June 25, 2020

FERRET: Fall-back to LTE Microservices for Low Latency Data Access

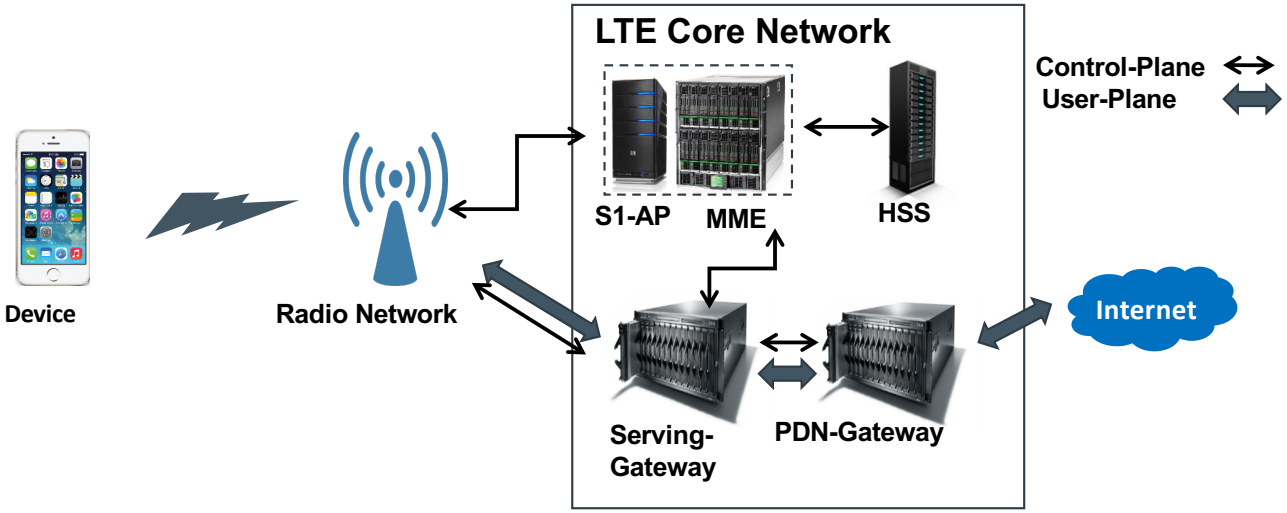
Presenter: Muhammad Taqi Raza, Ph.D.
The University of Arizona

Background - LTE

– LTE Evolved Packet Core (EPC) functionality is divided into **control-plane** and **user-plane**.

– **Control plane** logic performs device registration/deregistration, mobility, location update, paging, and many more.

– **User plane** forwards traffic to the next hop along the path to the selected destination network according to control plane logic.



