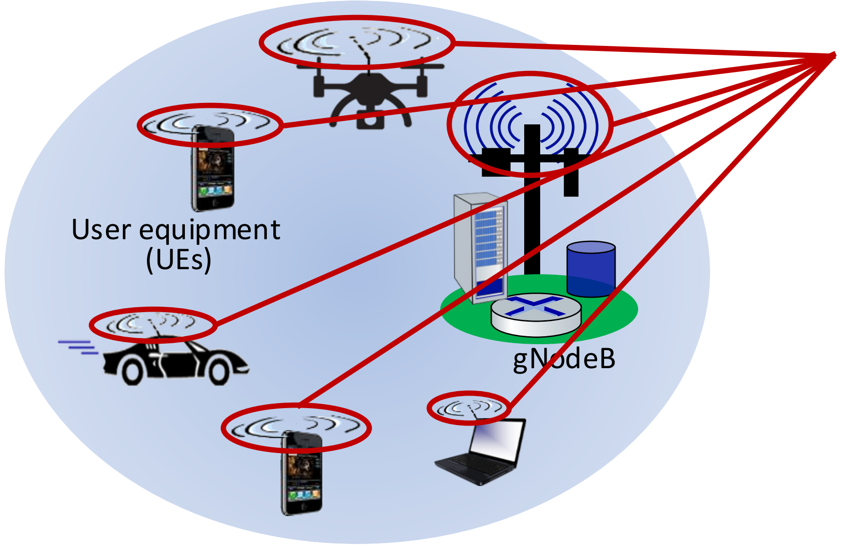
5G RAN components



5G User Equipment (UE):

- smartphone, tablet, laptop, IoT device
- UEs host, run applications
- devices attached in RAN
- may or may not be mobile

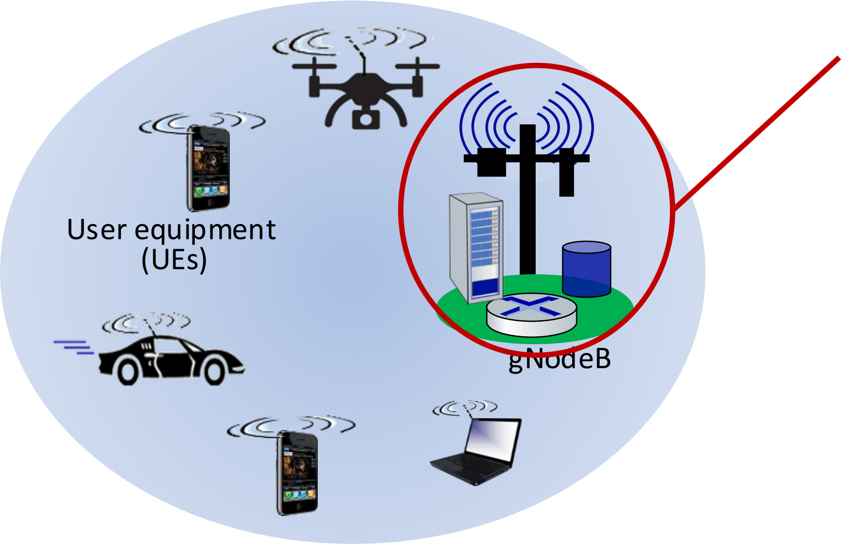
5G RAN components



Radio Channel (aka New Radio, NR, in 5G)

- Physical (PHY) layer: OFDMA
- UEs only communicate with base station
  - via uplink, downlink channels
  - UEs do not communicate directly with each other
- various uplink/downlink physical and logical channel defined

5G RAN components

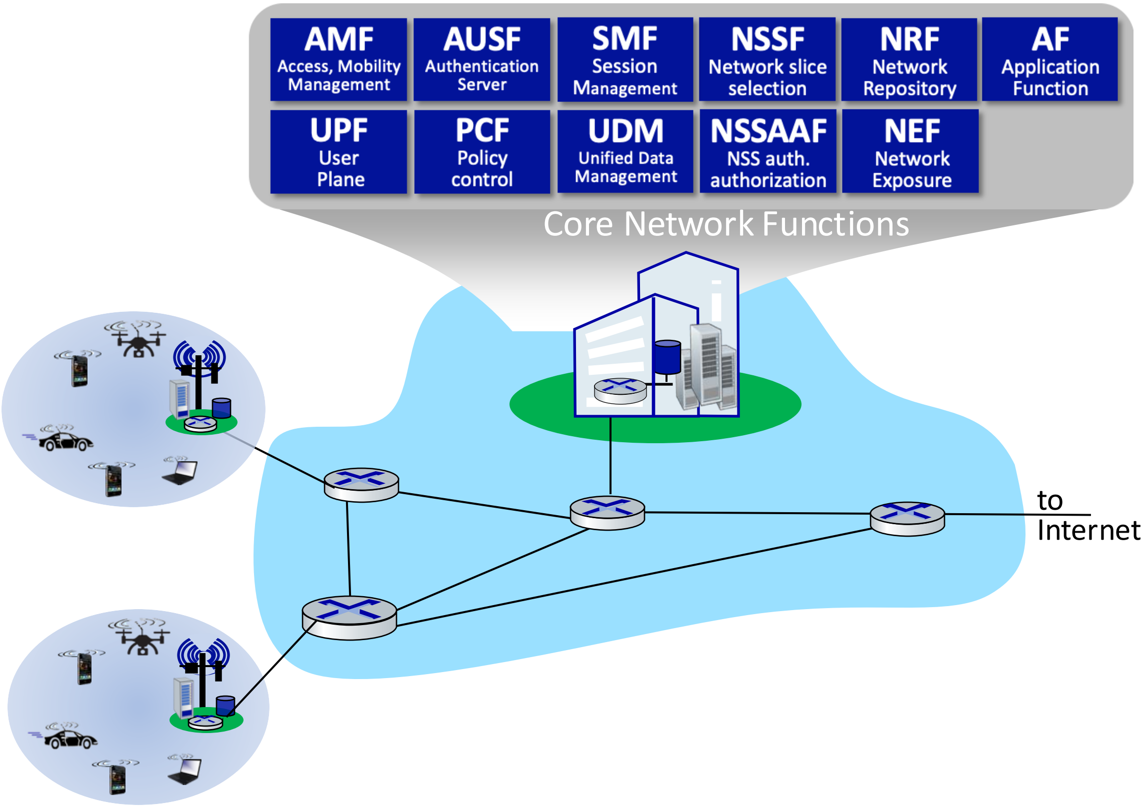


Base station (aka Next Generation Node B (gNodeB or gNB in 5G):

- central control point for RAN, role somewhat analogous to that of WiFi AP
- computing, storage may be located at gNB
- all communication between UEs and other endpoints (other UEs, 5G Core, Internet) go through gNB
- aka Extended Node B (eNB) in 4G

5G Core Network

- Core situated between RAN and other endpoints (Core, larger Internet)
  - single Core; multiple RANs
- consist of links, routers, servers, providing services to UEs and gNBs
  - “all Internet” Core, but very different services than traditional Internet apps
- clear logical separation between control-plane, user plane:
  - *CUPS:* Control-Plane and User-Plane Separation

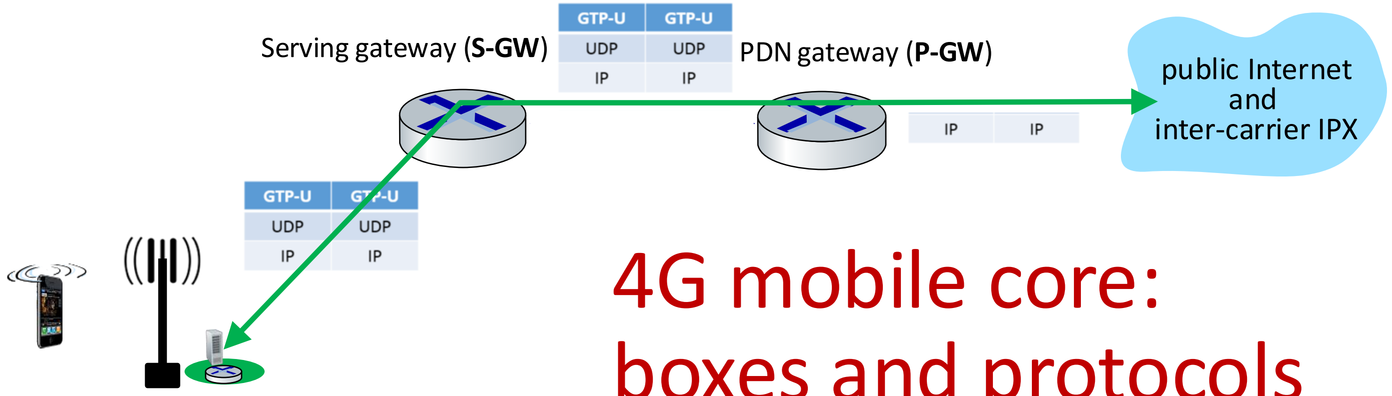


Overview of today's REU seminar

- Wireless (5G, nG) networks: review
- network software-i-zation
- Three 5G/nG network control/management challenges

