CAN Use cases

draft-liu-dyncast-ps-usecases-03

P. Liu, China Mobile
P. Eardley, British Telecom
D. Trossen, Huawei
M. Boucadair, France Telecom
LM. Contreras, Telefonica
C.Li,Huawei

CAN: Computing-Aware Networking

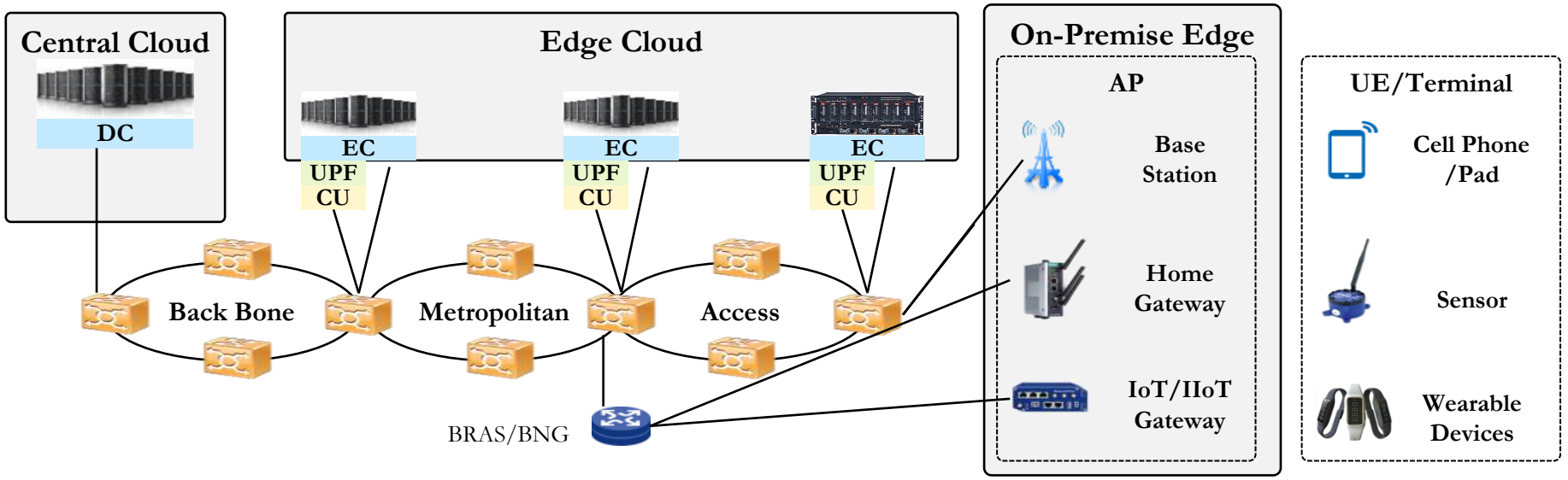
ITU-T and Industry wide initiative

CAN aims at computing and network resource joint optimization based on the awareness, control and management over network and computing resources, to determine the appropriate service node and dispatch the service request and provide a better user experience.

CAN IETF: Problem domain for IETF routing

CAN in IETF is to utilize computer resource metric and associated metadata for routers‘ path computation in order to optimally steer service requests from clients to optimal (wrt to the provided resource metrics) computing resources across different network locations.

ICT Infrastructure Redefinition



Facts in China Mobile

- CDN nodes in every city (**330+**) and major county (**250+**), with **25000+** servers installed
  - *These nodes can be upgrade to vCDN and then edge computing infrastructure*
  - *More diverse computing resource need to be provided；*
- More edge computing nodes will be setup in an on-demand manner
  - County aggregation **6000+,** Access aggregation **10,000+,** On-site **100,000+**

Service providers are offering the integrated computing and networking infrastructure.

What do users want?

- **The best user experience such as low latency and high reliability, etc .**
- **Stable application’s experience when moving to different areas.**

How to deploy in order to meet the requirements?