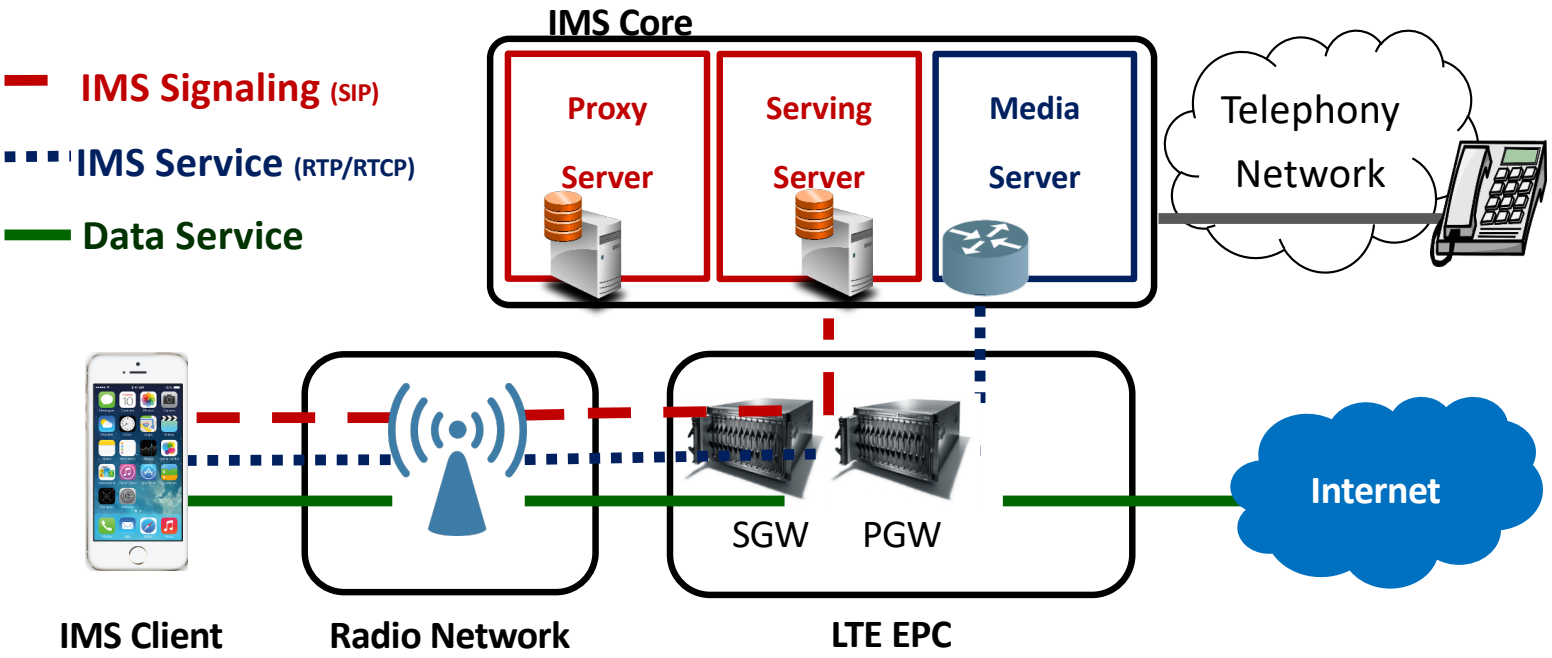
Background – IP Multimedia Subsystem (IMS) – Our Use Case Example

IMS supports real-time multimedia services.

– Voice call request (control plane)

– Voice call speech packets flow (user plane)

– Regular user data traffic (user plane)