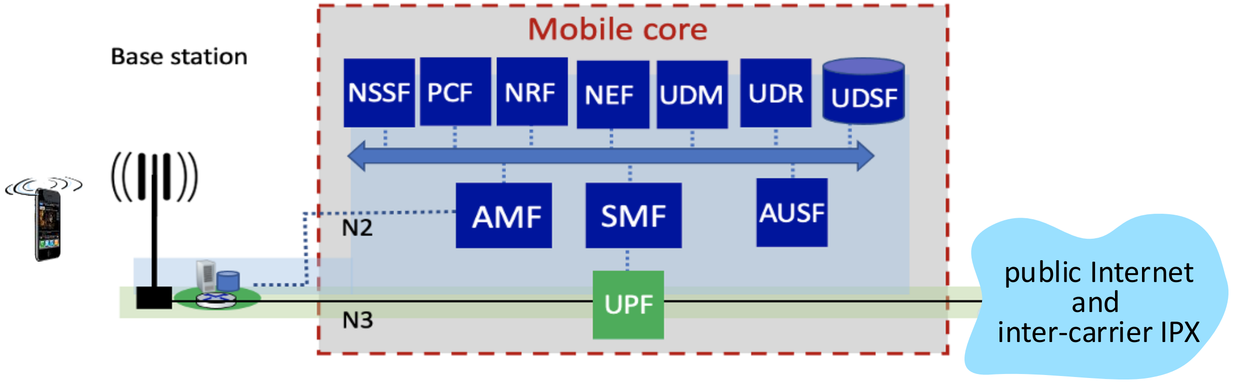
  ... and opportunities for AI/ML

5G: migration away from "protocols" *is* underway!



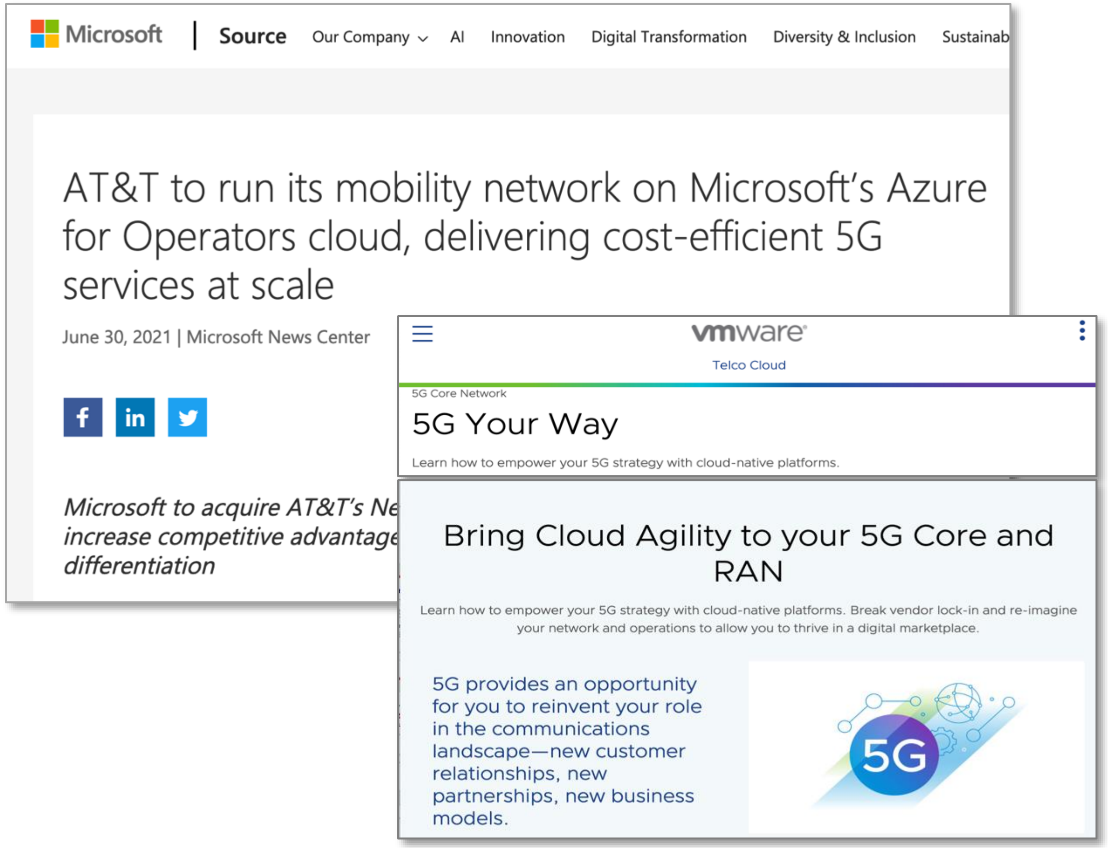
4G mobile core:
boxes and protocols

Ericsson SGSN-MME



5G: microservice-like architecture

5G: migration away from "protocols" *is* underway!


Microsoft | Source Our Company AI Innovation Digital Transformation Diversity & Inclusion Sustainab

AT&T to run its mobility network on Microsoft's Azure for Operators cloud, delivering cost-efficient 5G services at scale

June 30, 2021 | Microsoft News Center

Microsoft to acquire AT&T's Ne
increase competitive advantage
differentiation

vmware
Telco Cloud

5G Core Network

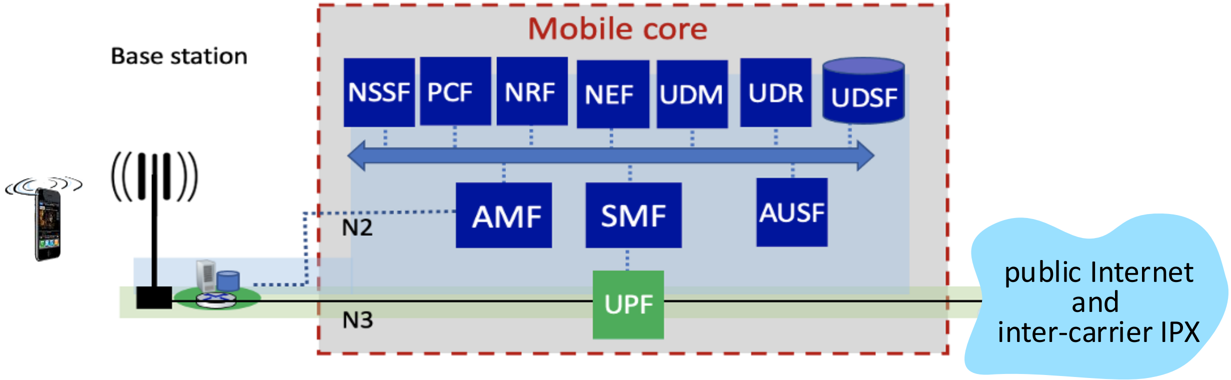
5G Your Way

Learn how to empower your 5G strategy with cloud-native platforms.

Bring Cloud Agility to your 5G Core and RAN

Learn how to empower your 5G strategy with cloud-native platforms. Break vendor lock-in and re-imagine your network and operations to allow you to thrive in a digital marketplace.

5G provides an opportunity for you to reinvent your role in the communications landscape—new customer relationships, new partnerships, new business models.

