Towards Reducing Mobile Network OPEX with Open RAN

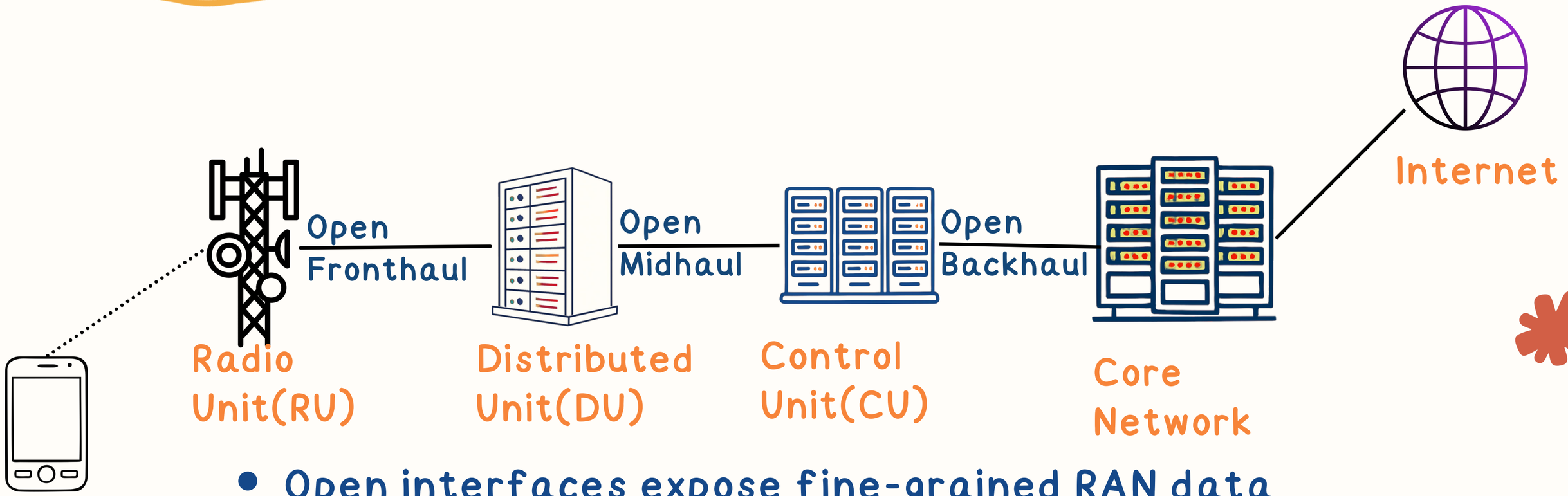
Ujjwal Pawar*,
Tianxin Wang, Jinyan Jiang, Mahesh Marina
University of Edinburgh

Telecom OPEX is a trillion-dollar problem

- $1.63tn global telecom OPEX
- $441bn network OPEX
- 20–40% of network OPEX is energy
- **Manual network management and energy cost are major recurring burdens for operators.**

[1] https://omdia.tech.informa.com/om136022/telcos-achieve-cost-control-milestones-as-labor-and-energy-efficiency-initiatives-deliver

OpenRAN: A Primer



- Open interfaces expose fine-grained RAN data
- Embraces Network function Virtualization
- Visibility across the layer

Manual network management / drive testing

- AT&T[1] estimated that drive testing 25% of its nationwide 4G LTE coverage would cost about $45M/year

[1] https://api.ctia.org/wp-content/uploads/2020/09/200908-CTIA-DODC-PFR-and-Comments.pdf

