THE FRAMEWORK PROGRAMME FOR RESEARCH AND INNOVATION

HORIZON 2020

T02: RAN Slicing & Control: Challenges, Technologies, and Tools

Navid Nikaein

Associate Professor - Eurecom

Coordinator of Mosaic-5G Initiative

Tutorial IEEE NoF SIGCOMM

19-21 November, 2019, Poznan, Poland

Provide a comprehensive guide on RAN slicing and data-driven

(1) highlight the importance and timeliness of softwarization, virtualization, and disaggregation of RAN to enable multiservice multi-tenant RAN toward So-RAN architecture

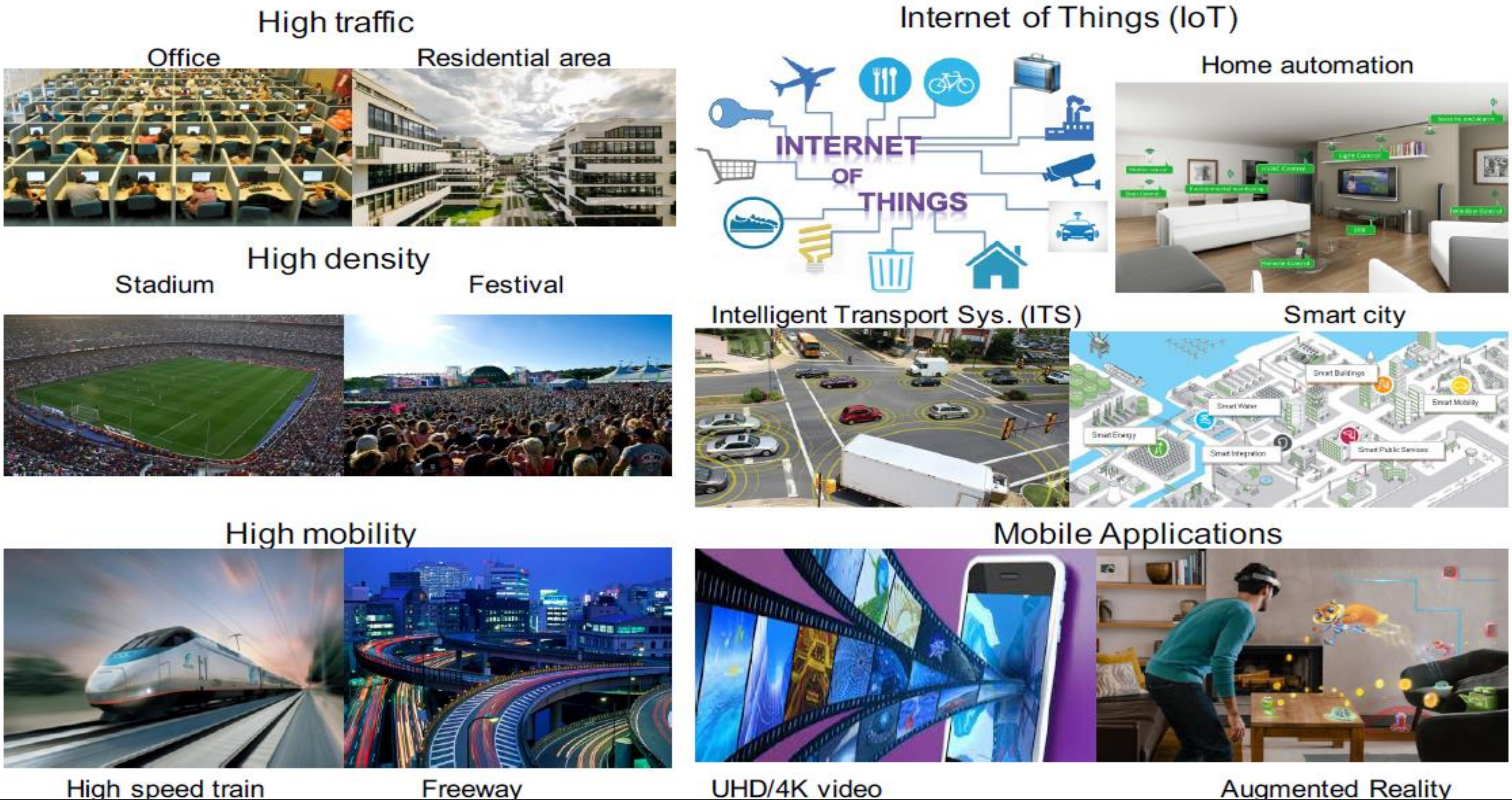
(2) Elaborate the benefits and implications of RAN data mining and analytics to enable data-driven RAN control loop and provide QoS

(3) Cover a well-balanced research and development topics including challenges, key technologies, and proof-of-concept prototyping

Tutorial Objectives


Connected, Controlled, and Flexible
Digital Society
Value Creation
Consistent experience
Sustainable business model

What is 5G?



Many 5G Use-cases

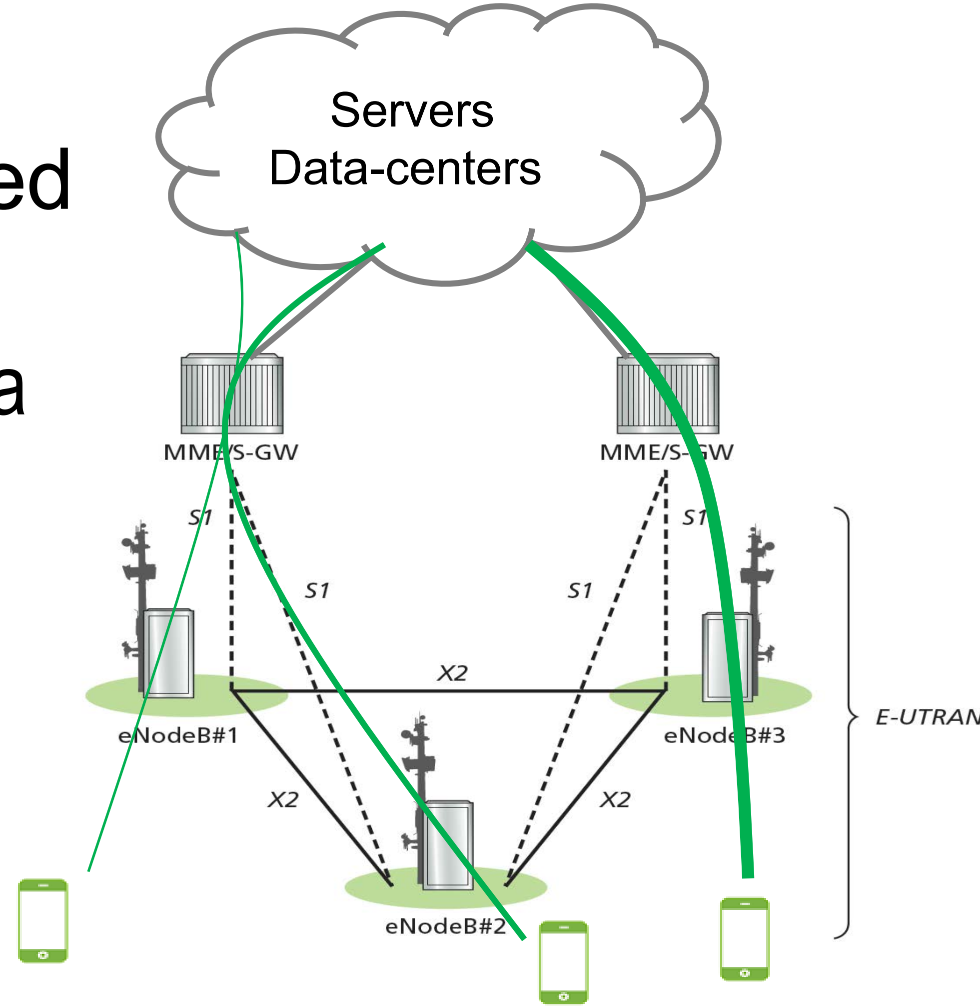
Source: Dohler

Moving towards Internet of Skills

Communication-oriented

Today's 4G is designed for a limited number of UCs

- Throughput-optimized
- Fixed
- Rigid

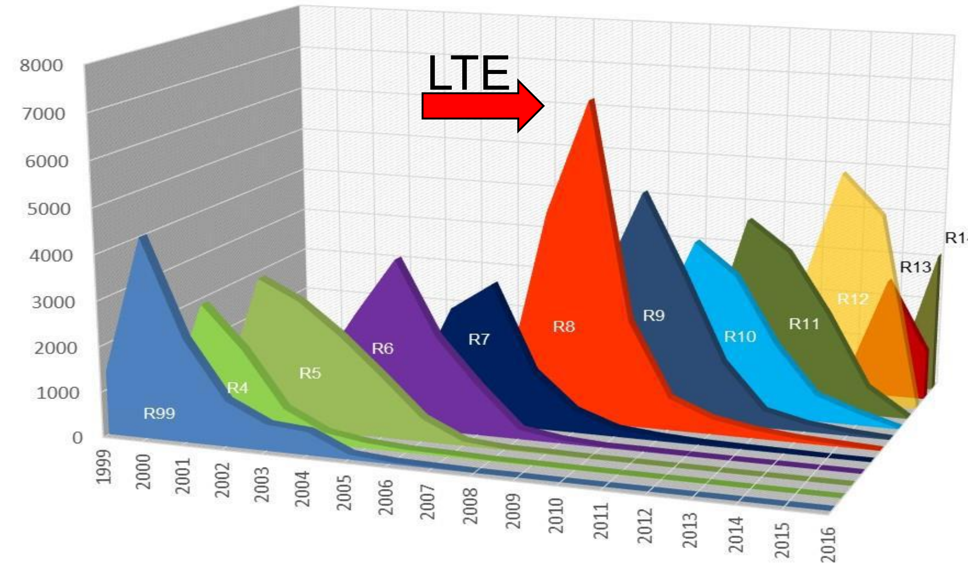


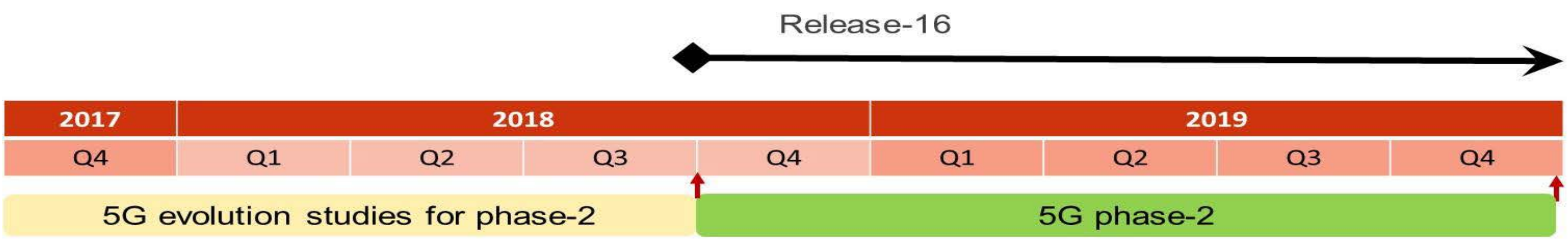
Is 4G enough?

Mindful about

3GPPP facts and figures

514 Companies from 45 Countries
50,000 delegate days per year
40,000 meeting documents per year
1,200 specifications per Release
10,000 change requests per year

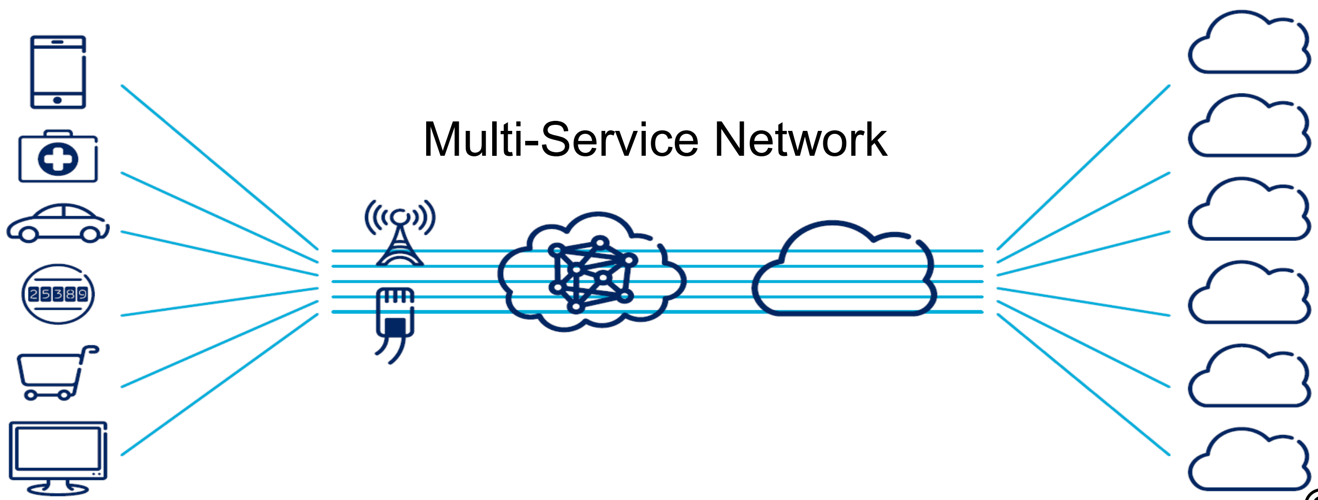




Communication-oriented 4G

Turn physical infrastructure into multiple logical networks, one per service instance: One-Network, Many-Service

NOT a one-size fits all architecture NOT a Dedicated Network

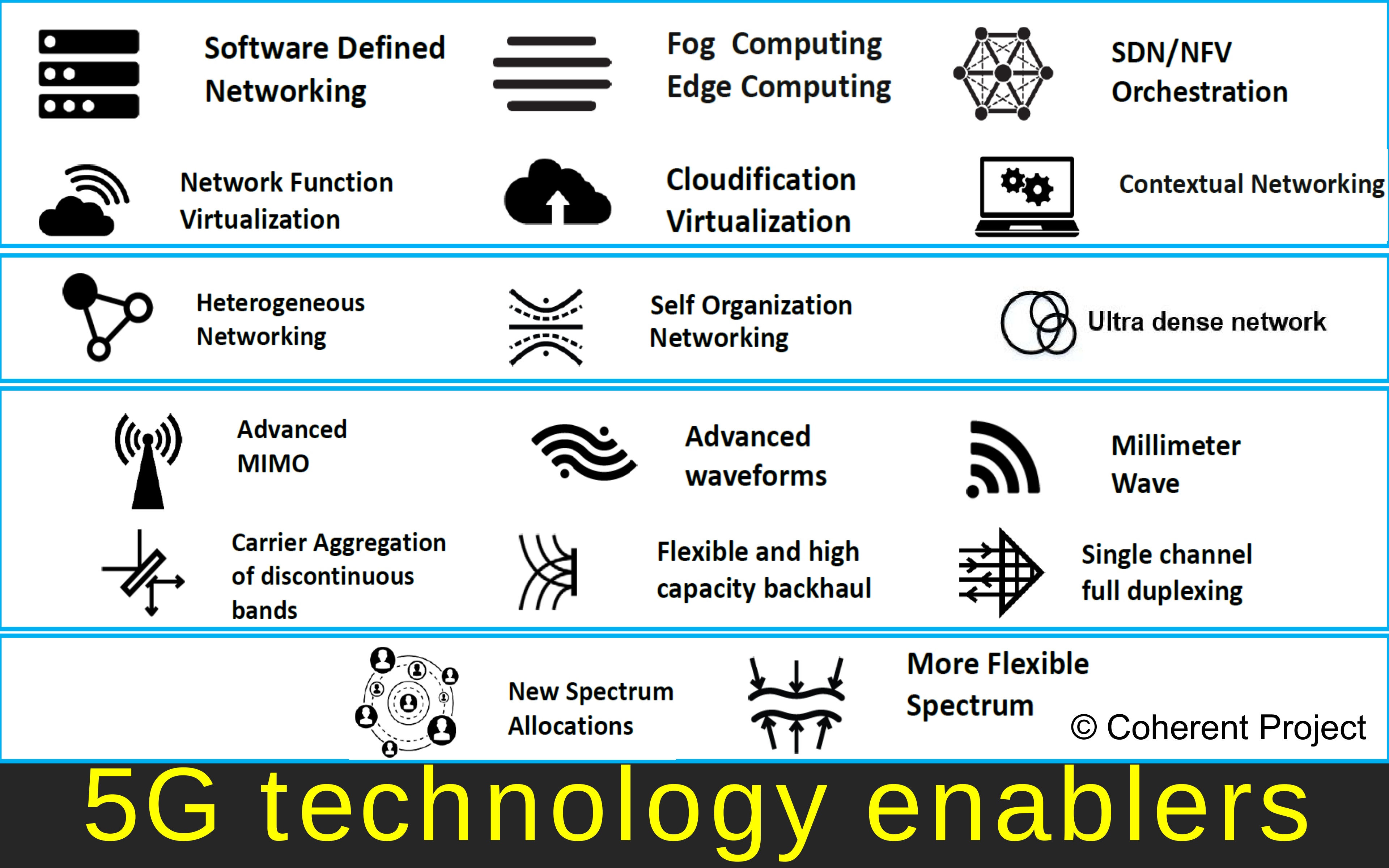


Service-oriented 5G

Different aspects of network slicing have been already prototyped both Opensource and commercials platforms

Industry is currently providing network slicing by means of
(a) Local/dedicated services enabled by MEC platform
(b) Dedicated core networks and RAN sharing

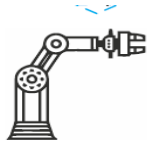
Next steps : SO-CN and SO-RAN


Software Defined Networking
Fog Computing
Edge Computing
SDN/NFV Orchestration
Network Function Virtualization
Cloudification Virtualization
Contextual Networking
Heterogeneous Networking
Self Organization Networking
Ultra dense network
Advanced MIMO
Advanced waveforms
Millimeter Wave
Carrier Aggregation of discontinuous bands
Flexible and high capacity backhaul
Single channel full duplexing
New Spectrum Allocations
More Flexible Spectrum
© Coherent Project
5G technology enablers