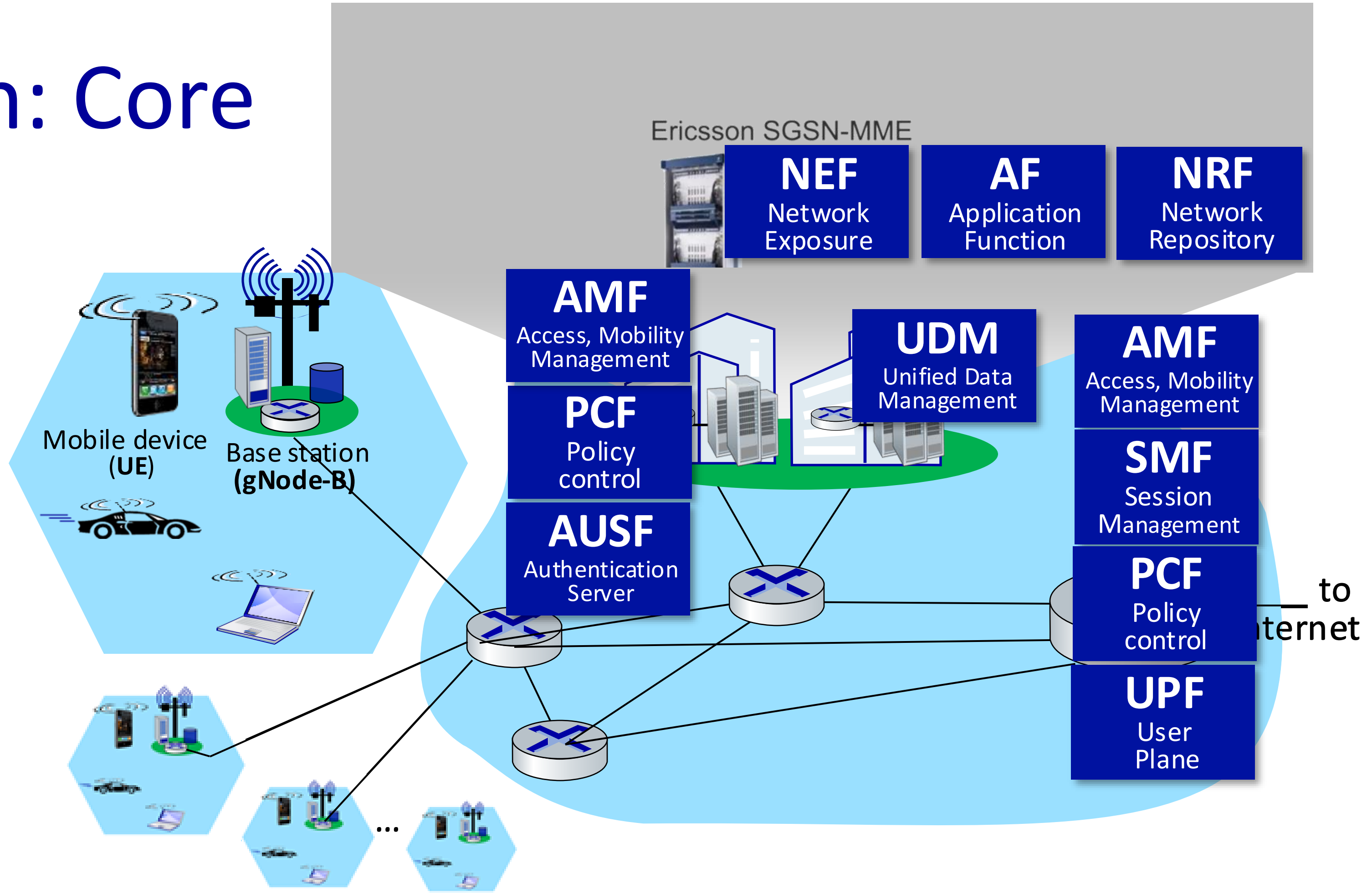


5G: microservice-like architecture

Note: 4G/5G operators manage *intradomain* services in their mobile network

5G: software-i-zation: Core

- from 4G boxes to 5G functions and services
- cloud infrastructure within the mobile 5G network
- software defined networking (SDN): 5G microservices can be implemented in data centers

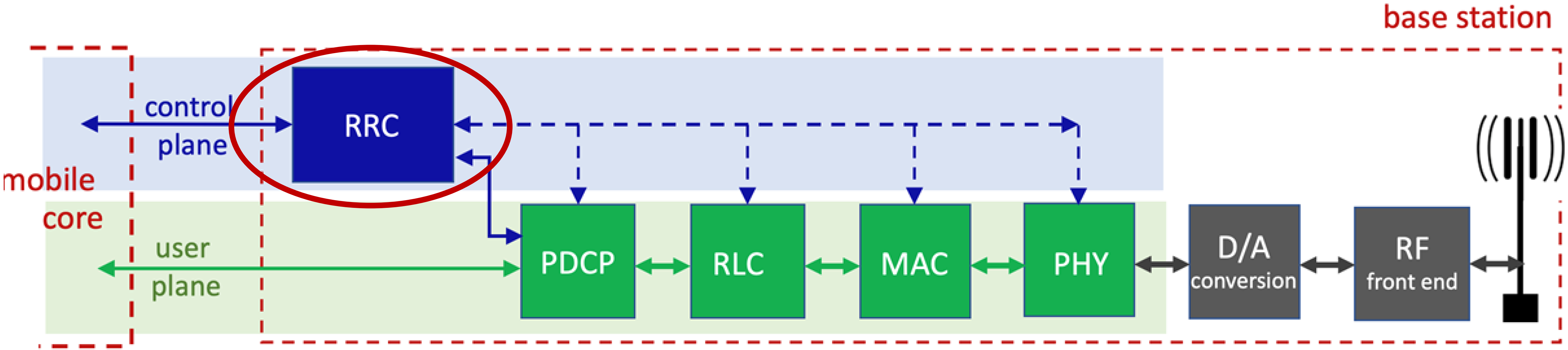


5G: software-i-zation: RAN

RAN: transfers datagrams between mobile core and UEs

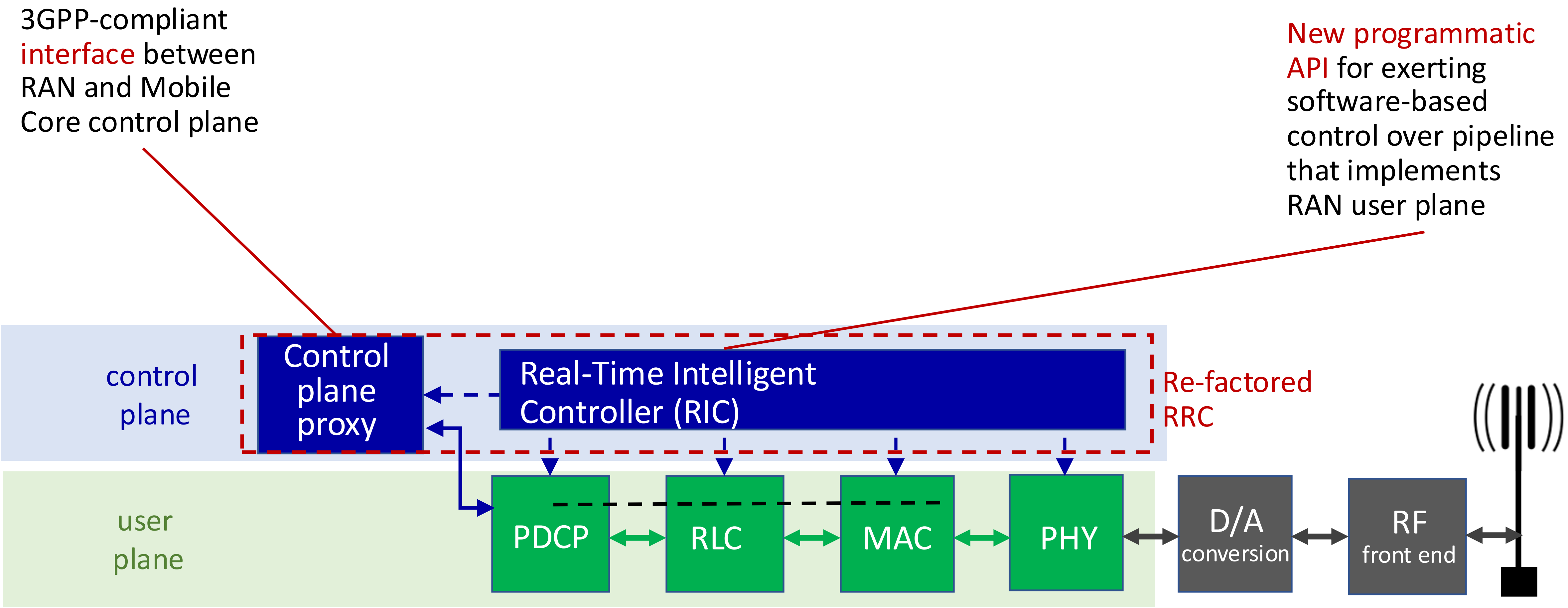
Recall our earlier description of traditional RAN base station (below)