Network Side Drive Test

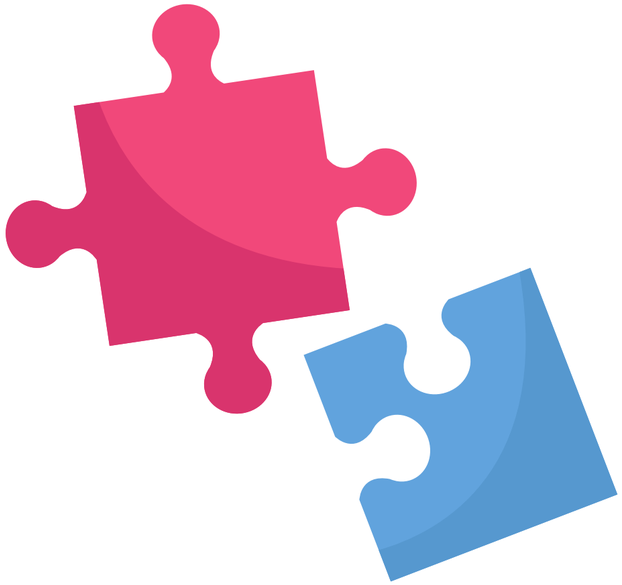
Device Positioning + Measurement Report = Coverage Map