Network Slicing

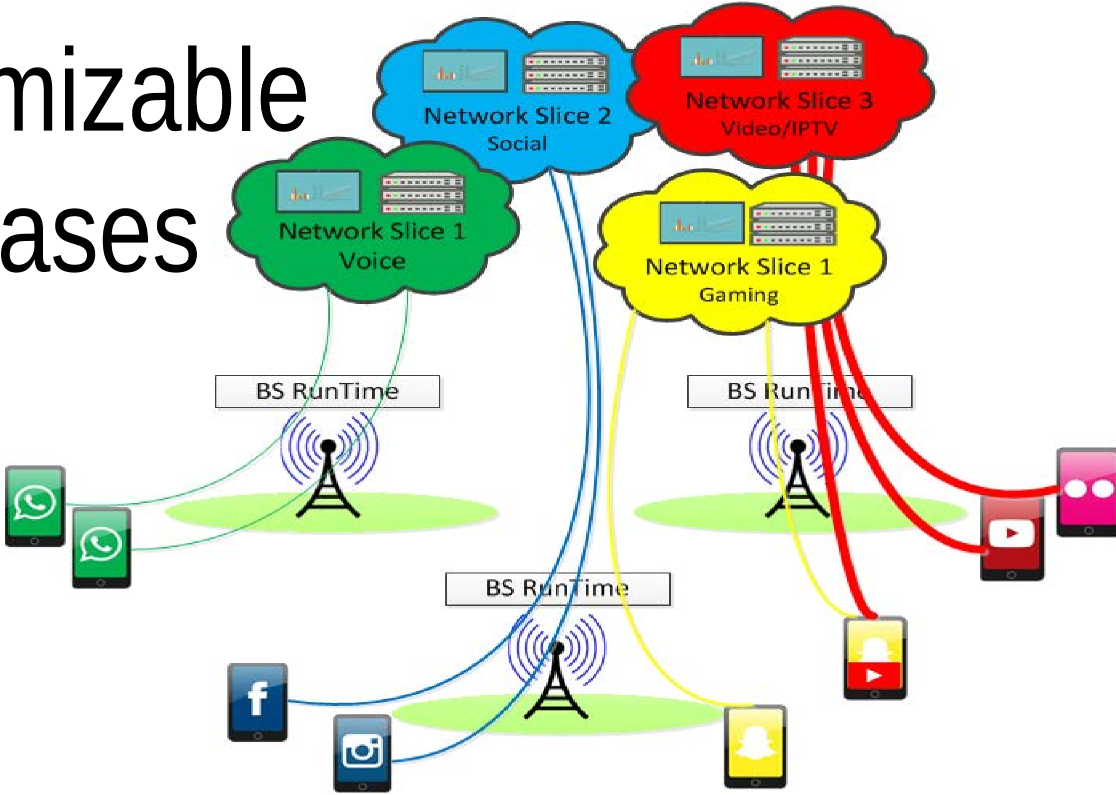
Flexible & Customizable for each use-cases

xMBB



uRLLC

mMTC



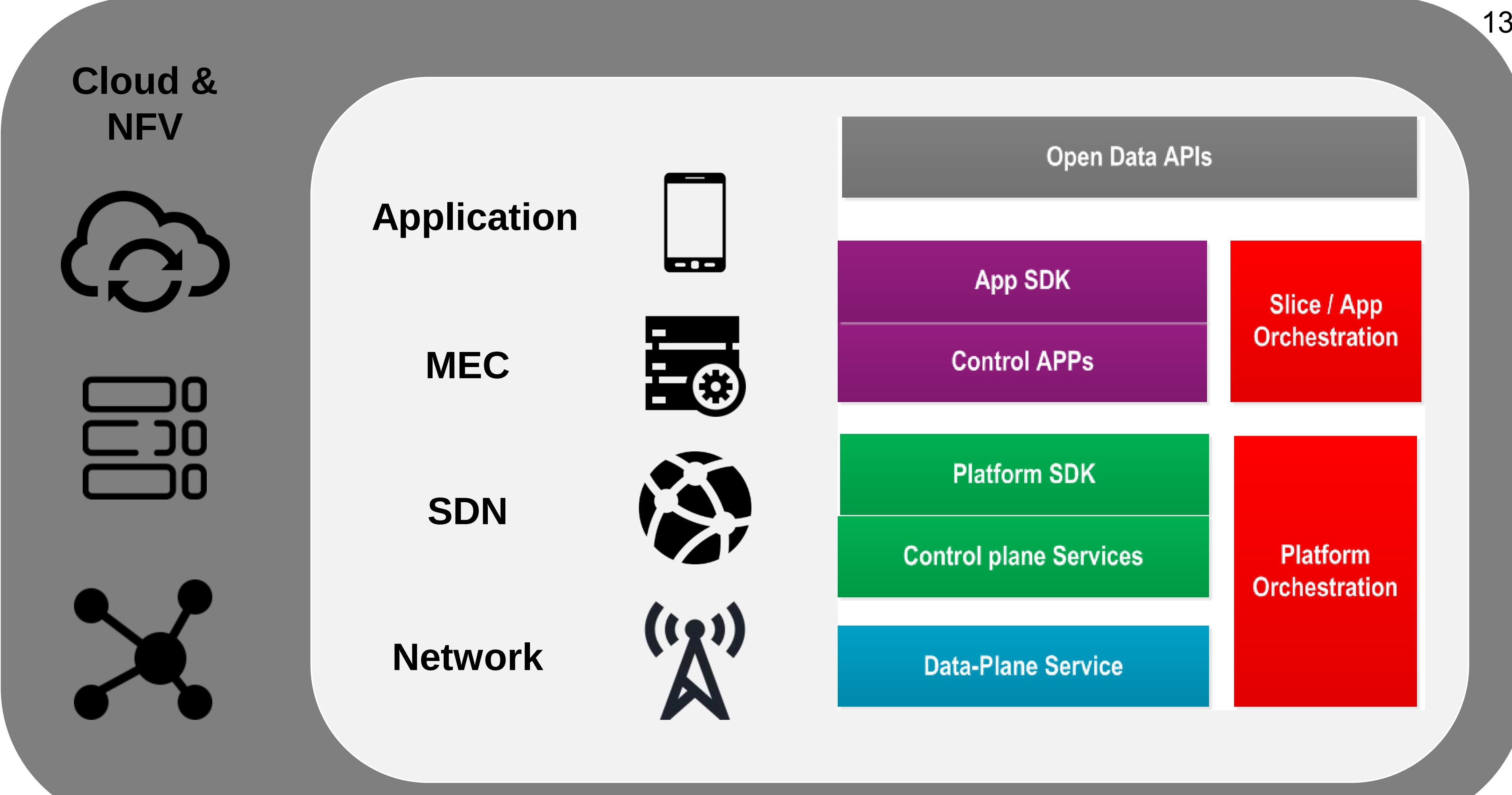
Service-oriented 5G

Slicing Technology Enablers:

- Softwarization
- Virtualization
- Disaggregation

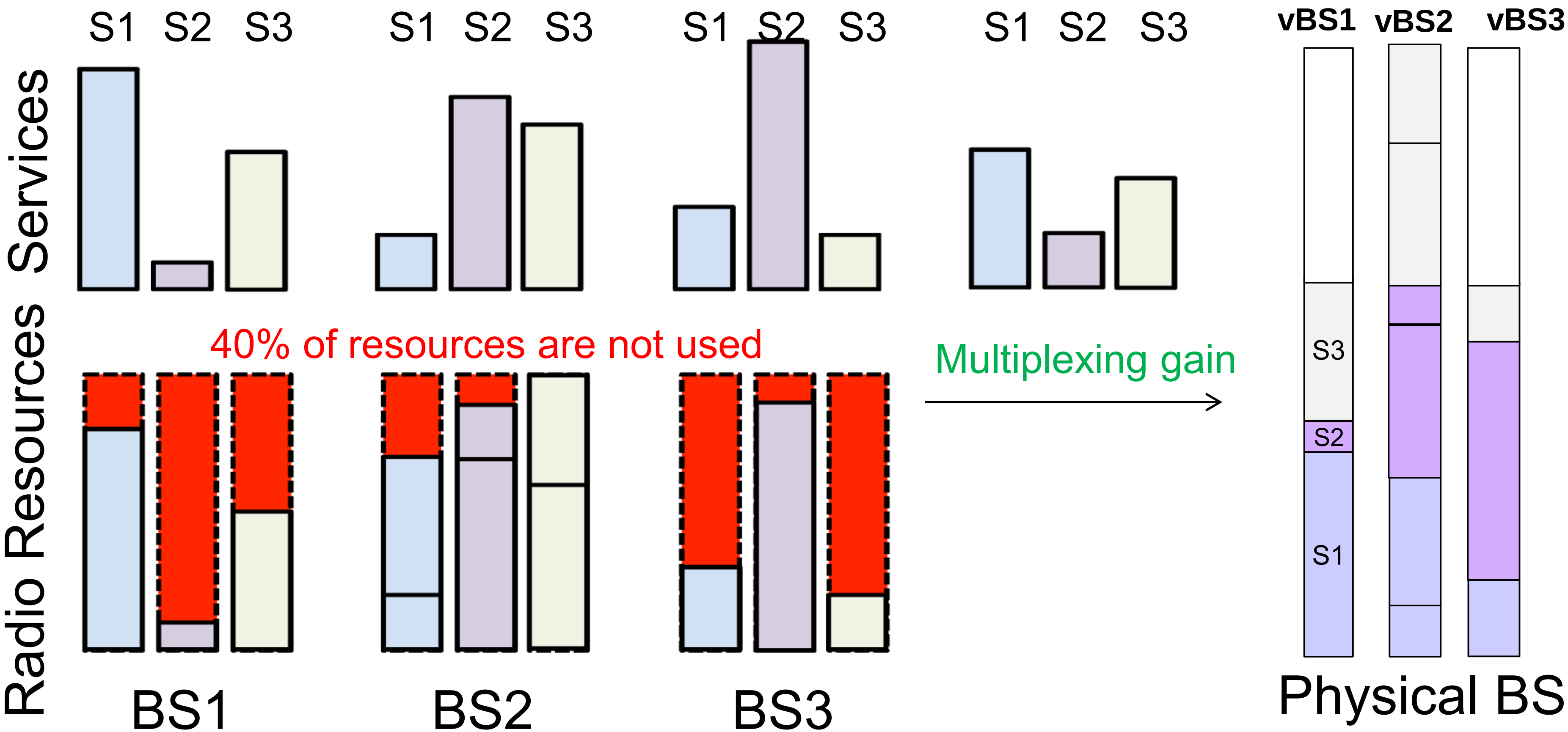
Multi-service multi-tenant network

Service-oriented 5G



Slicing Technology Enablers

Disaggregation

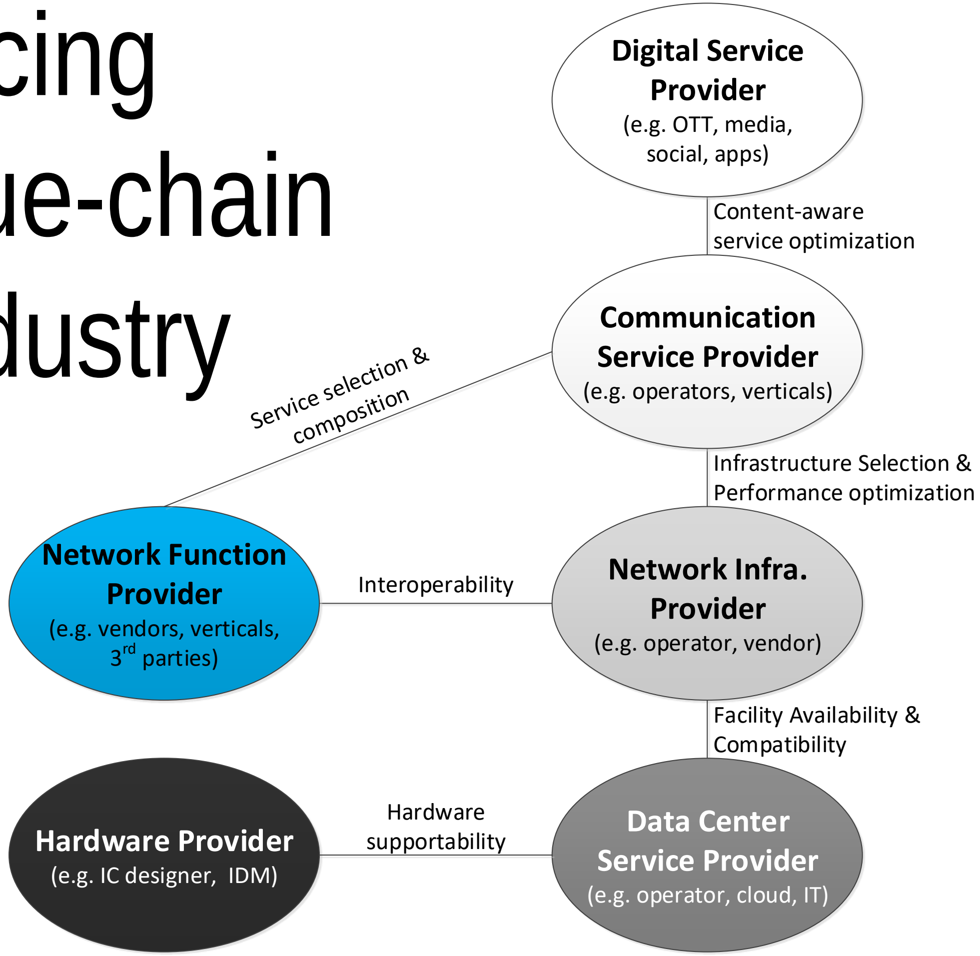


Slicing Technology Enablers

Why will it happen?

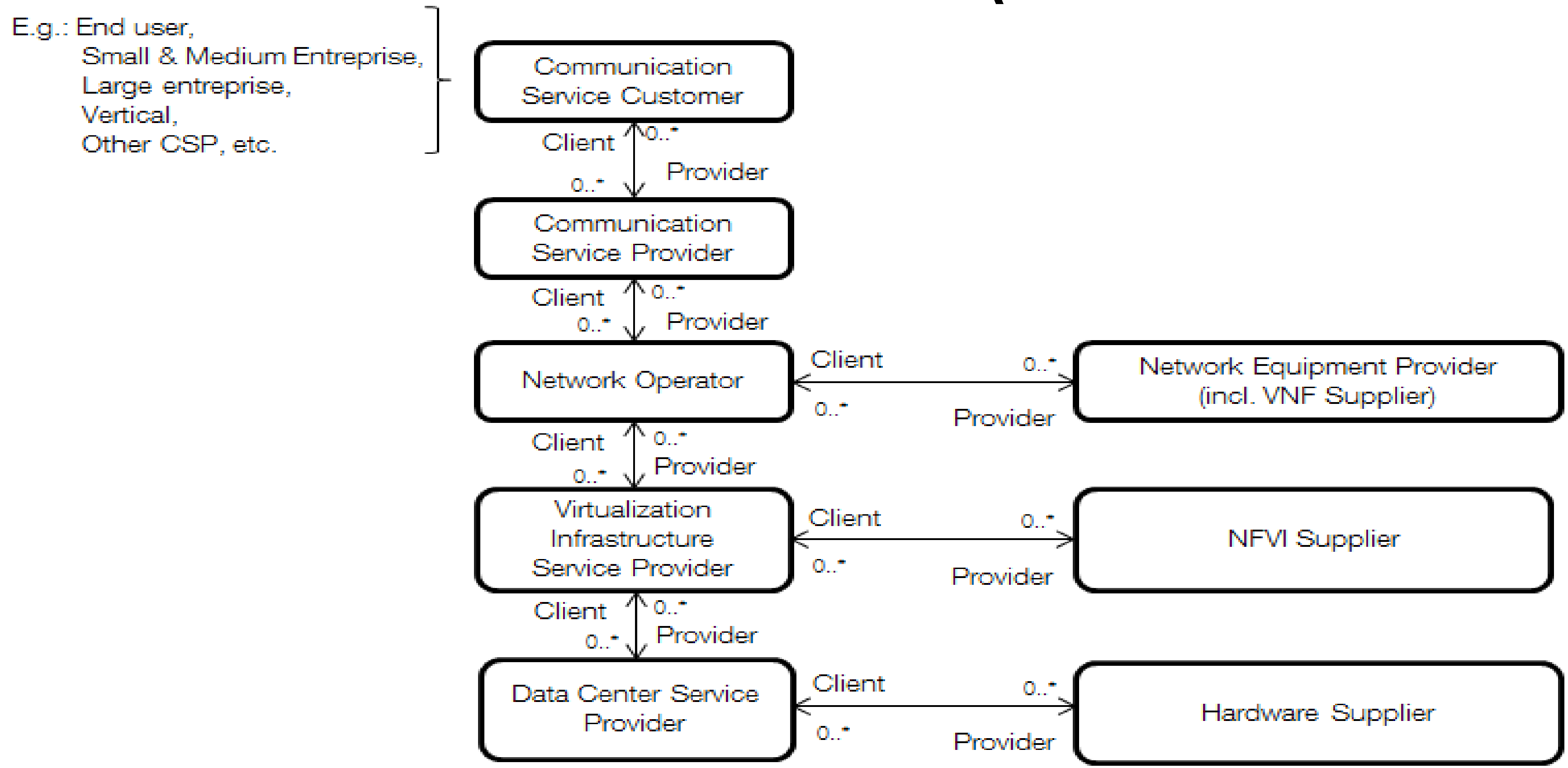
Extreme network flexibility and elasticity

Network Slicing Evolves the value-chain of telecom industry



Service-oriented 5G

3GPP Role Model (3GPPP 28.801)



3GPP re-architects mobile networks

	3G	4G	5G
Downlink waveform	CDMA	OFDM	OFDM, SCFDMA
Uplink waveform	CDMA	SCFDMA	OFDMA, SCFDMA
Channel coding	Turbo	Turbo	LDPC (data) / Polar (L1 contr.)
Beamforming	No	Only data	Full support
Spectrum	0.8 – 2.1 GHz	0.4 – 6 GHz	0.4 – 90 GHz
Bandwidth	5 MHz	1.4 – 20 MHz	Up to 100 MHz (400MHz for >6GHz)
Network slicing	No	No	Yes
QoS	Bearer based	Bearer based	Flow based
Small packet support	No	No	Connectionless
In-built cloud support	No	No	Yes

Service-oriented 5G

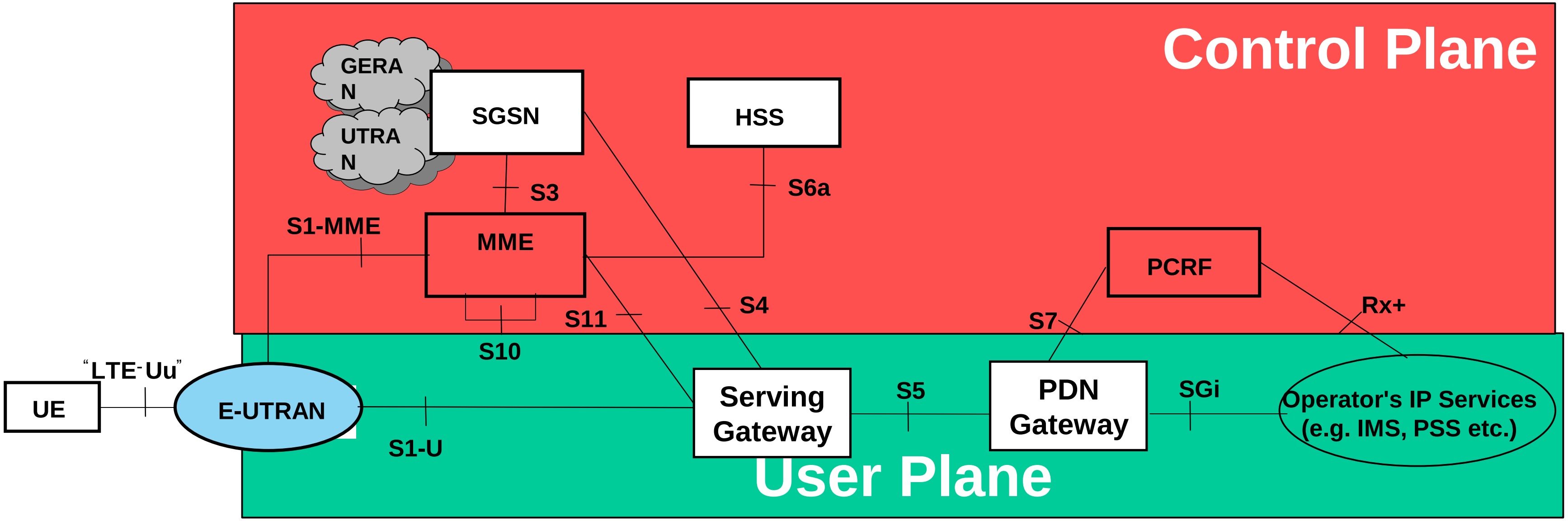
Monolithic BS

Stateful network entities

Transactional communication mode

Certain level of CP and UP separation

Common entity for user mobility and session management



Communication-oriented 4G

3GPP Network Sharing Models

- Multi-operator RAN(MORAN)
  - Shared RAN nodes, dedicated spectrum, but separated CN per operator
- Multi-operator CN (MOCN)
  - Shared RAN nodes and spectrum, but separated CN per operator with proprietary services
- Gateway CN (GWCN)
  - shared RAN and part of core networks