- tightly coupled control and data planes
- let's focus on control / management: RRC *implementation*



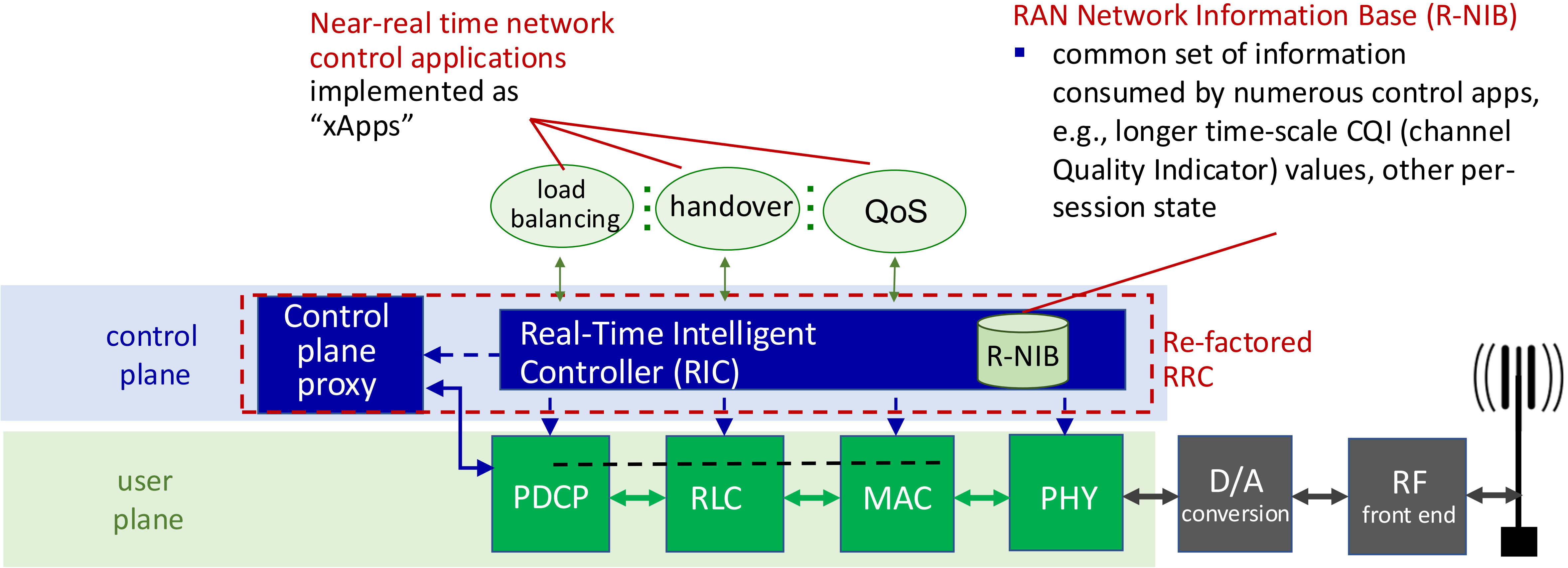
Software-defined RAN

- *SD-RAN:* implementing RAN using SDN approach



Software-defined RAN

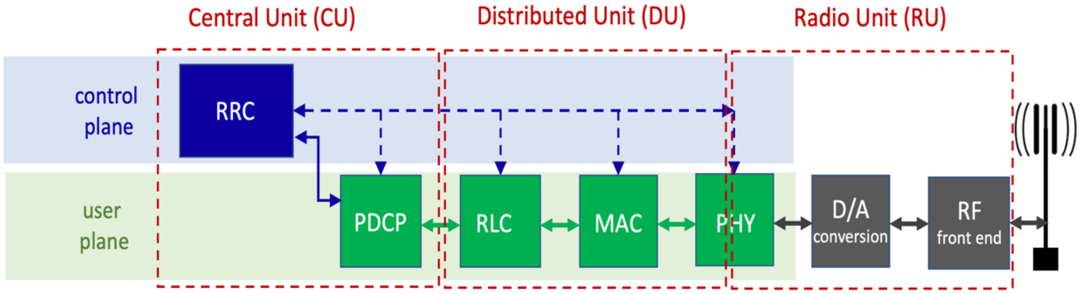
- *SD-RAN:* implementing RAN using SDN approach



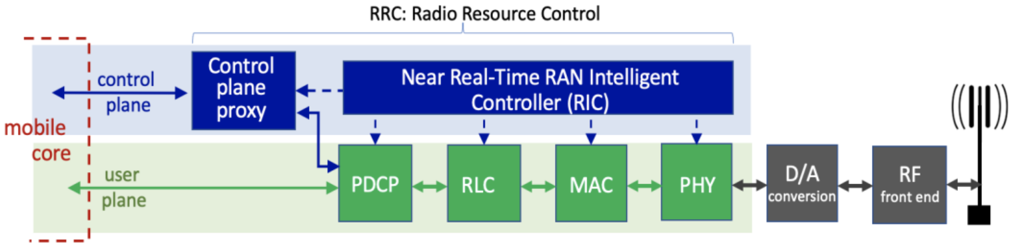
Software-defined RAN

- SDN-inspired refactoring of RAN is free both to move functionality around and to introduce new module boundaries, as long as the original 3GPP-defined interfaces are preserved

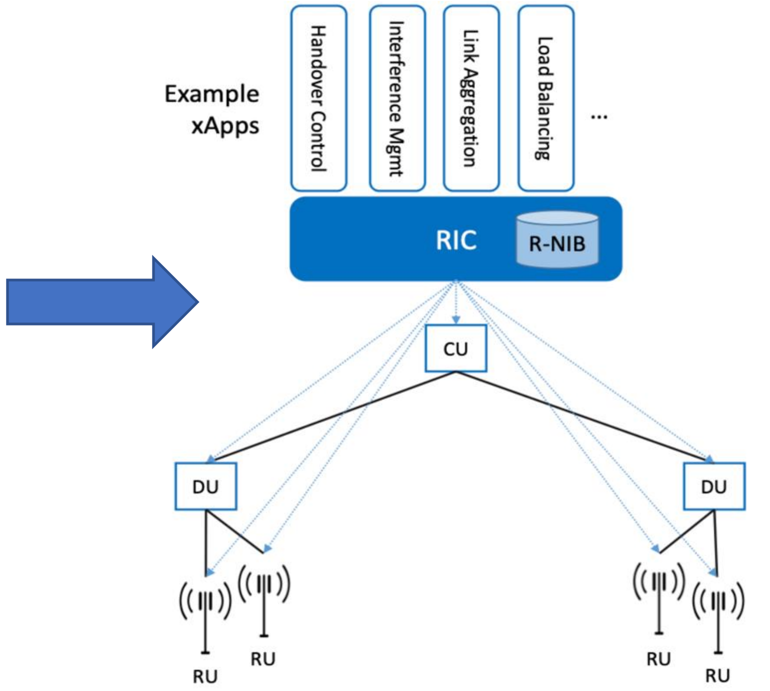
O-RAN re-partitioning of functional units



preserved interface to 3GPP mobile core, new software-based control



Near-RT RIC implemented as SDN Controller. hosting SDN control apps (O-RAN)