Measurement report from CU


Omni Dir Antenna
Dir Antenna
Google

Challenges in Realizing 5G Positioning

- Cell Availability for Positioning
- Non-Ideal Antenna
- Diverse mobile device

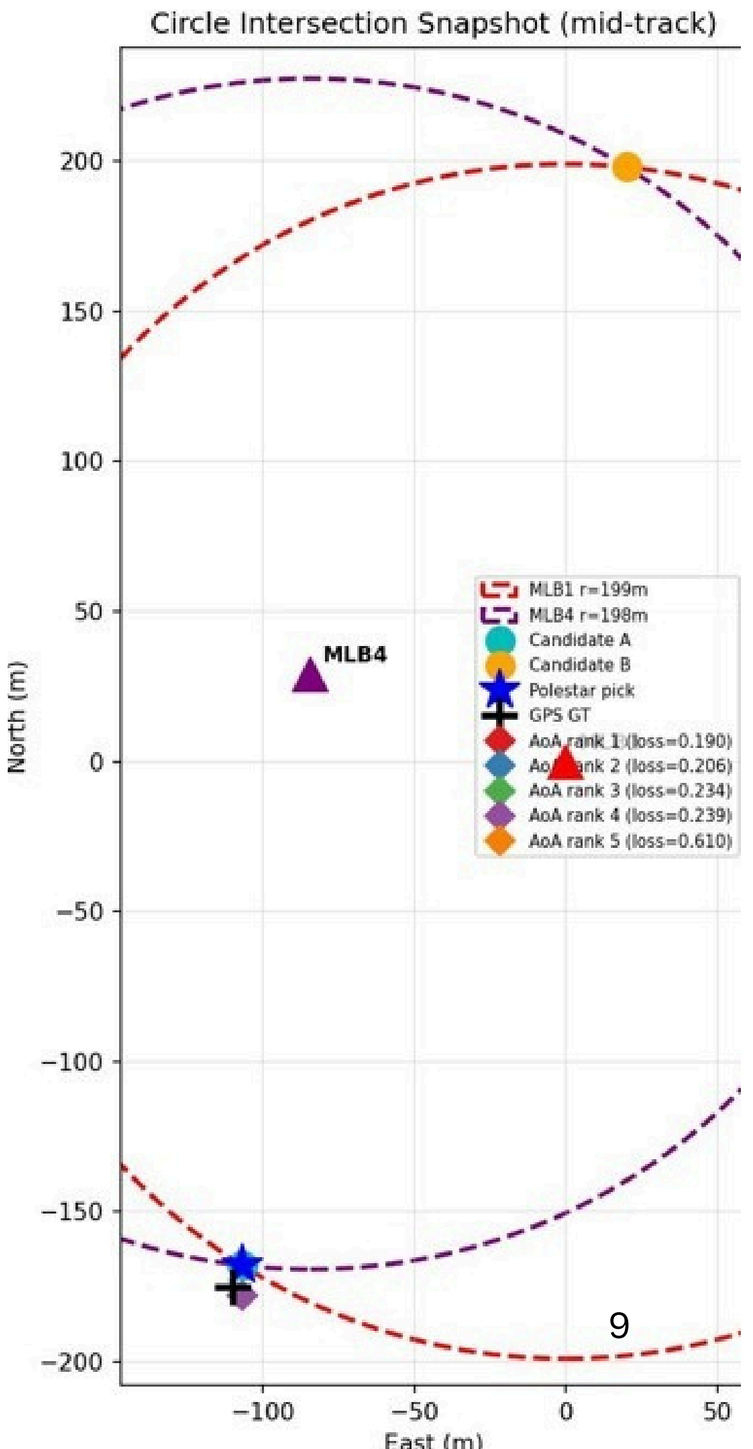


Cell Availability-Aware 5G Positioning

- 3 or more use Triangulation
- 2 or 1 use AoA with ToF
- 0 cell ??

Resolving AoA Ambiguity with ToF and Tracking

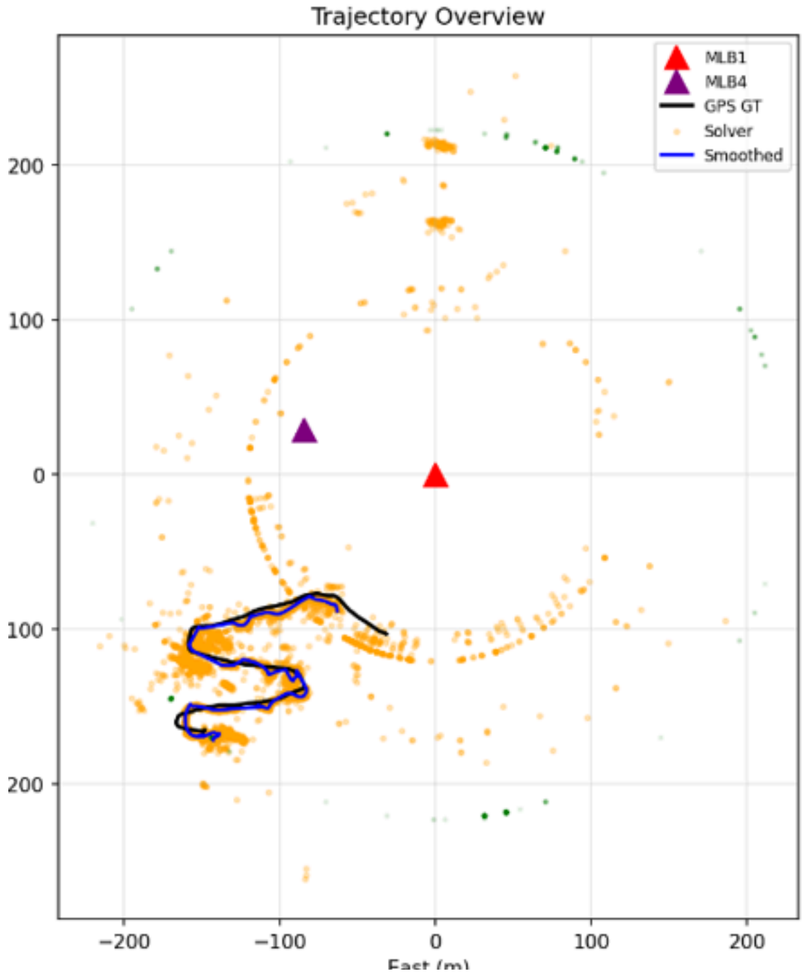
- Two-cell case
  - ToF measurements from two cells create two range circles.
  - The intersection of these circles gives two possible UE locations.
- Single-cell case
  - A single ToF and AoA measurement may still be insufficient to uniquely determine the true UE location when AoA is ambiguous. Use trajectory tracking.