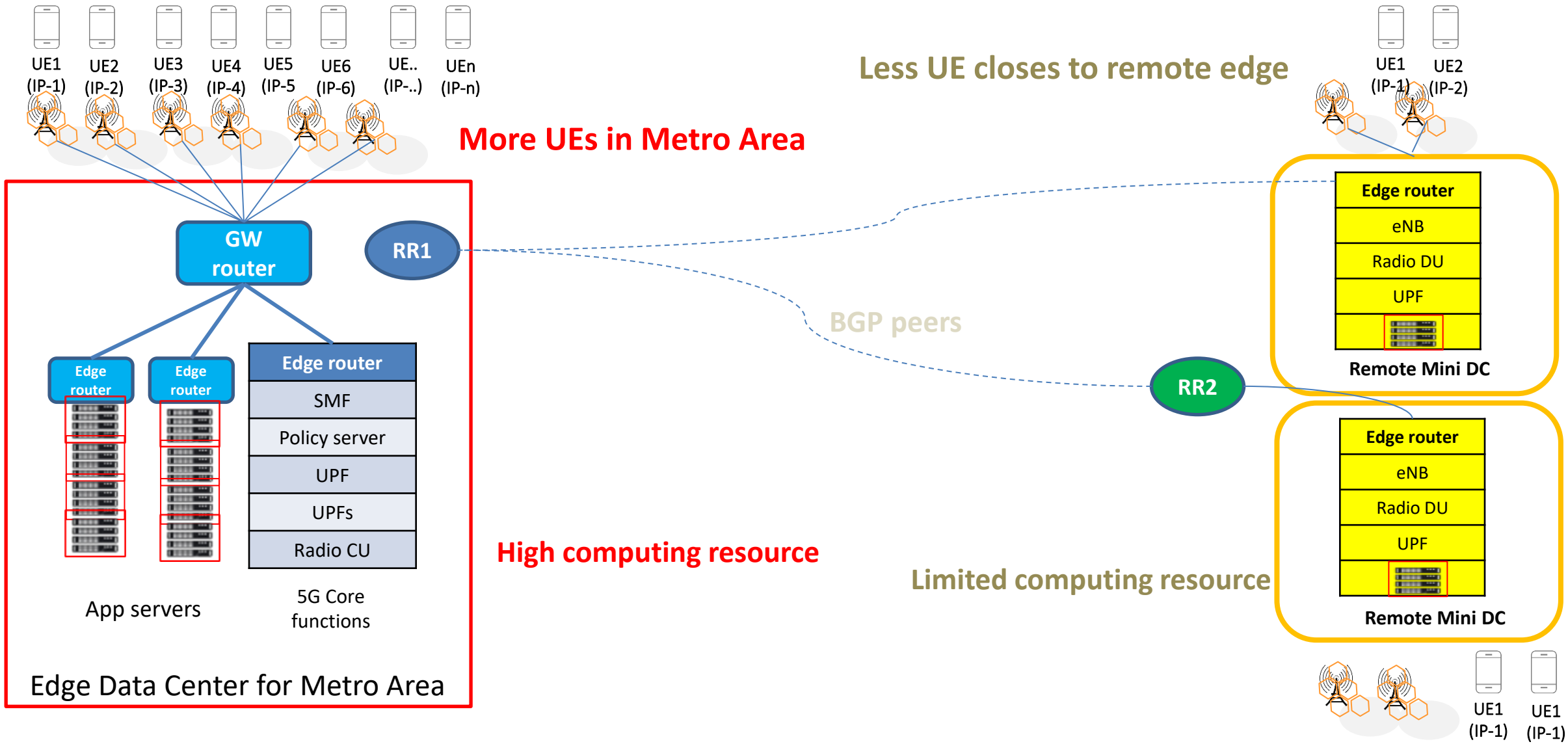
- **Providing Functional Equivalency**
  - Deploy instances for the same service across edge sites for better availability
- **Keep the load balance for both static and dynamic scenarios**
  - by both upper layer and network layer solutions
- **Steer traffic dynamically to the “Best” Service Instance**
  - Traffic is delivered to optimal edge sites according to more status,e.g,computing (defining what ‘best’ is for each service)