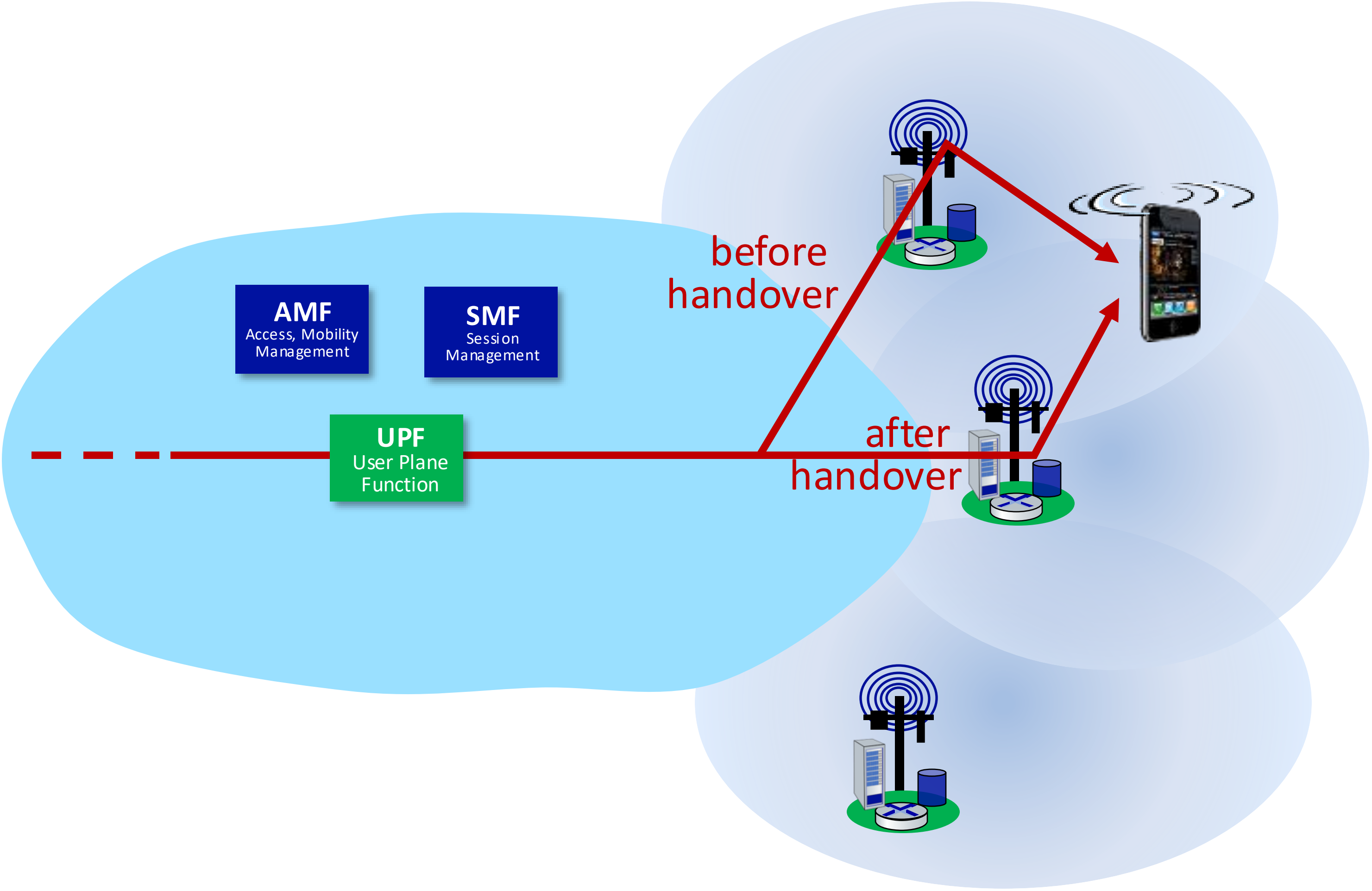
separating RAN control from disaggregated RAN components, logically centralizing control in apps running on SDN Controller

Overview of today's REU seminar

- Wireless (5G, nG) networks: review
- network software-i-zation
- Three 5G/nG network control/management challenges

  ... and opportunities for AI/ML

Challenge 1: Mobility in 5G networks: handover



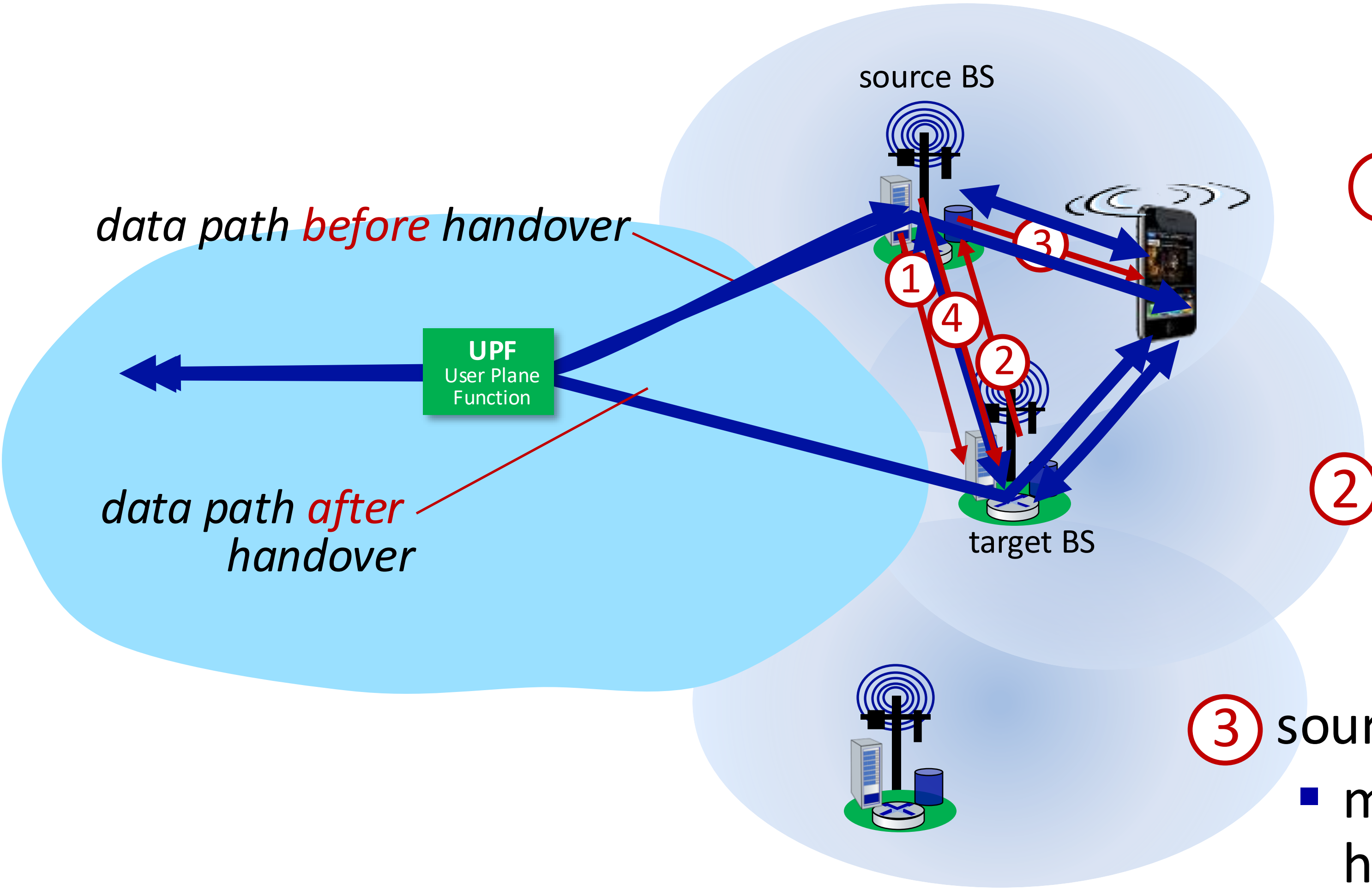
Handover:

- mobile device changes is point of attachment to the network
- data flow to device changes from *source* base station to *target* base station

Why perform handover?

- stronger signal from target base station
- target base station has less devices, less traffic