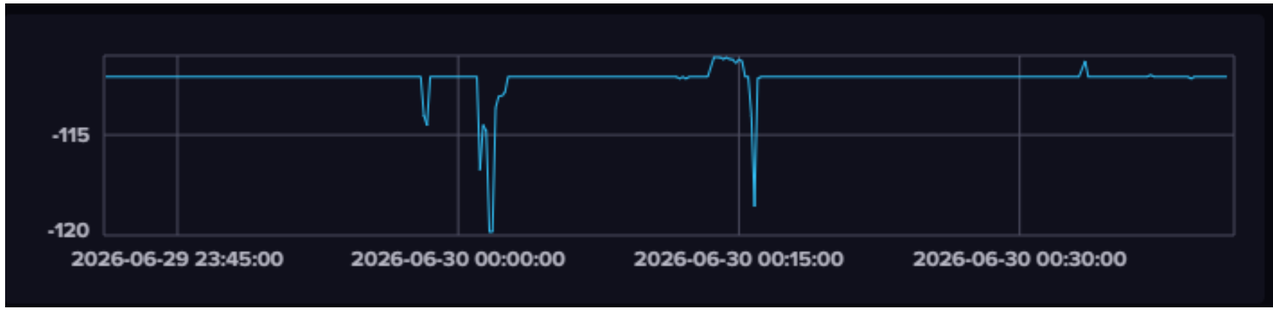


Use TDoA to find UE biases on the fly

- Time Difference of Arrival (TDoA) cancels out the UE biases
- Using TDoA and comparing that against ToF results gives the bias induced by UE