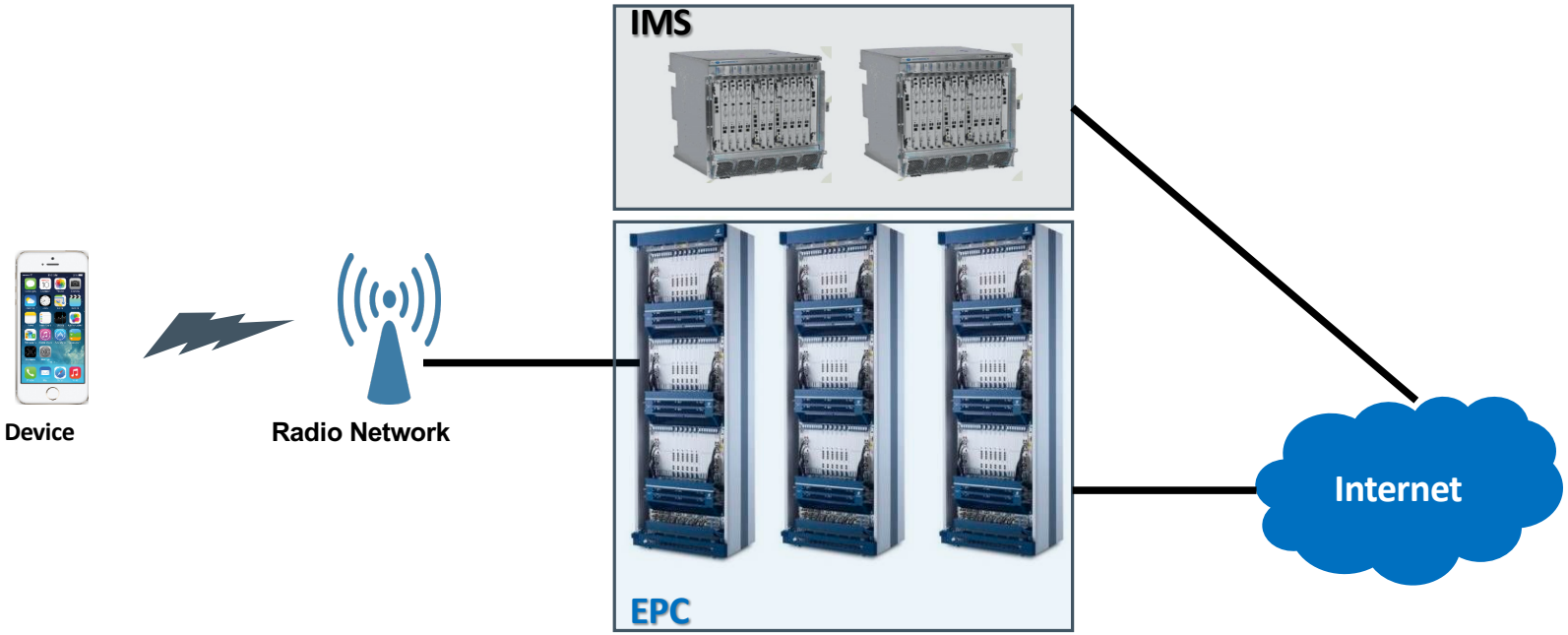


Existing IMS over LTE

– Purpose-built hardware platforms, such as Ericsson's Blade Systems (EBS), Alcatel-Lucent's Element Management System (EMS), etc.

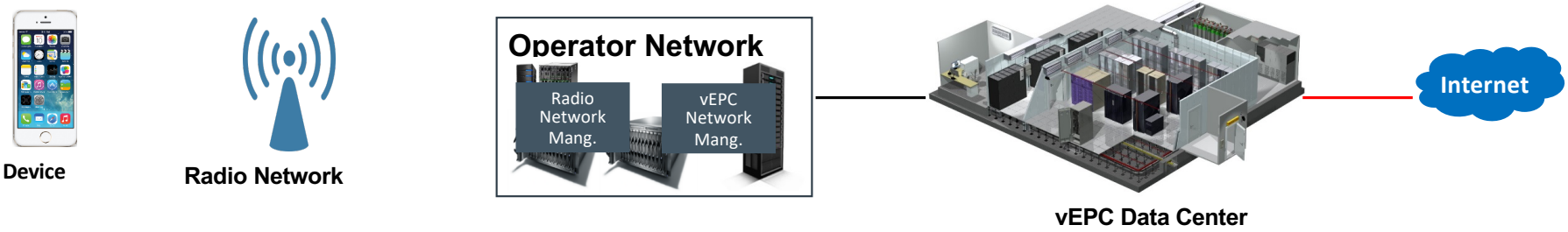
– Coupling between software and hardware, such as Ericsson's ERLANG, Alcatel-Lucent's NVP, etc.

– Dedicated network resources for each user.