- Dedicated core (DECOR)
  - Deploy multiple dedicated CNs (DCNs) within a single operator network
  - One or multiple MMEs and SGWs/PGWs, each element
- Evolved DECOR (eDECOR)
  - UE assisted DCN selection
  - Network Node Selection Function (NNSF) at RAN to select directly the proper DCN towards which the NAS signaling needs to be forwarded
  - Congestion control and load balancing among multiple DCN with shared MME

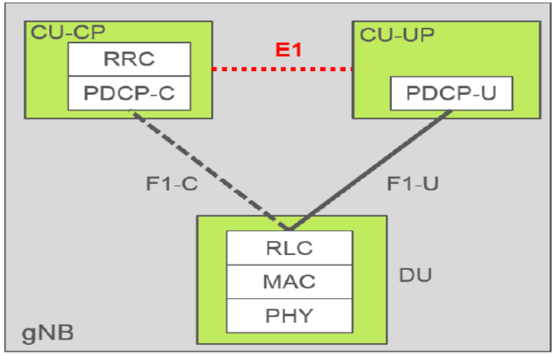
3GPP re-architects mobile networks

3 Tier RAN Node

CU0 → DU[0-n] → RRU[0-m]

Functions Split

CP UP split

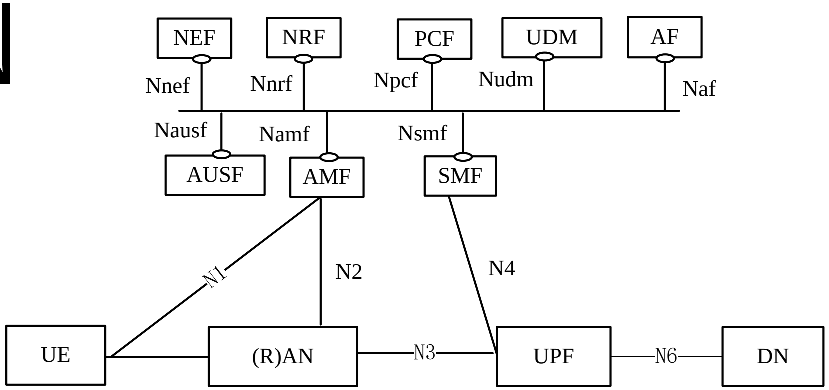


Service-oriented CN

service catalog and discovery

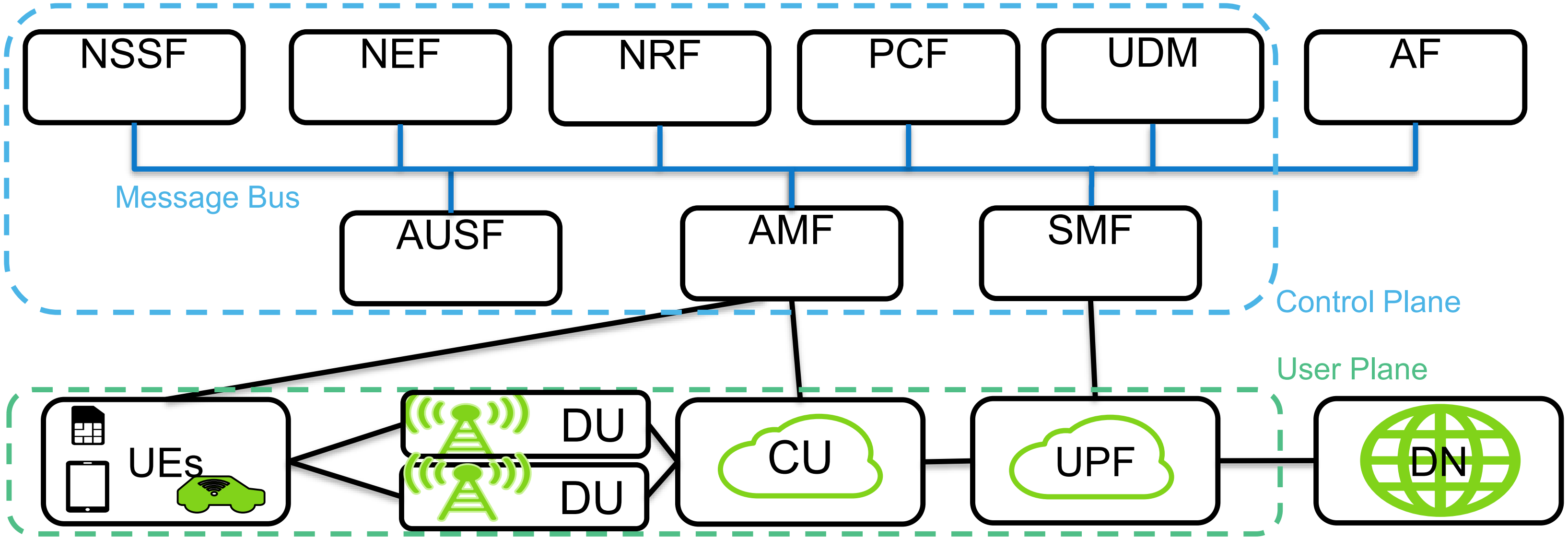
Slice selection function

CP and UP split



3GPP 5G RAN and CN

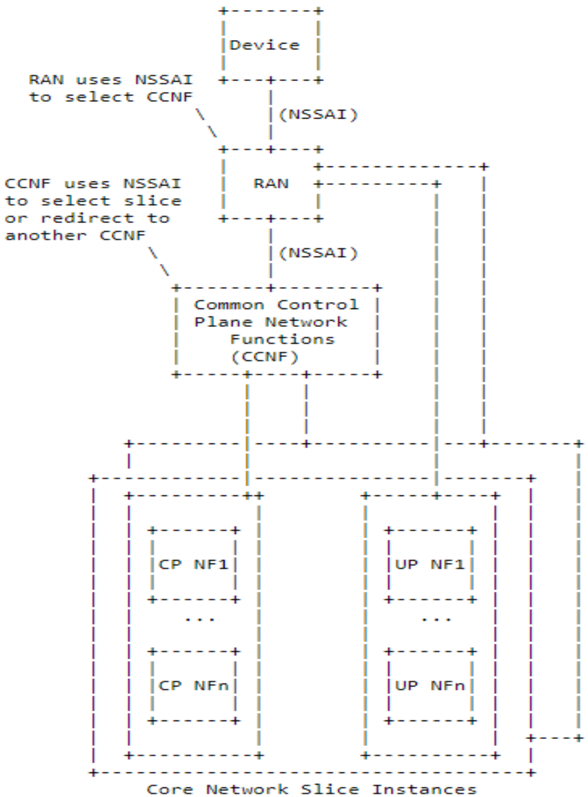
3GPP re-architects mobile networks

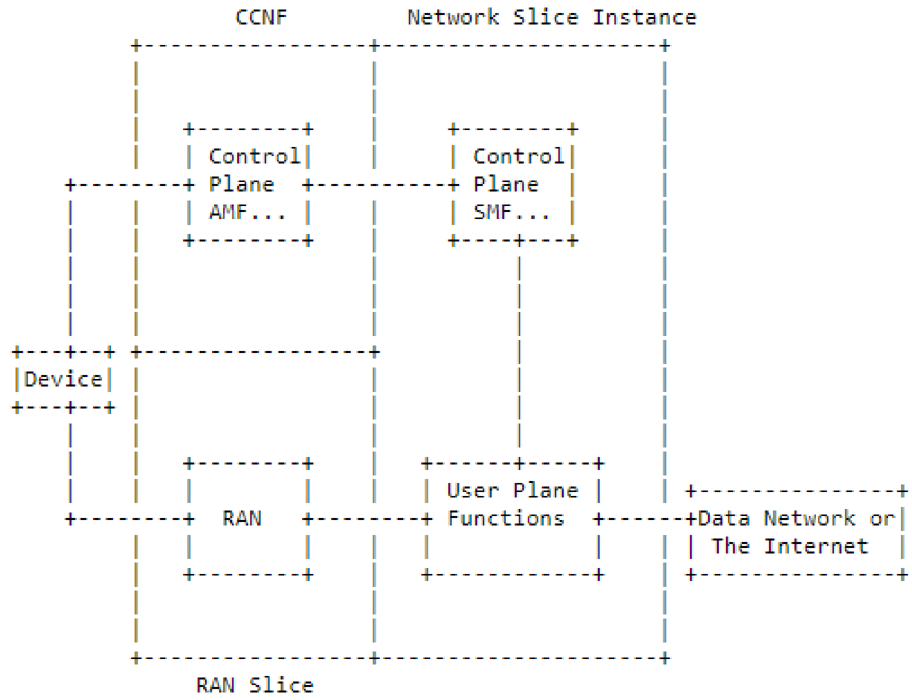


AMF	Access & Mobility Management Function	SMF	Session Management Function
AUSF	Authentication Server Function	UPF	User Plane Function
NRF	Network Repository Function	AF	Application Function
UDM	Unified Data Management	PCF	Policy Control Function
NSSF	Network slice selection function	NEF	Network Exposure Function

Service-oriented 5G

3GPP re-architects mobile networks





3GPP network slicing

Select the set of network slice instances serving the UE

Determine the allowed Network Slice Selection Assistance Information (NSSAI) and the mapping to the subscribed S-NSSAIs

Determine the configured NSSAI and the mapping to the subscribed S-NSSAIs

Determine the AMF set to be used to serve the UE or a list of candidate AMFs by querying the NRF

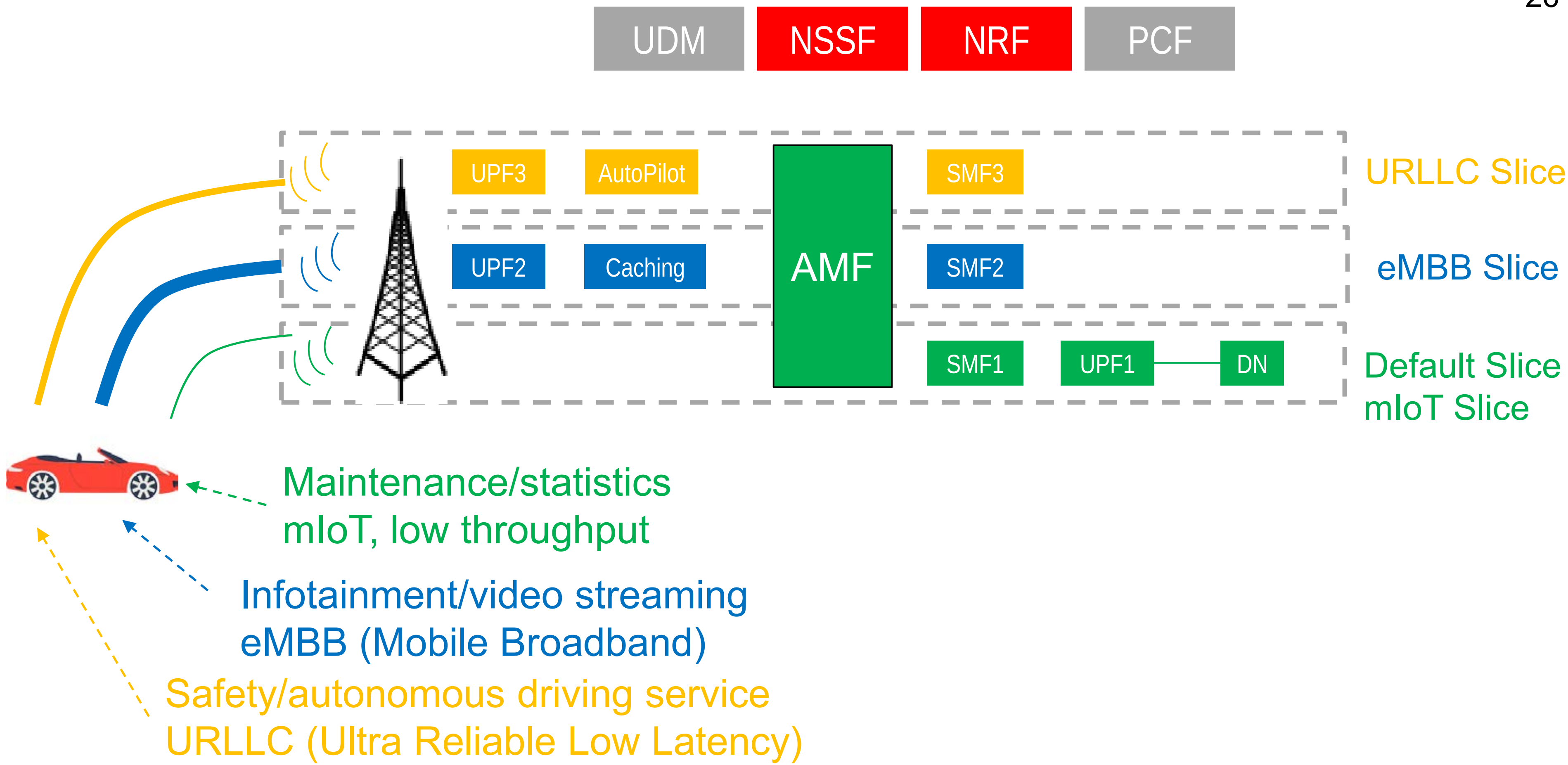
NSSF: Network slice selection function

Provides information on the discovered NF instances upon discovery requests

Maintains the NF profile of available NF instances and their supported services

NF Profile: instance ID, type, PLMN ID, Network Slice identifiers, IP address of NF, NF capacity information, NF specific service authorization information, names of supported services, endpoint addresses of supported services, identification of stored data information

NRF: network repository function

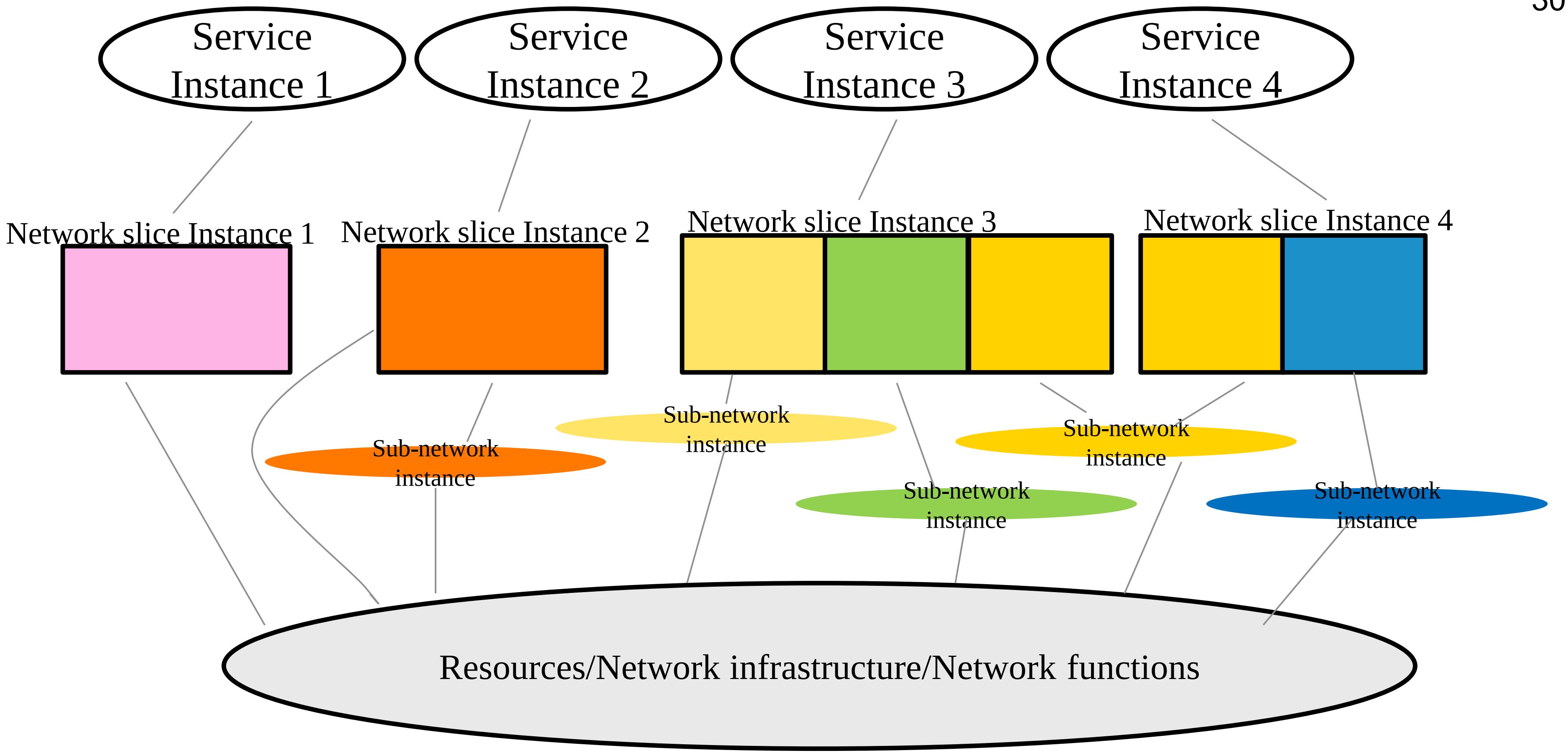


Dedicated or Shared Functions?

Dedicated or Shared Resources?

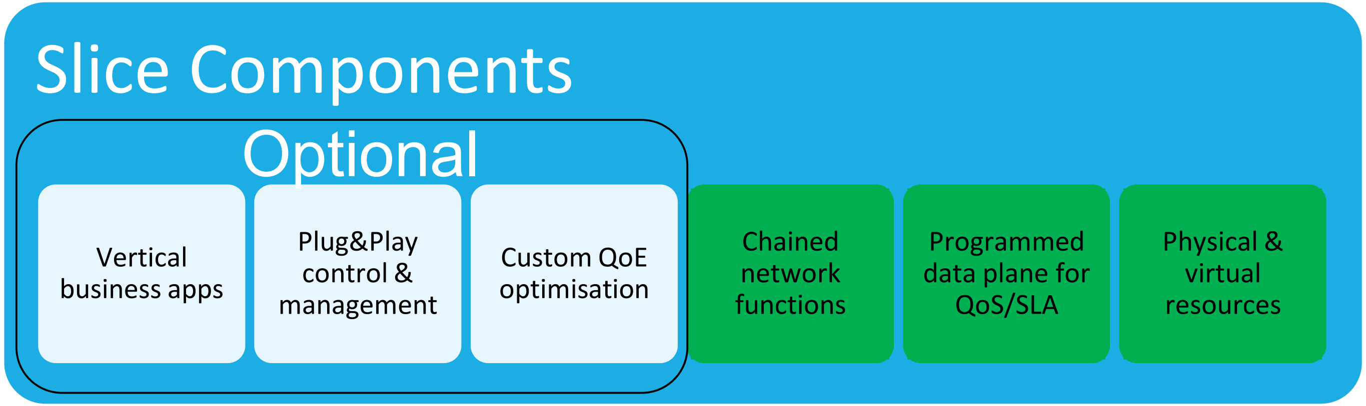
Dedicated or Shared Resources?

