"AI based Network Resource Management (1)"

Choong Seon Hong

Department of Computer Science and Engineering
Kyung Hee University, Republic of Korea.

KYUNG HEE UNIVERSITY

NETWORKING INTELLIGENCE LAB

Outline

- Introduction
- Network Slicing: The Concept
- Use Case 1: Virtual Reality
- Use Case 2: Chunk-Based Resource Allocation
- Use Case 3: Energy Efficient Communication and Computation Resource Slicing for eMBB and URLLC
- Concluding Remarks

Introduction

- Diverse Requirements
- Requirements in
  - Manufacturing Industry
  - Transportation Industry
  - Health Sector
- Evolution of Cellular Systems
- Challenges to realize 5G Networks

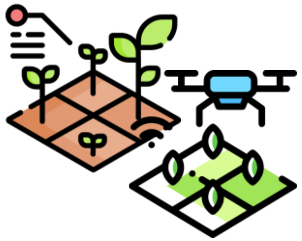
Introduction : Diverse Requirements

Smart Factory

Smart Entertainment

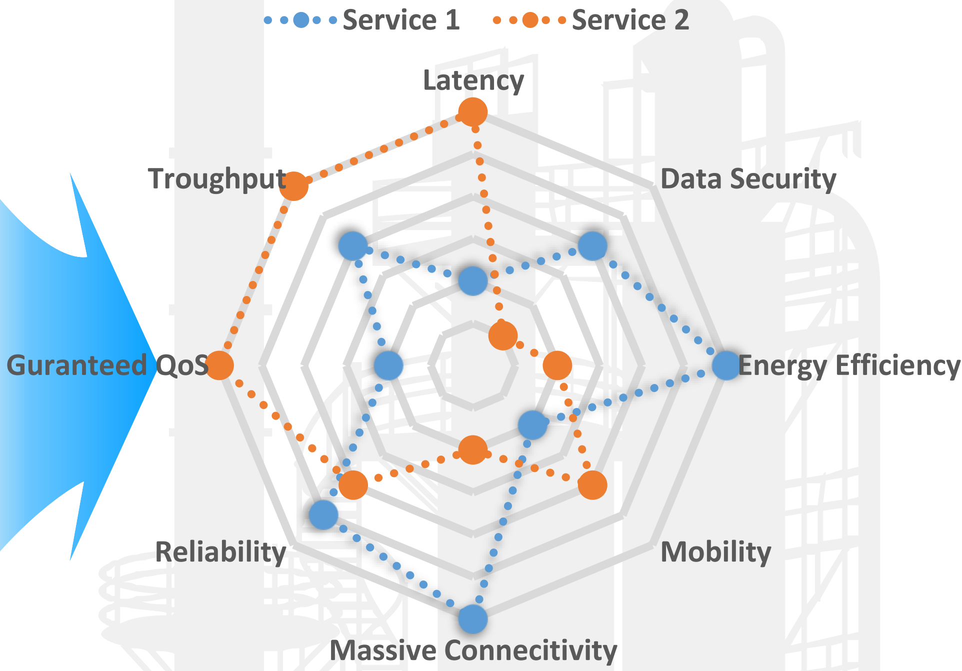
INDUSTRY

Smart City



Smart Agriculture

Service-Specific Requirements



✓ The network resources management becomes more complex because of the very diverse requirements.

Introduction : Requirements in Manufacturing Industry

- The manufacturing industry requires
  - high-quality,
  - time-sensitive,
  - automated,
  - intelligent and
  - flexible industrial control.
- So that materials, products and processes can be monitored, optimized and controlled in real time.

Smart Factory

Example applications

- **Low-latency**, **high-reliability** and **high-availability connectivity**, **mobility** and **precise positioning** of all the devices (e.g. sensors and actuators) for **real-time monitoring** and **control of processes**, and **end-to-end** logistics and asset tracking
- Connectivity and local processing for **real-time video capture** and video-based applications
- **Connectivity of massive numbers of sensors**, and platforms for collation and **processing of large amounts of data**
- **Augmented reality to optimize** and improve maintenance tasks.

https://www.ngmn.org/wp-content/uploads/NGMN-5G-White-Paper-2.pdf

- The road transportation industry expects to provide efficient, safe, environmentally-friendly and comfortable transportation, especially by exploiting the potential of artificial intelligence, to achieve connected and automatic driving through perception, decision and control.

Example applications

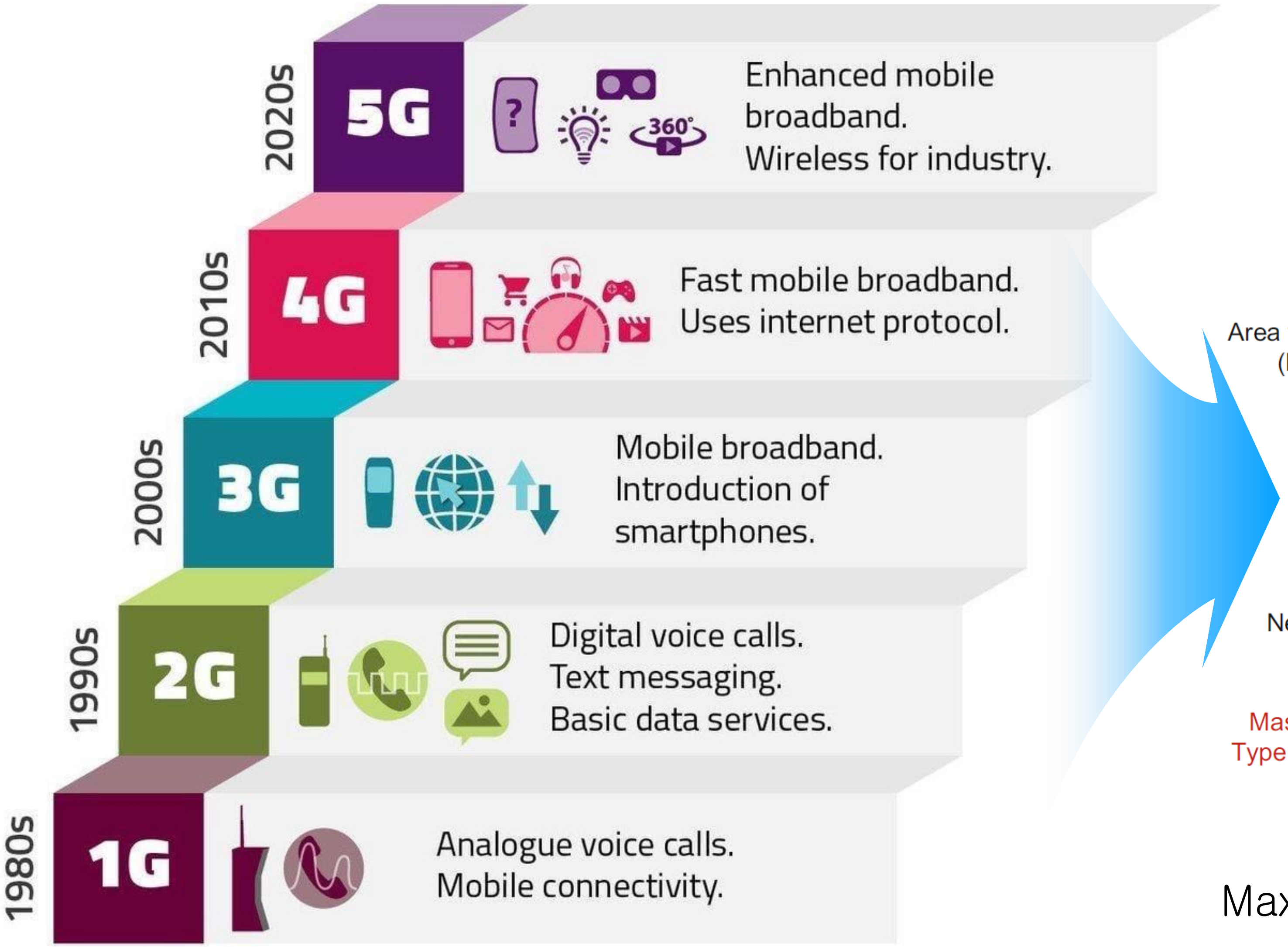
- **Transmission of high quality video** or images of **road condition and roadside facilities** to help navigation, remote and automatic driving, as well as identification of blind zones and other vulnerabilities for vehicles
- **Real-time communication among vehicles and road infrastructures**, coupled with precise vehicle positioning and local (edge) computing capabilities, enabling identification of potential dangers, which can help decision making including route planning and updating, emergency braking, and intelligent car collision avoidance

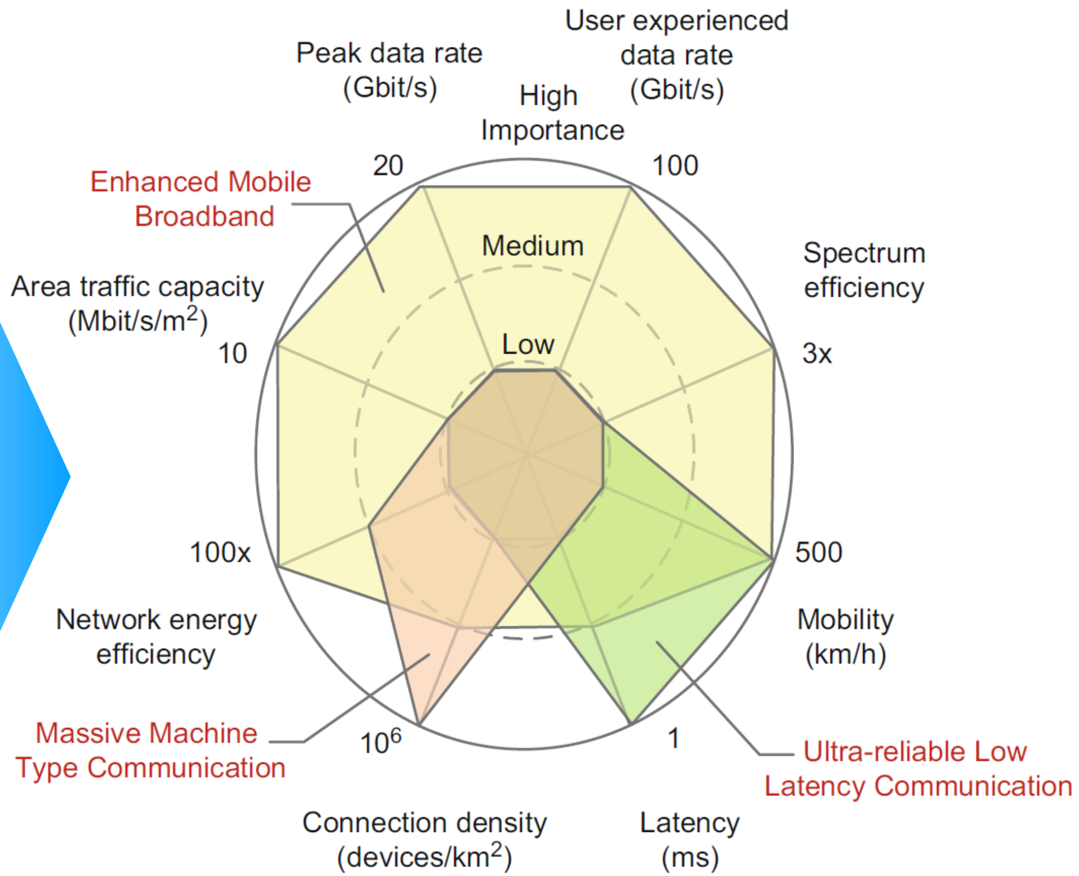
- The health industry requires a balanced allocation of medical resources, portable and intelligent medical equipment, improved medical vehicle treatment capabilities, and the transformation of surgical operations from the operating room to multiple regions.

Example applications

- Wide-area continuous coverage for ambulances, including **sending live high quality video and patient vital signs** in **real time** to the command center in the hospital
- **Sensors collecting vital signs from the wearable devices** of patients or the elderly, wherever they are, helping remote medical staff make timely treatment decisions and administer medication remotely

Introduction : Evolution of Cellular Systems





Maximum 5G capabilities for different use cases

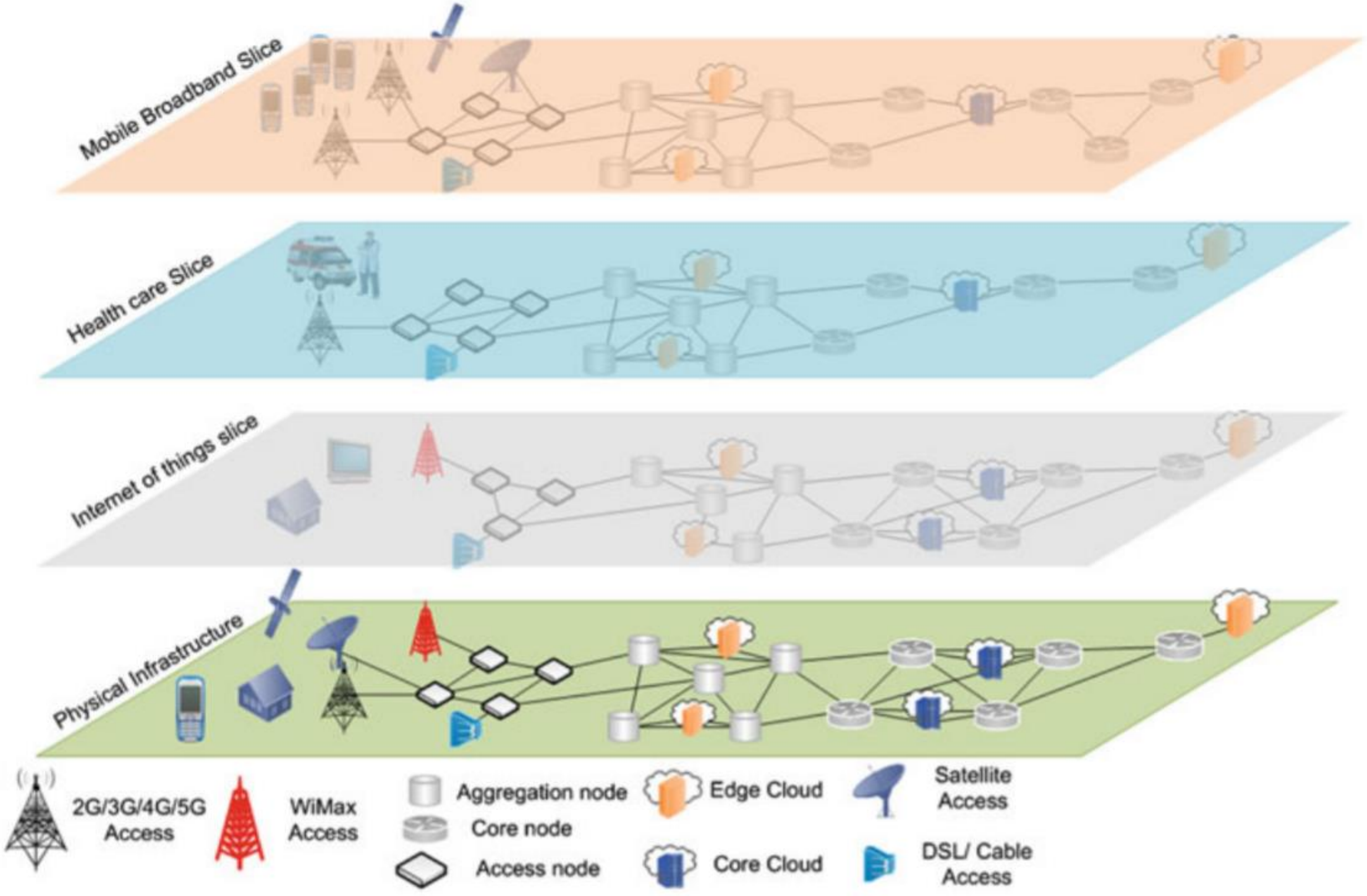
- Generally, to realize diverse 5G use cases there is need to resolve the given challenges:
  - Scalability and Reliability
  - Interoperability
  - Sustainability
  - Network Slicing
  - Security
  - Integration of AI

Network Slicing: The Concept

- Network Slicing
- Key Enablers
- Network Slicing: Industrial Efforts

Network Slicing (1)

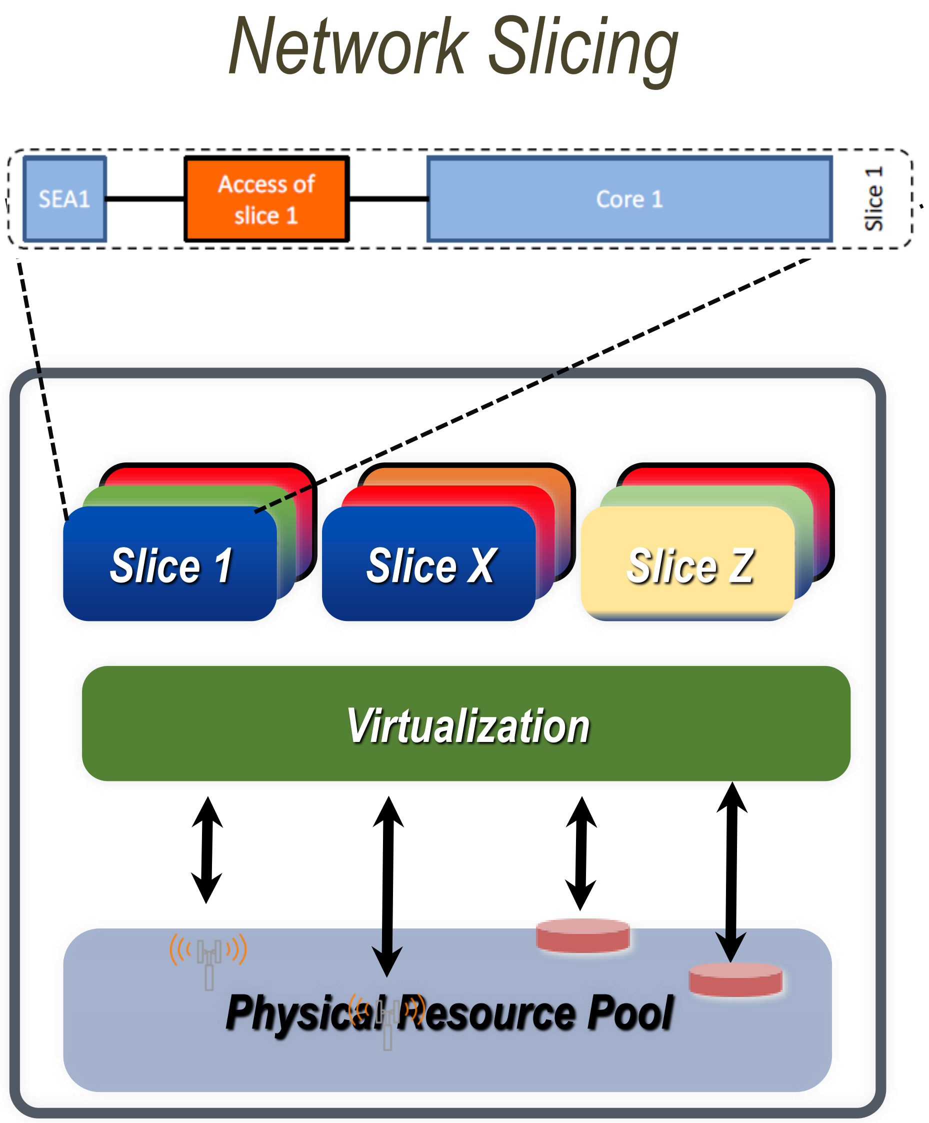
Network slicing is a new network approach that can provide highly tailored services to specific customer groups and even individual customers.