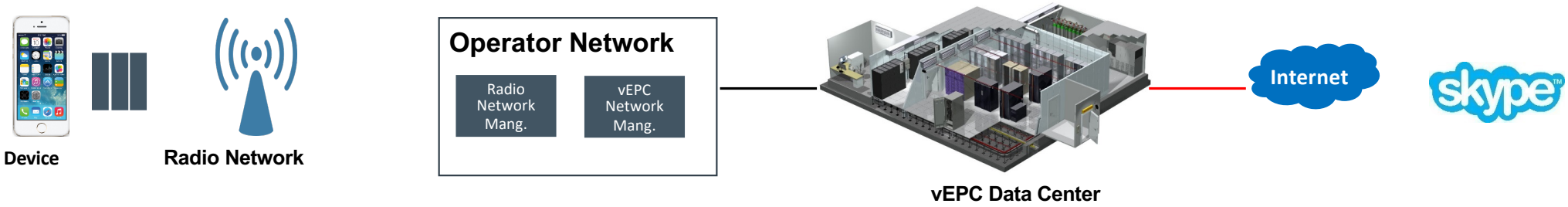
LTE-NFV Architecture

– LTE NFV virtualizes LTE core network functions over off-the-shelf boxes.
  – This reduces CAPEX and OPEX for operators.



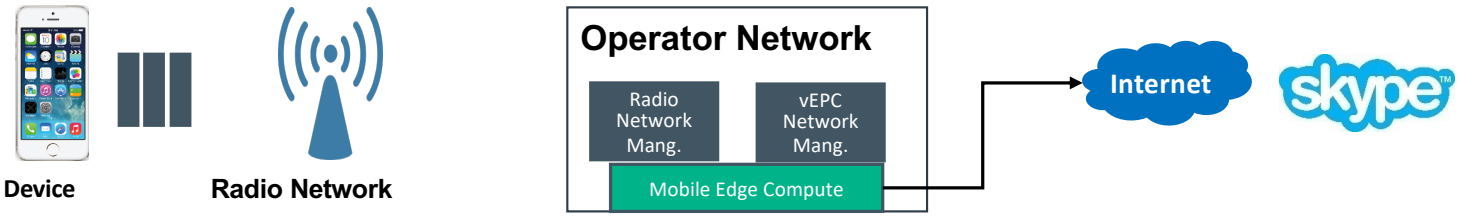
Why LTE-NFV Architecture is bad for IMS?

– **Control plane** signaling messages are executed on the cloud
  – **Good**: Different vNFs coordinate with each other to facilitate an event (e.g. mobility)
– User **plane** traffic is forwarded to Internet through vEPC Gateways (i.e. Serving Gateway, PDN Gateway, IMS NFs).
  – **Bad**: Voice traffic **unnecessarily** goes through the data center.



Data Forwarding through Mobile Edge Compute NF

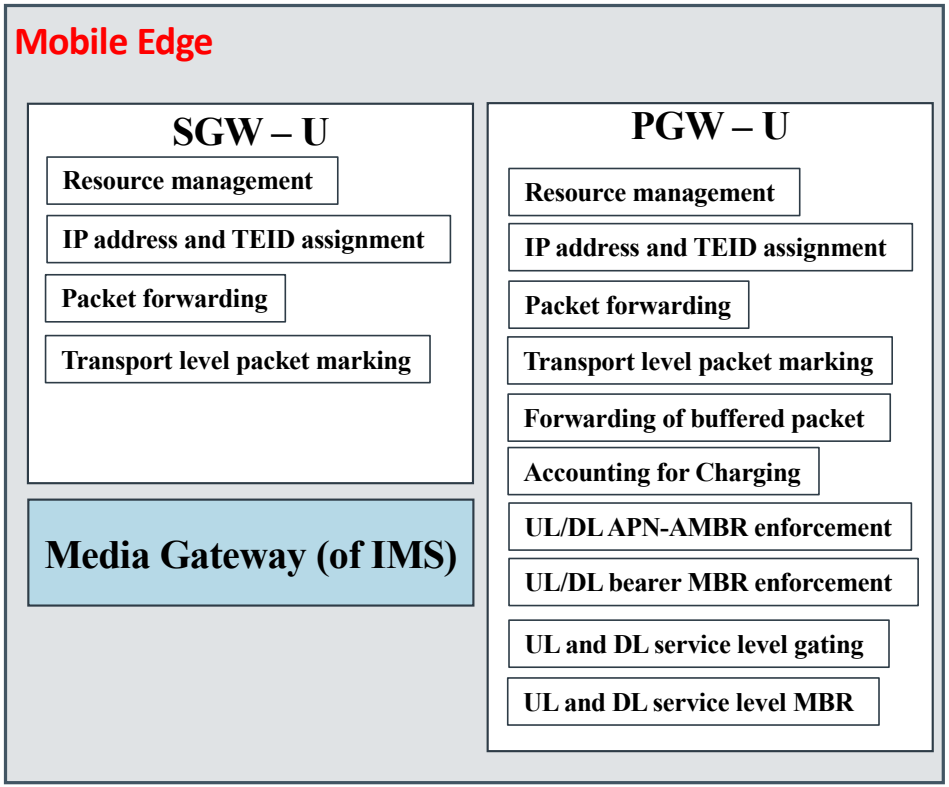
– Let's decouple control-plane and user-plane
  – Control-plane traffic still goes to vEPC and vIMS
  – User-plane traffic is routed through mobile edge compute NFs.