RAN Slicing



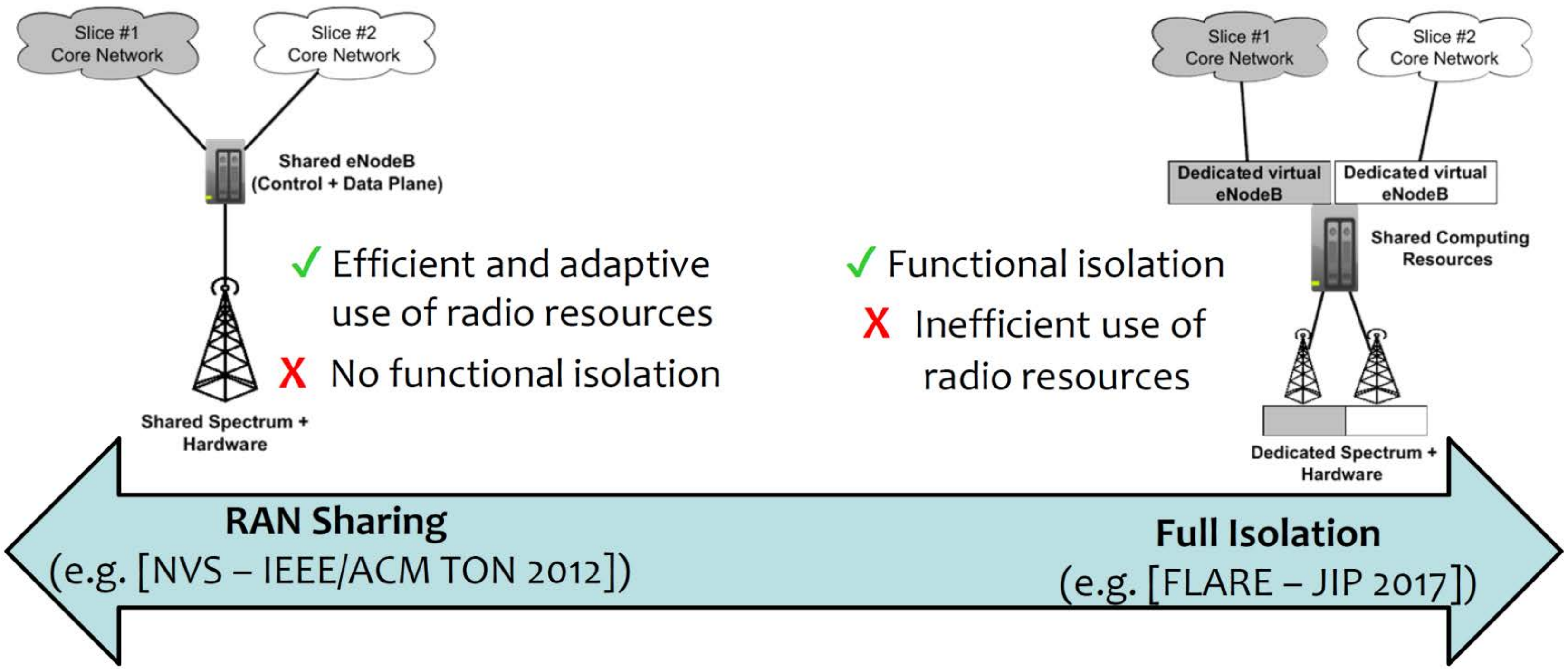
Network Slicing Concept

Composition and deployment of multiple E2E logical networks tailors to a service over a shared infrastructure, and their delivery as a slice



What is a slice?

RAN Slicing



Dedicated and Shared

FlexRAN : a SD-RAN platform enabling RAN sharing (Foukas et al., 2016)

Fully isolation platform with vBSs as different slices (Nakao et al., 2017)

Separated radio resources for intra/inter-slice scheduler (Rost et al., 2017)

RRM is enforced using a resource visor per slice (Ksentini et al., 2017)

ORION: BS hypervisor isolate slice-specific control logics and share the virtualized radio resources (Foukas et al., 2017)

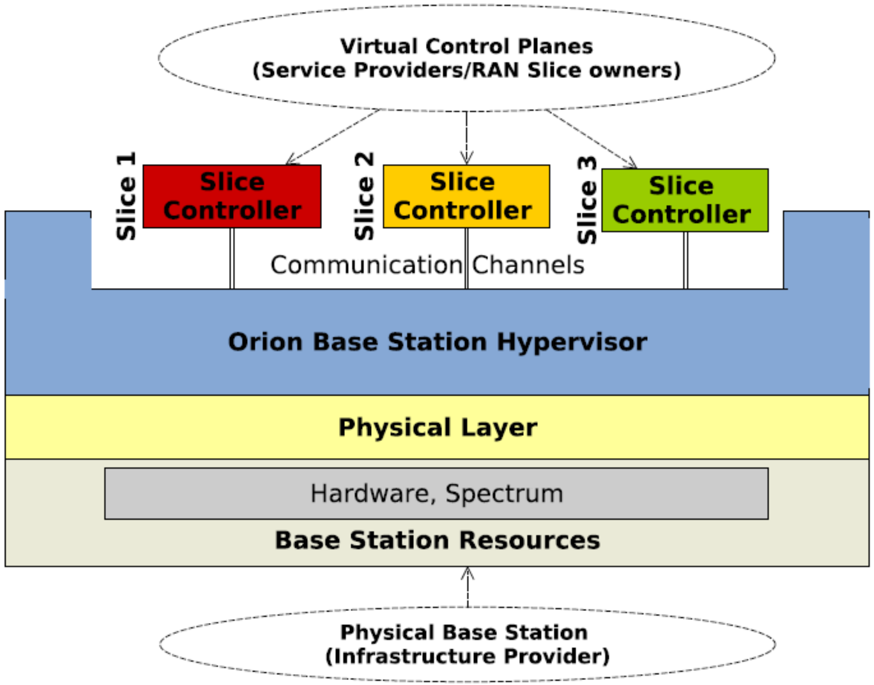
RAN runtime targets customization and multiplexing in several aspects over disaggregated RAN (Chang et al., 2017)

State of the Art

RAN slicing system

(1) Isolate slice-specific control logics while keeping common CP/UP functions

(2) Share radio resources in virtualized or physical form



ORION

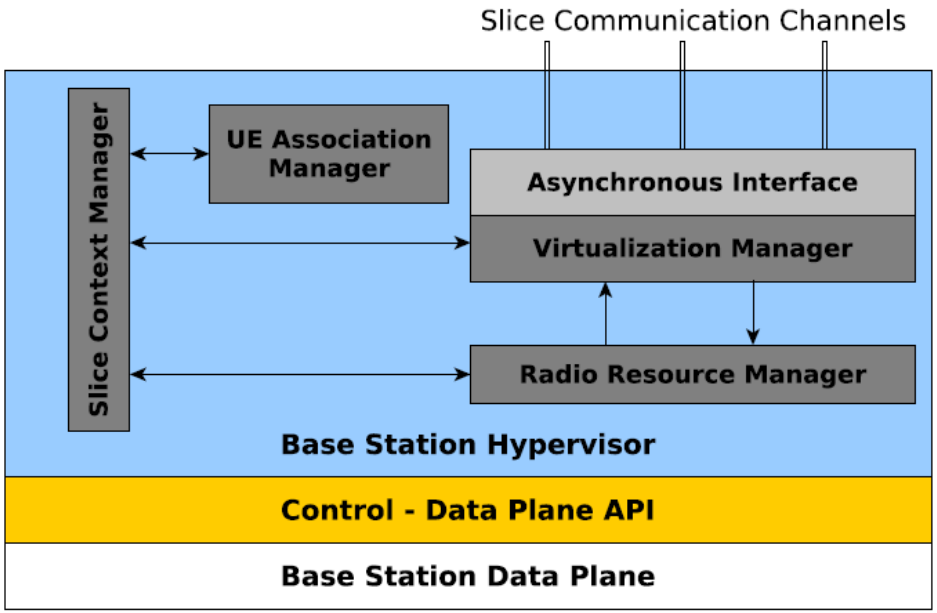
Components

(1) Slice context manager performs lifecycle management of each slice (SLA, active UEs, admission control)

(2) Virtualization manager
- provides a generic form of abstraction for virtualizing radio resources and data plane state
- presents a virtual/isolated view to each slice virtual control plane

(3) Radio resource manager allocates physical resources among slices

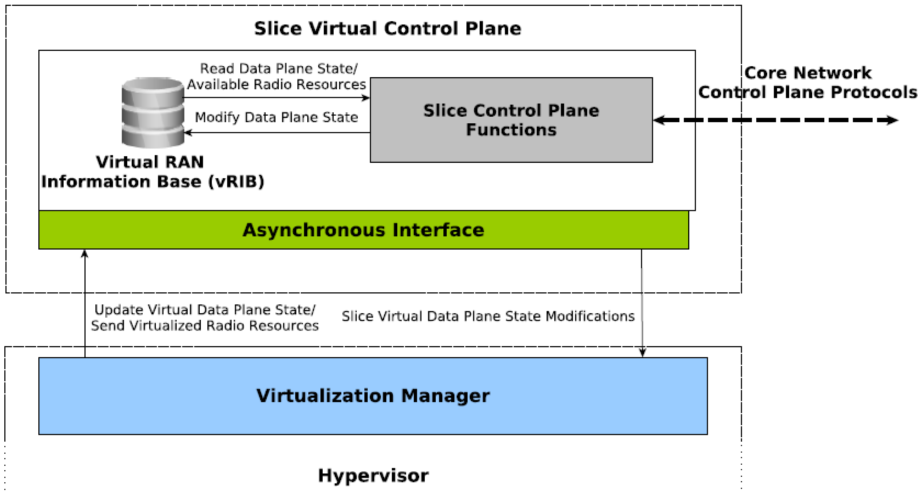
(4) UE association manager handles slice discovery by UEs and maps UEs to slices



Virtual Control Plane

(1) Interacts with the underlying infrastructure via the virtualization Manager of the Hypervisor

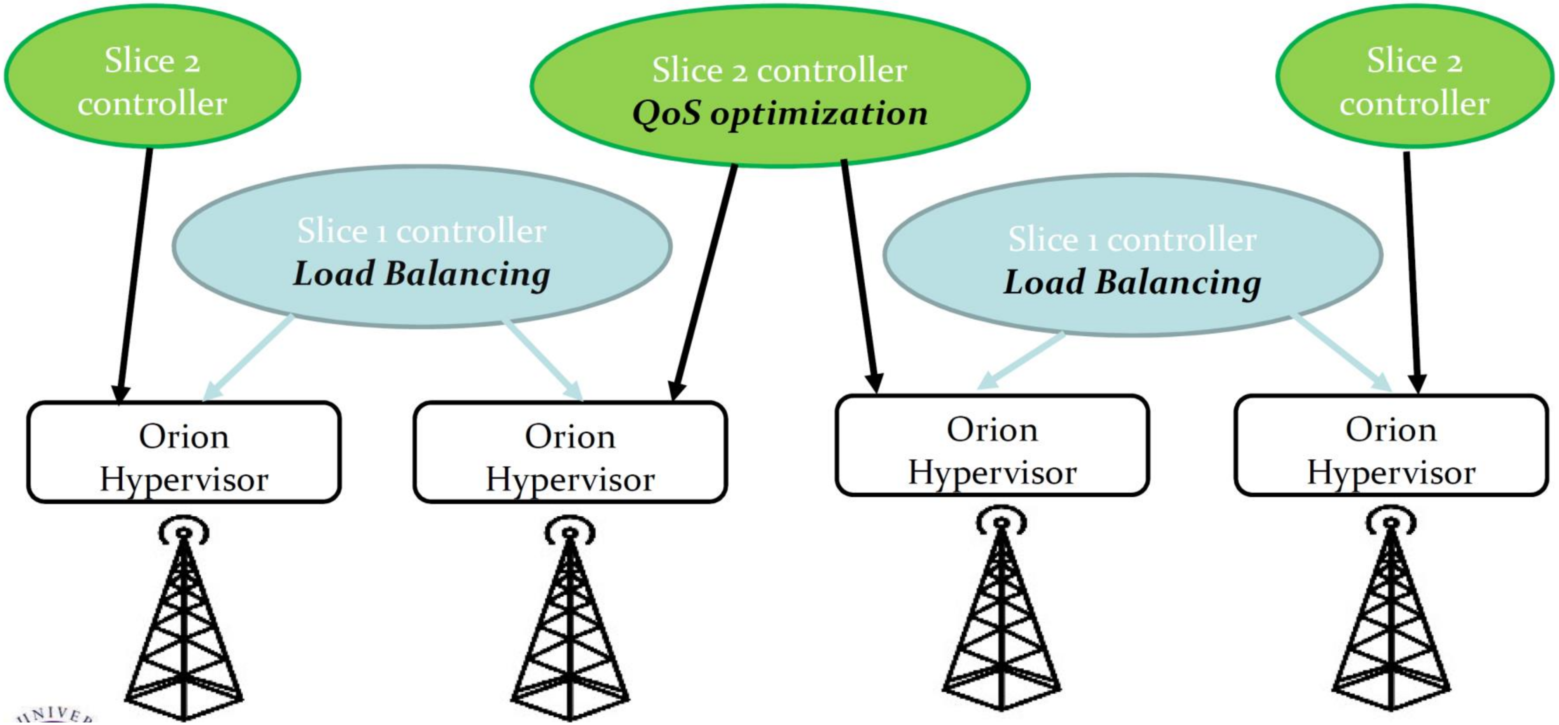
- translated into control-data APIs



(2) Operates over vRIB, the locally maintained state of virtual radio resources and data plane

- Slice network view and state

ORION RAN Slicing System

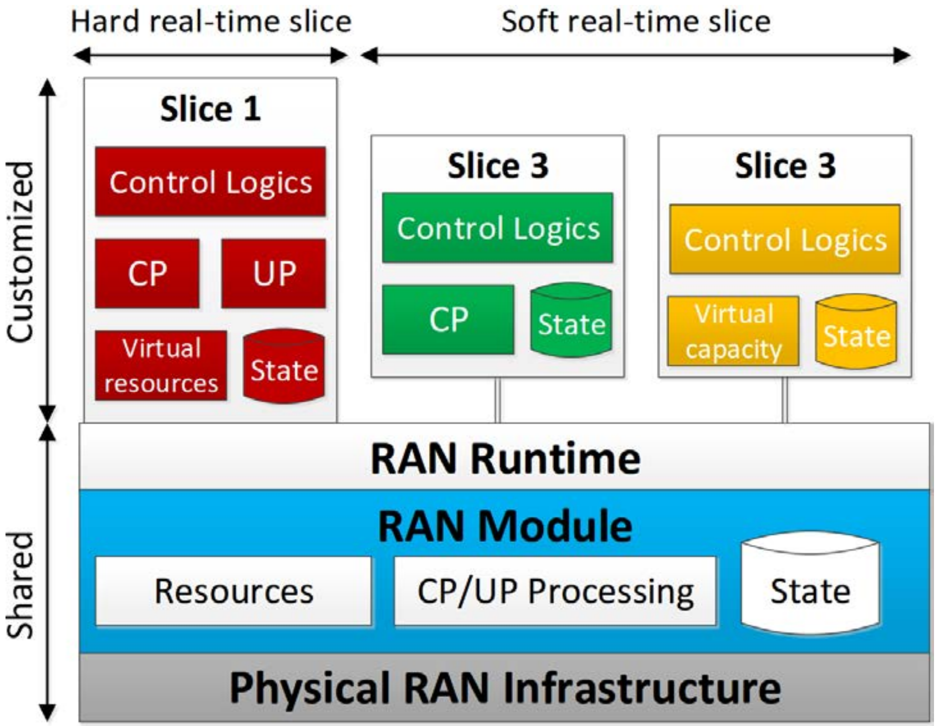


RAN Slicing Execution Env.