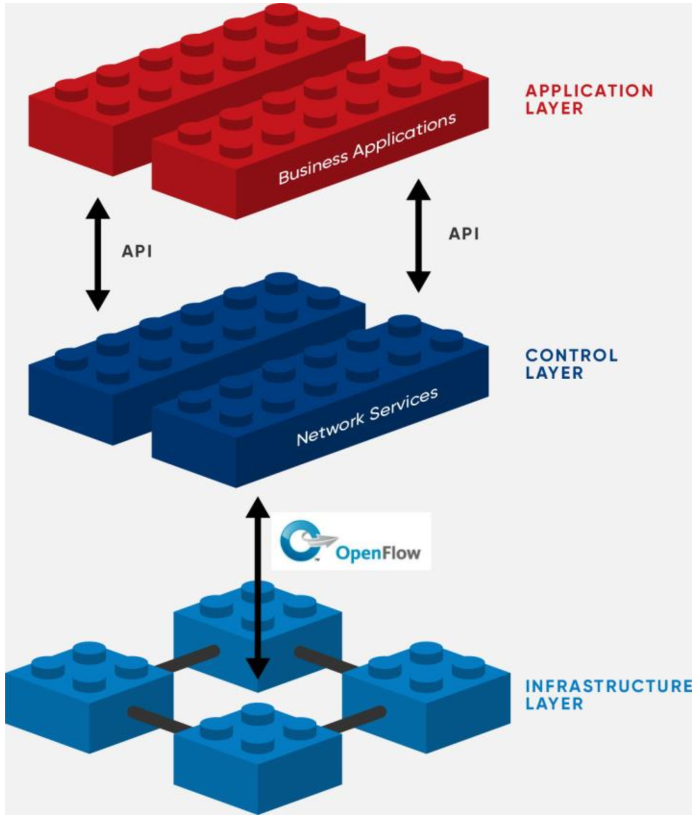
S.M. Ahsan Kazmi, Latif U. Khan, Nguyen H. Tran, Choong Seon Hong, "Network Slicing for 5G and Beyond Networks," ISBN 978-3-030-16169-9, Springer

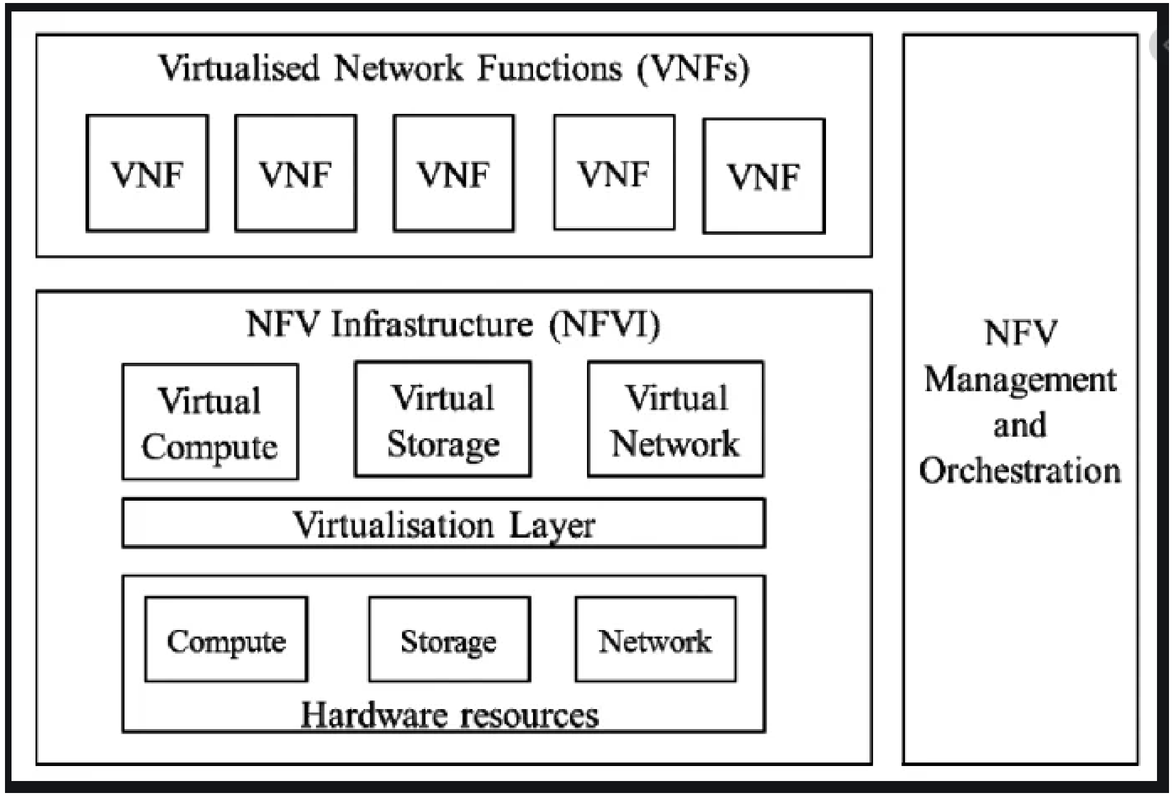
Network Slicing

- Network slicing can fulfil the diverse requirements of these novel network services
- Network slicing enables one physical network into multiple, virtual, end-to-end (E2E) networks, each logically isolated including device, access, transport and core network
- A slice is dedicated for different types of service with different characteristics and requirements given to a Service End-point Agent (SEA)
- Enforce strong isolation between slices, i.e., actions in one slice do not affect another



- Network Slicing enablers: How to do it ?
  - Software-defined networking (SDN)
  - Network Functions Virtualization (NFV)

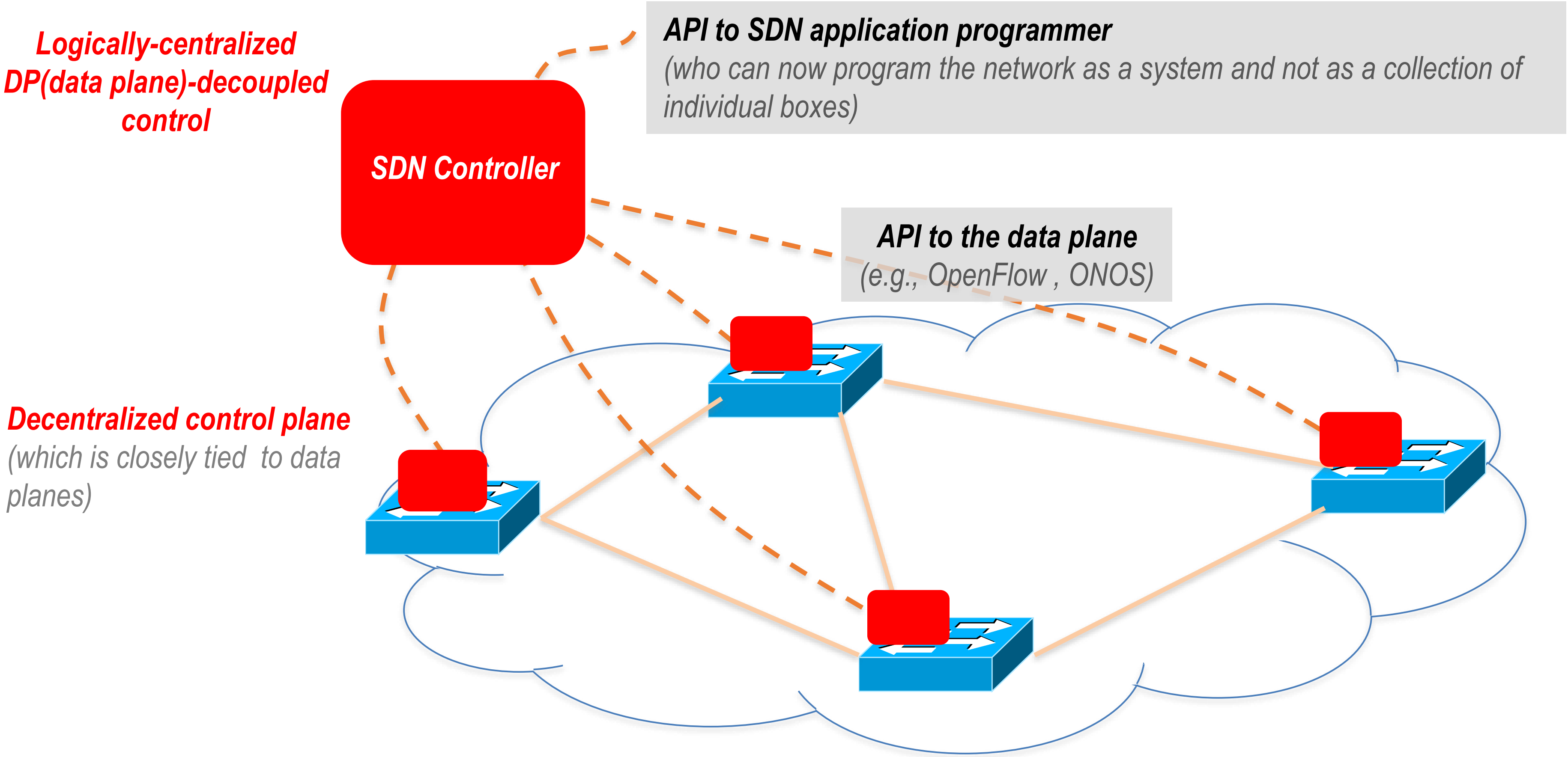




High-level NFV framework. Source: ETSI

https://www.opennetworking.org/sdn-definition/

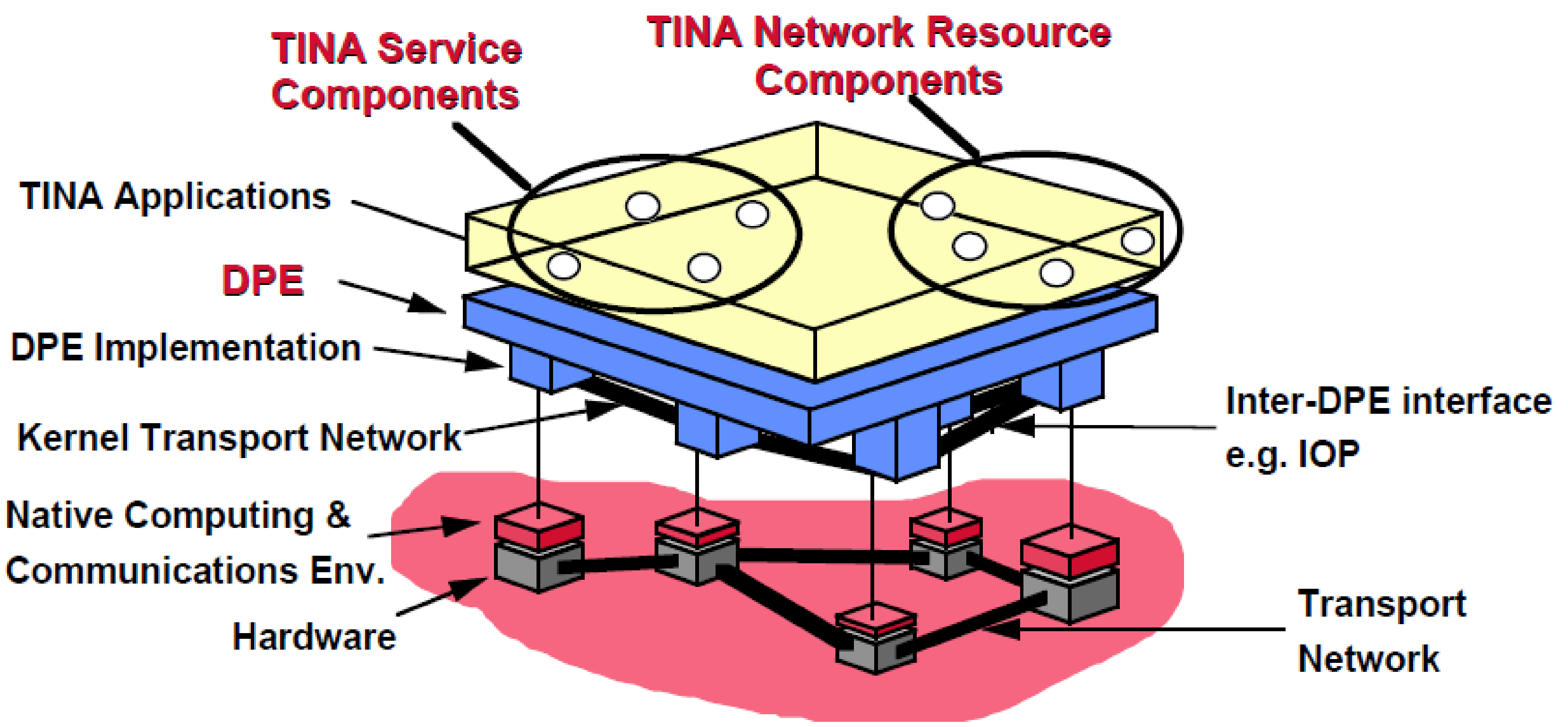
Network slicing enablers: Software defined network (SDN)

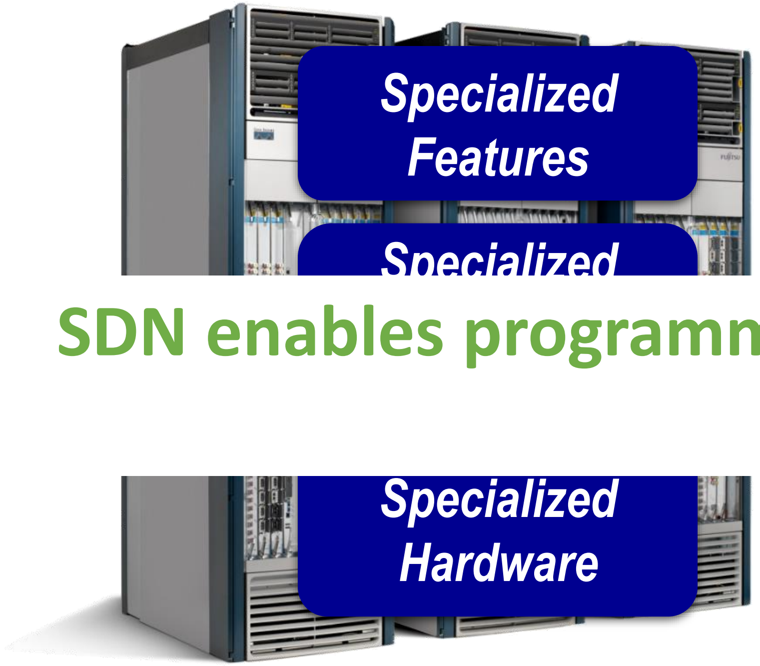


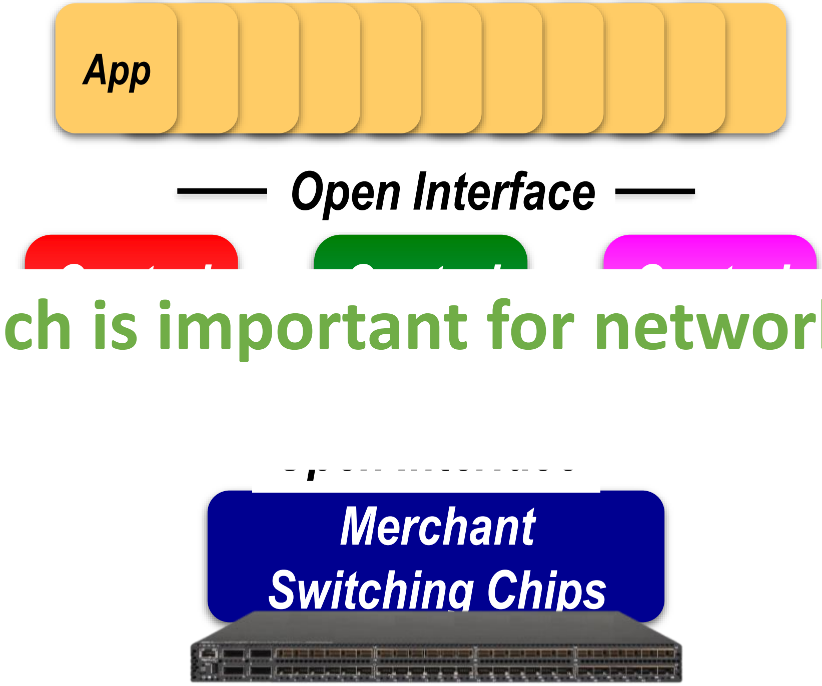
`At the highest level, the SDN movement is an effort to build networks you can program at a higher level of abstraction— just as you can program a computer.`