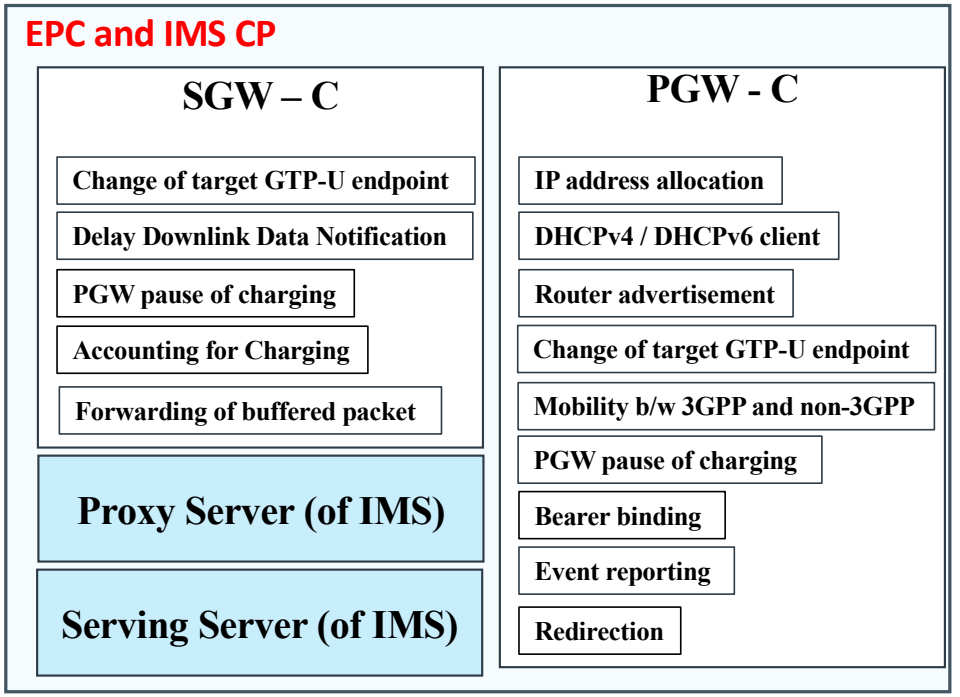


– We decompose SGW and PGW user plane functionalities and install them at the edge
– We also install MGW of IMS at the edge.

Design Considerations

– Decouple LTE and IMS CP and UP and install them at core and edge, respectively.