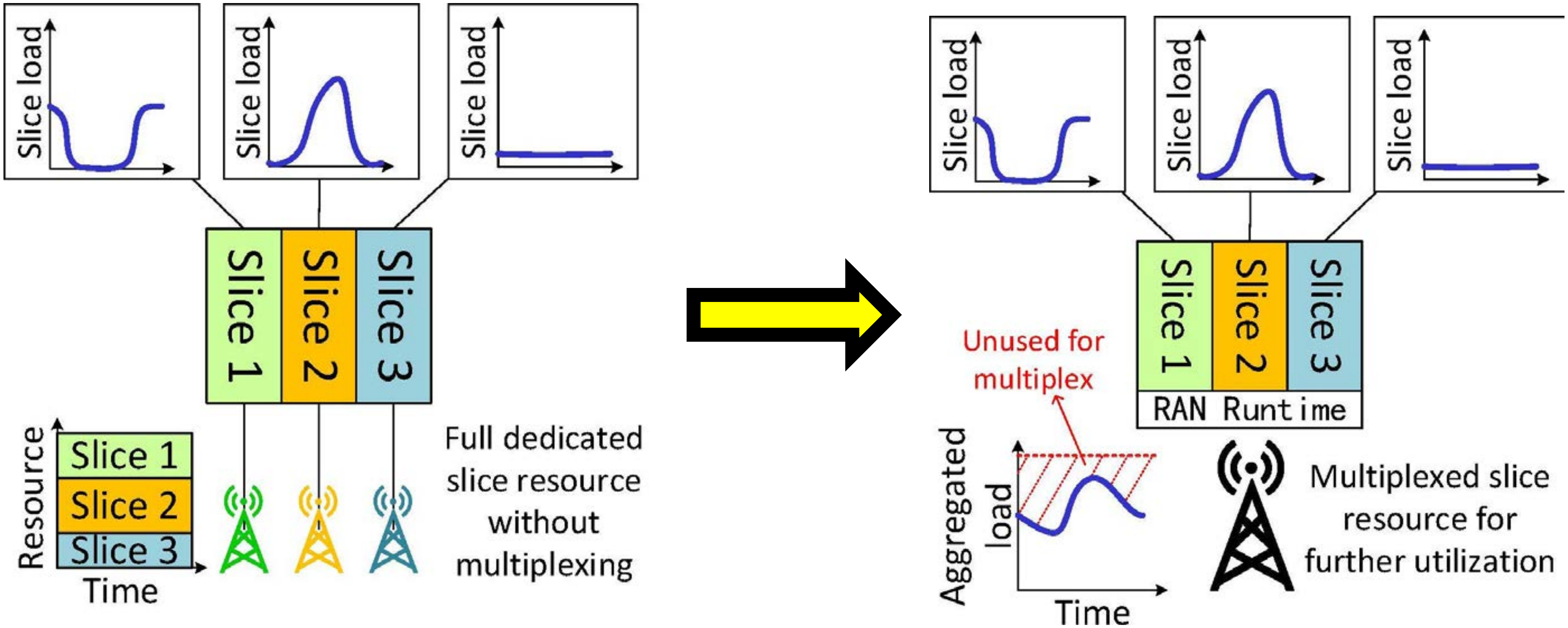
(1) run multiple virtualized RAN module instances with different level of isolation and sharing

(2) Pipeline RAN functions to either via multiplexed or customized CP/UP functions

(3) Share radio resources in virtualized or physical form



Multiplexing Gain



RAN Runtime

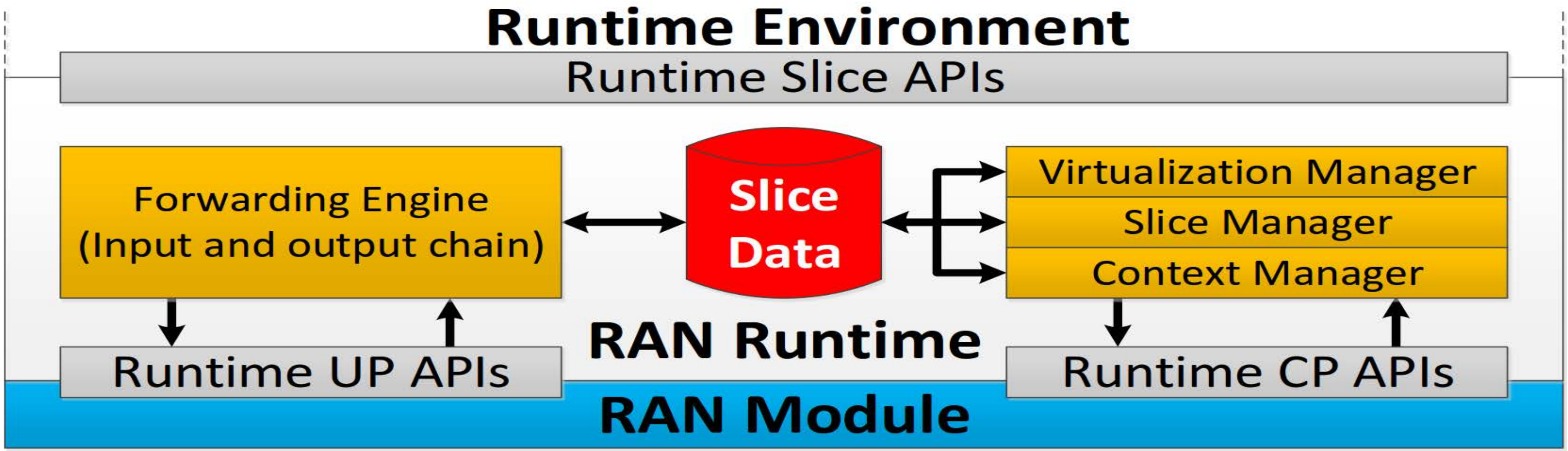
(1) Slice data: Slice context and RAN module context

(2) Context manager: Manage slice data and perform CRUD operation

(3) Slice manager: slice life-cycle, program forwarding engine, conflict resolution

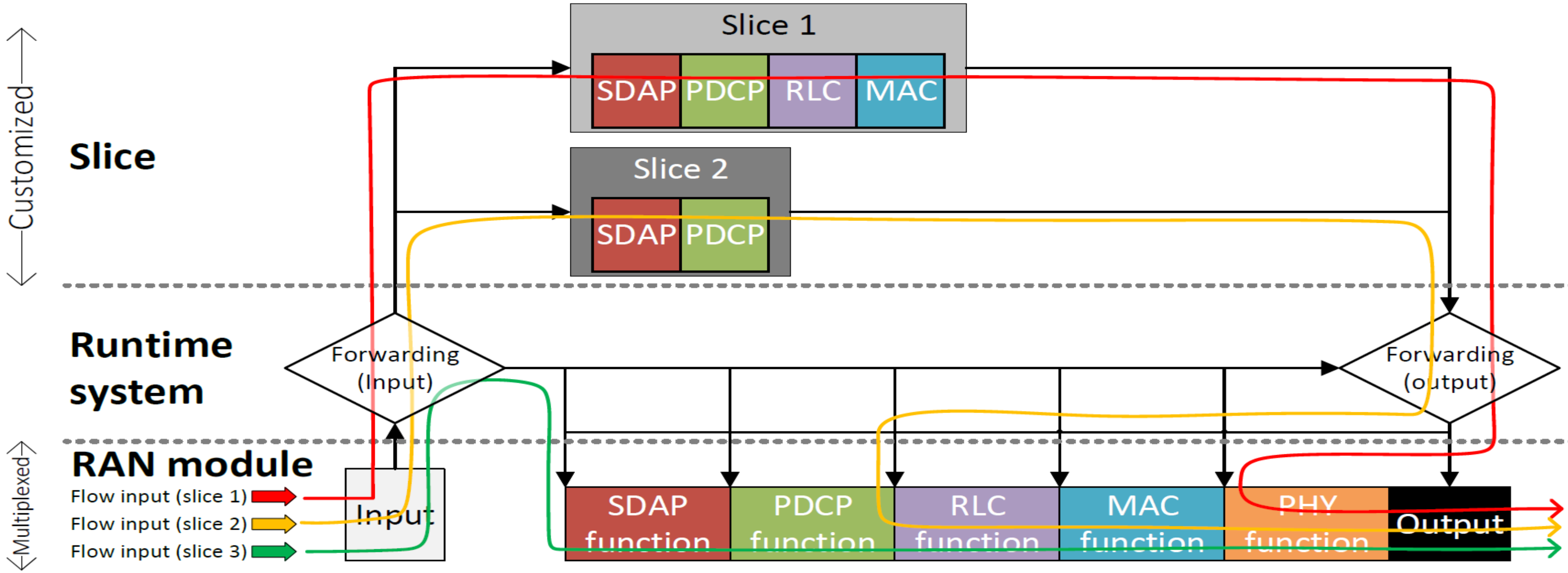
(4) Virtualization manager: resource abstraction, partitioning, and accommodation

(5) Forwarding engine: establish slice-specific UP path

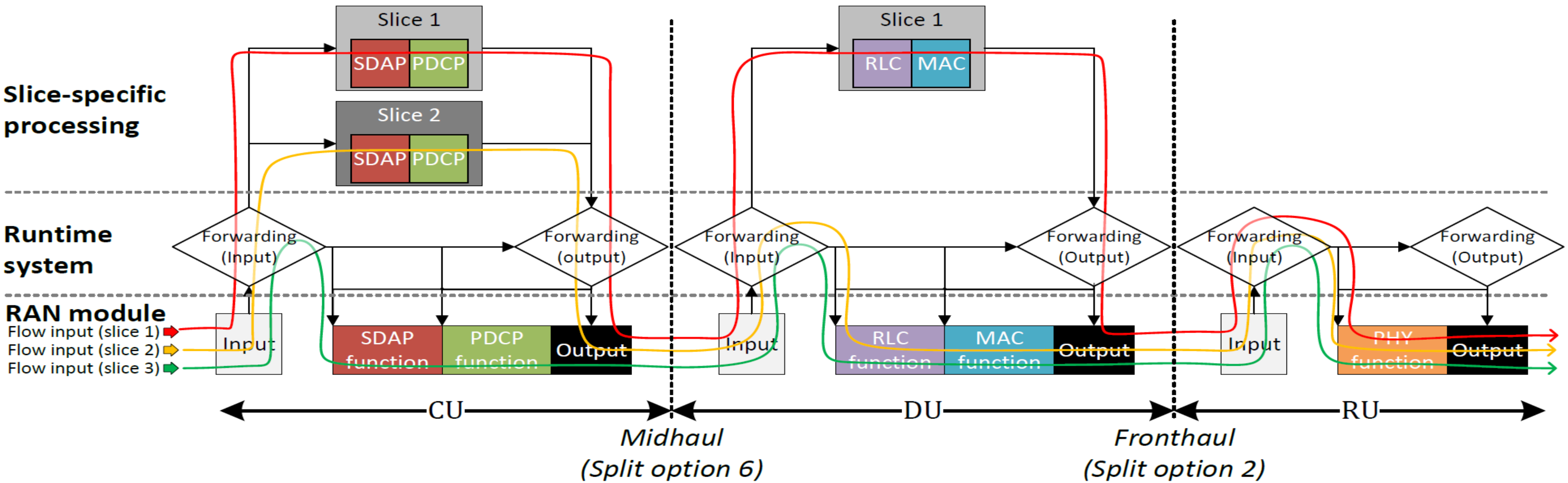


RAN Runtime

Function customization in Monolithic BS



Disaggregated BS



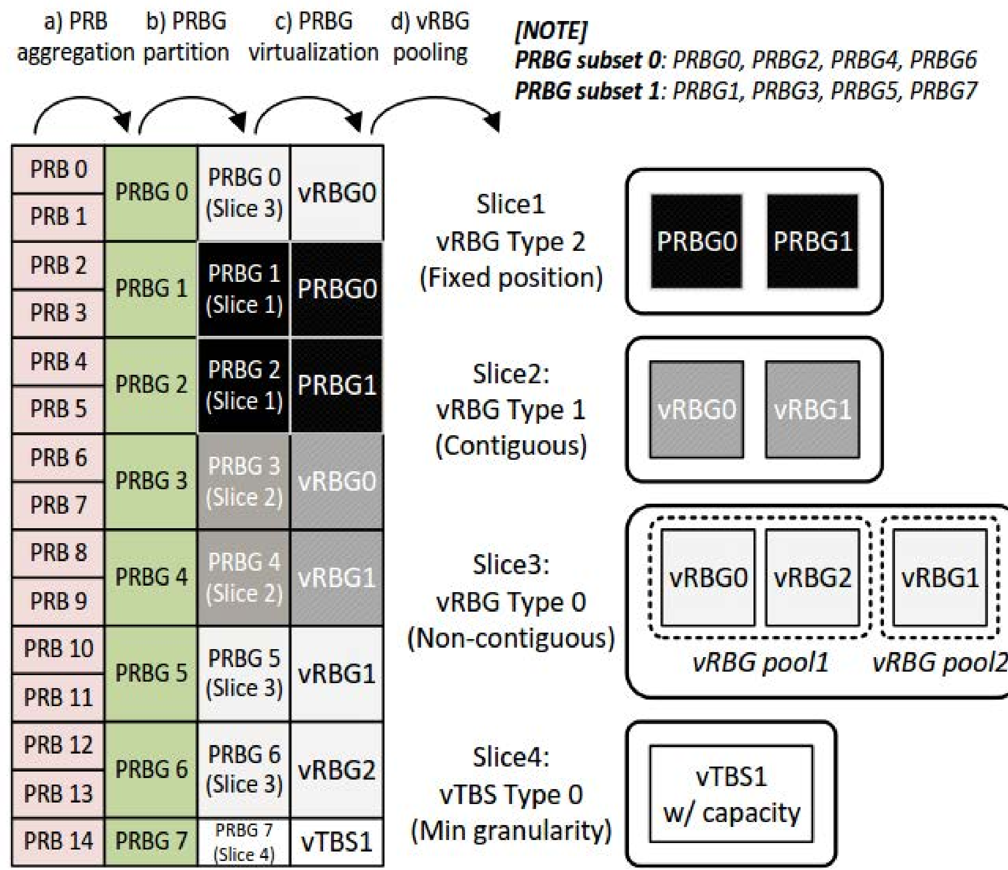
RAN runtime

Resource Abstraction

Requested resources	Abstraction types (Resource granularity)	DL resource allocation type	UL resource allocation type
Resource Block	vRBG Type 0 (Non-contiguous)	Type 0, Type 1, Type 2 distributed	Type 1
	vRBG Type 1 (Contiguous)	Type 0, Type 2 localized	Type 0
	vRBG Type 2 (Fixed position)	Type 2 localized	Type 0
Capacity	vTBS Type 0 (Min RBG granularity)	All Types	All Types

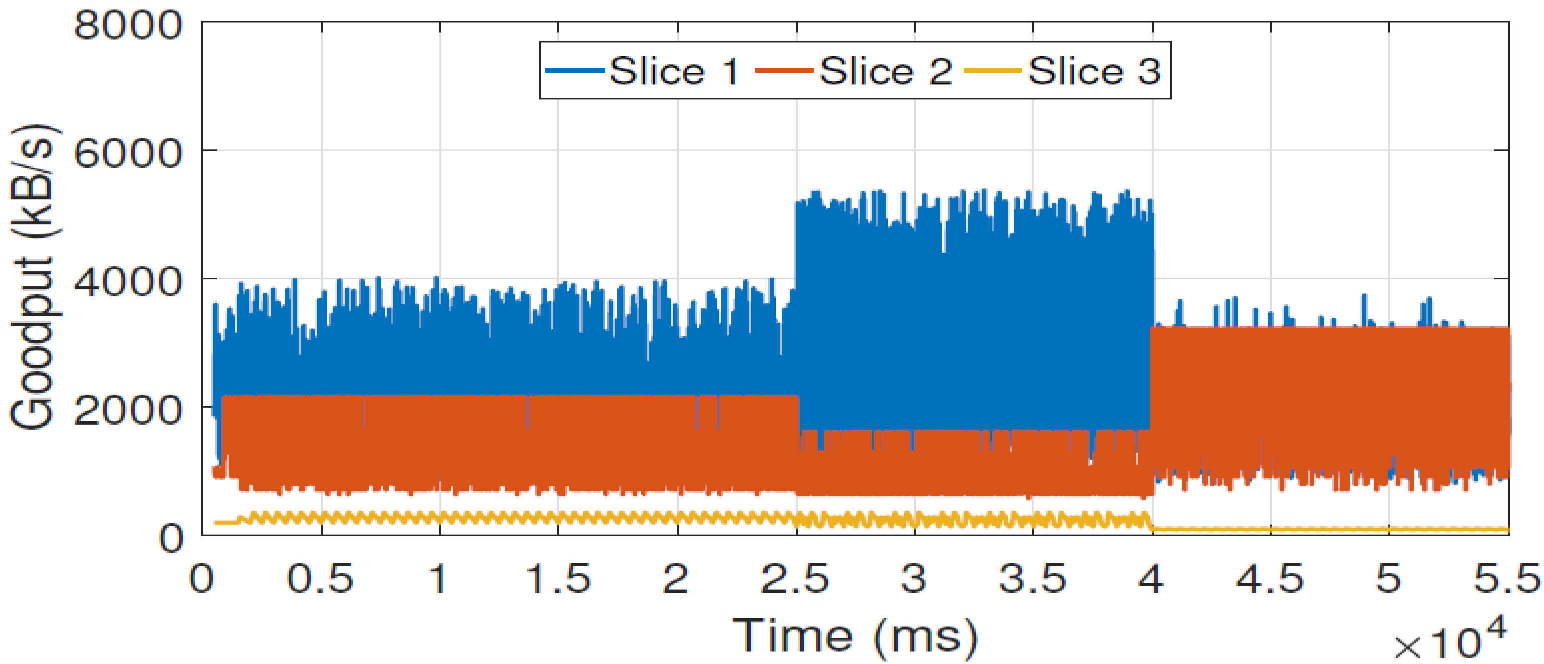
4 Steps to radio resources abstraction:

(1) Aggregation
(2) Partitioning
(3) Virtualization
(4) Polling
(5) Slice resource allocation
(6) Slice Scheduling & Accommodation
(7) Multiplexing/preemption