TINA Architectural Separations

Telecommunications Information System







SDN enables programmability which is important for network slicing

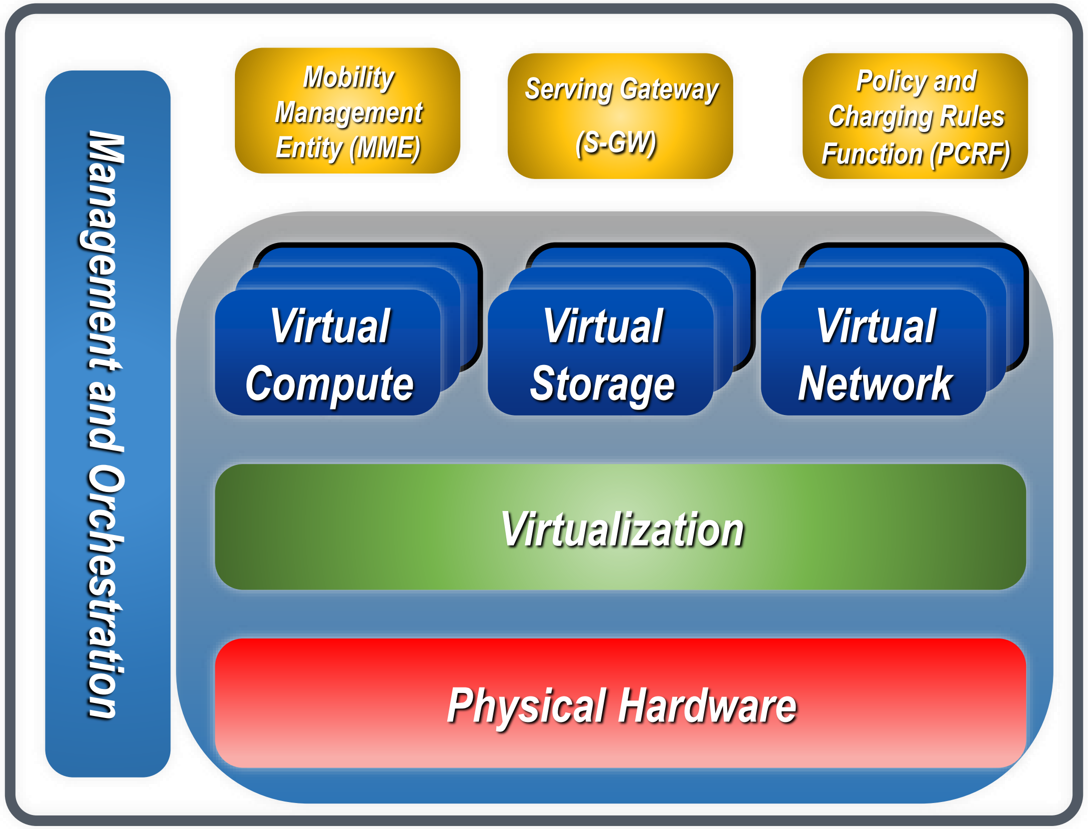
Vertically integrated
Closed, proprietary
Slow innovation

Horizontal
Open interfaces
Rapid innovation

Network slicing enablers: Network function virtualization (NFV)

- A network architecture concept that uses the technologies of *IT virtualization* to virtualize entire classes of network node functions that may connect, or chain together, to create communication services

- NFV is envisioned to play a crucial role in network slicing as it will be responsible to build isolated slices based on user service requirements



Wireless Network Virtualization

- Due to massive success of NFV and SDN in wired domain, a number of studies are being conducted to adopt them both in the core and radio access networks (RANs) for future cellular networks such as:
  - CORD (Central Office Re-architected as a Datacenter) [1]
  - Radisys M-CORD [2]

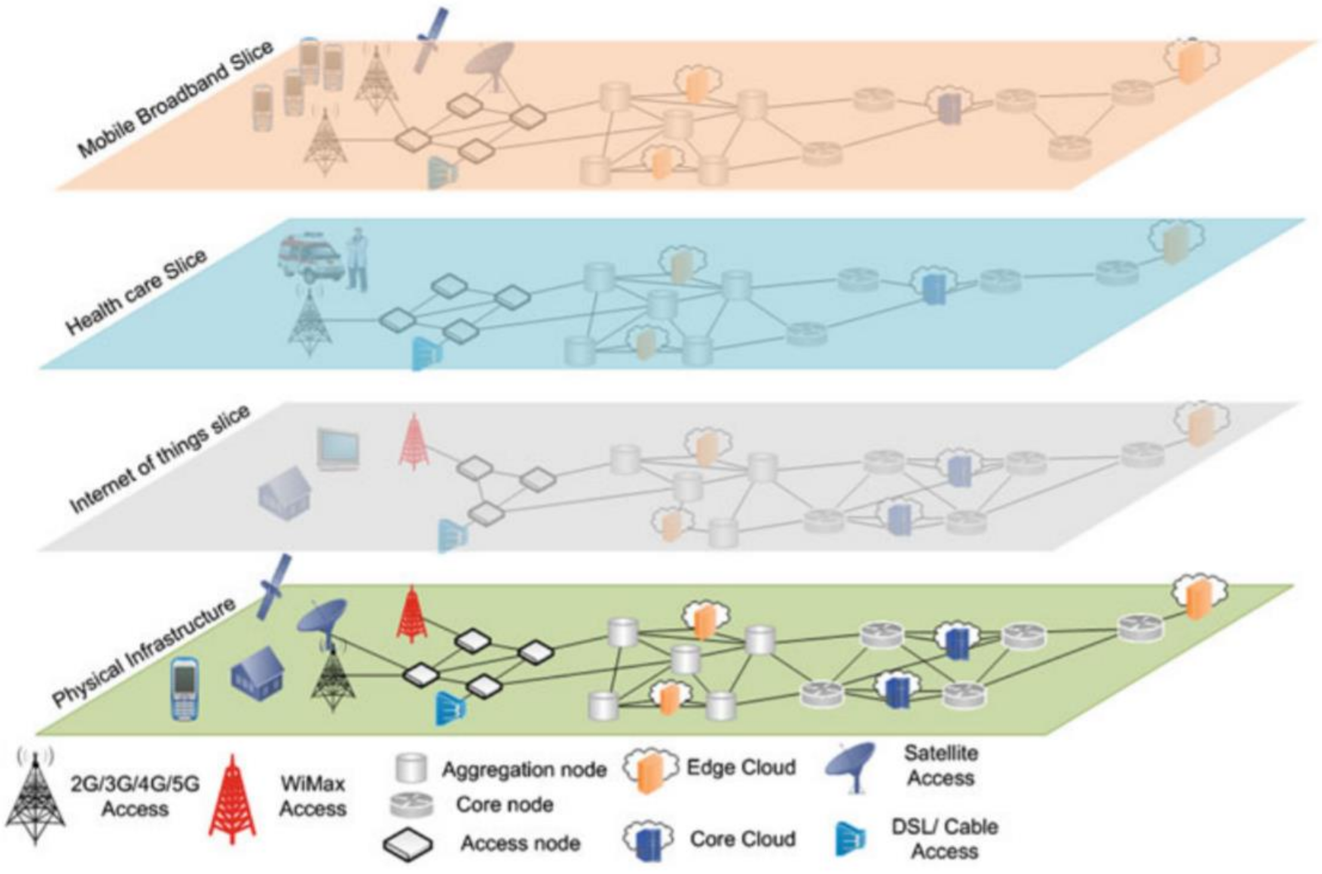
- Wireless network virtualization (WNV) is a novel concept for *virtualizing the RANs* of future cellular networks

- WNV has a very broad scope ranging from spectrum sharing, infrastructure virtualization, to air interface virtualization

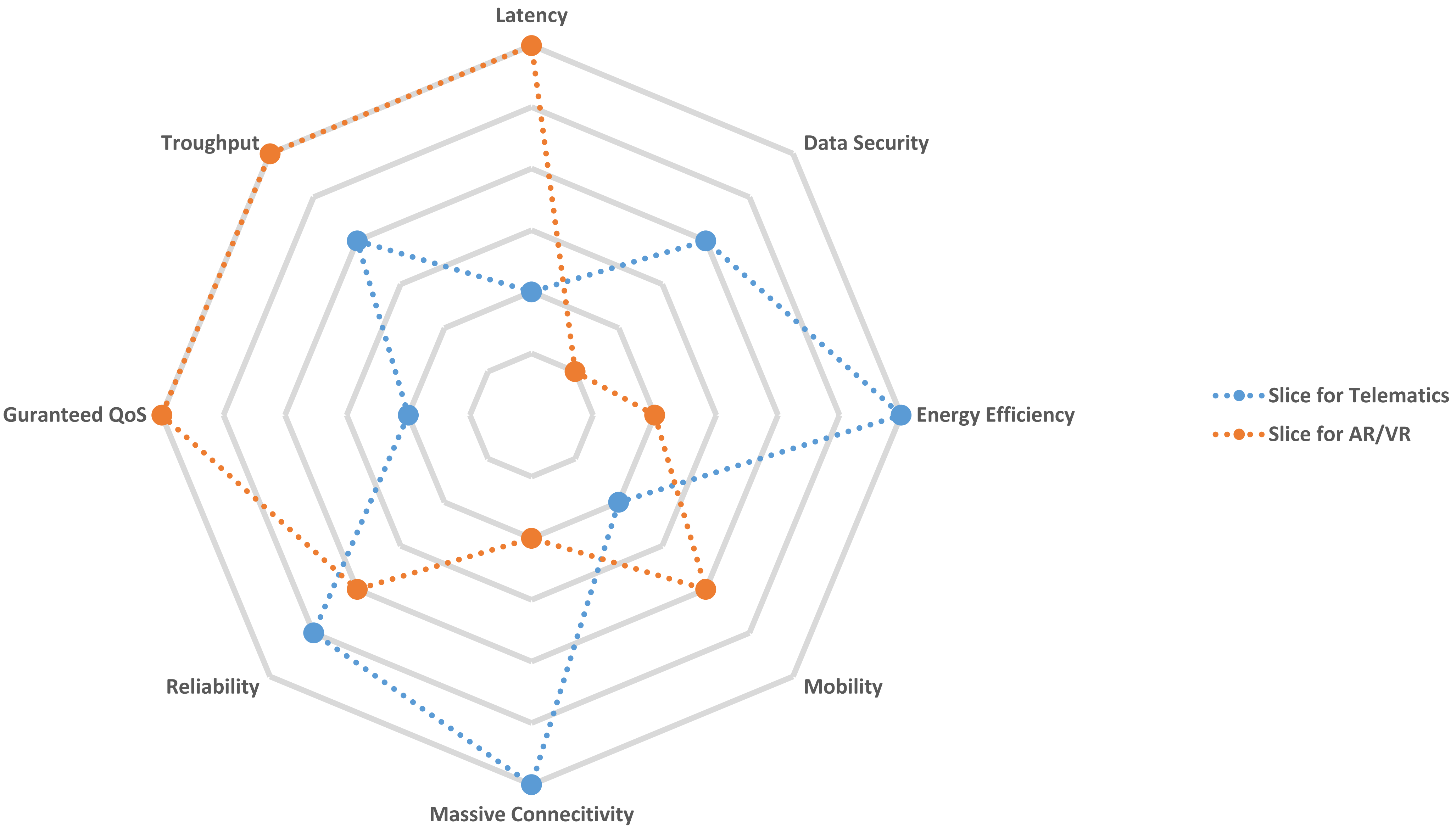
[1] https://opencord.org/
[2] http://www.radisys.com/radisys-m-cord-open-platform-emerging-5g-applications

Network Slicing (4)

- Network Slicing Principles
  - Slice Isolation
  - Elasticity
  - End-to-End Customization



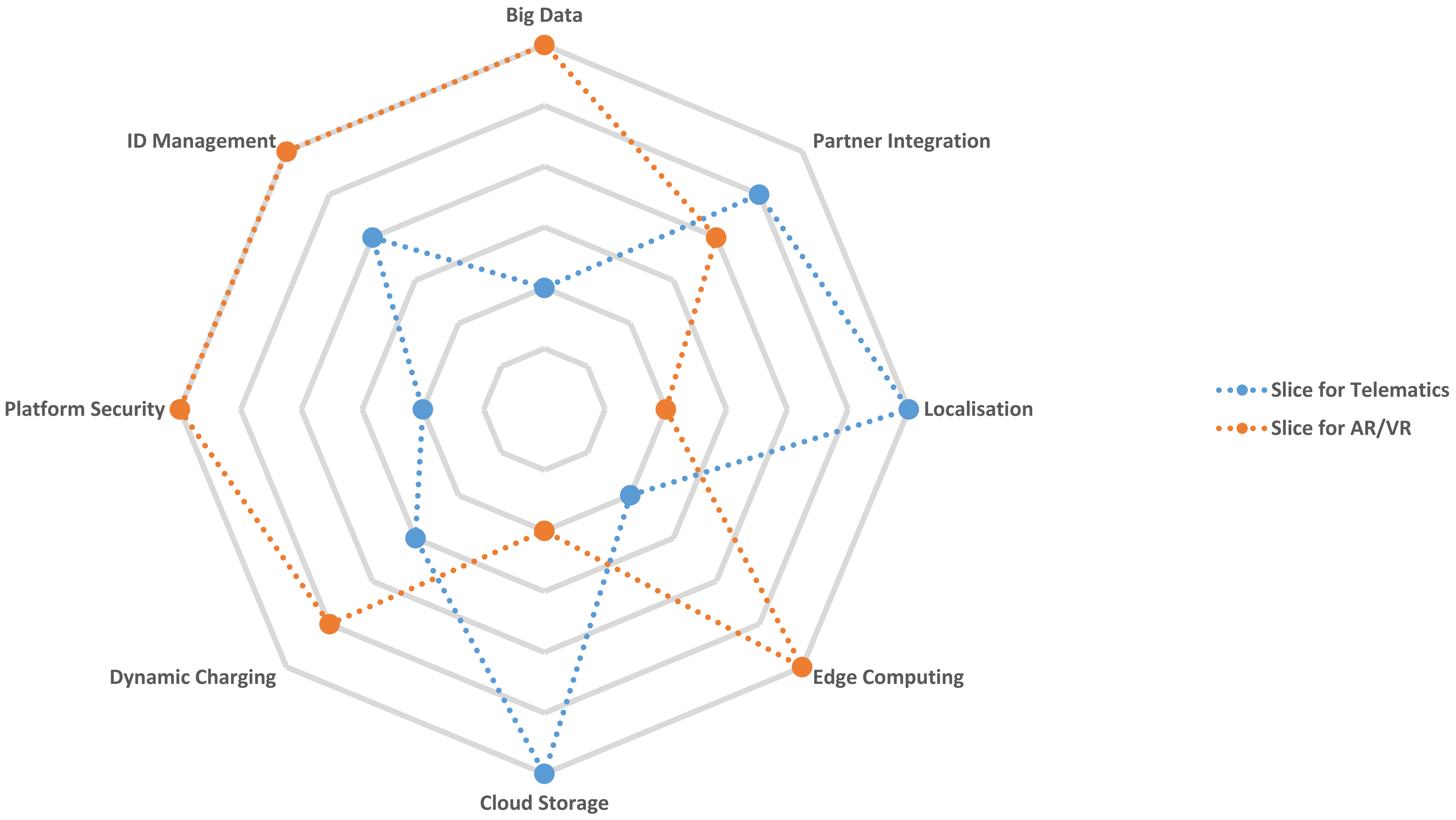
Network Slicing based on Network Capability