Cross layer visibility via OpenRAN

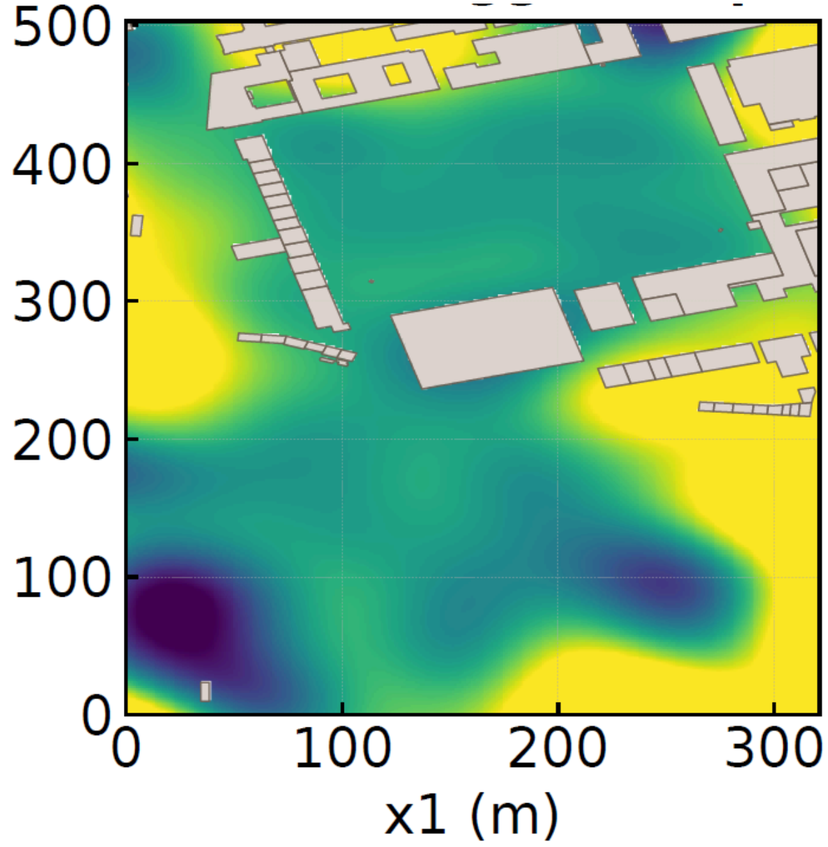




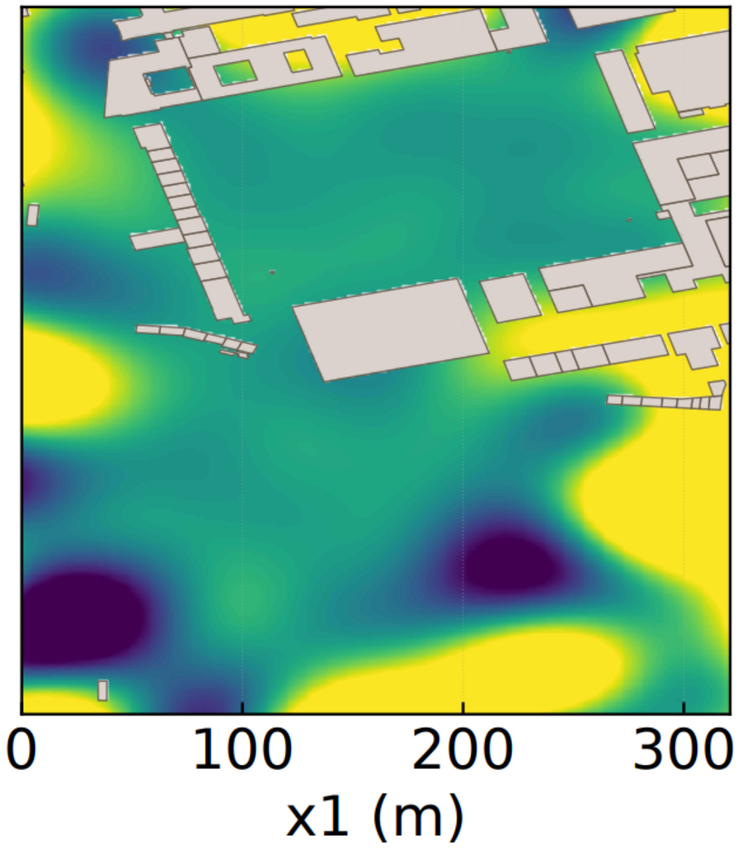
Measurement Reports from CU

iq data from RU → Localization Results

Results: a Coverage Map



GPS based coverage map



5G localization coverage map

RAN energy burden

- RAN accounts for about 73% of total mobile-network energy consumption[2]
- A 5G Huawei BBU[3] can consume about 2,000 W at maximum power.

[2] https://www.5gamericas.org/wp-content/uploads/2023/12/Energy-Efficiency-and-Sustainability-in-Mobile-Communications-Networks-WP.pdf
[3]https://carrier.huawei.com/~/media/CNBG/Downloads/Spotlight/5g/5G-Power-White-Paper-en.pdf

Minimizing Active Servers to Reduce RAN Energy OPEX

- At low load, workloads can be consolidated onto fewer servers. Goal is to find the minimum number of servers
- This is Bin Packing Problem -> NP-Hard
- This is a well-studied RAN optimization problem but limited to simulations.

The Gap Between RAN Compute Optimization and Real Deployment

- Current deployments often use static DU-to-server mapping.
- Servers are typically provisioned for worst-case load, leading to poor utilization during normal operation.
- Prior systems such as **Concordia**[4], **CloudRIC**[5], **vrAIn**[6], **RENC**[7] and **CRT**[8] etc. show that RAN compute can be optimized in practice.
- However, these systems mainly focus on **single-server** optimization.

Why Dynamic DU Placement is Hard ?

- The DU is a stateful, real-time network function.
- Requires Migration while meeting sub-millisecond deadlines
- Can be done [ATLAS[9]]
- For Atlas Resilience was the goal not load balancing

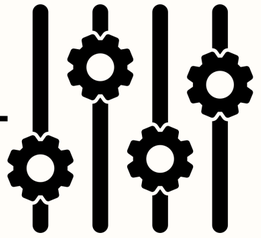
Why Resilience-Based Migration is Not Enough

- Not designed for load balancing
- Using Atlas based migration with 100 UE gave only ~80% successful attach.