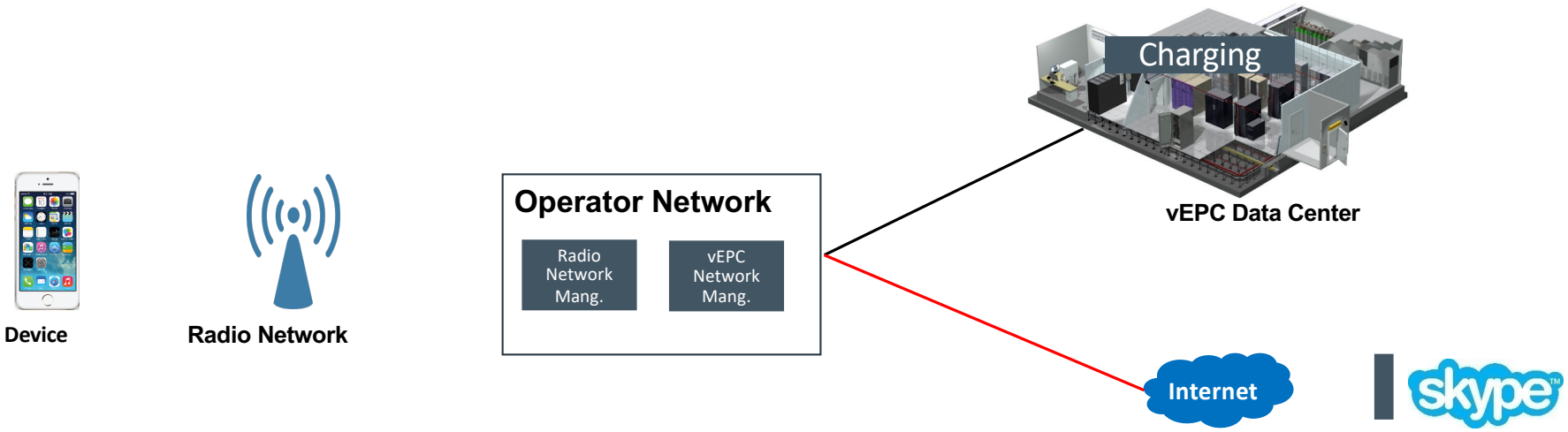


Design Considerations: Pure UP and CP Separation

– All UP functions are moved to edge, whereas all CP functions (including PCRF, LCS and others) are placed at the core.

Issue:

– The UP traffic will steer to CP for charging. This adds latencies and bandwidth wastage for voice call operation.



Provide copy of all common CP NFs at the edge

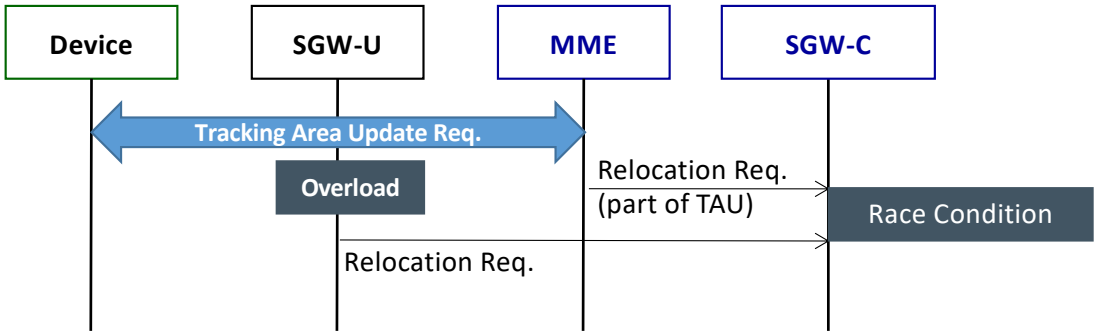
Design Considerations: Share Common Functions between CP and UP

– Those functions (e.g. PCRF or charging function) which both CP and UP rely should be shared (i.e. copied).

– Let the CP, being central entity, handle the NF allocation.

Issue:

– Creates race conditions. Example, SGW-U requests SGW-C for relocating SGW-U, while the TAU procedure is ongoing at the core, e.g. MME performs SGW relocation by sending same request to SGW-C.