KYUNG HEE UNIVERSITY

NETWORKING INTELLIGENCE LAB

Network Slicing based on Network Services



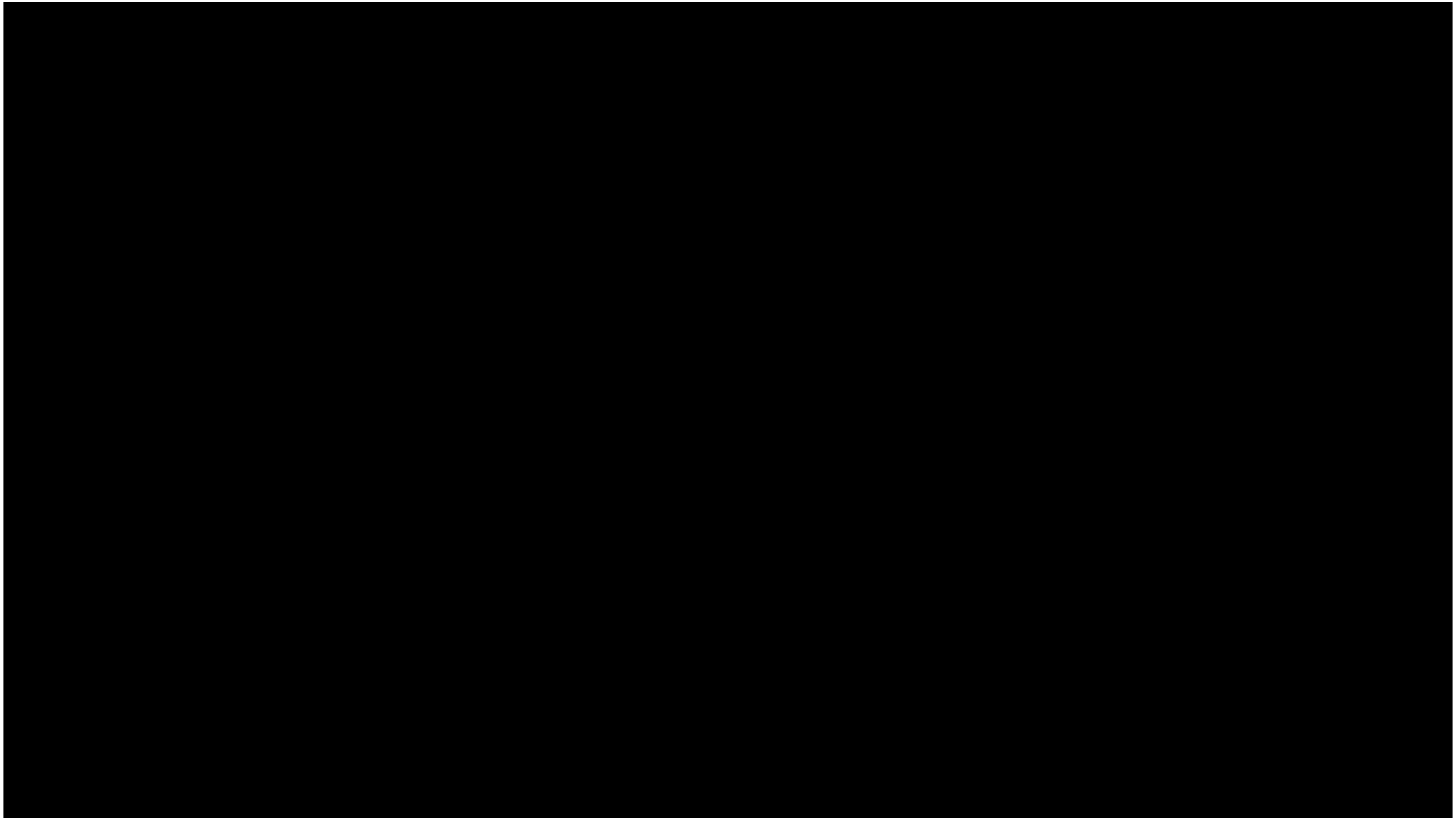
Network Slicing: Industrial Efforts (1)

- **Network Slicing** certainly is one of the most discussed technologies these days. Network operators like KT, SK Telecom, China Mobile, DT, KDDI and NTT, and also vendors like Ericsson, Nokia and Huawei are all recognizing it as an ideal network architecture for the coming 5G era.

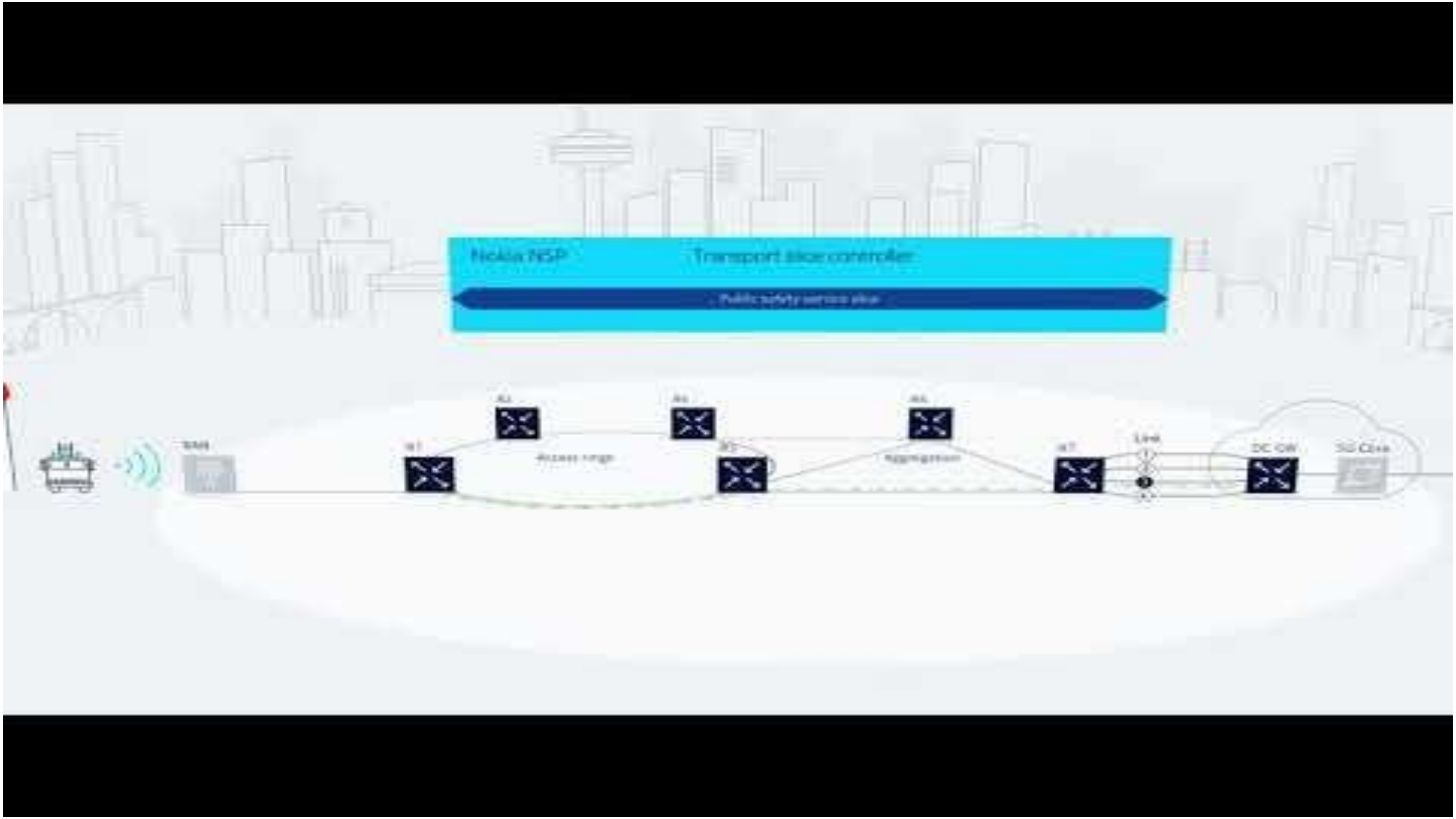
- **Ericsson** has been working on network slicing with **NTT DOCOMO** since 2014. In 2016 the two announced a successful **proof of concept of dynamic network slicing technology** for 5G core networks.
  - They created a slice management function and network slices based on requirements such as latency, security or capacity.

- **Samsung** and **KDDI** Complete 5G **End-to-End Network Slicing** Demonstration in **September 2020**

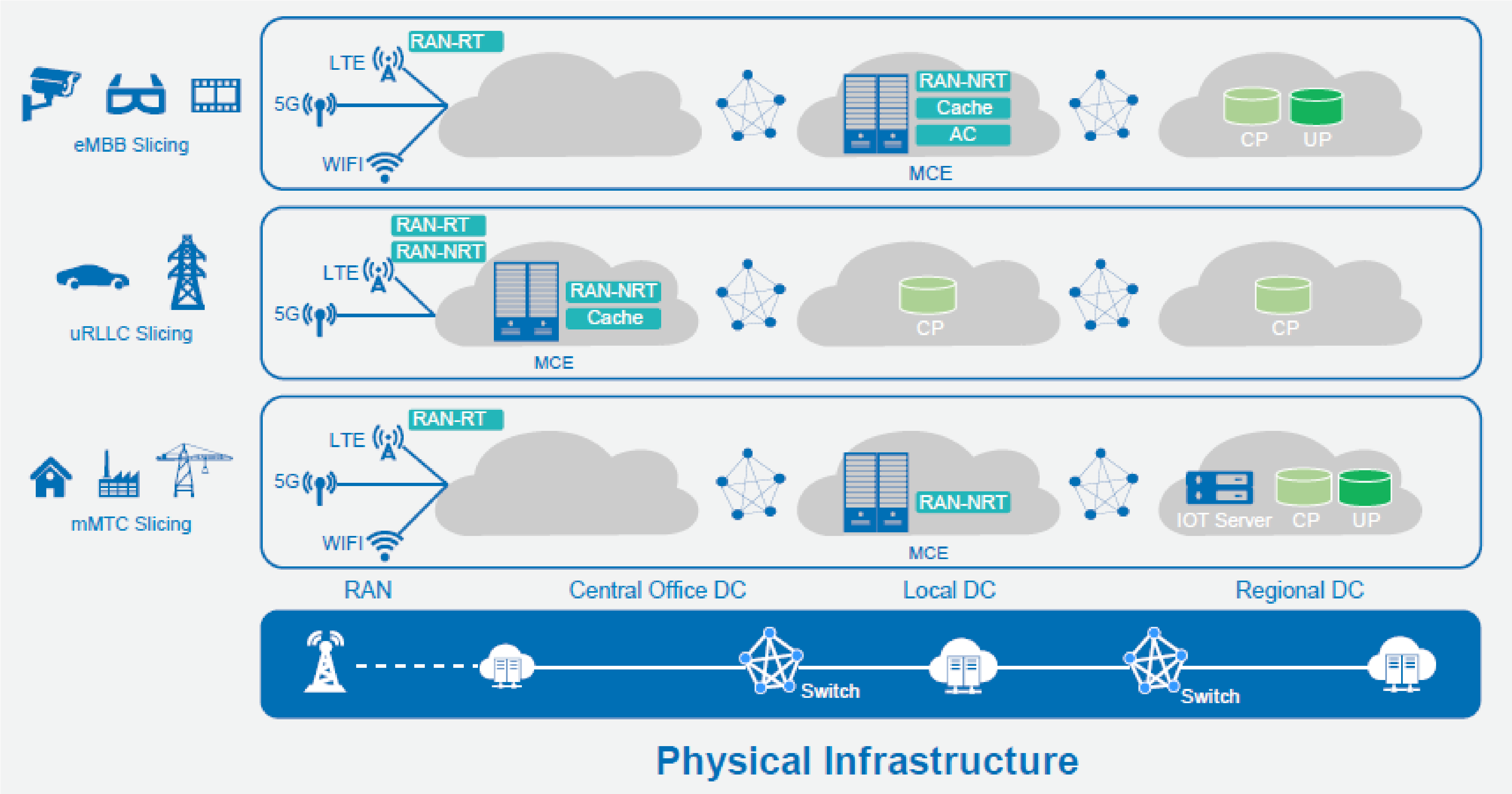
Network Slicing: Industrial Efforts (2)



Network Slicing: Industrial Efforts (3)



Network Slicing: Industrial Efforts - Huawei Technologies (4)



RAN-RT: Radio Access Network-Real Time

RAN-NRT: Radio Access Network-non Real Time

AC: Access Cloud

CP: Control Plane

UP: User Plane

MCE: Mobile Cloud Engine

DC: Data Center

http://www.huawei.com/minisite/5g/img/5G_Nework_Architecture_A_High_Level_View_en.pdf

Network Slicing Key Resources

- Radio resources
  - Access network resources
  - Core network resources
- Caching
  - On-device caching
  - Edge caching
  - Core network caching
- Edge Computing Servers
  - Cloudlets
  - Fog servers
  - Multi-access edge computing servers

Use Case 1: Virtual Reality

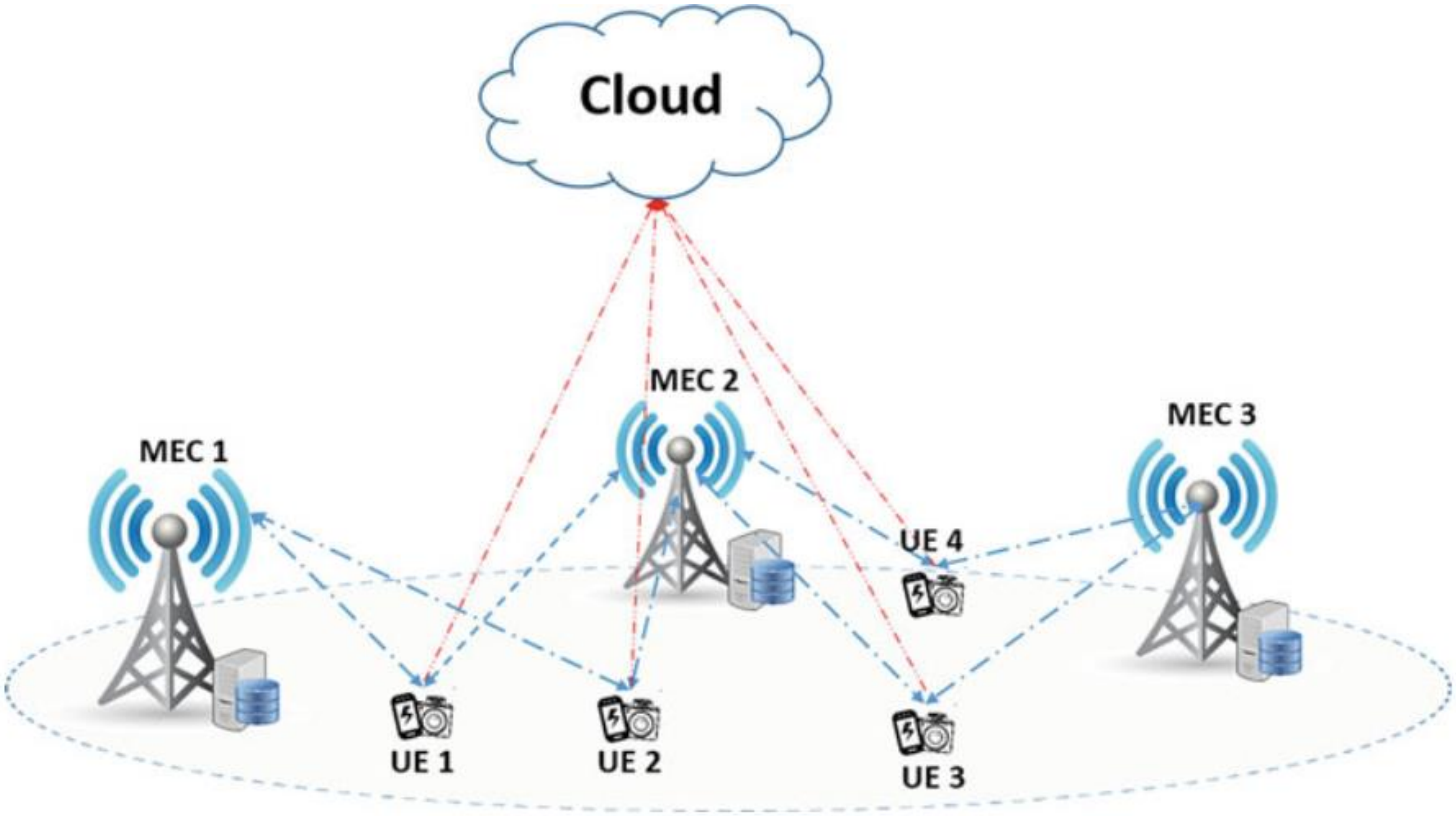
- Introduction
- System Model
- Problem Formulation
- Solution Approach
- Simulation Results

Virtual Reality: A Use Case

- Virtual Reality Applications
  - Smart Health-Care
  - Smart Industries
  - Smart Gaming

- Virtual Reality Challenges
  - High Computational Power for Processing Complex Algorithms
  - Strict-Latency Constraints

Virtual Reality: System Model (1)



$\mathcal{N} = \{1, \dots, N\}$ be the set of N mobile users with VR capability

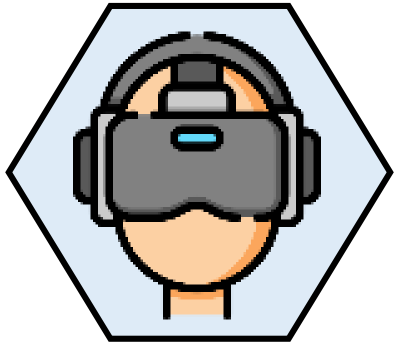
$\mathcal{M} = \{1, \dots, M\}$ be the set of M MEC servers.