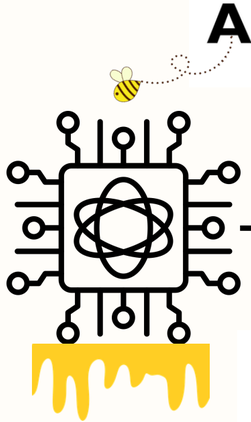
OpenRAN to the rescue

Distributed Unit(DU)

Open-FAPI

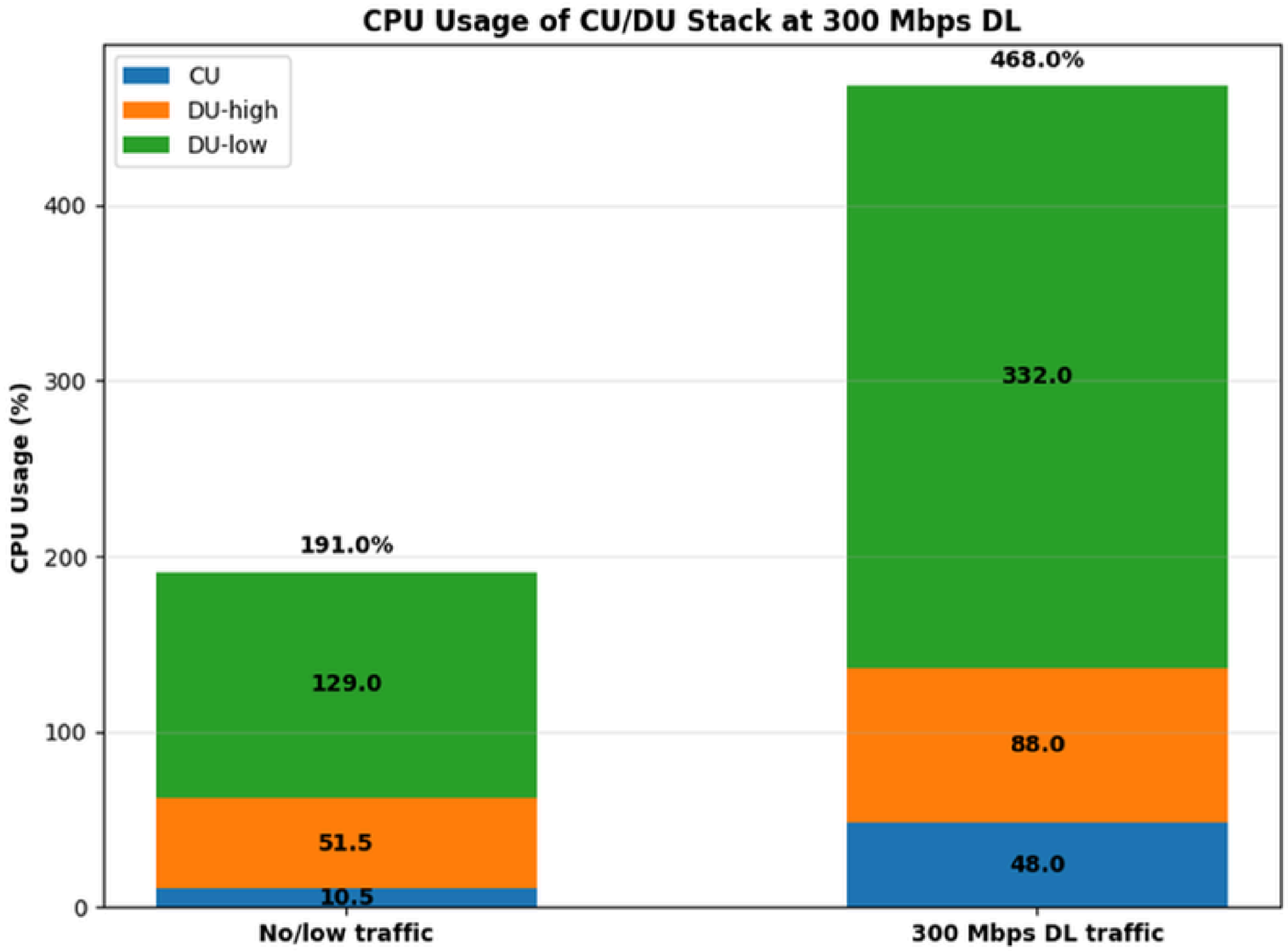


DU-High (MAC, RLC)