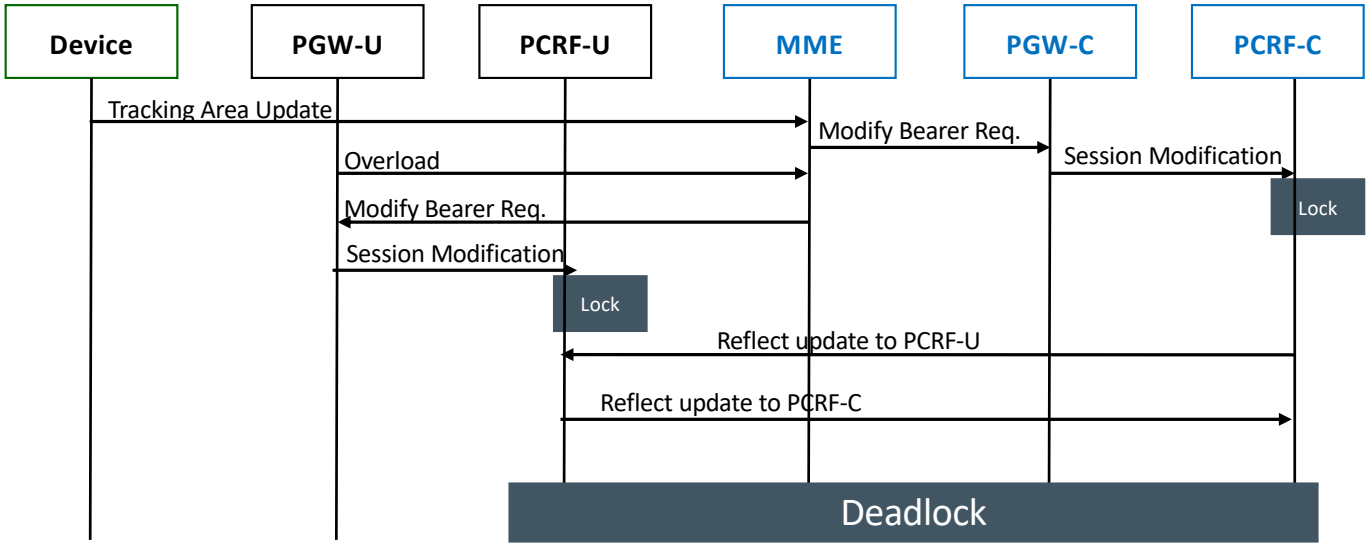
Let UP manage its own resources based on policy provided by CP

Design Considerations: Share Common Functions between CP and UP

Those functions (e.g. PCRF or charging function) which both CP and UP rely should be shared (i.e. copied).

Issue:

– Creates deadlocks. Example, when ModifyBearer request at CP and UP locks their respective PCRF copies.



Use blocking mode operations

Realizing IMS Application over MEC through FERRET

– VoLTE call requires:

(1) Call establishment phase (control plane), and

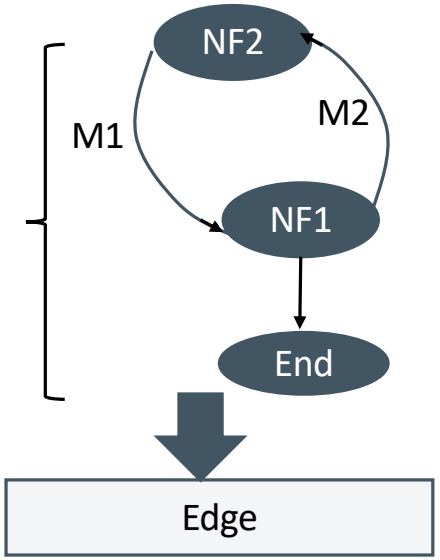
(2) Speech data packets flow (user plane).

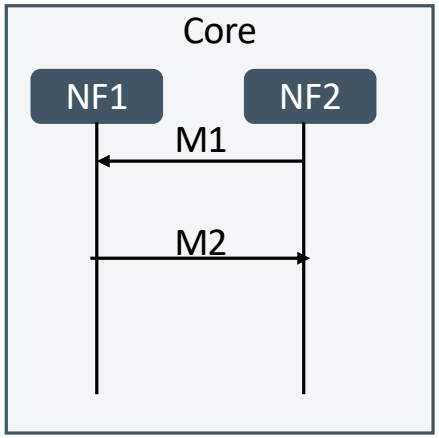
FERRET Key Idea

– Let the CP perform all control-plane operations at the core.