S.M. Ahsan Kazmi, Latif U. Khan, Nguyen H. Tran, Choong Seon Hong, "Network Slicing for 5G and Beyond Networks," ISBN 978-3-030-16169-9, Springer

Virtual Reality: System Model (2)

Computational task to offload for each mobile user → $\tau_i = \{S_i, T_i, C_i\}$

- Video size → S_i
- Worst case execution time of the task → T_i
- Budget → C_i

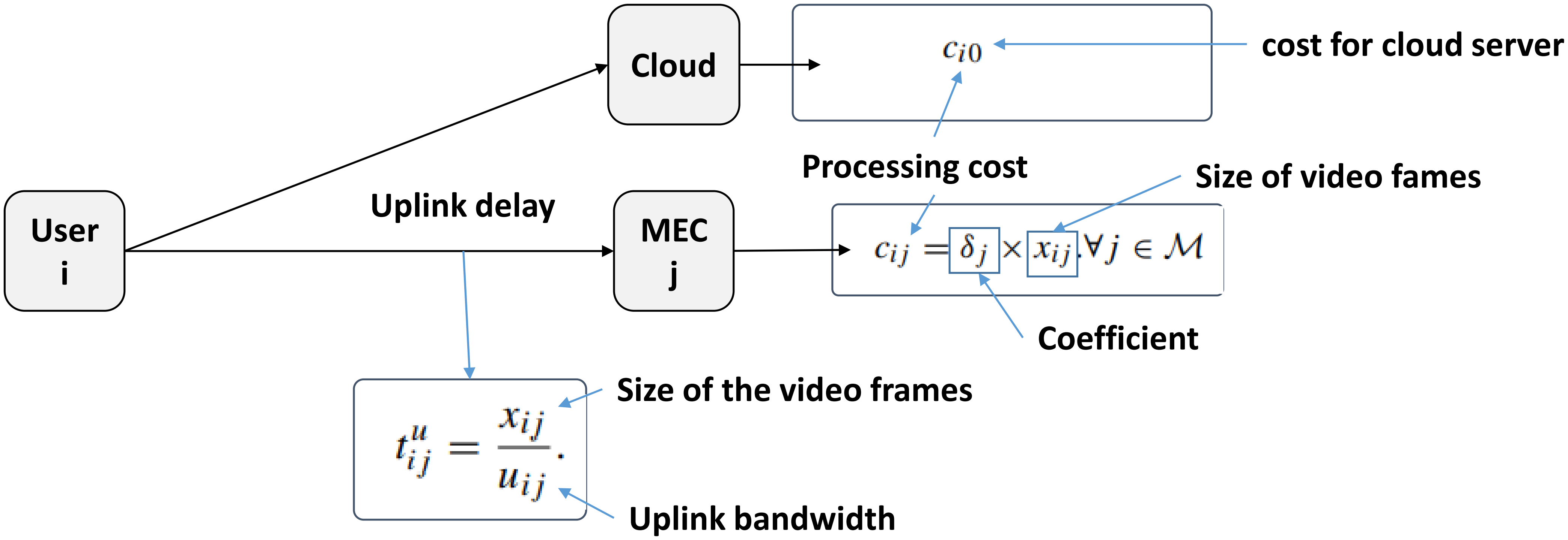


Video size → $$S_i = \frac{H \times V \times F \times L \times 3}{1024 \times 1024}$$

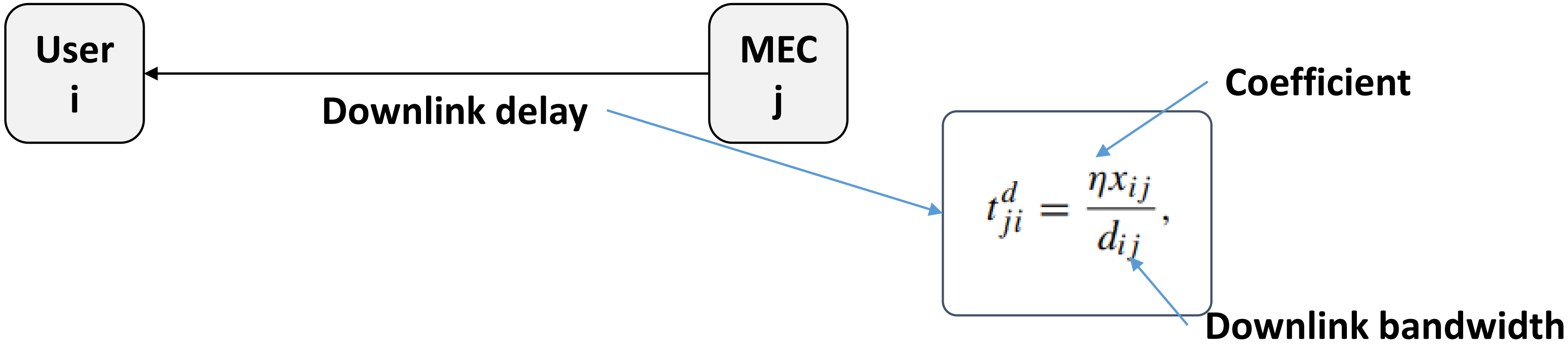
- Horizontal pixels → H
- Vertical pixels → V
- Frames per second → F
- Video length → L
- Color depth → 3
- Bit to bytes division → 1024×1024

S.M. Ahsan Kazmi, Latif U. Khan, Nguyen H. Tran, Choong Seon Hong, "Network Slicing for 5G and Beyond Networks," ISBN 978-3-030-16169-9, Springer

Virtual Reality: System Model (2)



Virtual Reality: System Model (2)



$$t_{ji}^{d} = \frac{\eta x_{ij}}{d_{ij}},$$

S.M. Ahsan Kazmi, Latif U. Khan, Nguyen H. Tran, Choong Seon Hong, "Network Slicing for 5G and Beyond Networks," ISBN 978-3-030-16169-9, Springer

Total transmission and processing cost minimization problem.

Uplink delay — Downlink delay — Processing cost

$$\underset{x}{minimize}: \sum_{i \in N} \alpha \left(\frac{\sum_{j \in M} t^{u}_{ij}}{t^{u}_{i0}} + \frac{\sum_{j \in M} t^{d}_{ji}}{t^{d}_{i0}} \right) + (1-\alpha) \frac{\sum_{j \in M} c_{ij}}{c_{i0}} \quad (3.10)$$

$$subject\ to: \sum_{j \in M} x_{ij} = \frac{H_i \times V_i \times F_i \times L_i \times 3}{1024 \times 1024}, \forall i \in N, \quad (3.11)$$

Guarantees that all users are served by network.

$$\sum_{i \in N} x_{ij} \le \Gamma_j, \forall j \in M, \quad (3.12)$$

Maximum MEC server processing capacity

$$\sum_{j \in M} t_i \le T_i, \forall i \in N, \quad (3.13)$$

Latency limit constraint

$$\sum_{j \in M} c_i \le C_i, \forall i \in N, \quad (3.14)$$

Budge cost must not be greater than processing cost

$$t'_i < 1, \forall i \in N, \quad (3.15)$$

Total time constraint variable

$$c'_i < 1, \forall i \in N, \quad (3.16)$$

Total cost constraint variable

$$x_{ij} \ge 0, \forall i \in N, \forall j \in M. \quad (3.17)$$

Video frames size variable

ADMM-Based Solution (1)

Re-write the objective function.

$$(3.10) = \alpha \left(\sum_{j \in M} \left(\frac{x_{ij}}{u_{ij} t_{i0}^{u}} + \frac{\eta x_{ij}}{d_{ij} t_{ji}^{d}} \right) \right) + (1-\alpha) \left(\frac{\sum_{j \in M} \delta_j x_{ij}}{c_{i0}} \right) \quad (3.18)$$

$$= \alpha \left(\sum_{j \in M} \left(\frac{1}{u_{ij} t_{i0}^{u}} + \frac{\eta}{d_{ij} t_{ji}^{d}} \right) \right) \boxed{x_{ij}} + (1-\alpha) \left(\frac{\sum_{j \in M} \delta_j}{c_{i0}} \right) \boxed{x_{ij}} \quad (3.19)$$

$$= \left(\alpha \sum_{j \in M} \left(\left(\frac{1}{u_{ij} t_{i0}^{u}} + \frac{\eta}{d_{ij} t_{ji}^{d}} \right) + (1-\alpha) \left(\frac{\sum_{j \in M} \delta_j}{c_{i0}} \right) \right) \right) \boxed{x_{ij}} \quad (3.20)$$

$$= \boxed{f_i(\mathbf{x}_i)} \quad (3.21)$$

S.M. Ahsan Kazmi, Latif U. Khan, Nguyen H. Tran, Choong Seon Hong, "Network Slicing for 5G and Beyond Networks," ISBN 978-3-030-16169-9, Springer

ADMM-Based Solution (2)

Modified Problem

$$\underset{x}{minimize} : \sum_{i \in N} f_i(\mathbf{x}_i) \quad (3.22)$$

$$subject\ to : \mathbf{1}^T \mathbf{x}_i = S_i, \forall i \in N \quad (3.23)$$

Guarantees that all users are served by network.

$$\mathbf{1}^T \mathbf{x}_j \leq \Gamma_j, \forall j \in M \quad (3.24)$$

Maximum MEC server capacity constraint

$$\sum_{j \in M} t_i \leq T_i, \forall i \in N \quad (3.25)$$

Latency limit constraint

$$\sum_{j \in M} c_i \leq C_i, \forall i \in N \quad (3.26)$$

Budge cost must not be greater than processing cost

$$t_i' < 1, \forall i \in N \quad (3.27)$$

Total time constraint variable

$$c_i' < 1, \forall i \in N \quad (3.28)$$

Total cost constraint variable

$$x_{ij} \geq 0, \forall i \in N, \forall j \in M \quad (3.29)$$

Video frames size variable

S.M. Ahsan Kazmi, Latif U. Khan, Nguyen H. Tran, Choong Seon Hong, "Network Slicing for 5G and Beyond Networks," ISBN 978-3-030-16169-9, Springer

ADMM-Based Solution (3)

For ADMM-based solution new variable z is introduced.

$$\begin{aligned} &\underset{x}{minimize} : \sum_{i \in N} f_i(\mathbf{x}_i) + h(z) \\ &subject\ to : \mathbf{x}_i = z \\ &\qquad\qquad\quad \mathbf{x}_i \in \mathcal{X}, \forall i \in N \end{aligned} \tag{3.31}$$

where $h(z) = 0$ when $\mathbf{x}_i \in \mathcal{X}$.

$$h(z) = I_{\mathcal{X}}(z) = \begin{cases} 0, & \mathbf{x}_i \in \mathcal{X} \\ \infty, & otherwise \end{cases} \tag{3.32}$$

ADMM-Based Solution (3)

For ADMM-based solution new variable z is introduced.

Then, the augmented Lagrangian function of (3.31) is as follows:

Lagrangian — **Lagrangian penalty term** — **Augmentation**

$$\mathcal{L}(\mathbf{x}, z, \lambda) = \sum_{i \in N} \left(f_i(\mathbf{x}_i) + \lambda_i^T (x_i - z) + \frac{\rho}{2} ||x_i - z||_2^2 \right) \quad (3.33)$$

Lagrange multiplier — **penalty parameter**

Based on the solution from [43], the resulting ADMM variables update are the following:

$$x_i^{k+1} = \arg\min \left(f_i(x_i) + \lambda_i^{kT} (x_i - z^k) + \frac{\rho}{2} ||x_i - z^k||_2^2 \right) \quad (3.34)$$

$$z^{k+1} = \arg\min \left(h(z) + \sum_{i=1}^{N} \left(-\lambda_i^{kT} z + \frac{\rho}{2} ||x_i^{k+1} - z||_2^2 \right) \right) \quad (3.35)$$

Lagrange multiplier →

$$\lambda_i^{k+1} = \lambda_i^k + \rho (x_i^{k+1} - z^{k+1}) \quad (3.36)$$

[43] S.M. Ahsan Kazmi, Latif U. Khan, Nguyen H. Tran, Choong Seon Hong, "Network Slicing for 5G and Beyond Networks," ISBN 978-3-030-16169-9, Springer

ADMM-Based Task Offloading Algorithm.

Algorithm 1 ADMM-based task offloading

input:Initialization for $\mathcal{N}$, $\mathcal{M}$, $\mathbf{D}$, $\mathbf{U}$
Output:Minimal offloading cost
Initialization
$max\_iteration = 1000$, $\rho = 0.5$, $\alpha = 0.5$ $\mathbf{x}_i^0 \geq 0$, $\lambda_i^0 \geq 0$, $z \geq 0$, t_{i0}^u, t_{i0}^d, c_{i0}, $\forall i \in N$
for $k \in max\_iteration$ **do**
    Each user $i \in \mathcal{N}$ update its offloading decision by Eqs. (3.34), (3.35), (3.32) respectively, parallelly
    After getting updated values from all users each MEC will update λ using (3.36), parallelly
    After all variable updated, update the objective function (3.10)
end for
    return Optimal value of objective (3.10)

$$x_i^{k+1} = \arg\min \left(f_i(x_i) + \lambda_i^{kT}(x_i - z^k) + \frac{\rho}{2}||x_i - z^k||_2^2 \right) \qquad (3.34)$$

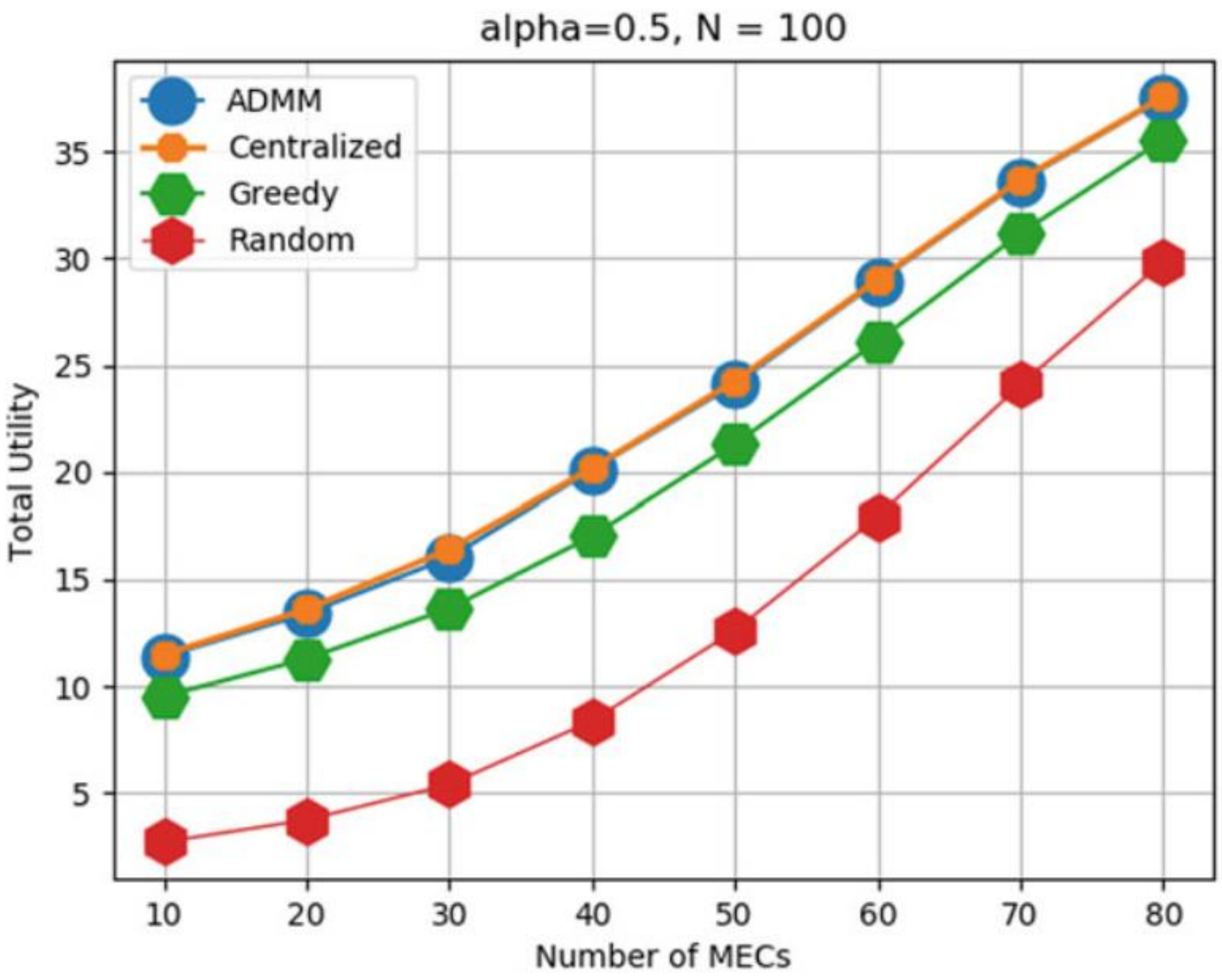
$$z^{k+1} = \arg\min \left(h(z) + \sum_{i=1}^{N} \left(-\lambda_i^{kT} z + \frac{\rho}{2}||x_i^{k+1} - z||_2^2 \right) \right) \qquad (3.35)$$

$$h(z) = I_{\mathcal{X}}(z) = \begin{cases} 0, & \mathbf{x}_i \in \mathcal{X} \\ \infty, & otherwise \end{cases} \qquad (3.32)$$