Use Case

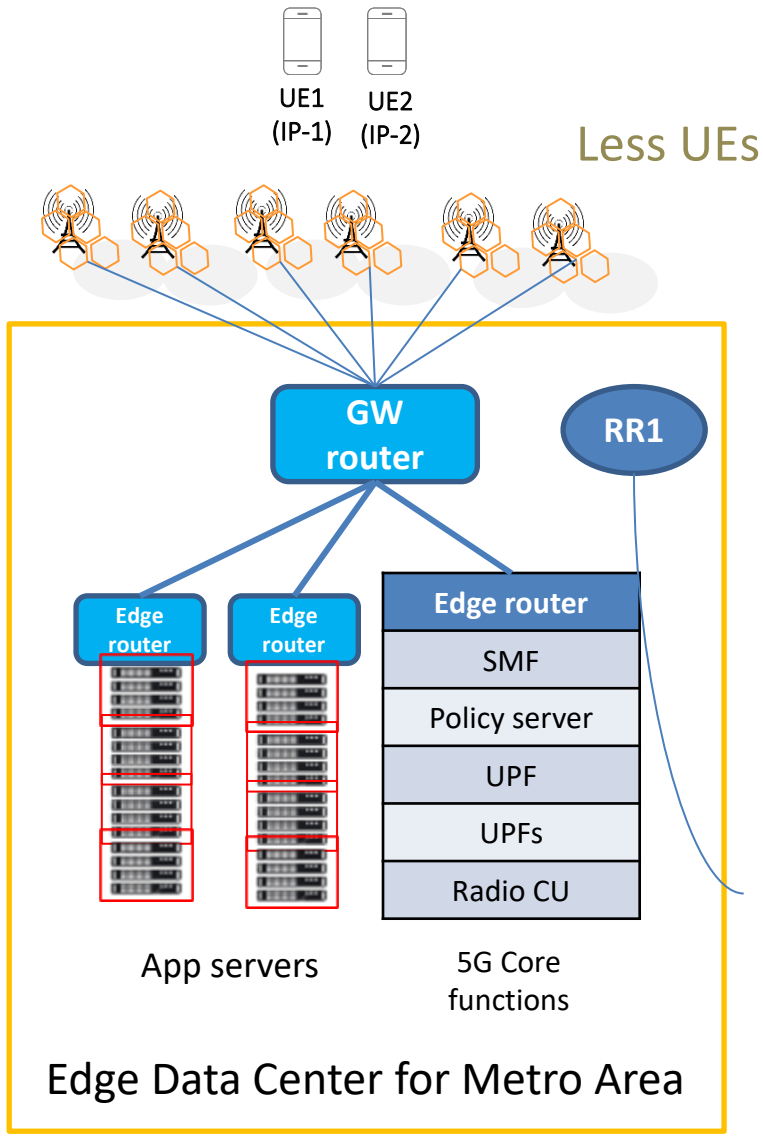
Edge computing has the advantage of ‘closest’, but in some cases, **the ‘closest’ is not the ‘best’.**

High computing resources needed by UEs at a remote site for short period of time, which is not long enough to justify adding more computing resources at the remote site.

High compute resources allocated at Metro Edge DC for large numbers of UEs (for working time)



Weekend events at a remote site require high compute usage (only for 1~2 days, can't justify adding servers to the remote site)

