T01: RAN Slicing Challenges, Technologies, and Tools

Navid Nikaein
Associate Professor - Eurecom
Coordinator of Mosaic-5G Initiative

Tutorial ACM SIGCOMM
20-25 August 2018, Budapest, Hungary
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
Connected, Controlled, and Flexible Digital Society

Value Creation

Consistent experience

Sustainable business model

What is 5G?
Many 5G Use-cases
Internet of Skills

Source: Dohler
E-Health: When Robotics meets 5G
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

© 3GPP

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

© Ericsson WP
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**
Network Slicing
Flexible & Customizable
for each use-cases

Service-oriented 5G
Slicing Technology Enablers:

- Softwarization
- Virtualization
- Disaggregation

Multi-service multi-tenant network
Slicing Technology Enablers

- Application
  - RAN Control & Data Plane separation
  - Abstraction & Virtualized Control Functions
  - RESTful API
  - RAN Optimization
  - Network slicing
  - IoT Gateway
  - Content caching

- MEC
  - LTE eNB
  - OpenFlow switch
  - Ethernet switch

- SDN

- Network

Open Data APIs

- App SDK
- Control APPs
- Platform SDK
- Control plane Services
- Data-Plane Service
- Slice / App Orchestration
- Platform Orchestration
Slicing Technology Enablers

Disaggregation

Radio Resources Services

BS1  BS2  BS3

40% of resources are not used

Multiplexing gain

Physical BS

vBS1  vBS2  vBS3
Why will it happen?

Extreme network flexibility and elasticity
Network Slicing Evolves the value-chain of telecom industry

Service-oriented 5G
3GPP Role Model (3GPPPP 28.801)

E.g.: End user, Small & Medium Enterprise, Large enterprise, Vertical, Other CSP, etc.

Service-oriented 5G
<table>
<thead>
<tr>
<th>Service-oriented 5G</th>
<th>3G</th>
<th>4G</th>
<th>5G</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Downlink waveform</strong></td>
<td>CDMA</td>
<td>OFDM</td>
<td>OFDM, SCFDMA</td>
</tr>
<tr>
<td><strong>Uplink waveform</strong></td>
<td>CDMA</td>
<td>SCFDMA</td>
<td>OFDMA, SCFDMA</td>
</tr>
<tr>
<td><strong>Channel coding</strong></td>
<td>Turbo</td>
<td>Turbo</td>
<td>LDPC (data) / Polar (L1 contr.)</td>
</tr>
<tr>
<td><strong>Beamforming</strong></td>
<td>No</td>
<td>Only data</td>
<td>Full support</td>
</tr>
<tr>
<td><strong>Spectrum</strong></td>
<td>0.8 – 2.1 GHz</td>
<td>0.4 – 6 GHz</td>
<td>0.4 – 90 GHz</td>
</tr>
<tr>
<td><strong>Bandwidth</strong></td>
<td>5 MHz</td>
<td>1.4 – 20 MHz</td>
<td>Up to 100 MHz (400MHz for &gt;6GHz)</td>
</tr>
<tr>
<td><strong>Network slicing</strong></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>QoS</strong></td>
<td>Bearer based</td>
<td>Bearer based</td>
<td>Flow based</td>
</tr>
<tr>
<td><strong>Small packet support</strong></td>
<td>No</td>
<td>No</td>
<td>Connectionless</td>
</tr>
<tr>
<td><strong>In-built cloud support</strong></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
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
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 re-architects mobile networks

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

**Service-oriented 5G**
3GPP re-architects mobile networks
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
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?

- Safety/autonomous driving service
- URLLC (Ultra Reliable Low Latency)
- Infotainment/video streaming
- eMBB (Mobile Broadband)
- Maintenance/statistics
- mIoT, low throughput

- URLLC Slice
- eMBB Slice
- Default Slice
- mIoT Slice

- Dedicated functions for safety/autonomous driving
- Shared functions for infotainment/video streaming, eMBB, maintenance/statistics, mIoT
Dedicated or Shared Resources?
Dedicated or Shared Resources?
RAN Slicing
Composition and deployment of multiple E2E logical networks tailors to a service over a shared infrastructure, and their delivery as a slice.

Slice Components

- Vertical business apps
- Plug&Play control & management
- Custom QoE optimisation
- Chained network functions
- Programmed data plane for QoS/SLA
- Physical & virtual resources

What is a slice?
RAN Slicing

- Efficient and adaptive use of radio resources
- No functional isolation

- Functional isolation
- Inefficient use of radio resources

RAN Sharing (e.g. [NVS – IEEE/ACM TON 2012])

Full Isolation (e.g. [FLARE – JIP 2017])

© M. Marina
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)
RAN slicing system

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

(2) Share radio resources in virtualized or physical form
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

Slice 2 controller

Slice 2 controller
**QoS optimization**

Slice 1 controller
**Load Balancing**

Orion Hypervisor

Orion Hypervisor

Orion Hypervisor

Orion Hypervisor
RAN Slicing Execution Env.