$$\lambda_i^{k+1} = \lambda_i^k + \rho(x_i^{k+1} - z^{k+1}) \qquad (3.36)$$

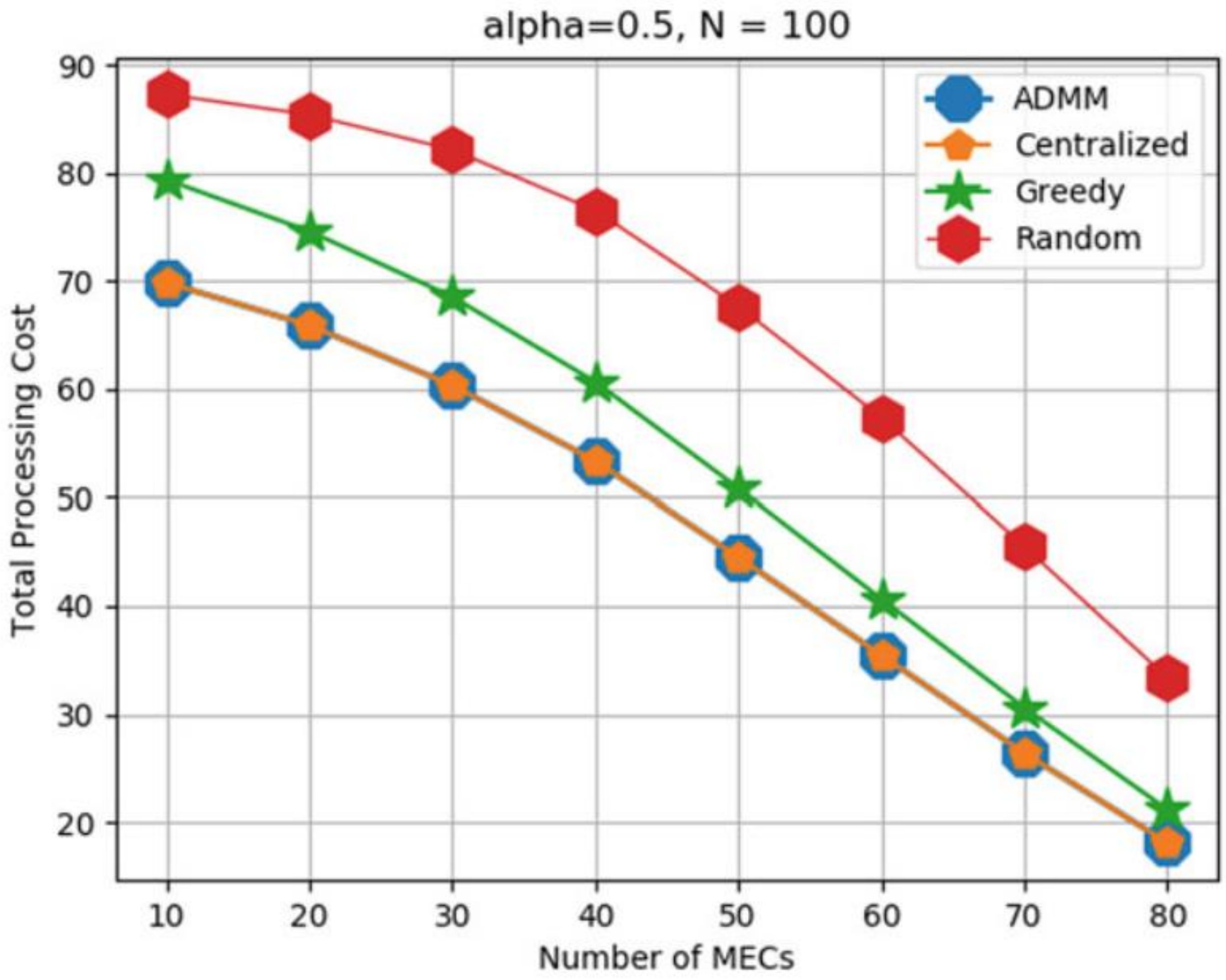
$$\underset{x}{minimize} : \sum_{i \in N} \alpha \left(\frac{\sum_{j \in M} t_{ij}^u}{t_{i0}^u} + \frac{\sum_{j \in M} t_{ji}^d}{t_{i0}^d} \right) + (1 - \alpha) \frac{\sum_{j \in M} c_{ij}}{c_{i0}} \qquad (3.10)$$

Performance Evaluation (1)



Total utility vs. number of MEC servers

S.M. Ahsan Kazmi, Latif U. Khan, Nguyen H. Tran, Choong Seon Hong, "Network Slicing for 5G and Beyond Networks," ISBN 978-3-030-16169-9, Springer



Total Processing cost vs. number of MEC servers

S.M. Ahsan Kazmi, Latif U. Khan, Nguyen H. Tran, Choong Seon Hong, "Network Slicing for 5G and Beyond Networks," ISBN 978-3-030-16169-9, Springer

Summary

- An overview of resource management for network slicing has been presented in this lecture.
- Numerous key resources for network slicing are discussed.
- A use case of virtual reality is along with its ADMM-based solution is presented.

Use Case 2: Chunk-Based Resource Allocation

- Introduction
- System Model
- Problem Formulation
- Solution Approach
- Simulation Results

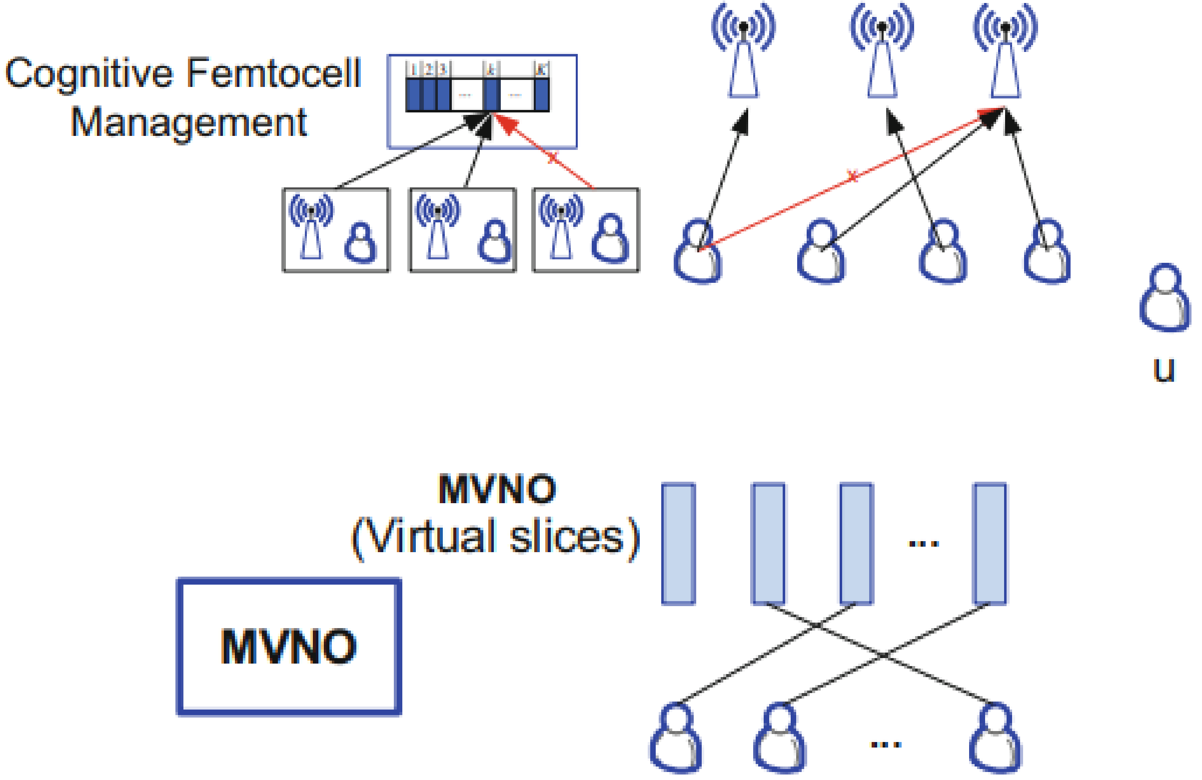
- Traditionally in cellular networks radio resources were only considered as a performance bottleneck
- A number of solutions were devised to only cater the radio resource allocation challenge
- The proliferation of end users and novel applications have also imposed limitations on other network resources such as backhaul and cache spaces

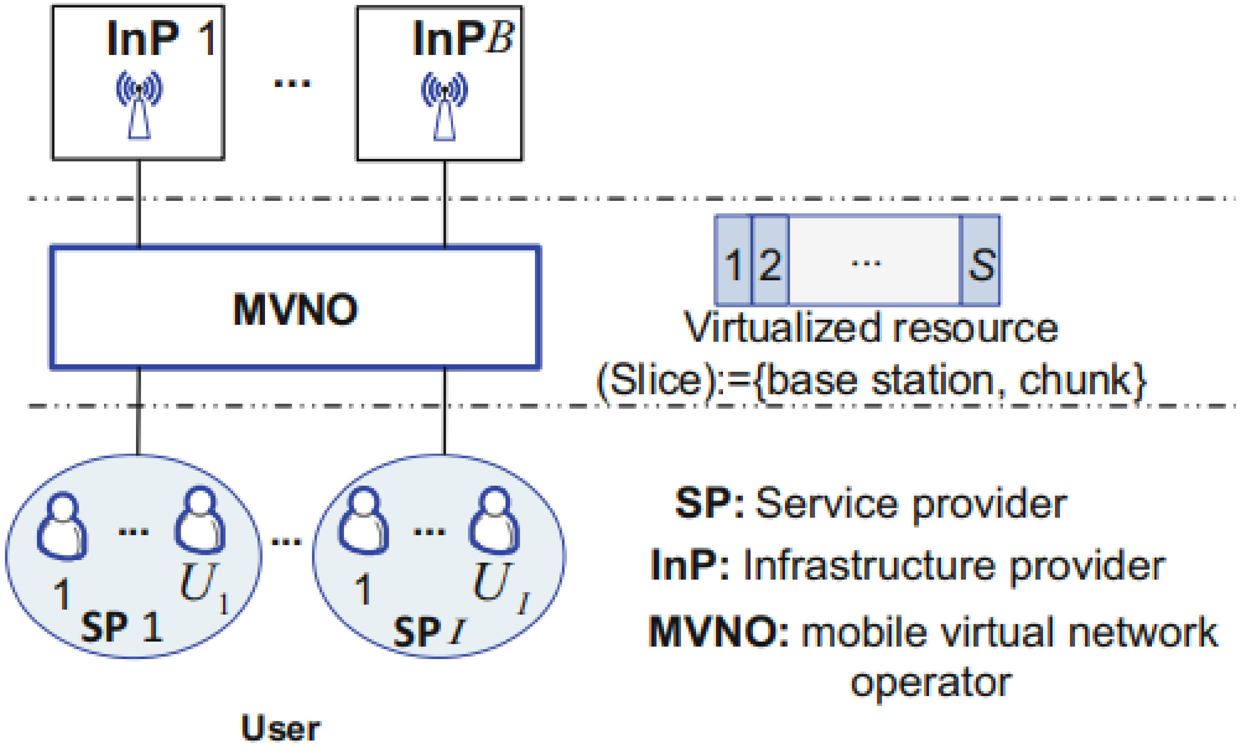
System Model

Set of Infrastructure provider $\mathcal{B} = \{1, 2, \ldots, B\}$

Set of Service provider $\mathcal{I} = \{1, 2, \ldots, I\}$

Set of user $\mathcal{U}^i = \{1, 2, \ldots, U_i\}$ $i \in \mathcal{I}$



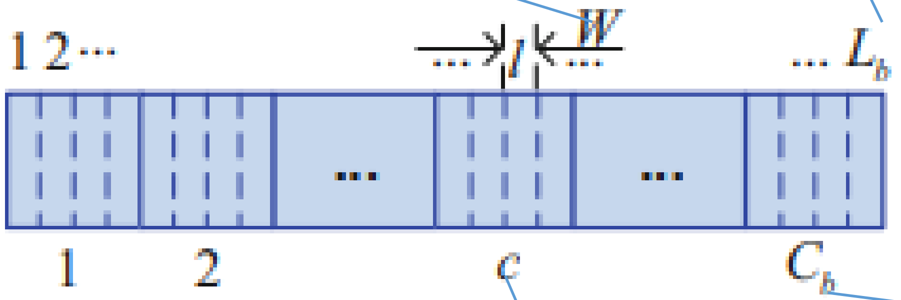


Virtualized cellular network hierarchical model

S.M. Ahsan Kazmi, Latif U. Khan, Nguyen H. Tran, Choong Seon Hong, "Network Slicing for 5G and Beyond Networks," ISBN 978-3-030-16169-9, Springer

The bandwidth of narrowband orthogonal subcarriers is W

Subcarriers set $\mathcal{L}_b = \{1, 2, \ldots, L_b\}$.



$\mathcal{C}_b = \{1, 2, \ldots, C_b\}$ chunks

Chunk-based radio resources of InP b

Every chunk is formed by aggregating $\mathcal{L}_{b,c} = \{1, 2, \ldots, L_b/C_b\}$ subcarriers

Data rate of a user u associated with service provider i

$\mathcal{B}$	Set of InPs
$\mathcal{I}$	Set of virtual resources given by MVNO to SPs
$\mathcal{U}_m$	Set of customers connected with SPs
$\mathcal{L}_b$	Set of radio resources subcarriers owned by InPs
$\mathcal{C}_b$	Set of subcarriers chunks
α	Slice allocation matrix
$\alpha^{\mu}_{b,c}$	Binary indicator matrix
R^i_{μ}	Data rate
σ^2_b	Background noise
W	Bandwidth
$P^{\mu}_{b,c}$	Transmission power vector
$Z_{b,bh}$	Predefined Backhaul capacity
ϕ^{sp}_i	The payment (in units/Mbps) of each SP i to the MVNO
ϕ^{bh}_b	Unit price (in units/Mbps) of the Backhaul set by InP
$L_{\alpha,P,\lambda,\beta}$	Lagrangian function
λ	Lagrangian nonnegative multiplier
β	Lagrangian nonnegative multiplier
μ	Lagrangian nonnegative multiplier
$\omega^{\mu}_{b,c}$	Used in Lagrangian dual function
ϕ^k_u	Utility function

Binary indicator matrix

$$R^i_u(\boldsymbol{\alpha}, \boldsymbol{P}) = \sum_{b\in\mathcal{B}} \sum_{c\in\mathcal{C}_b} \alpha^u_{b,c} r^u_{b,c}(\boldsymbol{P}^u_{b,c}),$$

instantaneous channel gain

$$\gamma^{u,l}_{b,c} = \frac{g^{u,l}_{b,c}}{\sigma^2_b};$$

Data rate $r^u_{b,c}(\boldsymbol{P}^u_{b,c}) = \sum_{l\in\mathcal{L}_{b,c}} W \log_2(1+\gamma^{u,l}_{b,c} P^{u,l}_{b,c})$

transmit power

$$U_{\text{MVNO}}(\boldsymbol{\alpha}, \boldsymbol{P}) = U^{\text{rev}}(\boldsymbol{\alpha}, \boldsymbol{P}) - U^{\text{cost}}(\boldsymbol{\alpha}, \boldsymbol{P}), \quad (6.4)$$

How to maximize the Profit of MVNO?

Problem Formulation

$$\max_{(\boldsymbol{\alpha}, \boldsymbol{P})} \quad U_{\mathrm{MVNO}}(\boldsymbol{\alpha}, \boldsymbol{P}) \tag{6.5}$$

s.t.

$$R_u^i(\boldsymbol{\alpha}, \boldsymbol{P}) \geq R_{u,\min}^i, \ \forall u \in \mathcal{U}, \forall i \in \mathcal{I}. \tag{6.2}$$

(Guaranteeing the required minimum rate)

$$\sum_{i \in \mathcal{I}} \sum_{u \in \mathcal{U}_i : \alpha_{b,c}^u = 1} R_u^i(\boldsymbol{\alpha}, \boldsymbol{P}) \leq Z_{b,\mathrm{bh}}, \ \forall b \in \mathcal{B}, \ \forall c \in \mathcal{C}_b, \tag{6.3}$$

(Aggregated data rate of users)

$$\sum_{b \in \mathcal{B}} \sum_{c \in \mathcal{C}_b} \alpha_{b,c}^u \sum_{l \in \mathcal{L}_c} P_{b,c}^{u,l} \leq \bar{P}_u, \ \ \forall u \in \mathcal{U}, \tag{6.6}$$

(Total transmit power)

$$P_{b,c}^{u,l} \geq 0, \ \forall b \in \mathcal{B}, \forall c \in \mathcal{C}_b, \forall u \in \mathcal{U}, \tag{6.7}$$

$$\alpha_{b,c}^u \in \Pi_{\boldsymbol{\alpha}}, \ \forall b \in \mathcal{B}, \forall c \in \mathcal{C}_b, \forall u \in \mathcal{U}, \tag{6.8}$$

$$\sum\nolimits_{u \in \mathcal{U}} \alpha_{b,c}^u \leq 1, \ \ \forall c \in \mathcal{C}_b, \ \forall b \in \mathcal{B}, \tag{6.9}$$

(Restricts the allocation of a slice to at most user)

$$\sum\nolimits_{b \in \mathcal{B}} \sum\nolimits_{c \in \mathcal{C}_b} \alpha_{b,c}^u \leq 1, \ \ \forall u \in \mathcal{U}, \tag{6.10}$$

(Isolation of the slices)

$$\sum\nolimits_{u \in \mathcal{U}} \sum\nolimits_{c \in \mathcal{C}_b} \alpha_{b,c}^u \leq 1, \ \ \forall b \in \mathcal{B}, \tag{6.11}$$

(Isolation of the slices)

$$\sum\nolimits_{u \in \mathcal{U}} \sum\nolimits_{b \in \mathcal{B}} \alpha_{b,c}^u \leq 1, \ \ \forall c \in \mathcal{C}_b, \tag{6.12}$$

(Isolation of the slices)

$$^1\alpha_{b,c}^u = \{0, 1\}, \forall u \in \mathcal{U}, \ \forall b \in \mathcal{B}, \ \forall c \in \mathcal{C}_b. \tag{6.13}$$

NETWORKING INTELLIGENCE LAB

The Lagrangian function of (6.5) is as follows

$$L(P,\lambda,\mu,\beta) = U_{MVNO}(\alpha,P) + \lambda_i\left(R_u^i(\lambda,\beta) - R_{u,min}^i\right) + \beta_i\left(Z_{b,bh} - \sum_{i\in I}\sum_{u\in U} R_u^i(\alpha,\beta)\right) + \mu_i(\bar{P}_u - \sum_{b\in B}\sum_{c\in C_b}\alpha_{b,c}^u \sum_{l\in L_c} P_{b,c}^{u,l})$$

Solution Approach-1

Algorithm 1 JSPA-HSA: JSPA with Hungarian-based slice allocation

Initialization: $\mathcal{I}, \mathcal{B}, \mathcal{C}_b, \mathcal{U}_i$, $\boldsymbol{P}^{(0)}$, $\boldsymbol{\lambda}^{(0)}$, $\boldsymbol{\mu}^{(0)}$, and $\boldsymbol{\beta}^{(0)}$.
Repeat:
Power allocation phase:
*At the subscribed user u:
Update λ_u as: (Transmission rate)

$$\lambda_u(t+1) = [\lambda_u(t) - s_1(t)(R_u^i(\boldsymbol{\alpha}, \boldsymbol{P}) - R_u^{\min})]^+; \quad (6.18)$$

Update μ_u as: (Transmit power)

$$\mu_u(t+1) = \left[\mu_u(t) - s_2(t)\left(\sum_{b\in\mathcal{B}}\sum_{c\in\mathcal{C}_b}\alpha_{b,c}^u \sum_{l\in\mathcal{L}_c} P_{b,c}^{u,l} - \bar{P}_u\right)\right]^+; \quad (6.19)$$

Update transmit power $P_{b,c}^{u,l}(t+1)$ by $P_{b,c}^{u,l*} = \left[\frac{\varphi_i^{\text{sp}} - \varphi_b^{\text{bh}} + \lambda_u - \beta_b}{(\ln 2/W)\mu_u} - \frac{1}{\gamma_{b,c}^{u,l}}\right]^+, \quad (6.16)$
*At the SBS b:
Update congested backhaul link price $\beta_b(t+1)$:

(Backhaul data rate)

$$\beta_b(t+1) = \left[\beta_b(t) + s_3(t)\left(\sum_{i\in\mathcal{I}}\sum_{u\in\mathcal{U}_i} R_u^i(\boldsymbol{\alpha}, \boldsymbol{P}) - Z_{b,\text{bh}}\right)\right]^+; \quad (6.20)$$

Slice allocation phase:
*At the MVNO:
Update $\alpha_{b,c}^u(t+1)$ using the Hungarian algorithm to maximize

$$\max_{(\boldsymbol{\alpha},\boldsymbol{P})} \sum_{i\in\mathcal{I}}\sum_{u\in\mathcal{U}^i}\sum_{b\in\mathcal{B}}\sum_{c\in\mathcal{C}_b}\alpha_{b,c}^u\left[R_{b,c}^u(\boldsymbol{P}_{b,c}^{u*}) - \varphi_{b,c}^{\text{slice}}\right] \quad (6.17)$$

Until $|\lambda_u(t+1) - \lambda_u(t)| \le \epsilon_1$, $|\mu_u(t+1) - \mu_u(t)| \le \epsilon_2$, and $|\beta_b(t+1) - \beta_b(t)| \le \epsilon_3$ are simultaneously satisfied.