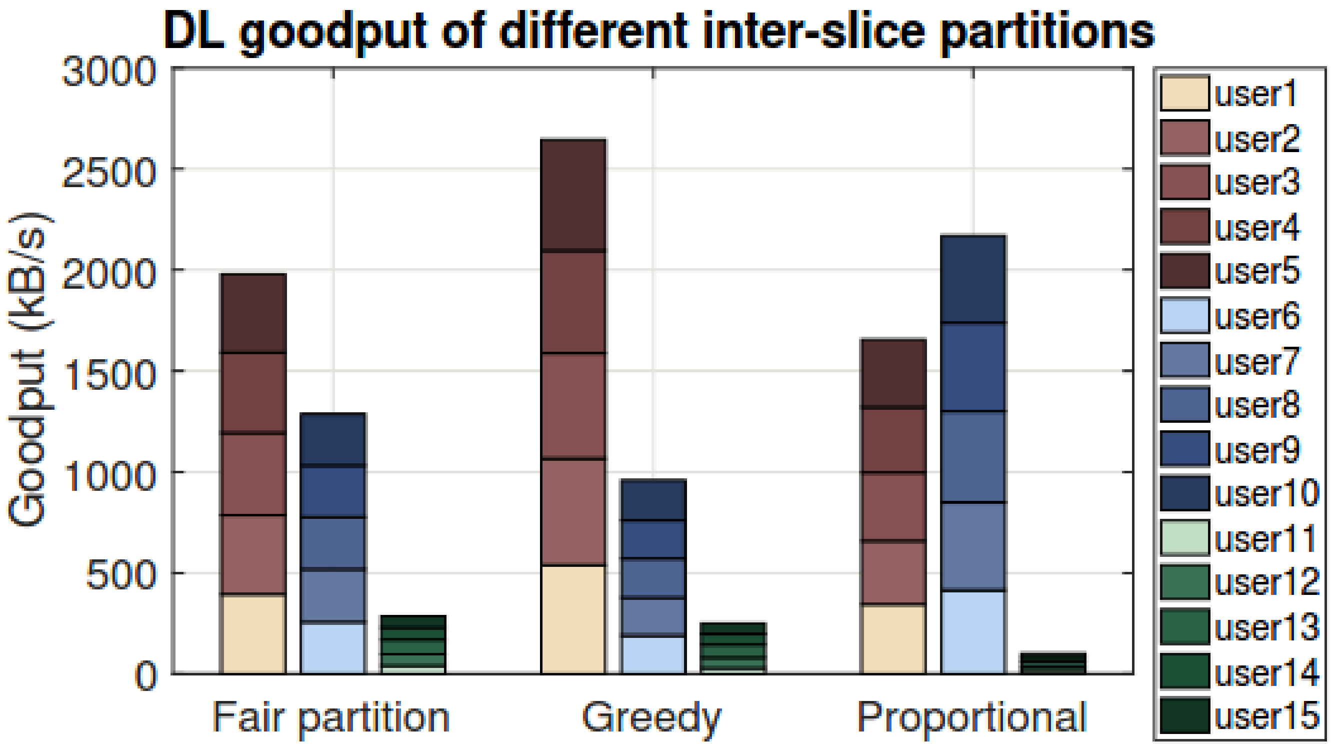
RAN Runtime

Inter-Slice Resource Partitioning and Polling

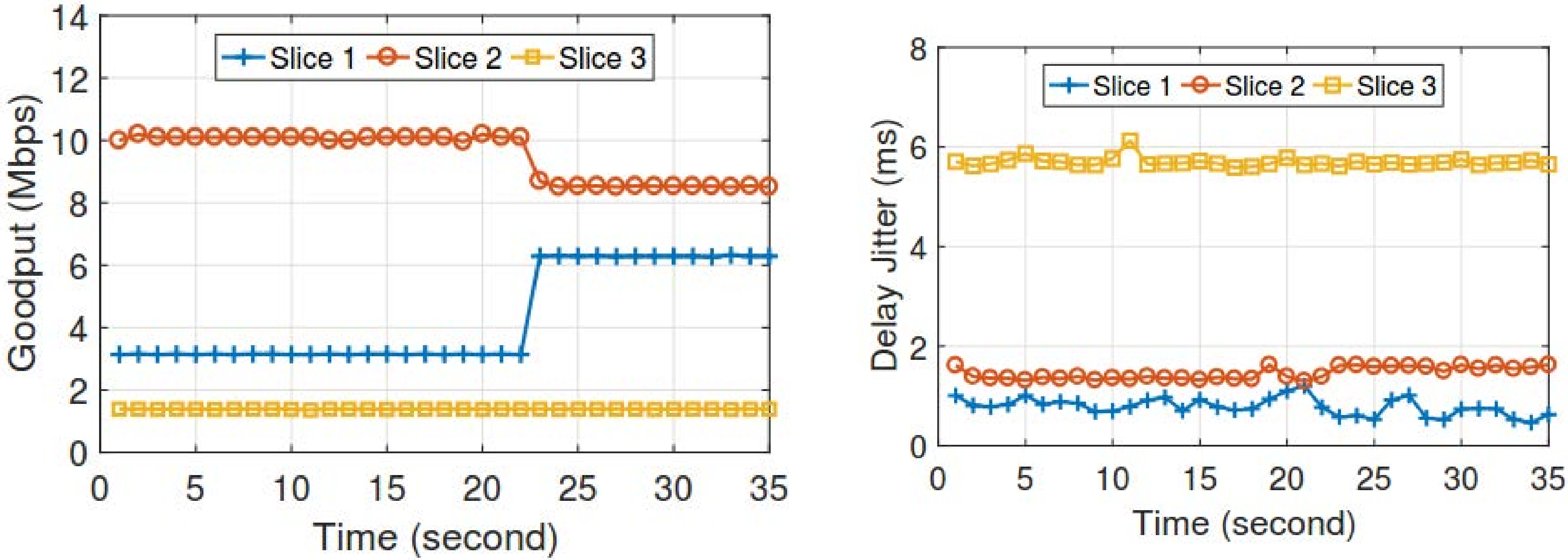


RAN Runtime

Decouple resource partitioning and accommodation from resource allocation

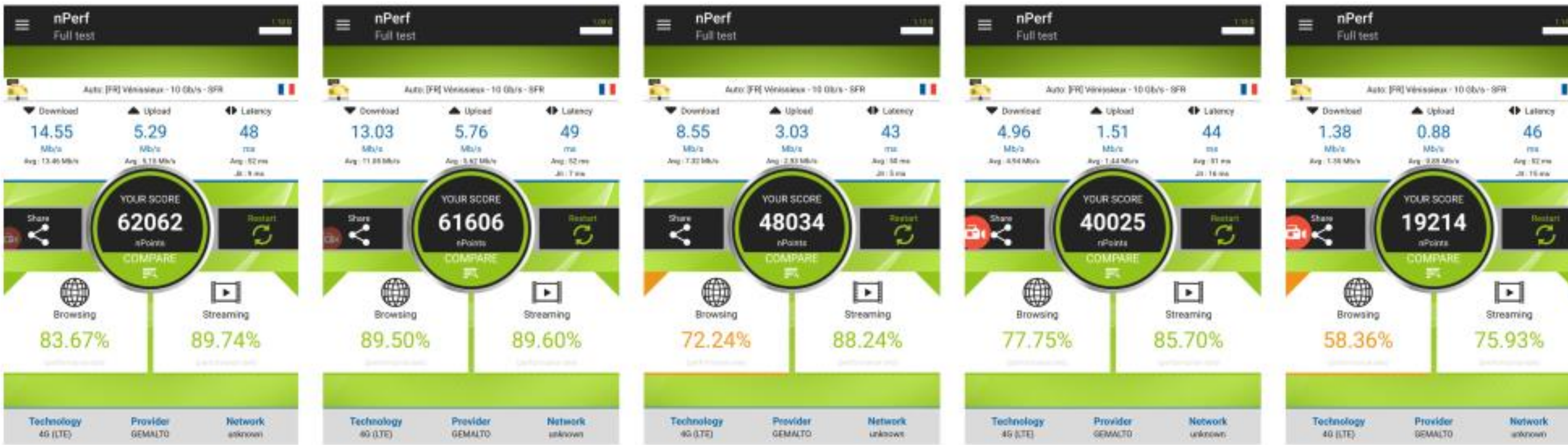


Slice QoS: Multiplexing/Preemption



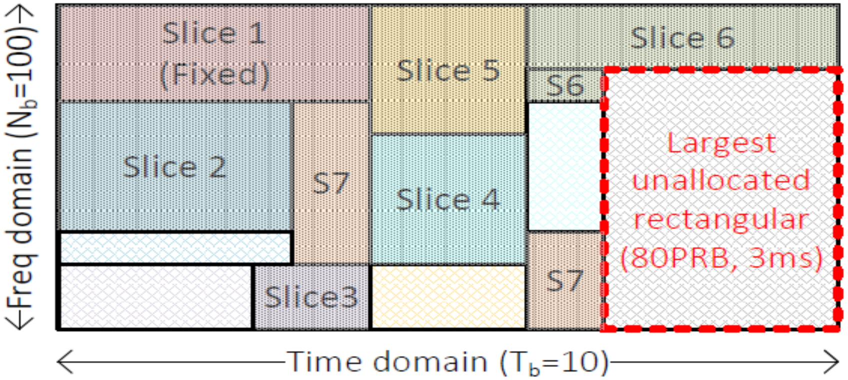
RAN Runtime

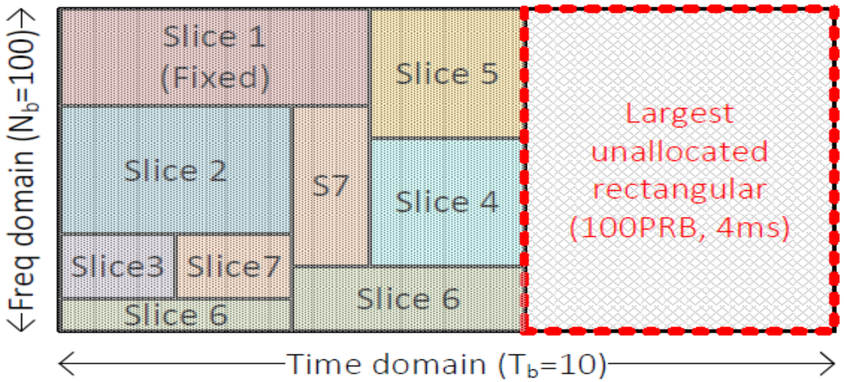
Slice programmability: Service differentiation via RRM policy enforcement

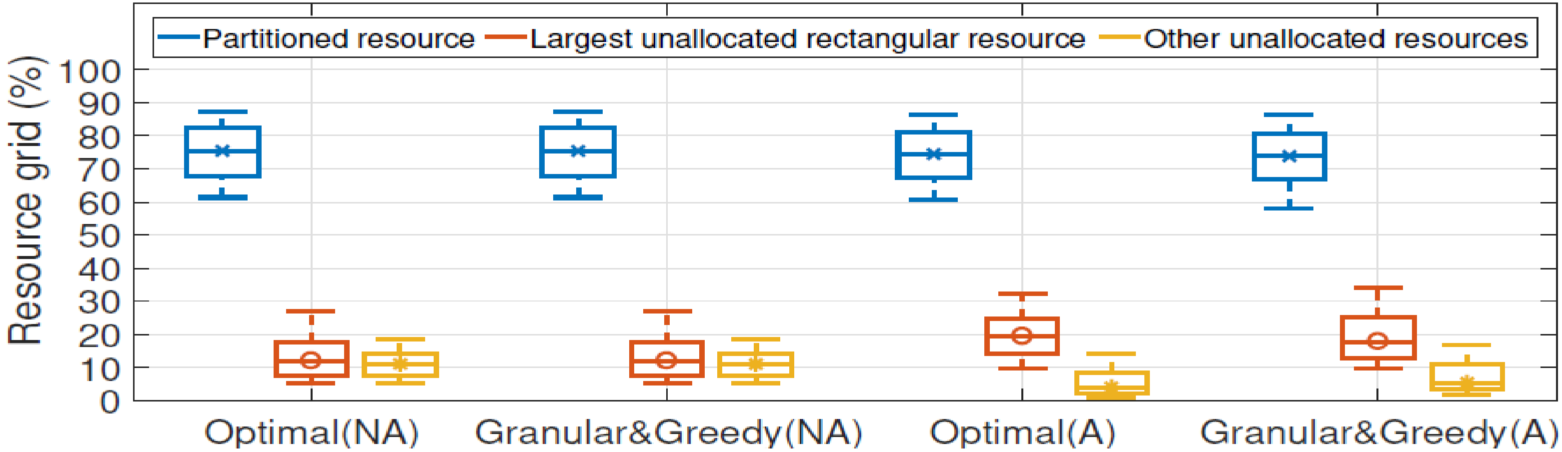


Super user | Platinum user | Gold user | Silver user | Bronze user

RAN Runtime

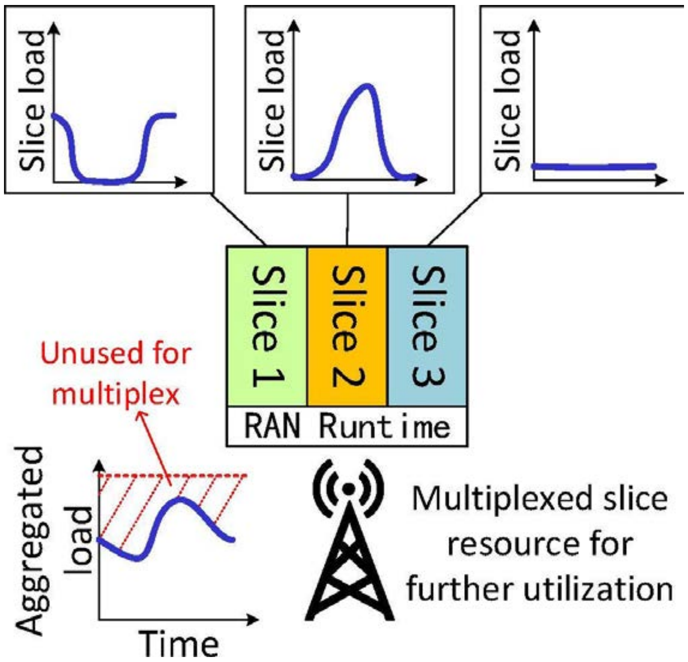

Slice 1
(Fixed)
Slice 5
Slice 6
S6
Slice 2
S7
Slice 4
Largest
unallocated
rectangular
(80PRB, 3ms)
Slice3
S7
←Freq domain (N_b=100)→
←Time domain (T_b=10)→

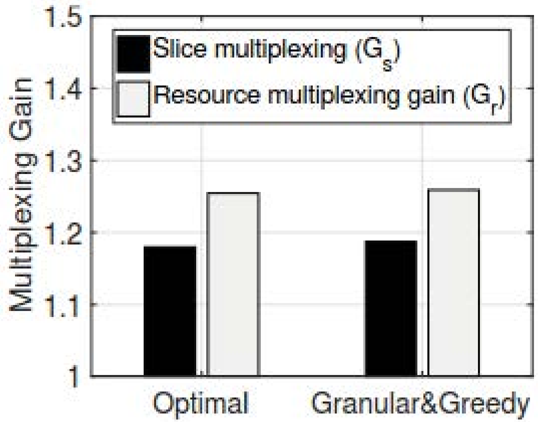

Slice 1
(Fixed)
Slice 5
Largest
unallocated
rectangular
(100PRB, 4ms)
Slice 2
S7
Slice 4
Slice3
Slice7
Slice 6
Slice 6
←Freq domain (N_b=100)→
←Time domain (T_b=10)→


Partitioned resource
Largest unallocated rectangular resource
Other unallocated resources
Resource grid (%)
100
90
80
70
60
50
40
30
20
10
0
Optimal(NA)
Granular&Greedy(NA)
Optimal(A)
Granular&Greedy(A)

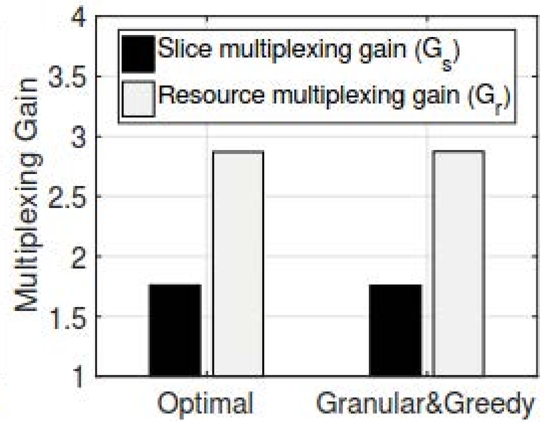
RAN Runtime

Multiplexing Gain





High traffic arrival rate



Low traffic arrival rate

RAN Runtime

Maximize the multiplexing gain

Isolate tenants resources

Customize tenant service

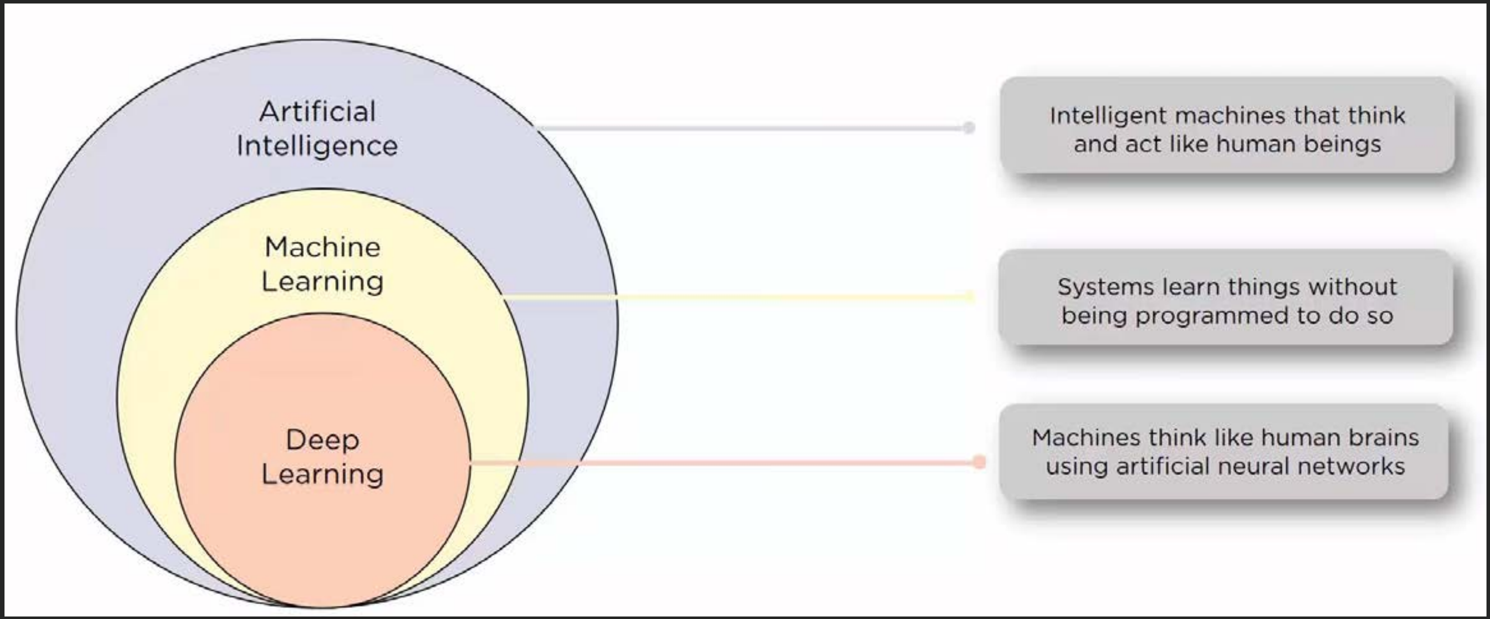
Benefit of Slicing

What is the typical number of slices?

What is the typical lifetime of a slice?

Two numbers in Slicing

Data-driven Network Control



RAN Slicing brings network flexibility and resource elasticity

(1) Openup the interfaces with the help of SDN

(2) Customized control Apps for monitoring, reconfigurability, and programmability

But, modern networks are too complex to be controlled and optimized by means of rule-based Alg.

Why do we need to evolve 5G?

Flexibility to generalize and comprehend: Never seen Z before, but it is similar to X, so do Y, but adjust as needed

Scale to automate control and management to meet the required QoS/QoE

Dynamicity to constantly adapt and anticipate for different workloads and use cases

Abstraction and multi-layering to combine sources with different semantics

Why do we need to evolve 5G?