(1) run multiple virtualized RAN module instances with different levels 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

Full dedicated slice resource without multiplexing

Unused for multiplex

Multiplexed slice resource for further utilization
(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
Function customization in Monolithic BS

RAN Runtime
Disaggregated BS

RAN runtime
## Resource Abstraction

<table>
<thead>
<tr>
<th>Requested resources</th>
<th>Abstraction types (Resource granularity)</th>
<th>DL resource allocation type</th>
<th>UL resource allocation type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource Block</td>
<td>vRBG Type 0 (Non-contiguous)</td>
<td>Type 0, Type 1, Type 2 distributed</td>
<td>Type 1</td>
</tr>
<tr>
<td></td>
<td>vRBG Type 1 (Contiguous)</td>
<td>Type 0, Type 2 localized</td>
<td>Type 0</td>
</tr>
<tr>
<td></td>
<td>vRBG Type 2 (Fixed position)</td>
<td>Type 2 localized</td>
<td>Type 0</td>
</tr>
<tr>
<td>Capacity</td>
<td>vTBS Type 0 (Min RBG granularity)</td>
<td>All Types</td>
<td>All Types</td>
</tr>
</tbody>
</table>
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
Inter-Slice Resource Partitioning and Polling

![Graph showing Goodput (kB/s) over time for three slices: Slice 1, Slice 2, and Slice 3. The graph illustrates the resource partitioning and polling in a RAN runtime.](Image)
Decouple resource partitioning and accommodation from resource allocation

RAN Runtime
Slice QoS: Multiplexing/Preemption

RAN Runtime
Slice programmability: Service differentiation via RRM policy enforcement
Multiplexing Gain

RAN Runtime

High traffic arrival rate
Low traffic arrival rate
Maximize the multiplexing gain

Isolate tenants resources

Customize tenant service
Two numbers in Slicing

What is the typical number of slices?

What is the typical lifetime of a slice?
Data-driven Network Control

- **Artificial Intelligence**
- **Machine Learning**
- **Deep Learning**

- Intelligent machines that think and act like human beings
- Systems learn things without being programmed to do so
- Machines think like human brains using artificial neural networks
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 “Reasoning-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

Why Data-driven Network Control?
**Objective:**
- maximize video quality
- minimize stall time
- Maintain service continuity

**Policy:**
- maintain SLA (e.g. minimum average throughput)

**Data:**
- Link quality
- sustainable TCP throughput

**Control (beyond just ABR, joint UE and BS):**
1. Adapt the video bit rate through video optimizer
2.1 Add/provision a new BS through SMA+ORCH
2.2 increase the BW of the current BS (SMA)
3. Interference coordination through RRM
4. Update frequency and power through SMA

---

**Use-Case: Video Streaming**

*No buffer freezes*  
*Smoother bitrate adjustment*
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
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
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?

Network actions
handover user admit/reject user inc/dec tx power

change

Network state
cell: bandwidth, #users, demand etc. user: serving cell, link quality etc.

causes

Network effects
throughput resource utilization contention

Data-driven Control

© S. Katti
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

• Realtime
• Batch
• streaming
• Posts

• Exabytes
• Transactional
• Records, files

• Structured
• Semi-structured
• Unstructured

• Availability
• Accountability
• Trustworthiness

Velocity

Volume

Variety

Veracity

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.

<table>
<thead>
<tr>
<th>Name</th>
<th>Health</th>
<th>Status</th>
<th>Primaries</th>
<th>Replicas</th>
<th>Docs count</th>
<th>Storage size</th>
</tr>
</thead>
<tbody>
<tr>
<td>enb_config-2018-07-31</td>
<td>yellow</td>
<td>open</td>
<td>5</td>
<td>1</td>
<td>8418</td>
<td>3.7mb</td>
</tr>
<tr>
<td>mac_stats-2018-07-28</td>
<td>yellow</td>
<td>open</td>
<td>5</td>
<td>1</td>
<td>559236</td>
<td>30.2mb</td>
</tr>
<tr>
<td>enb_config-2018-07-23</td>
<td>yellow</td>
<td>open</td>
<td>5</td>
<td>1</td>
<td>7643</td>
<td>2.2mb</td>
</tr>
<tr>
<td>mac_stats-2018-07-25</td>
<td>yellow</td>
<td>open</td>
<td>5</td>
<td>1</td>
<td>1174152</td>
<td>118.1mb</td>
</tr>
</tbody>
</table>
Meta-data to increase the hit rate

Combine features to create complex queries in time and space

ID
Timestamp
Domain (Network or admin)
Source (Region, entity)
Access Control List
Measurement type (config, stats, events)

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

→ Pattern and Anomaly detection
**Prediction:** How the forecasted network states influence the future service KPI?

ML can find hidden patterns, detect anomalies, show forecast. Note: large volume of relevant 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.
OpenSource Platforms
Need for agile network service delivery platforms and use-cases for 4G-5G R&D

5G Innovations empowered by open-source
Agile network service delivery platforms

- **LL-MEC**: A Low Latency SDN-based MEC Platform
- **FlexRAN**: A Flexible & Programmable SD-RAN Platform
- **JOX**: An event-driven juju-based service orchestrator core
- **Store**: Network function & application distribution Repository
- **Open5G Lab**: Access to 4G/5G network facilities and perform experiments
- **My Project?**: Create a project and build your use-case

Mosaic-5G.io Ecosystem
Agile network service delivery platforms

Mosaic-5G.io Ecosystem
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
Data-driven network control is difficult

Reason-Predict-Control is a generic framework

Prediction performance is limited by the available computing resources
Why such a big complexity to support slicing?

How the net neutrality principles be retained?
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?
Who owns training data?
Who pays the cost of resources?
How and where is data stored?
How is data processed?
Who manages data? What are the potential implications to security and privacy?
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