Lagrangian multiplier

Unit price

Data rate function

Matching-Based Low-Complexity Algorithm

Algorithm 2 MSA: matching-based slice allocation

```
while $\sum_{\forall u,\{b,c\}} b_{u\to\{b,c\}} \neq 0$ or convergence not achieved do
    At the subscribed users:
        Send a bid for the slice $\{b,c\}^* = \arg\max_{\{b,c\}\in\succ_u} \phi_u(\{b,c\})$.
    At the MVNO:
        Construct $\succ_{\{b,c\}}$ based on (6.21).
        Update $\{b,c\}^* = \mu(\{b,c\})|u^* = \arg\max_{u\in\succ_{\{b,c\}}} \phi_{\{b,c\}}(u)\}$.
        Update the rejected user lists on the slices and the preference $\succ_u$.
end while
```

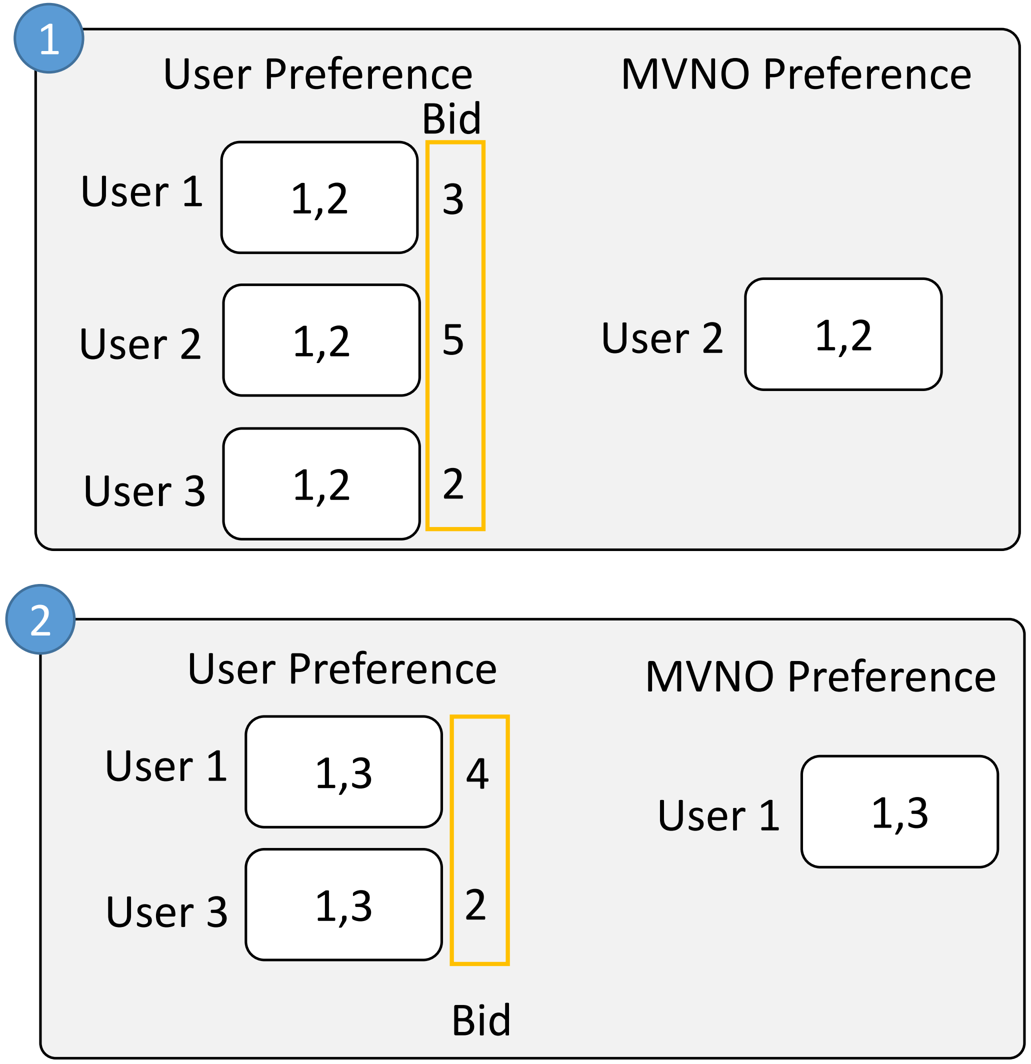
Users side

MVNO side

Unit price

$$\phi_u(k) = R^u_{b,c}(P^u_{b,c}) - \varphi^{\text{slice}}_{b,c}. \qquad (6.21)$$

Data rate function



S.M. Ahsan Kazmi, Latif U. Khan, Nguyen H. Tran, Choong Seon Hong, "Network Slicing for 5G and Beyond Networks," ISBN 978-3-030-16169-9, Springer

Simulation Results

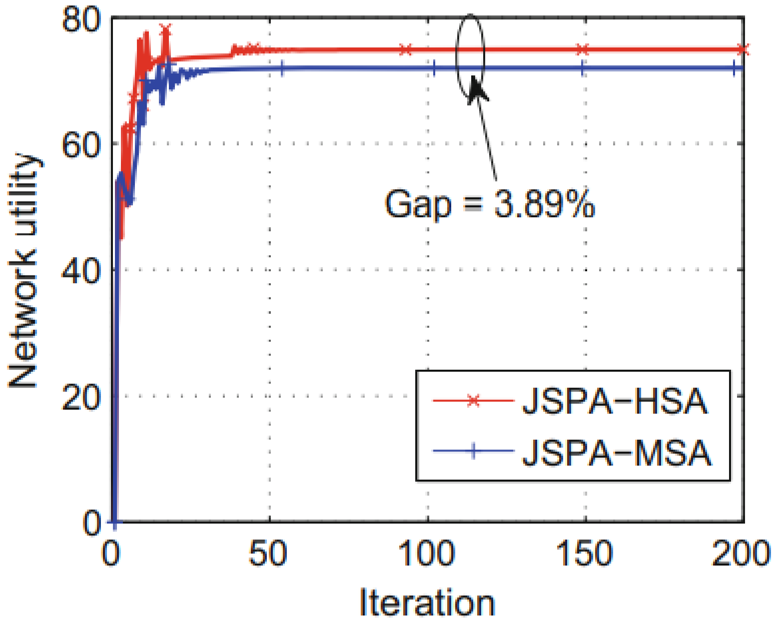


Fig. 6.3 Evaluation results of Network utility with $Z_{b,\text{bh}} = 10\,\text{Mbps}$

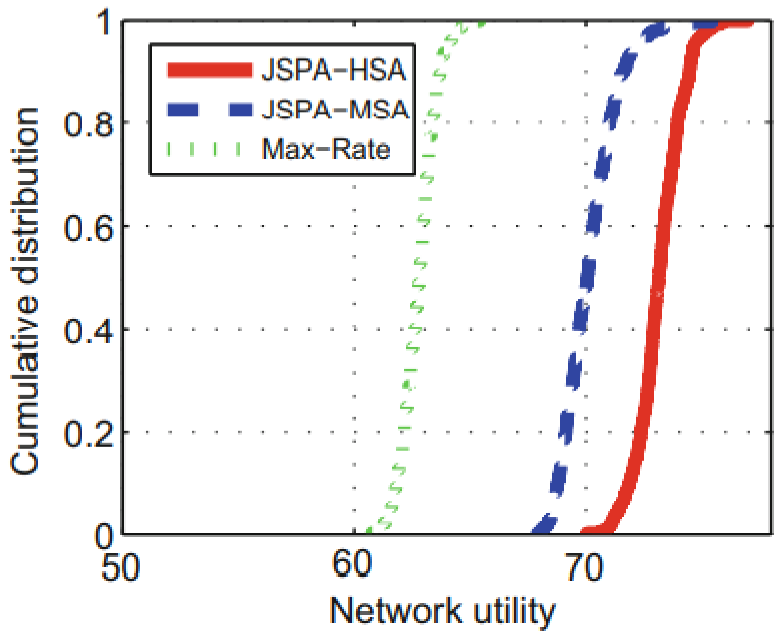


Fig. 6.4 CDF of network utilities

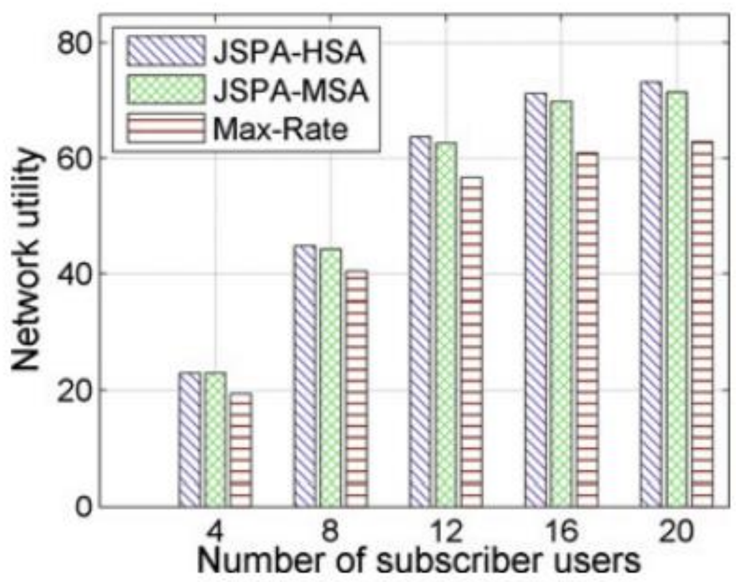


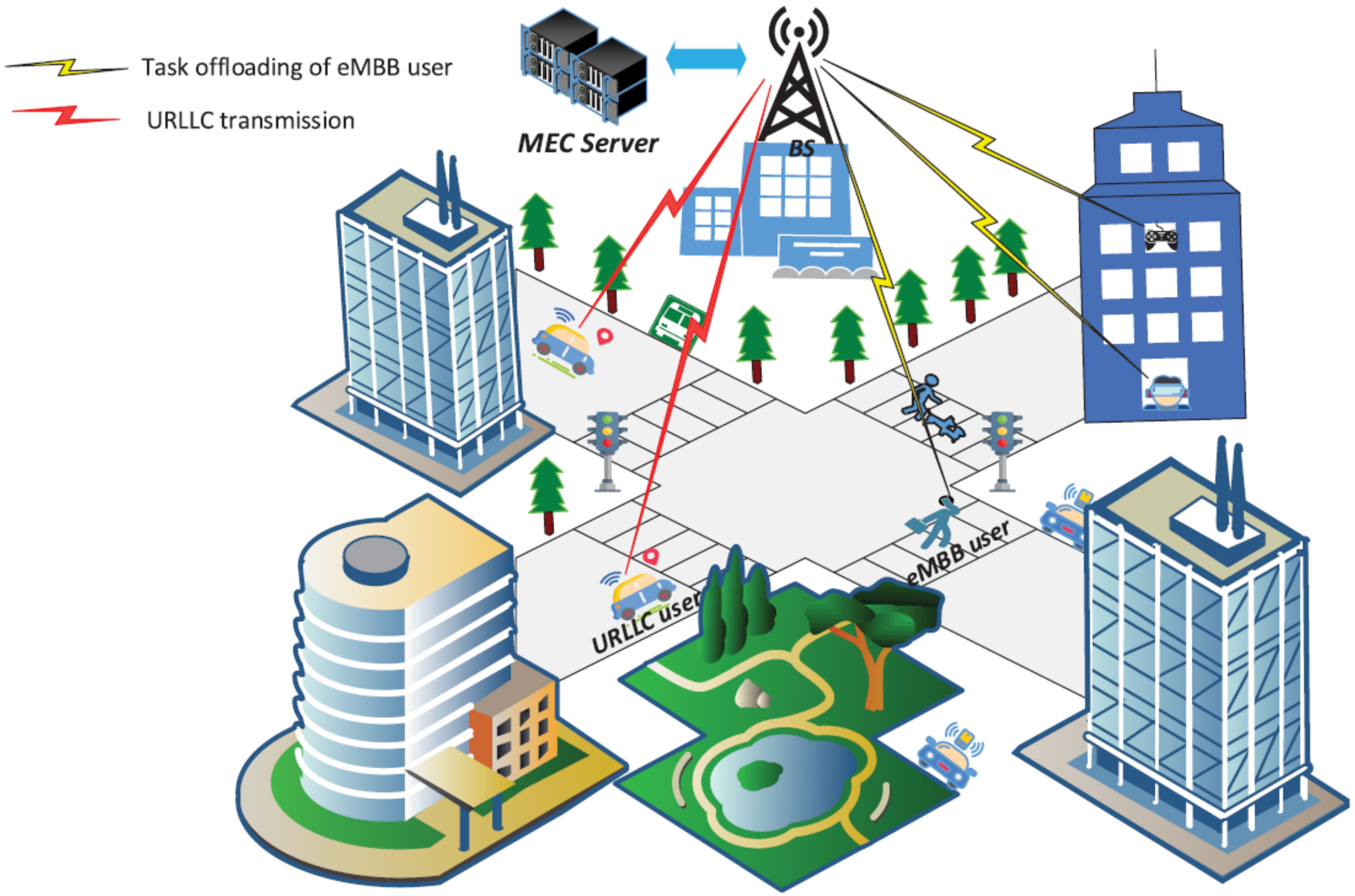
Fig. 6.5 Network utility versus number of subscriber users

S.M. Ahsan Kazmi, Latif U. Khan, Nguyen H. Tran, Choong Seon Hong, "Network Slicing for 5G and Beyond Networks," ISBN 978-3-030-16169-9, Springer

Use Case 3: Energy Efficient Communication and Computation Resource Slicing for eMBB and URLLC Coexistence in 5G and Beyond

- System Model
- Problem Formulation
- Solution Approach
- Simulation Results

Use Case 3 – System Model



Yan Kyaw Tun, Do Hyon Kim, Madyan Alsenwi, Nguyen H. Tran, Zhu Han, Choong Seon Hong, "Energy Efficient Communication and Computation Resource Slicing for eMBB and URLLC Coexistence in 5G and Beyond," IEEE Access, Vol.8, pp.136024-136035, Jul. 2020

Minimize the overall energy consumption

Notation	Definition
$\mathcal{U}$	Set of eMBB users, $\lvert\mathcal{U}\rvert = U$
F	Total system bandwidth
F_b	Fraction of system bandwidth allocated to eMBB users
F_s	Fraction of system bandwidth allocated to URLLC users
$\mathcal{B}$	Set of resource blocks, $\lvert\mathcal{B}\rvert = B$
d_u	Total input data size of eMBB user u
c_u	Required CPU cycles to accomplish one bit of the input data of eMBB user u
T_u	The execution deadline of the task of eMBB user u
l_u	The offloaded data size of the task of eMBB user u
t_u^L	The local computation execution time of eMBB user u
E_u^L	The local computation energy of eMBB user u
y_u^b	Resource block assignment variable
M	Number of minislots divided in each resource block
L_m	Traffic of URLLC users at minislot m
$L_{\max}$	Maximum traffic of URLLC users that can be served at a time slot
w_u	Weight of puncturing eMBB user u
g_u^b	Achievable channel gain of eMBB user u
P_u^b	Transmit power of eMBB user u
$R_{u,b}$	Achievable data rate of eMBB user u on resource block b
t_u^{up}	The uplink transmission delay experienced by eMBB user u
f^C	The total CPU capacity of the MEC server
f_u^C	The CPU capacity of the MEC server that is allocated to eMBB user u
E_u^{Off}	The energy consumption of eMBB user u for offloading data
R_s	The achievable data rate of URLLC user s
P_s	The transmit power of URLLC user s
g_s	The achievable channel gain of URLLC user s
R_{urllc}	The total achievable data rate of URLLC users