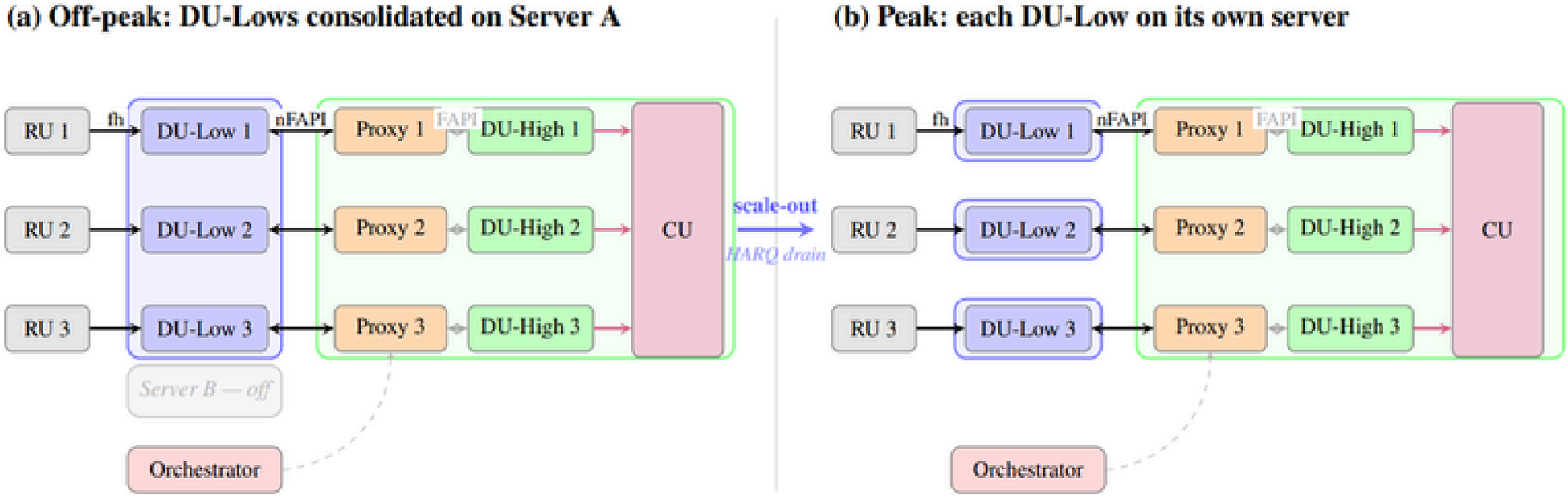
DU-Low (High-PHY)

Breakdown of Processing needed



Key insight

- The DU-low layer is near-stateless across slots: HARQ soft buffers are the only cross-slot state, and they are transient refilled each retransmission round.
- In our implementation DU low migration takes <10usec to switch



Conclusion: Open RAN as an OPEX Reduction Enabler

- Mobile network OPEX is not caused by a single factor; it emerges from manual operations, limited visibility, over-provisioning, and inefficient resource use.
- Open RAN exposes the data and control needed for smarter operations.
- The key challenge is no longer whether these optimizations are useful, but how to make them safe, automated, and operational at deployment scale.

Questions!!