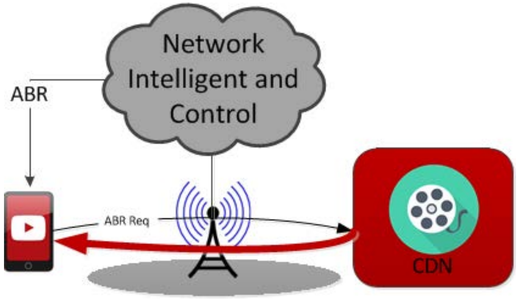
Pipelining by means of "Inference-Prediction-Control"

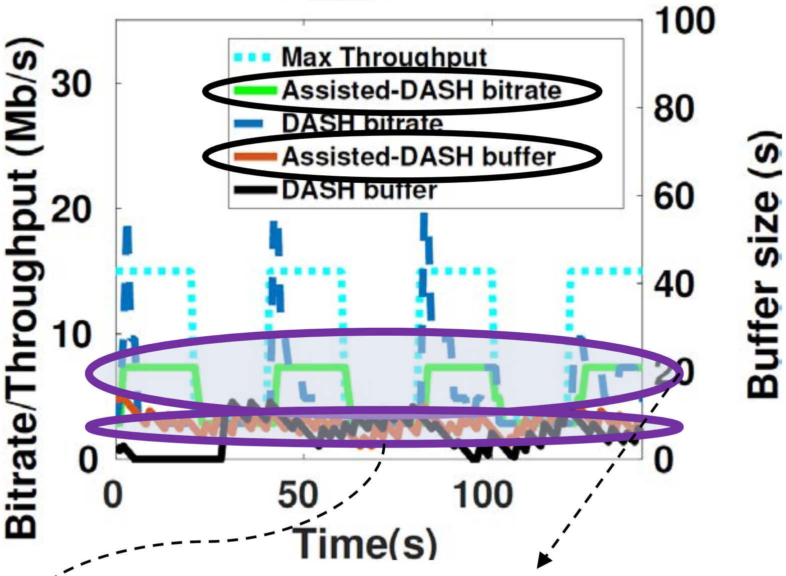
ML models to manage network and resources

(a) Comply better with slice SLAs

(b) Maximizes the revenue of physical network operators

(c) Robust against runtime issues





Smoother bitrate adjustment

No buffer freezes

Objective:

maximize video quality

minimize stall time

Policy:

maintain SLA (e.g. minimum average throughput)

Data:

Link quality

sustainable TCP throughput

Control:

Adapt the video bit rate

Use-Case: Video Streaming

Update the cell capacity to meet the workload demand

Offload users to less congested BS to balance the cell load while maintaining the QoS

Shutdown BS and handover users to the neighboring BS while maintaining the QoS

Other Use-case

ML Applications

- Prediction (per slice)
  - Focus: **predict user behavior**, time-based
  - Predict a time-based signal
  - Eg: users/cell, user BW, end-point location
  - To do: match to specific vertical UCs
- Dimension reduction (per NW / provider)
  - Focus: **optimize monitoring resources**, class-based
  - Select minimal metrics required to maintain SLA
  - Eg: remove dependent metrics, aggregate metrics
  - To do: define provider agnostic semantics
- Bayesian / correlation (per slice type)
  - Focus: **fault management**, self-healing, event-based
  - Correlate reported metrics to fault probability
  - Eg: health score, proactive failover, root cause analysis
  - To do: define specific faults
- Clustering (per slice type)
  - Focus: **identify context**, event-based, unsupervised
  - Find users with similar attributes (to each other)
  - For eg: prioritization, impact analysis, push down to NW
  - To do: define slice context semantics
- Classification (per slice type)
  - Focus: **QoE classification**, event-based, supervised
  - Given reported metrics output QoE class
  - Eg: find all QoS "combinations" that result in similar QoE
  - To do: match QoE to vertical UC perspective
- Anomaly detection (per NW)
  - Focus: **self-protection, self-healing**, event-based
  - Given a signal identify abnormal changes, unsupervised
  - Eg: IDS, detect failed components
  - To do: real-time

Need to predict the user and network performances in time and space with many unknown and/or dynamic variables

(1) Realtime control and coordination across cells

(2) Network Intelligence

Data-driven Control

MONITOR

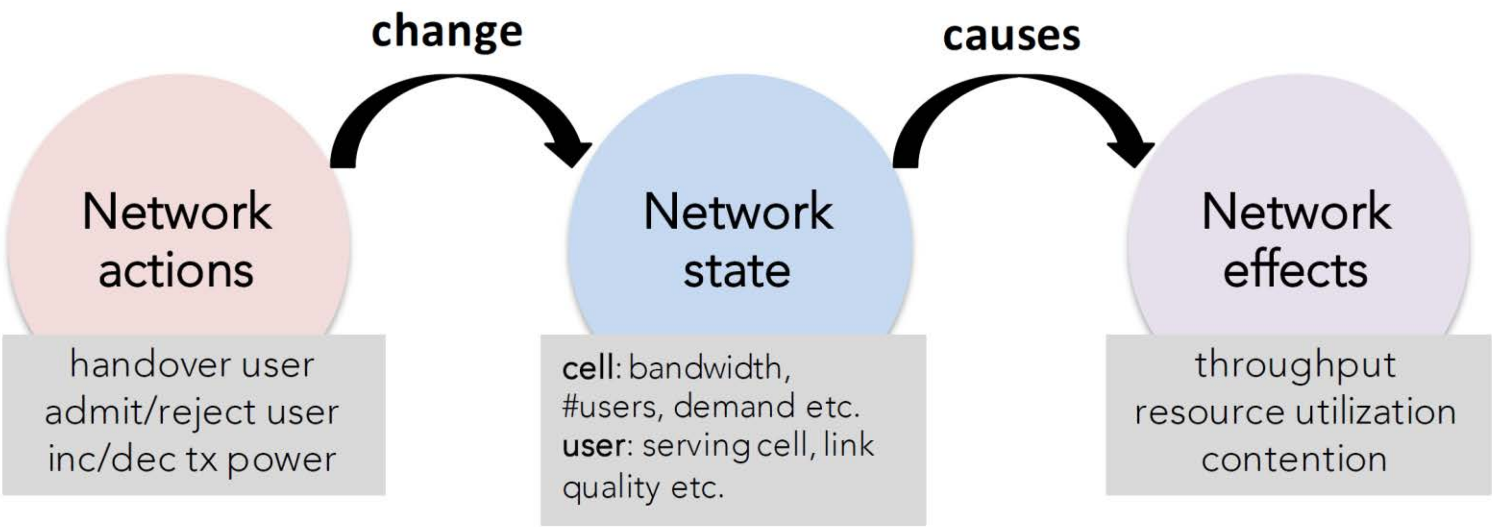
Right now, what is the
avg throughput of user/cell?
resource utilization of user/cell?
contention faced by user/cell?

FORECAST

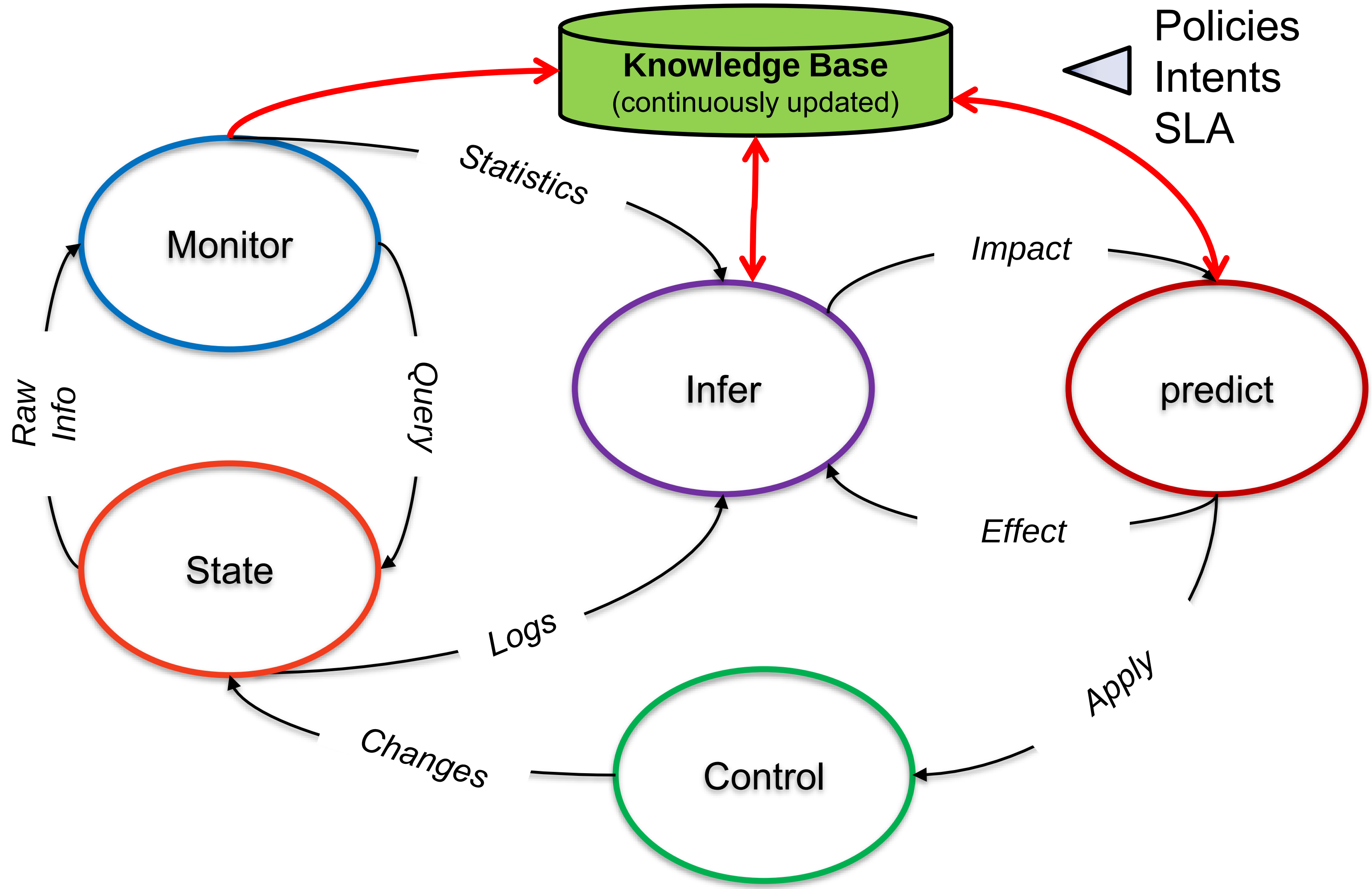
In the next 1s, what will be the:
avg throughput of user/cell?
resource utilization of user/cell?
contention faced by user/cell?

PREDICT IMPACT

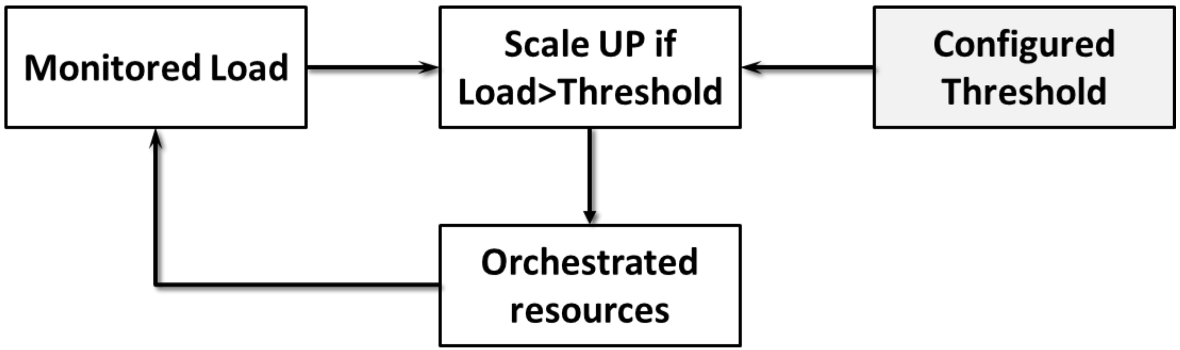
In the next 1s, what if:
handover users?
admit/reject new users?
increase/decrease tx power?