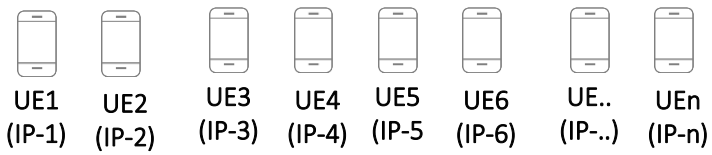


High computing resource

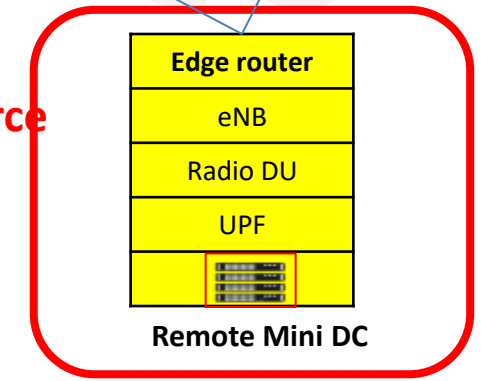
BGP peers

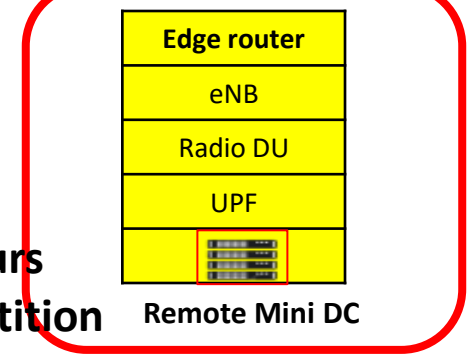
RR2

Limited computing resource



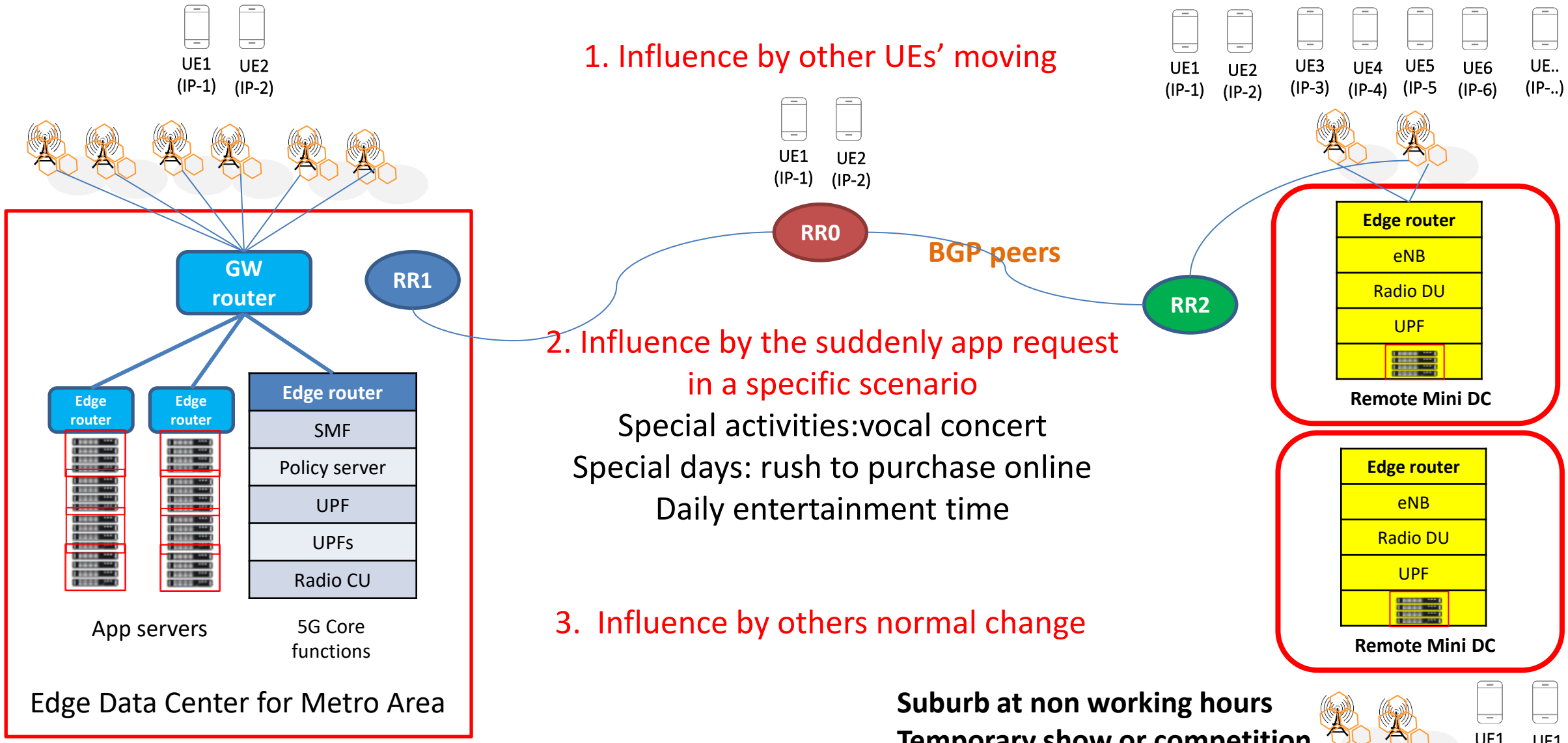
More UE closes to remote edge





Suburb at non working hours
Temporary show or competition

Variable network and computing resources
(For someone who is not moving but get a poor latency sometime)



Considerations

- Traffic needs to be steered among different edge sites.
- When steering traffic, it will be affected by many factors.

Use Case

Typical Applications

Some apps may require both low latency and high computing resources, which needs joint optimization of network and computing resource to guarantee the QoE.