References

[4] https://dl.acm.org/doi/10.1145/3452296.3472894

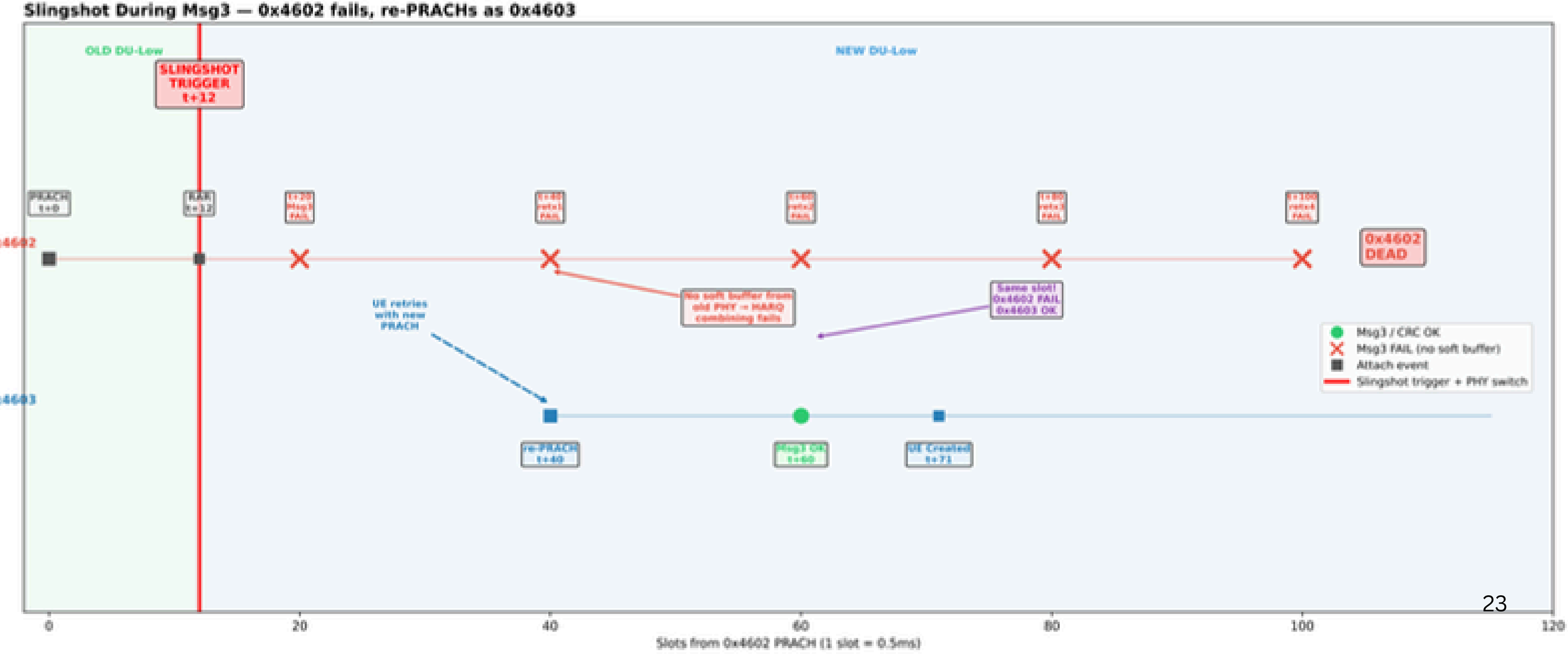
[5]https://dl.acm.org/doi/10.1145/3636534.3649381

[6] https://dl.acm.org/doi/10.1145/3300061.3345431

[7] https://ieeexplore.ieee.org/document/9685440/

[8] https://www.usenix.org/conference/nsdi25/presentation/kalia

[9]https://dl.acm.org/doi/10.1145/3570361.3613276


Slingshot During Msg3 — 0x4602 fails, re-PRACHs as 0x4603
OLD DU-Low
NEW DU-Low
SLINGSHOT TRIGGER t+12
PRACH t+0
RAR t+12
t+20 Msg3 FAIL
t+40 retx1 FAIL
t+60 retx2 FAIL
t+80 retx3 FAIL
t+100 retx4 FAIL
0x4602
0x4602 DEAD
No soft buffer from old PHY → HARQ combining fails
Same slot! 0x4602 FAIL 0x4603 OK
UE retries with new PRACH
0x4603
re-PRACH t+40
Msg3 OK t+60
UE Created t+71
Msg3 / CRC OK
Msg3 FAIL (no soft buffer)
Attach event
Slingshot trigger + PHY switch
0
20
40
60
80
100
120
Slots from 0x4602 PRACH (1 slot = 0.5ms)