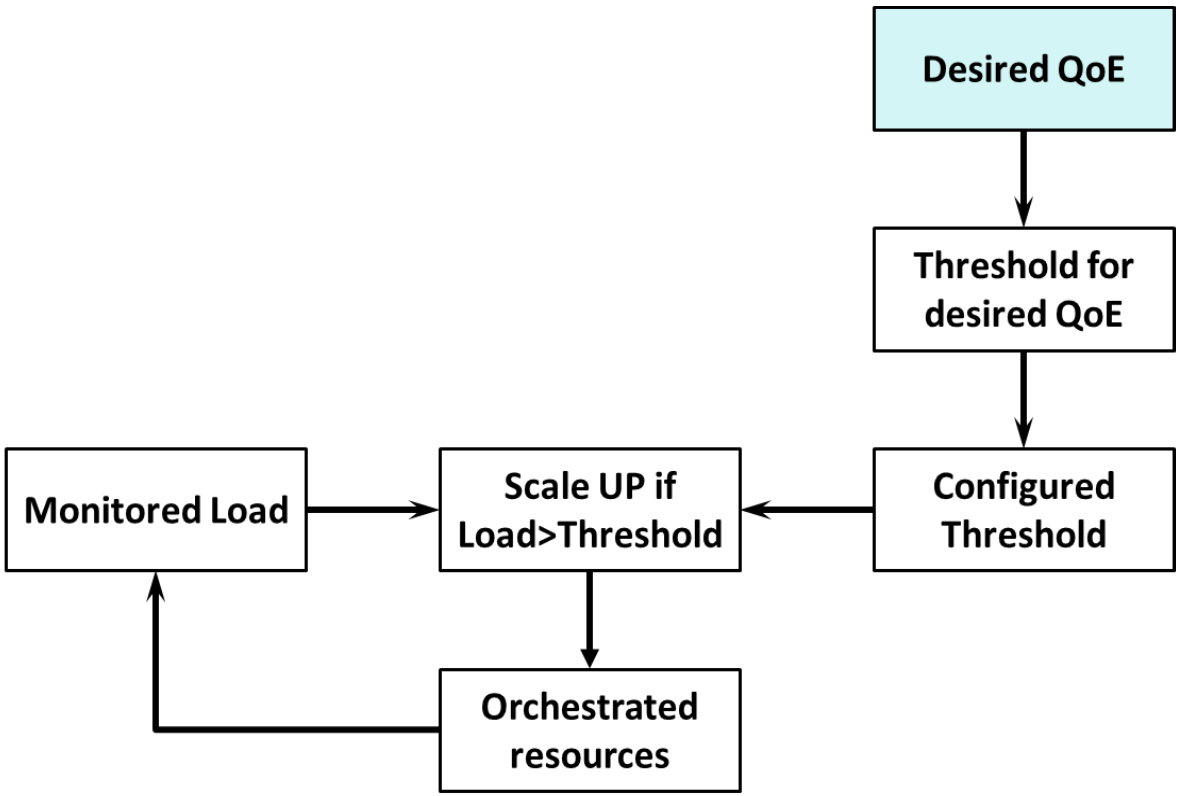
Data-driven Control



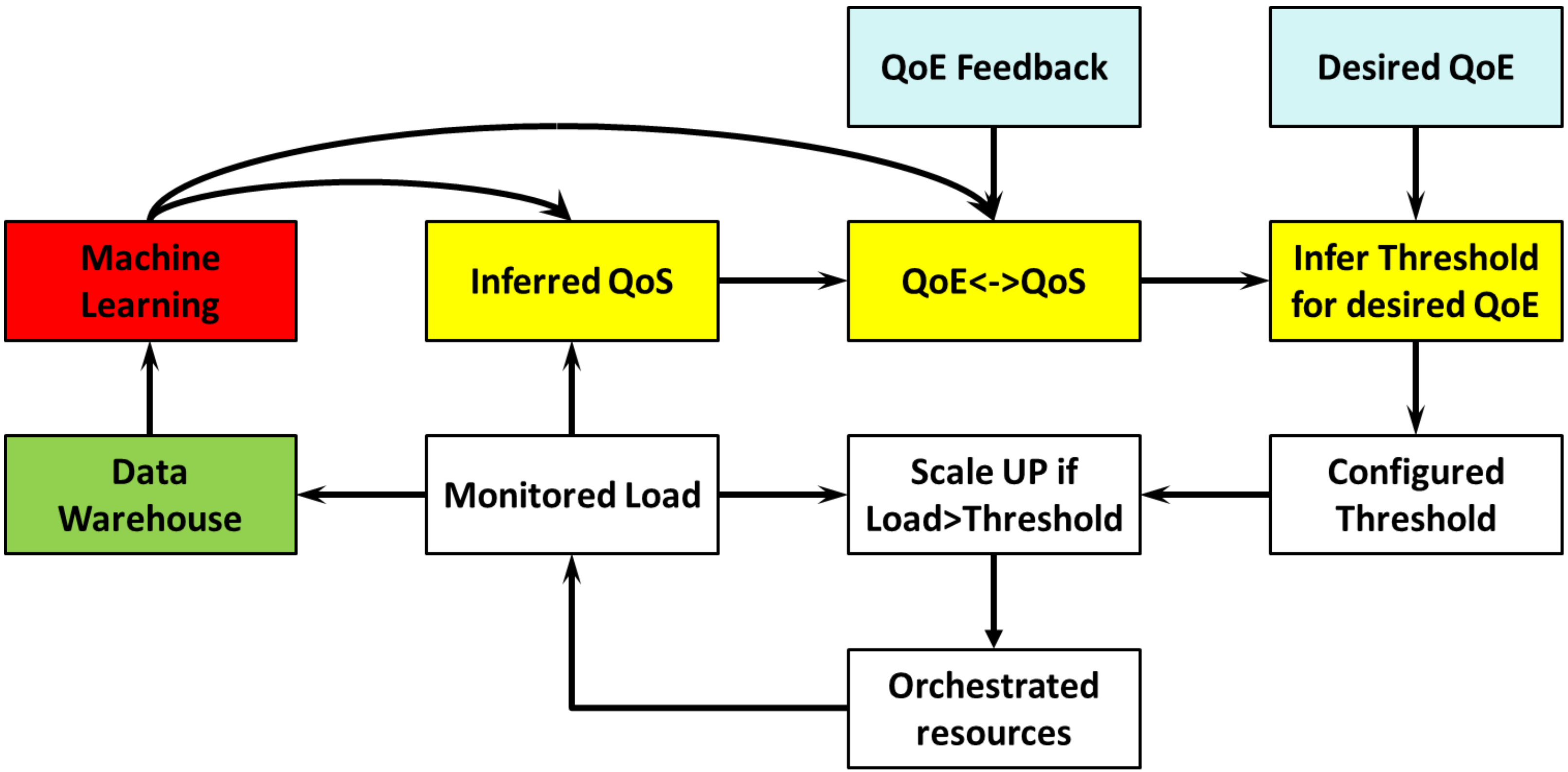
Proactive Control Scheme



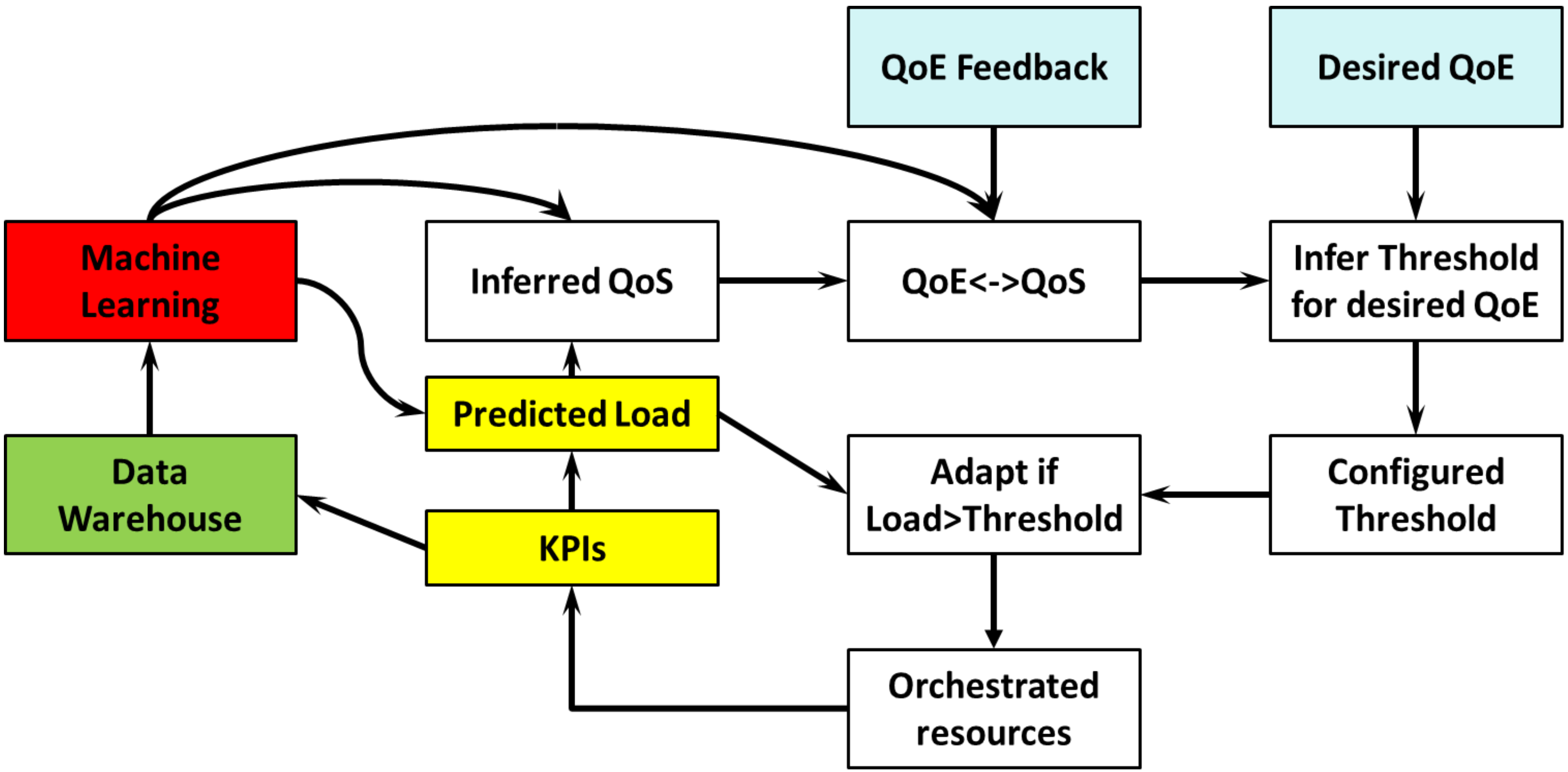
Example of QoE With PCS



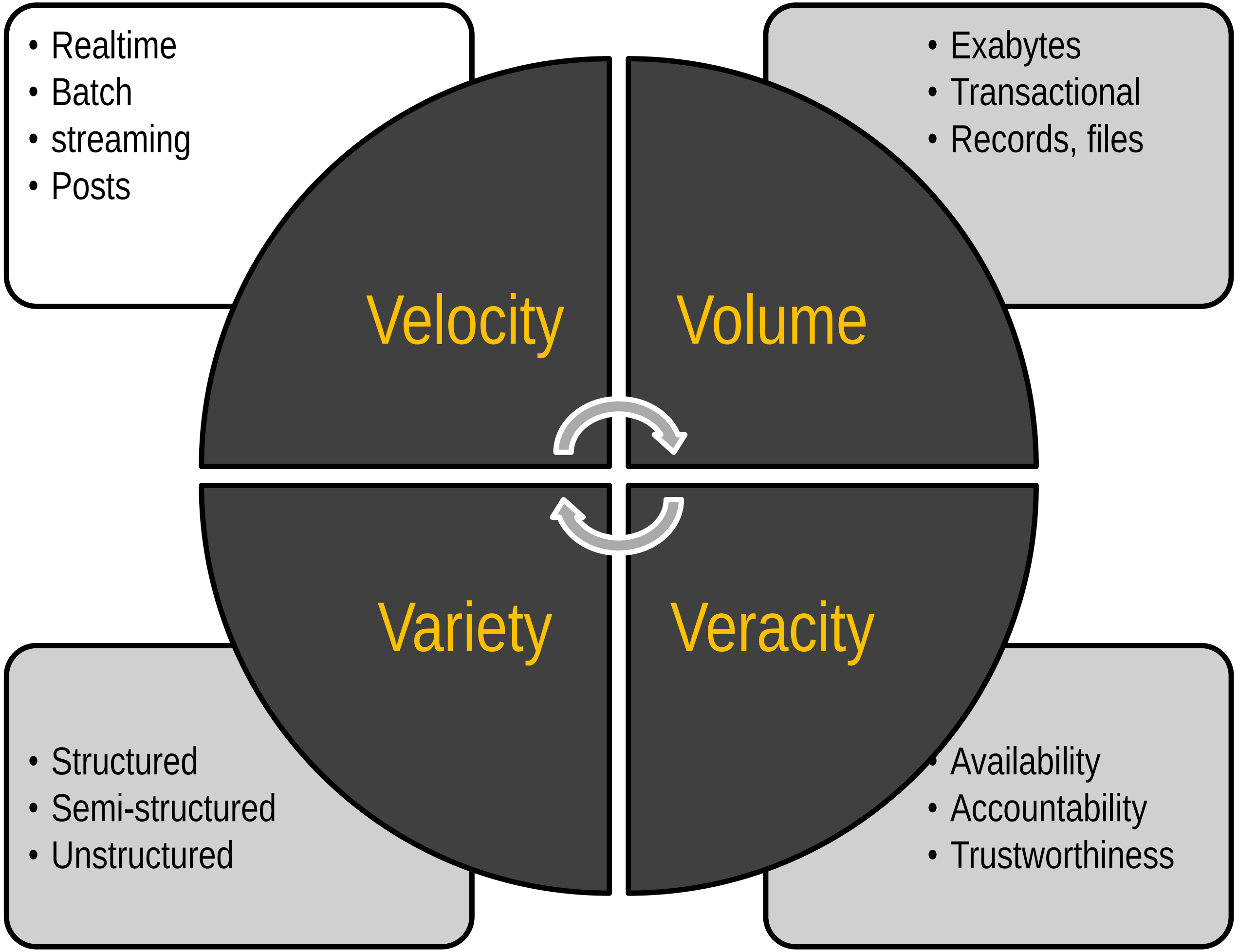
Example of QoE With PCS



Example of QoE With PCS



Example of QoE With PCS



Characterizing the Data

Indexing updates the typed document to make it searchable to suit the particular use case

Smart indexing prevents unnecessary resource usage and speed up the search procedure

Name	Health	Status	Primaries	Replicas	Docs count	Storage size
enb_config-2018-07-31	yellow	open	5	1	8418	3.7mb
mac_stats-2018-07-28	yellow	open	5	1	559236	30.2mb
enb_config-2018-07-23	yellow	open	5	1	7643	2.2mb
mac_stats-2018-07-25	yellow	open	5	1	1174152	118.1mb

Data Indexing

Meta-data to increase the hit rate

ID

Timestamp

Domain (Network or admin)

Source

Access Control List

Measurement type (config, stats, events)

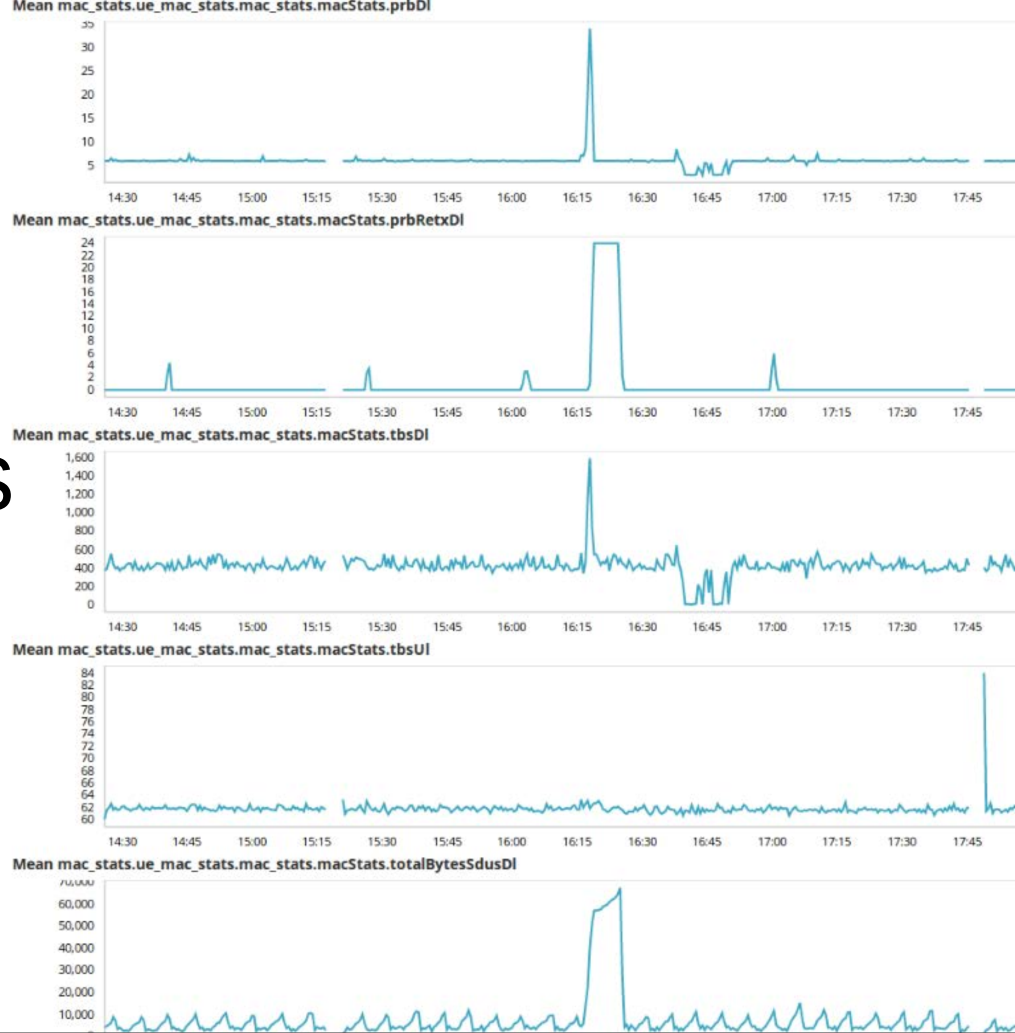
Combine features to create complex queries in time and space

```
"query": {
  "bool" : {
    "must" : {
      "term" : { "user" : "kimchy" }
    },
    "filter": {
      "term" : { "tag" : "tech" }
    },
    "must_not" : {
      "range" : {
        "age" : { "gte" : 10, "lte" : 20 }
      }
    },
    "should" : [
      { "term" : { "tag" : "wow" } },
      { "term" : { "tag" : "elasticsearch" } }
    ],
```

Efficient Search

Inference: how the current and past network states (CQI) affect the service KPIs (throughput)?

→ Pattern and Anomaly detection

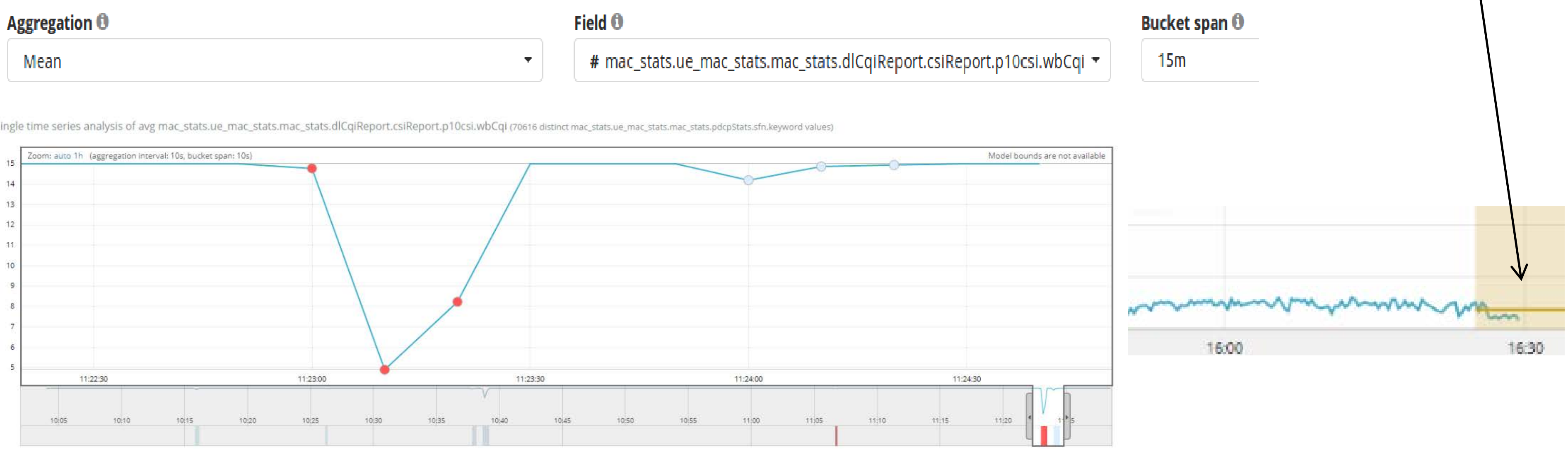


Proactive Control Scheme

Prediction: How the forecasted network states influence the future service KPI?

ML can find hidden patterns, detect anomalies, show forecast.

Note: A lot of data is needed to have a good model.



Proactive Control Scheme

Control: Given operators policies, SLA, client and app state, and the predicted KPI, what actions shall be enforced?

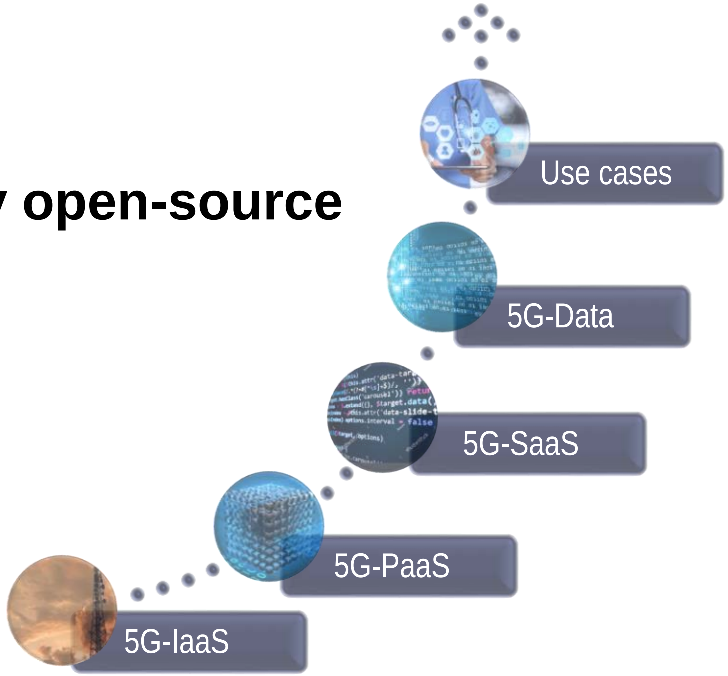
Example: handover user, change RRM policy, increase/decrease Tx power and/or BW and Frequency.

Proactive Control Scheme

OpenSource Platforms

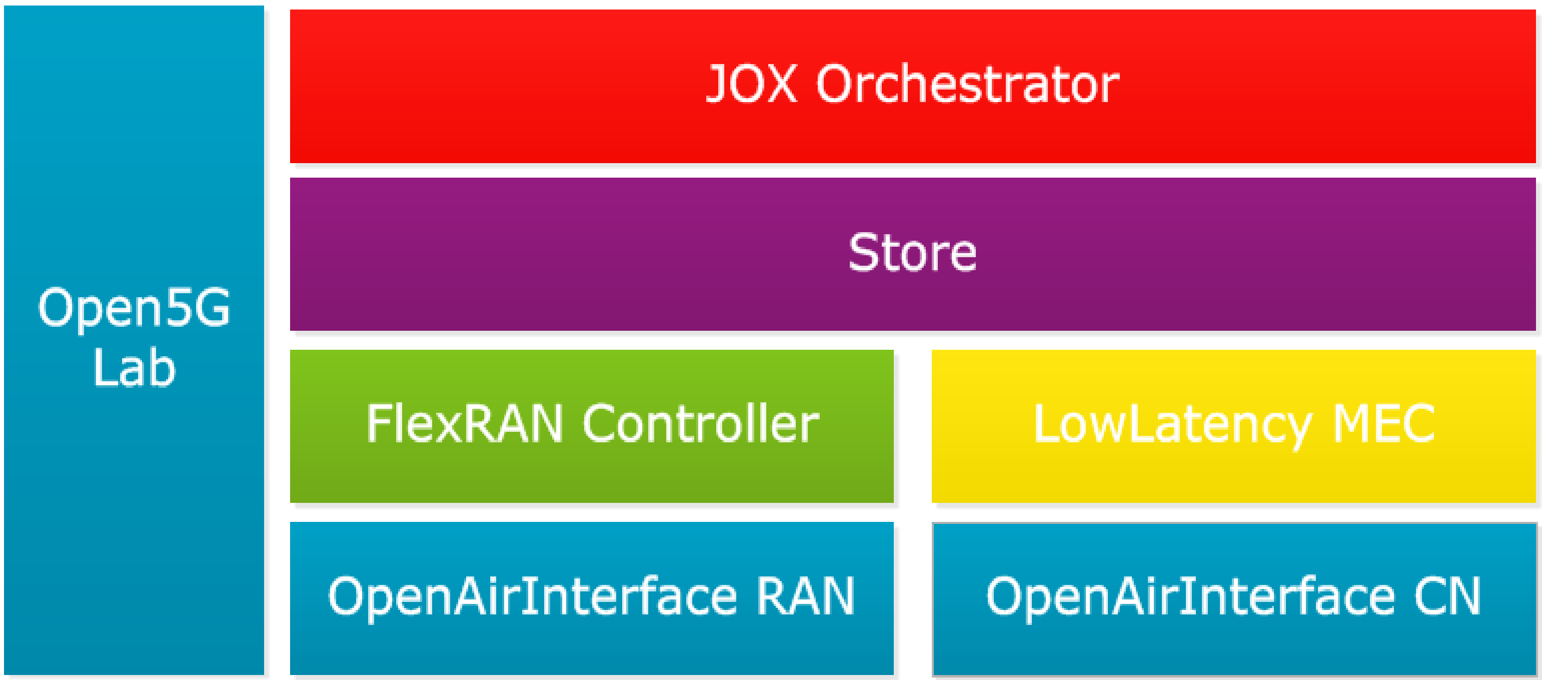
Need for agile network service delivery platforms and use-cases for 4G-5G R&D

5G Innovations empowered by open-source



Opensource Platforms

Agile network service delivery platforms



MOSAIC 5G

OPEN AIR INTERFACE

Mosaic-5G.io Ecosystem

MWC 2016, 2017

ITU, FG-13, 2016, 2017

ETSI 2016, 2017

EUCNS 2015, 2016, 2017

OPNFV 2016

Mobicom 2014,2016,2017

Success Stories

Conclusion

Fusion of Computing, Information and Cellular technologies

(a) 5G and beyond is not only New Radio and verticals, it is also an **evolution in General-Purpose computing for wireless networks**

(b) More and more software technologies (NFV,SDN,MEC) and Data (mining, analytics) jointly with radio signal processing

---

Conclusion

RAN slicing is an on-going research with several challenges Isolation, Sharing, Customization

Satisfy requirements from both slice owner and operator

Two main solutions: ORION and RAN runtime slicing systems

Conclusion

Data-driven network control is difficult

Reason-Predict-Control is a generic framework

Prediction performance is limited by the available computing resources

Conclusion

Can we predict user QoS/QoE per application in realtime?

Can we learn network-user-application dependencies across various network domains?

Can we automatically learn the right control to apply?

Questions


Q
A

Personal Info:

Email: navid.nikaein@eurecom.fr

Website: http://www.eurecom.fr/~nikaeinn/

Linkedin: https://www.linkedin.com/in/navidnikaein

Tel: +33.(0)4.93.00.82.11

Mosaic-5G.io :

Mail : contact@mosaic-5g.io

Website : http://mosaic-5g.io

Linkedin: https://www.linkedin.com/in/mosaic-5g

Twitter: @mosaic5g

Contact Information