Challenge 1: Mobility in 5G networks: handover



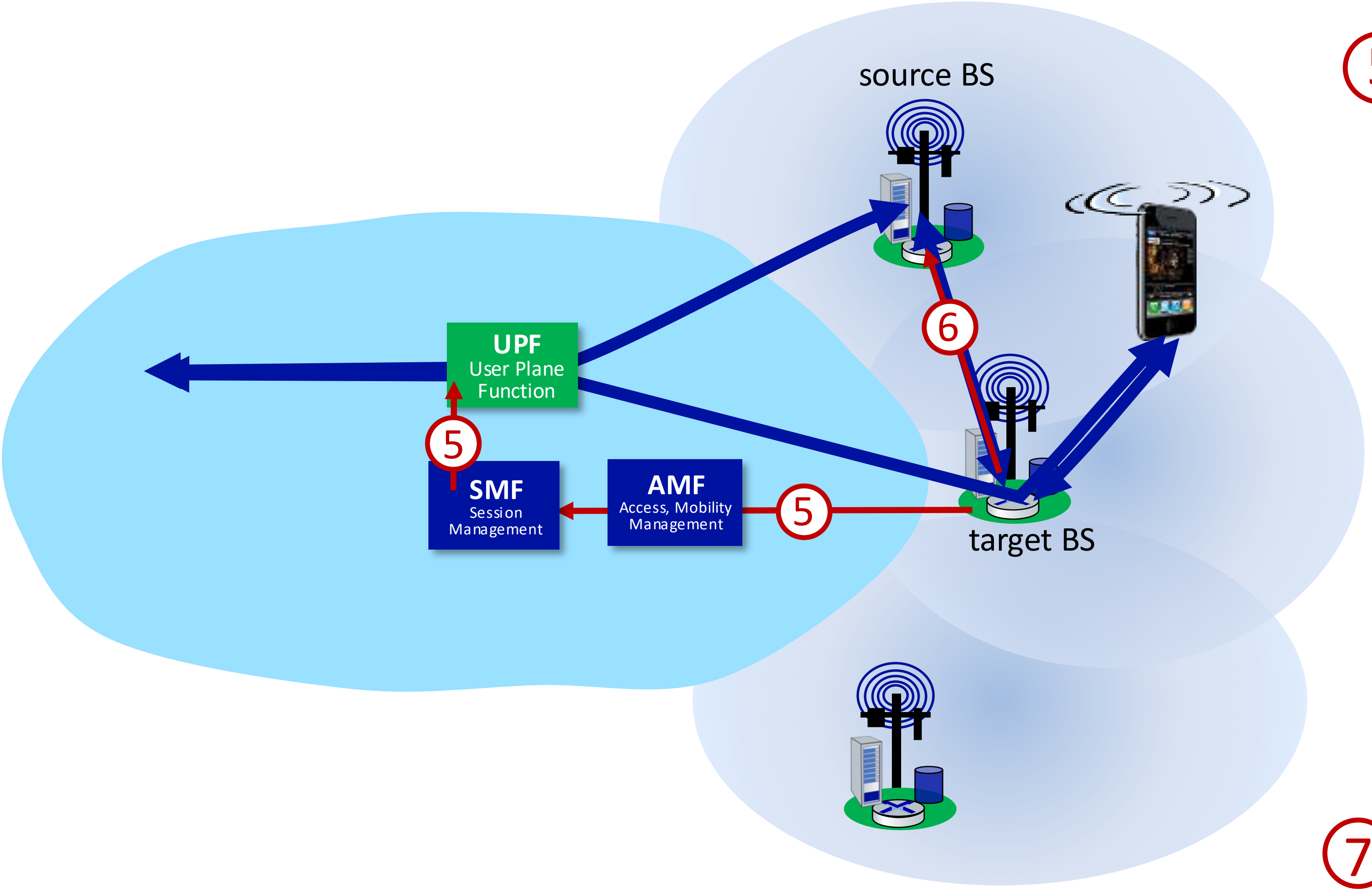
1. current (source) BS selects target BS, sends *Handover Request message* to target BS

2. target BS pre-allocates radio time slots, responds with HR ACK with info for mobile

3. source BS informs mobile of new BS
   - mobile can now send via new BS - handover *looks* complete to mobile

4. source BS stops sending datagrams to mobile, instead forwards to new BS (who forwards to mobile over radio channel)

Challenge 1: Mobility in 5G networks: handover



5 target BS informs AMF, SMF that it is new BS for mobile

- SMF instructs UPF to change tunnel endpoint to be (new) target BS

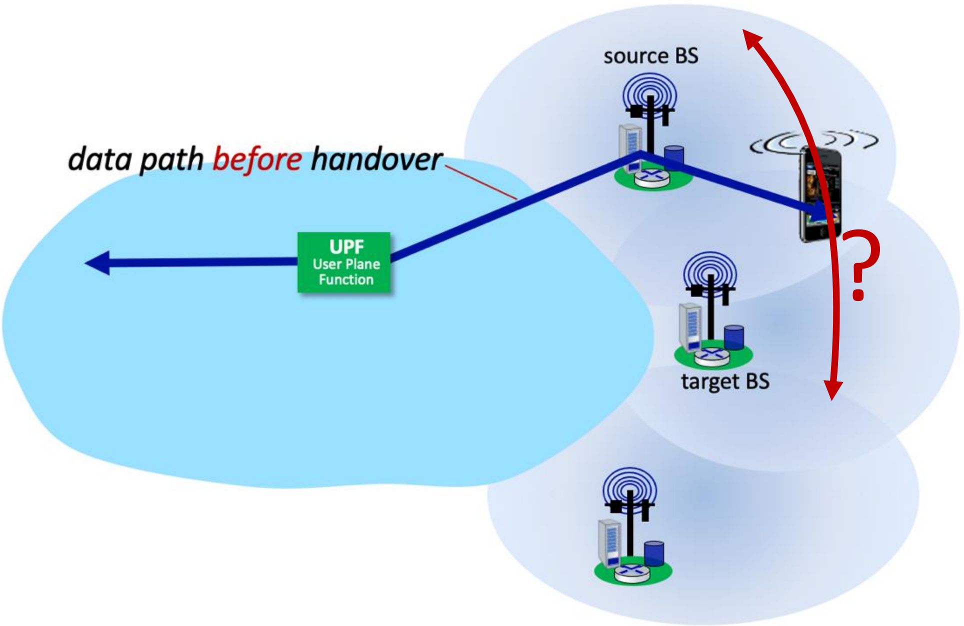
6 target BS ACKs back to source BS: handover complete, source BS can release resources

7 mobile's datagrams now flow through new tunnel from target BS to UPF

Handover in 5G networks: how AI/ML can help

User behavioral patterns can be learned/predicted:

- where are users likely to move in future
- future communication (bandwidth) requirements
- per-RAN predictions of bandwdith requirements of users in aggregate



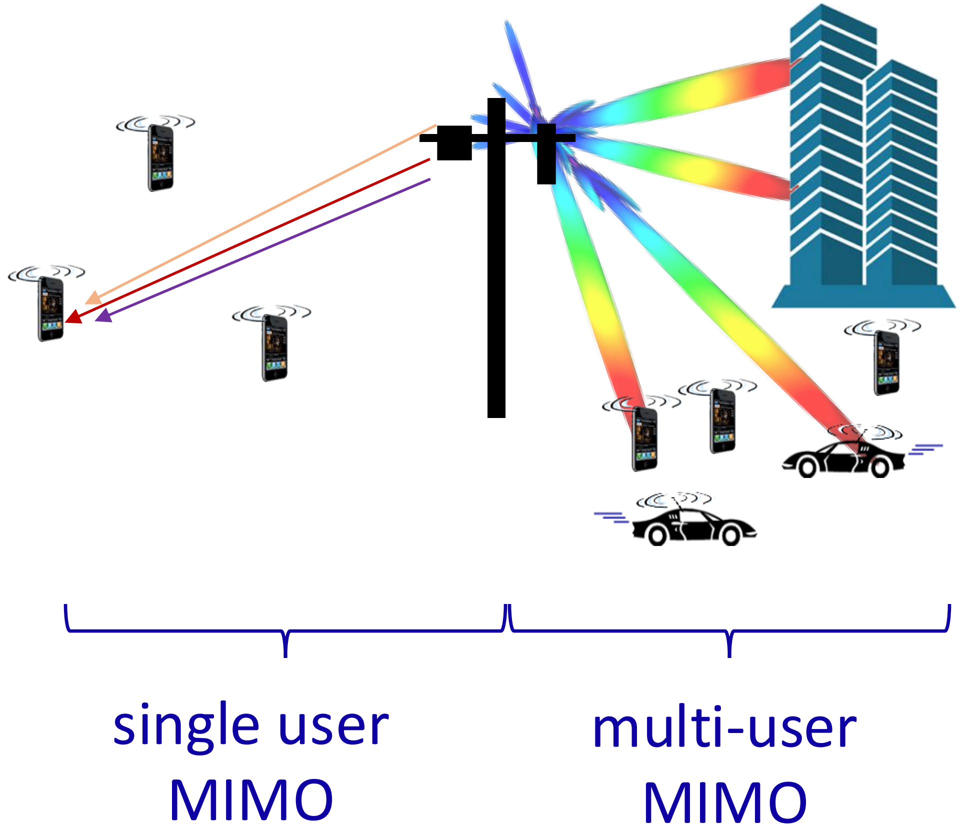
These predictions can be used in

- handover: where/when to perform handover?
- quality of service (QoS): what near-term-future quality of service guarantees can be made to a user

Challenge 2: beam (antenna) allocation/steering

- antenna performance to a given location is a complex, time-varying quantity due to
  - reflections (multipath)
  - Interference