– Once all operations are performed at CP, then replay them at the edge.

– Only transmit VoLTE call control plane packets once the call is established (before user-plane traffic starts).



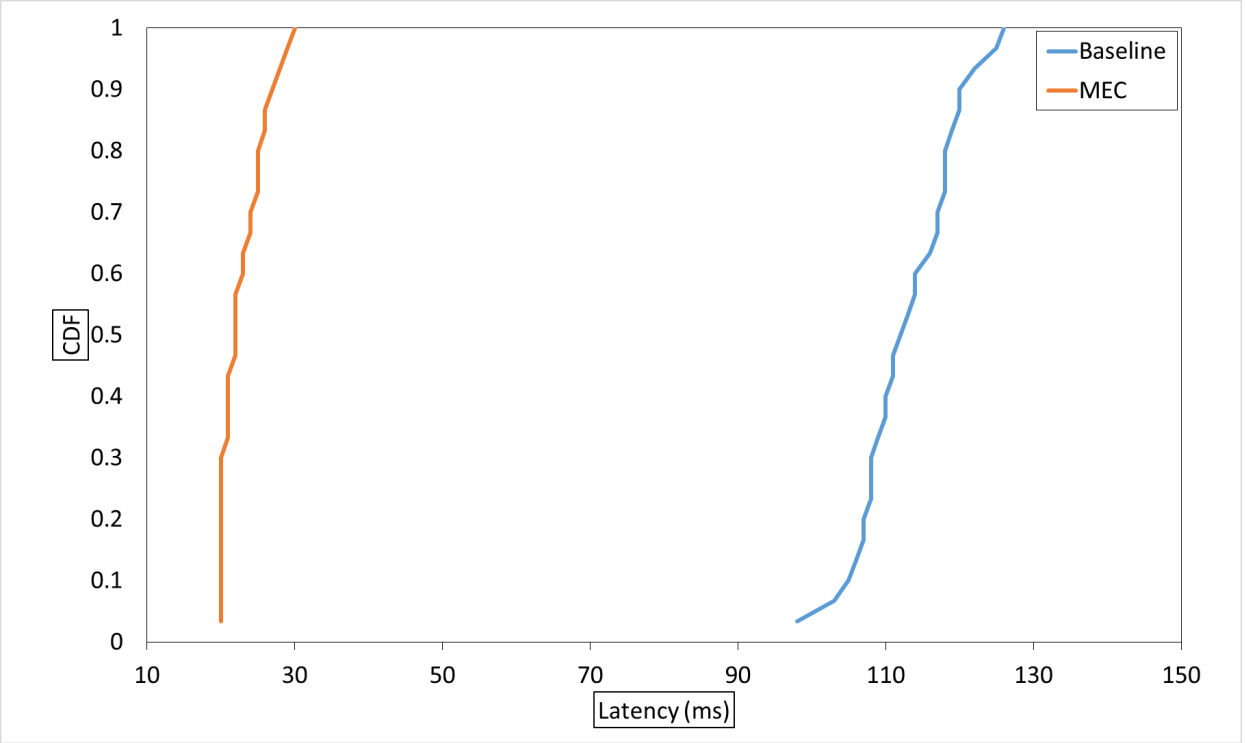


Reducing User Plane Latencies through FERRET

– VoLTE call packets are forwarded to the Internet without going to the core.

– Baseline results represent operational VoLTE network latencies.

  – This includes eNodeB <-> vEPC <-> IMS latencies for both caller and callee



Summary

– MEC is part of 5G agenda that requires important network components to be installed at the edge.

– Through IMS, we demonstrate MEC architecture that reduces upto 6X user-plane latencies.

– In the future work, we will measure MEC design from different system and networking aspects.
  – Bandwidth, Latency, Call Rate, Mobility and more.