Application 1 – AR/VR

Upper bound latency for motion-to-photon(MTP):includes frame rendering and requires less than **15-20 ms** to **avoid motion sickness**

client：

a) sensor sampling delay: <1.5ms

b) display refresh delay: ≈5 ms

Server：

c) frame rendering computing delay with **GPU≈ 5.5ms**

network：

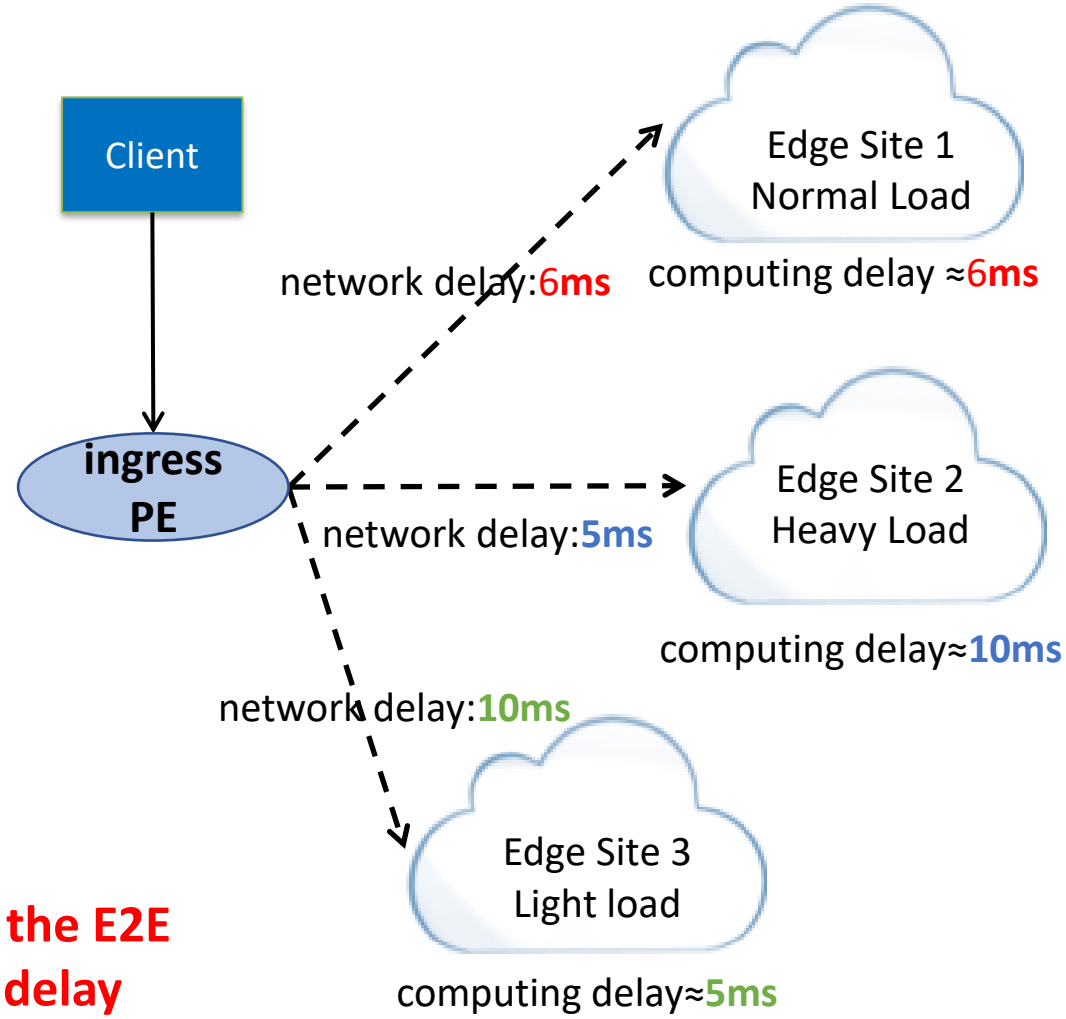
d) network delay =20-1.5-5.5-7.9 = **5.1ms**

The budgets for computing delay and network delay are almost equivalent!!

It can't meet the total delay requirements or find the best choice by either optimize the network or computing resource.