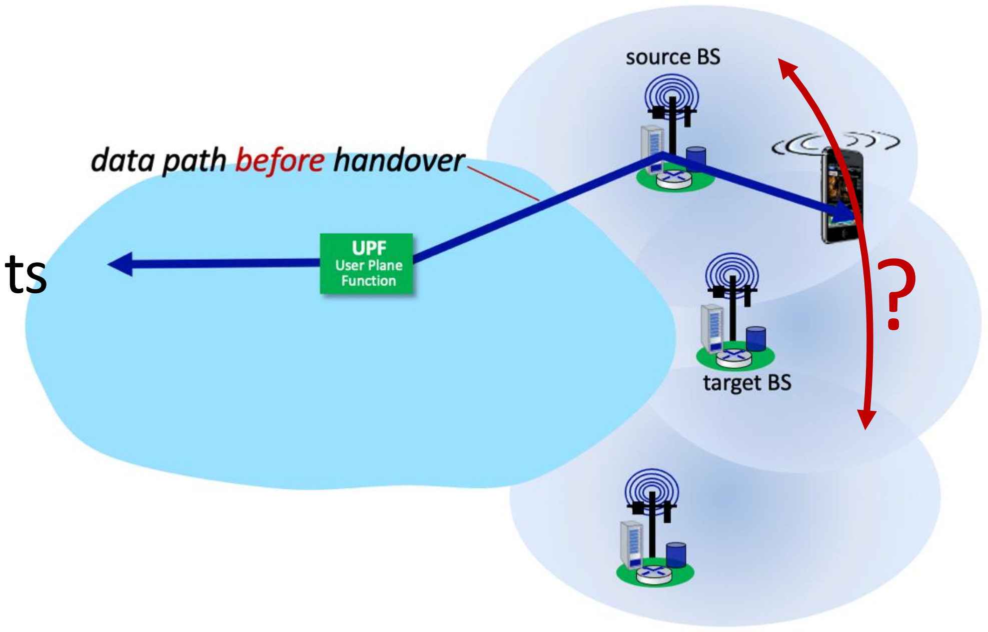
- antenna performance can be measured, future antenna performance predicted
- outputs from ML prediction model used to steer antennas to optimize performance



Antenna allocation/steering : how AI/ML can help

User behavioral patterns can be learned/predicted:

- where are users likely to move in future
- future communication (bandwidth) requirements
- per-RAN predictions of bandwith requirements of users in aggregate

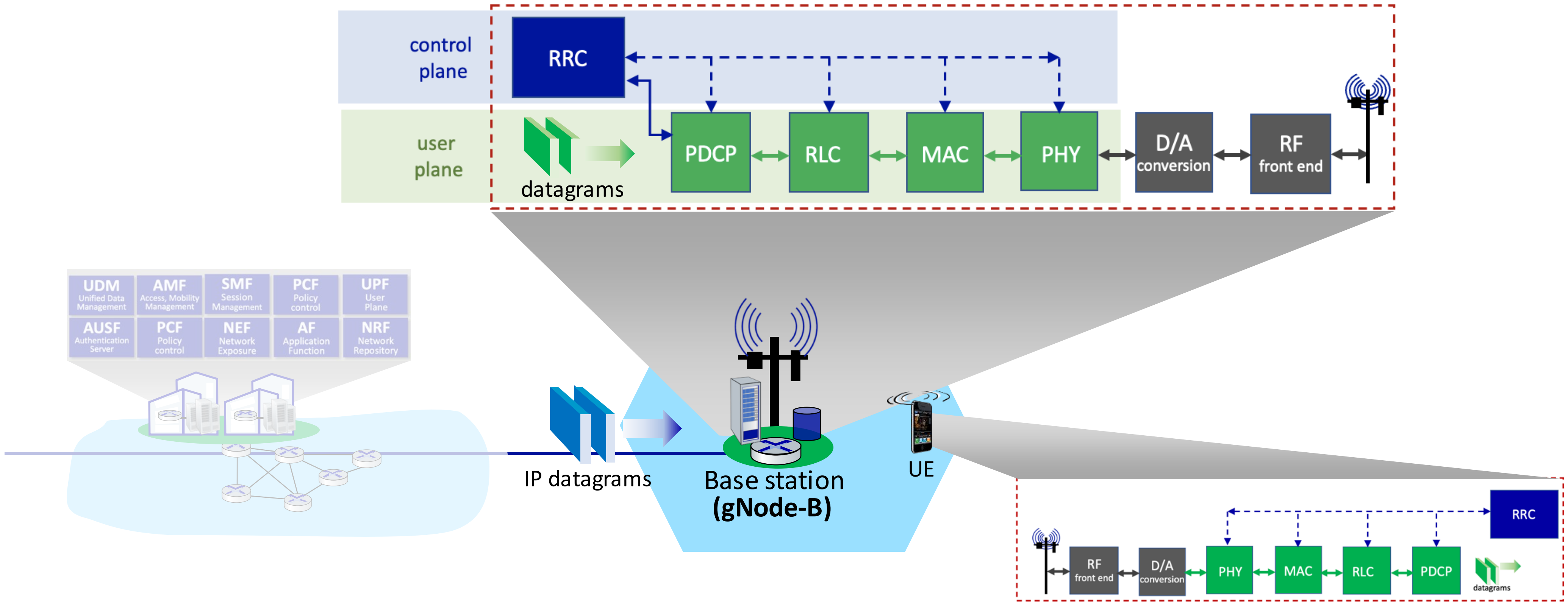


These predictions can be used in

- handover: where/when to perform handover?
- quality of service (QoS): what near-term-future quality of service guarantees can be made to a user