Require to dynamically steer traffic to the 'best' edge to meet the E2E delay requirements considering both network and computing delay

PS:Computing resources have a big difference in different edges



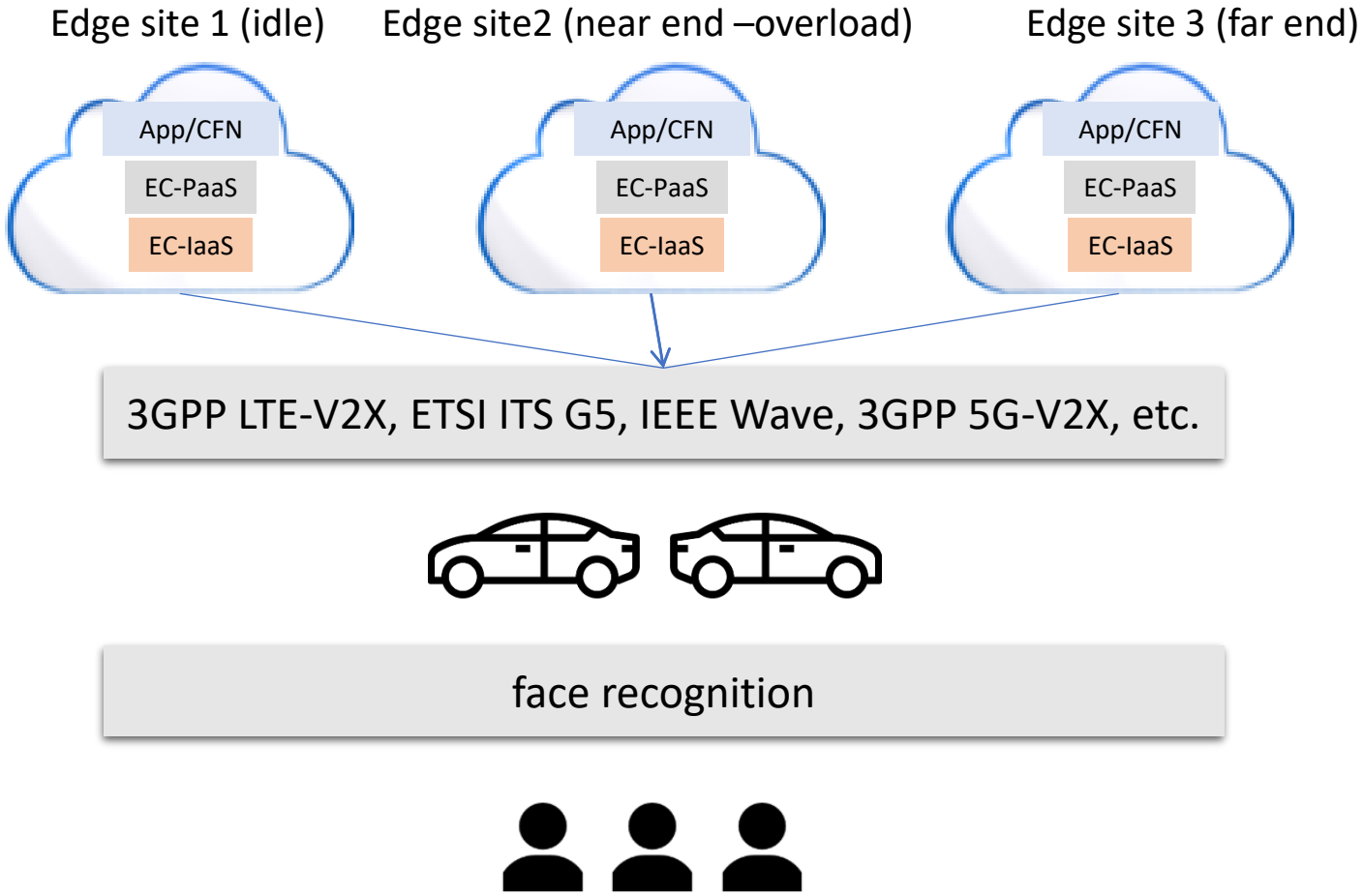
Application 2-Intelligent transportation

Connected Car

Function	Requirement
Driving-assist	Low Latency
HD and HP Map	High bandwidth

Video recognition at intersection

Function	Requirement
Safety Monitoring	Low Latency
Data analysis	High bandwidth



Considerations

Those use cases and apps need the flexibility scheduling of both network and computing resource.

Operators have the network resource and the edge resource(operator room), could steer the traffic through network.

Thank you!