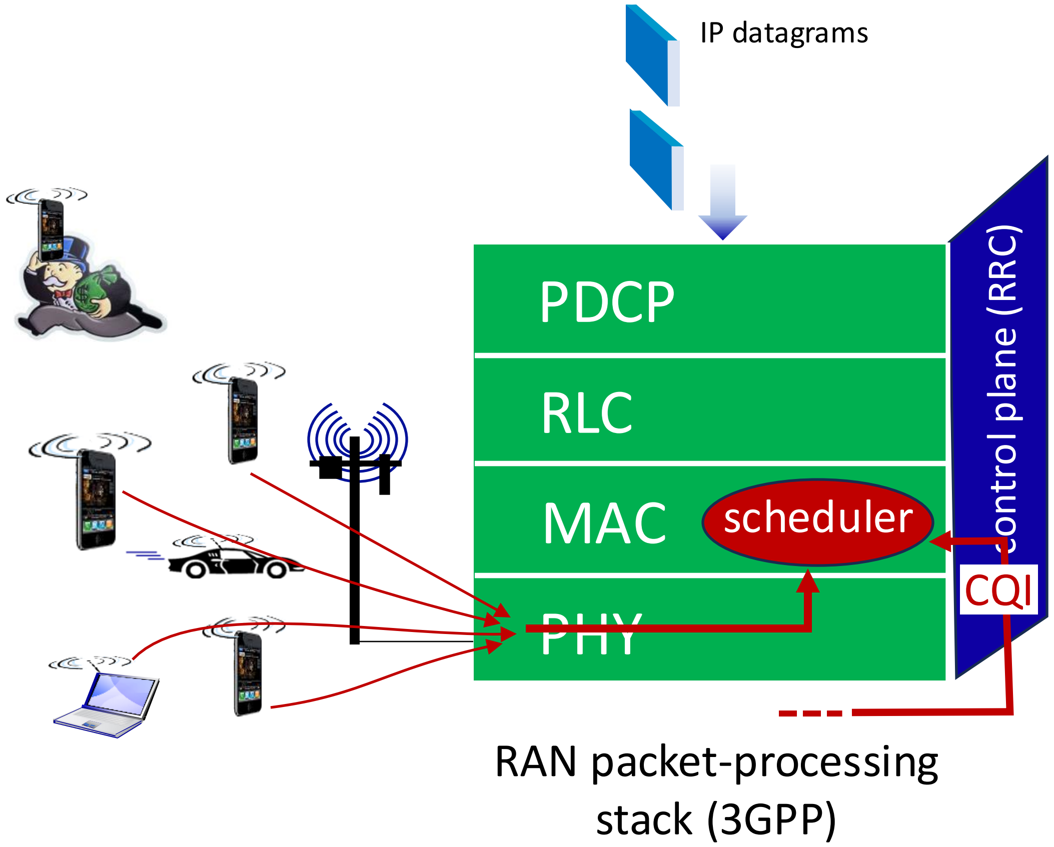
Challenge 3: packet scheduling in RAN

RAN: transfers datagrams between mobile core and UEs



protocol processing pipeline in UE

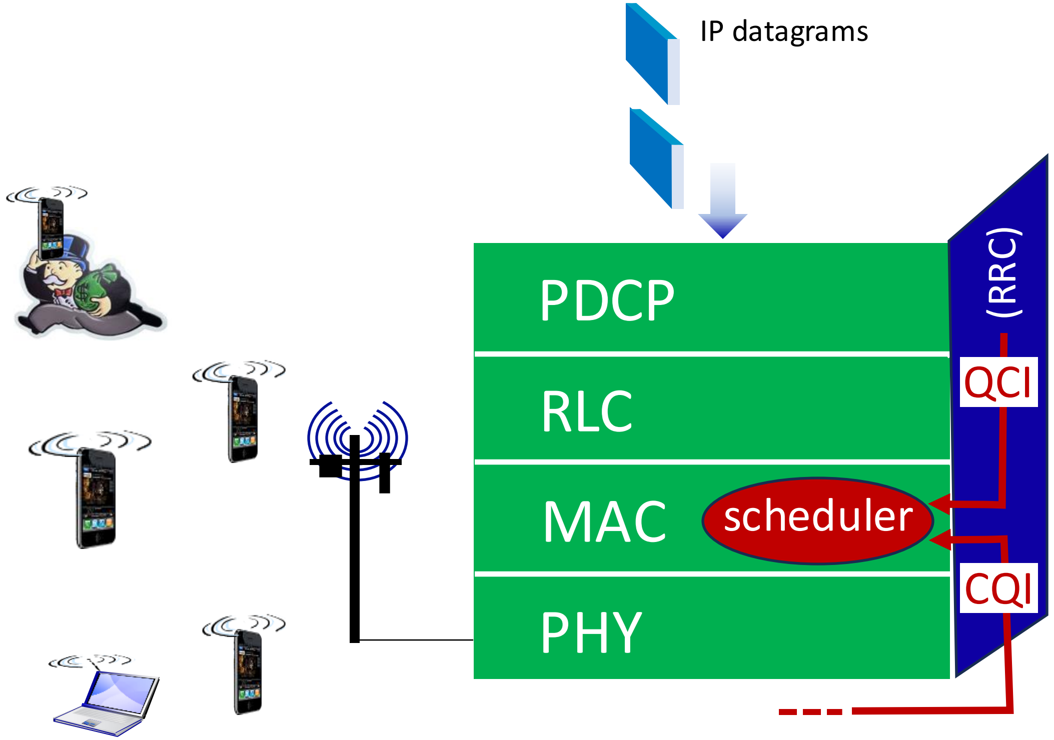
Challenge 3: packet scheduling in RAN



Channel Quality Indicator (CQI):

- UE measures quality of received reference signals (embedded in RBs), reports quality back to eNB
- 4-bit CQI value maps to modulation scheme to use (e.g., which QAM?) and expected throughput

Challenge 3: packet scheduling in RAN



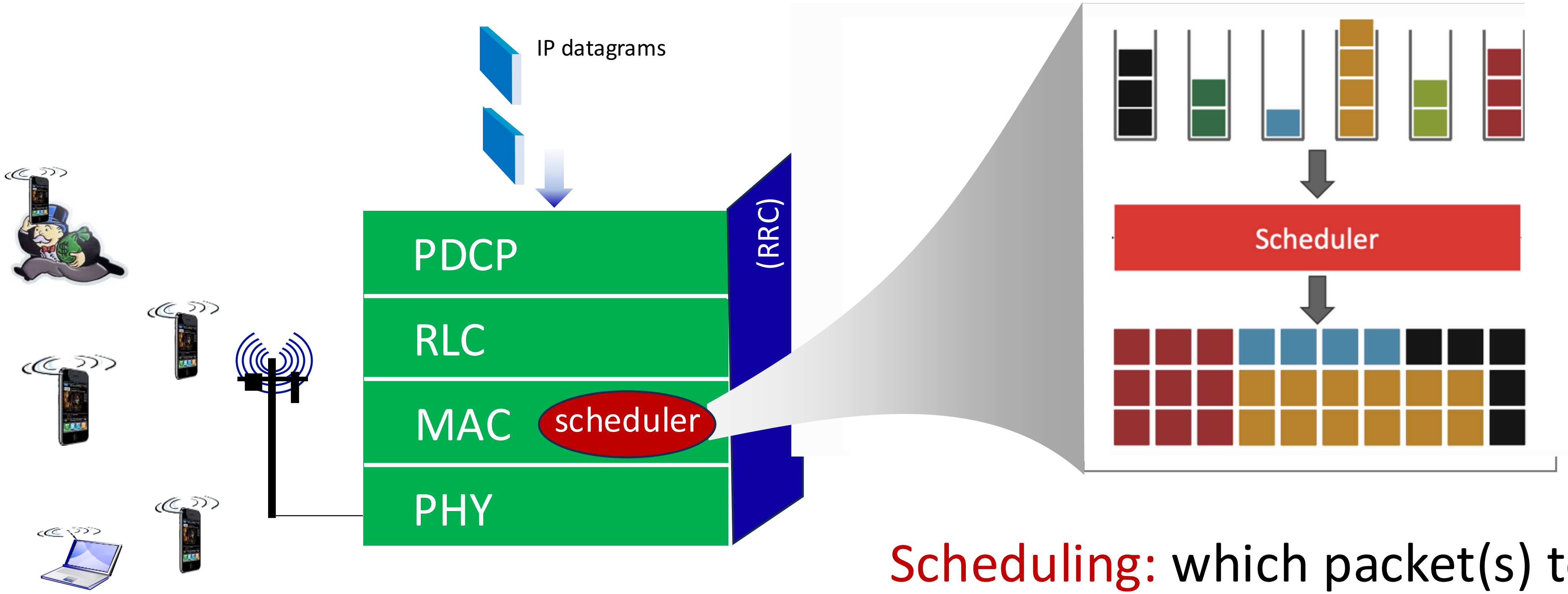
RAN packet-processing stack (3GPP)

QCI: QoS Class Indicator

- *QoS* network wants to provide to a particular UE (e.g., delay, guaranteed bit rate QoS guarantees)
- scheduler determines allocation of available radio spectrum to ensure all UEs meet their QoS requirements

Priority	QCI	Max delay	Max loss	Application
2	1	100ms	.01	Voice
3	4	50 ms	.001	Real-time gaming
6	7	300 ms	.000001	Streaming video
8	8	600	.000001	Web browsing, TCP

Challenge 3: packet scheduling in RAN



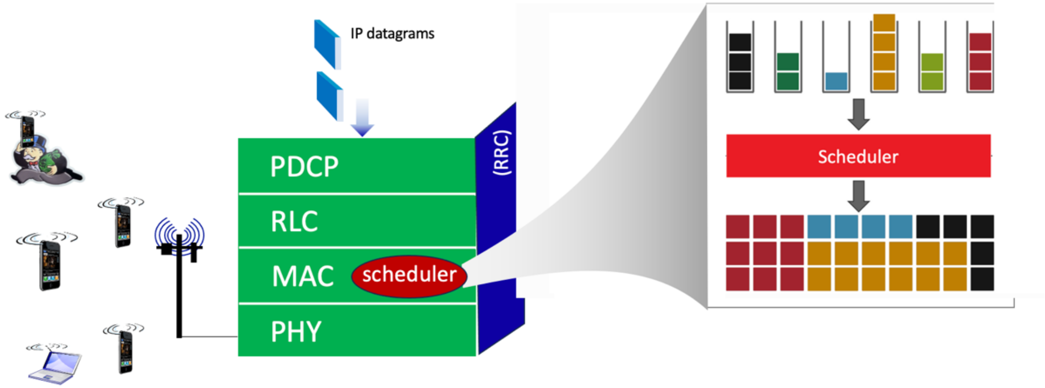
Scheduling: which packet(s) to schedule for transmission

- performance will depend on channel conditions, user device

Packet scheduling in RAN: how AI/ML can help

Transmission performance can be learned / predicted for each packet, depending on:

- channel conditions
- UE mobility



These predictions can be used to determine expected per-packet transmission performance

- actual packet selected depends on many factors: performance, fairness, quality of service

Summary of today's REU seminar

- wireless (5G, nG) networks: review
- network software-i-zation
- three 5G/nG network control/management challenges

  ... and opportunities for AI/ML

For more information:

- Networking: http://gaia.cs.umass.edu/kurose_ross
- Wireless Networking: https://gaia.cs.umass.edu/wireless_and_mobile_networking

THANKS!