Energy Consumption of eMBB users

weight

Data rate of eMBB user

$$\min_{y,l,w} \left(\sum_{u=1}^{U} E_u^{\text{Off}} + \sum_{u=1}^{U} E_u^{L}\right) - \phi \sum_{u=1}^{U}\sum_{b=1}^{B} y_u^b R_{u,b} \quad (18)$$

$$\text{s.t. C1}: \frac{l_u}{R_u} + \frac{c_u l_u}{f_u^C} \le T_u,\ u \in \mathcal{U}, \quad (19)$$

execution latency constraint of the eMBB users

$$\text{C2}: \frac{c_u(d_u - l_u)}{f_u^l} \le T_u, \quad u \in \mathcal{U}, \quad (20)$$

execution latency constraint of the eMBB users

$$\text{C3}: l_u \le d_u, \quad \forall u \in \mathcal{U}, \quad (21)$$

offloading data size of user u has to be less than the total input data size

$$\text{C4}: 0 \le w_u \le 1, \quad \forall u \in \mathcal{U}, \quad (22)$$

weight parameter for eMBB user u that can be punctured

$$\text{C5}: \text{CVaR}_\beta(R) \le \alpha, \quad (23)$$

reliability constraints of URLLC users

$$\text{C6}: Pr[R_{urllc} \le L] \le \epsilon, \quad (24)$$

reliability constraints of eMBB users

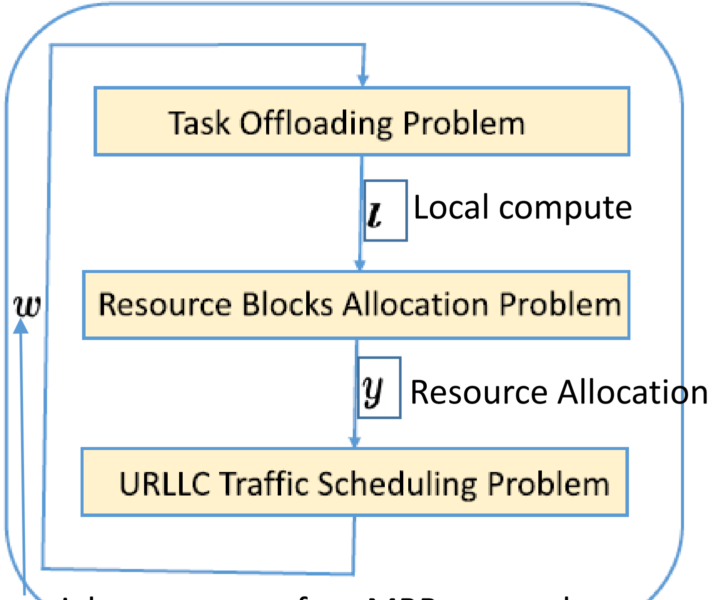
$$\text{C7}: \sum_{u=1}^{U} y_u^b \le 1, \quad \forall b \in \mathcal{B}, \quad (25)$$

one resource block can be allocated to only one eMBB user

$$\text{C8}: y_u^b \in \{0, 1\}, \quad \forall u \in \mathcal{U}, \forall b \in \mathcal{B}, (26) \quad (26)$$

Yan Kyaw Tun, Do Hyon Kim, Madyan Alsenwi, Nguyen H. Tran, Zhu Han, Choong Seon Hong, "Energy Efficient Communication and Computation Resource Slicing for eMBB and URLLC Coexistence in 5G and Beyond," IEEE Access, Vol.8, pp.136024-136035, Jul. 2020

BLOCK COORDINATE DESCENT BASED SOLUTION



weight parameter for eMBB user u that can be punctured by traffic of URLLC users

Optimization framework

Algorithm 1 Iterative Algorithm for the Relaxed Problem

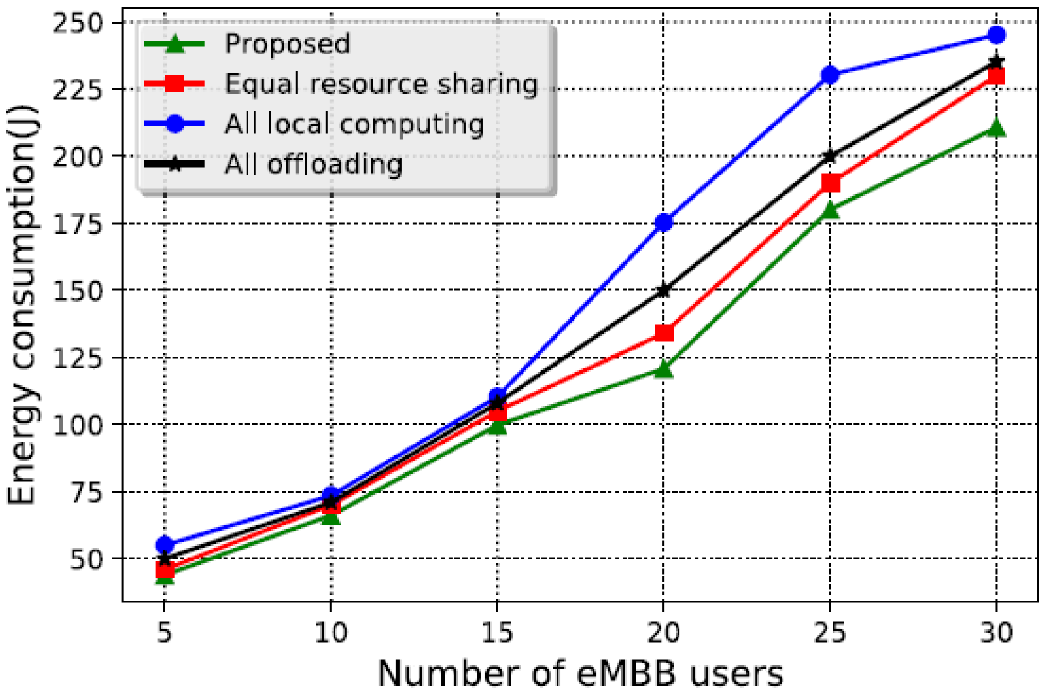
1: **Initialization:** Set $k = 0$, $\epsilon_1, \epsilon_2, \epsilon_3 > 0$, and initial solutions $(\boldsymbol{l}^{(0)}, \boldsymbol{y}^{(0)}, \boldsymbol{w}^{(0)})$;
2: **repeat**
3: Compute $\boldsymbol{l}^{(k+1)}$ from (P1) at given $\boldsymbol{y}^k$, and $\boldsymbol{w}^k$;
4: Compute $\boldsymbol{y}^{(k+1)}$ from (P2) at given $\boldsymbol{l}^{(k+1)}$, and $\boldsymbol{w}^k$;
5: Compute $\boldsymbol{w}^{(k+1)}$ from (P3) at given $\boldsymbol{l}^{(k+1)}$, and $\boldsymbol{y}^{(k+1)}$;
6: $k = k + 1$;
7: **until** $\| \boldsymbol{l}^{(k+1)} - \boldsymbol{l}^{(k)} \| \leq \epsilon_1$, $\| \boldsymbol{y}^{(k+1)} - \boldsymbol{y}^{(k)} \| \leq \epsilon_2$, and $\| \boldsymbol{w}^{(k+1)} - \boldsymbol{w}^{(k)} \| \leq \epsilon_3$;
8: Then, set $\left(\boldsymbol{l}^{(k+1)}, \boldsymbol{y}^{(k+1)}, \boldsymbol{w}^{(k+1)}\right)$ as the desired solution.

$$(\text{P1}): \min_{\boldsymbol{l}} \sum_{u=1}^{U} E_u^{\text{Off}} + \sum_{u=1}^{U} E_u^{L} \quad (30)$$
$$\text{s.t. (C1)-(C3),} \quad (31)$$

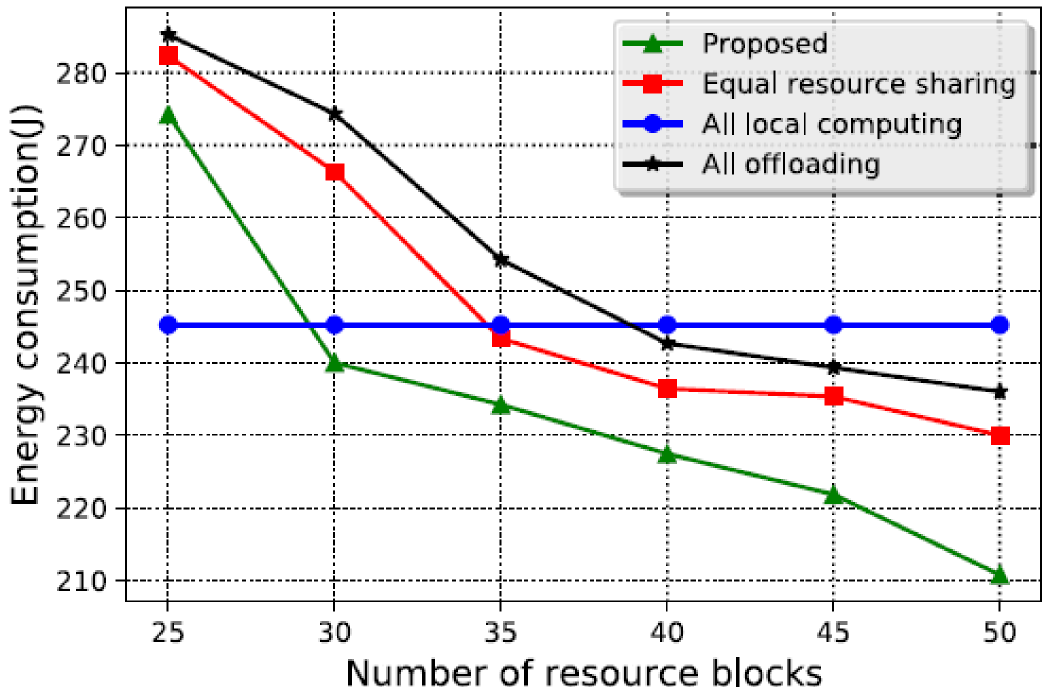
$$(\text{P2}): \min_{\boldsymbol{y}} \sum_{u=1}^{U} E_u^{\text{Off}} - \phi \sum_{u=1}^{U} \sum_{b=1}^{B} y_u^b R_{u,b} \quad (34)$$
$$\text{s.t. (C1), (C7), and (C8),} \quad (35)$$

$$(\text{P3}): \min_{\boldsymbol{w}} \sum_{u=1}^{U} E_u^{\text{Off}} - \phi \sum_{u=1}^{U} \sum_{b=1}^{B} y_u^b R_{u,b} \quad (41)$$
$$\text{s.t. (C4), (C5), and (C6),} \quad (42)$$

Yan Kyaw Tun, Do Hyon Kim, Madyan Alsenwi, Nguyen H. Tran, Zhu Han, Choong Seon Hong, "Energy Efficient Communication and Computation Resource Slicing for eMBB and URLLC Coexistence in 5G and Beyond," IEEE Access, Vol.8, pp.136024-136035, Jul. 2020

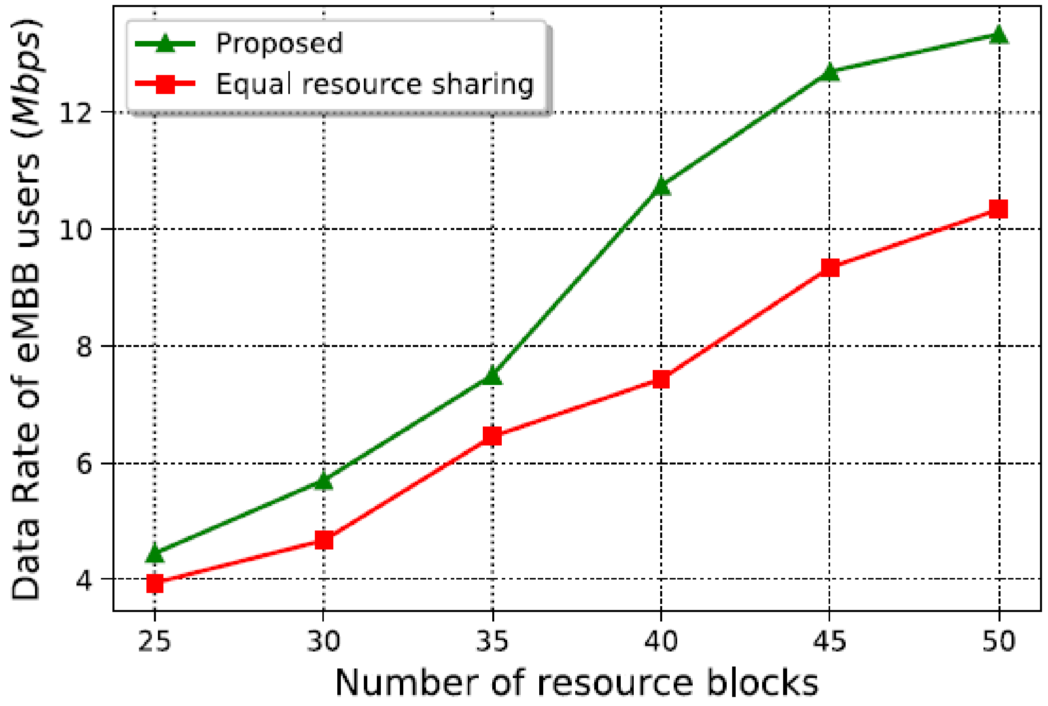


Energy consumption under different number of users.

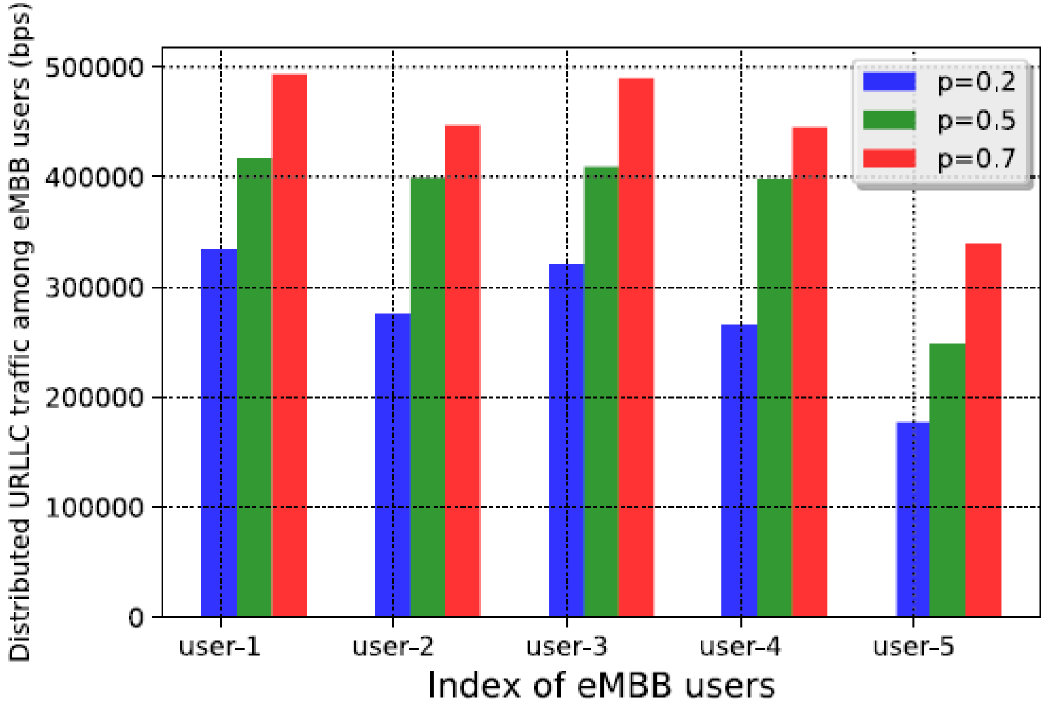


Energy consumption under different number of resource blocks

Yan Kyaw Tun, Do Hyon Kim, Madyan Alsenwi, Nguyen H. Tran, Zhu Han, Choong Seon Hong, "Energy Efficient Communication and Computation Resource Slicing for eMBB and URLLC Coexistence in 5G and Beyond," IEEE Access, Vol.8, pp.136024-136035, Jul. 2020



Data rate of eMBB users under different number of resource blocks.



Data rate of URLLC users on the puncturing resource of eMBB users.

p : probability of URLLC traffic arrival

Yan Kyaw Tun, Do Hyon Kim, Madyan Alsenwi, Nguyen H. Tran, Zhu Han, Choong Seon Hong, "Energy Efficient Communication and Computation Resource Slicing for eMBB and URLLC Coexistence in 5G and Beyond," IEEE Access, Vol.8, pp.136024-136035, Jul. 2020

Concluding Remarks

- Open Issues
- Conclusion

Open Issues

- Dynamic Slice Allocation
  - A practical system would have users arriving and leaving a system with different demands at different time slots.

- Mobility Aware Network slicing
  - The current approaches for network slicing are not designed to handle mobility in the network.
  - Handling and orchestrating the radio access and core network will be very challenging in case of mobility.
  - Cosidering the migration of services for slicing from one point to other points in the network.

Conclusion

- This lecture is mainly focus to understand a full view of the resource management problem in 5G networks.

- We learned
  - The requirements and enabling technologies of 5G networks.
  - A detailed overview of network slicing that can be adapted to fulfill the 5G deliverables.
  - The recent research works' motivation, issues, challenges, and solutions.
  - Some open issues for future research and their potentials.

Conclusion

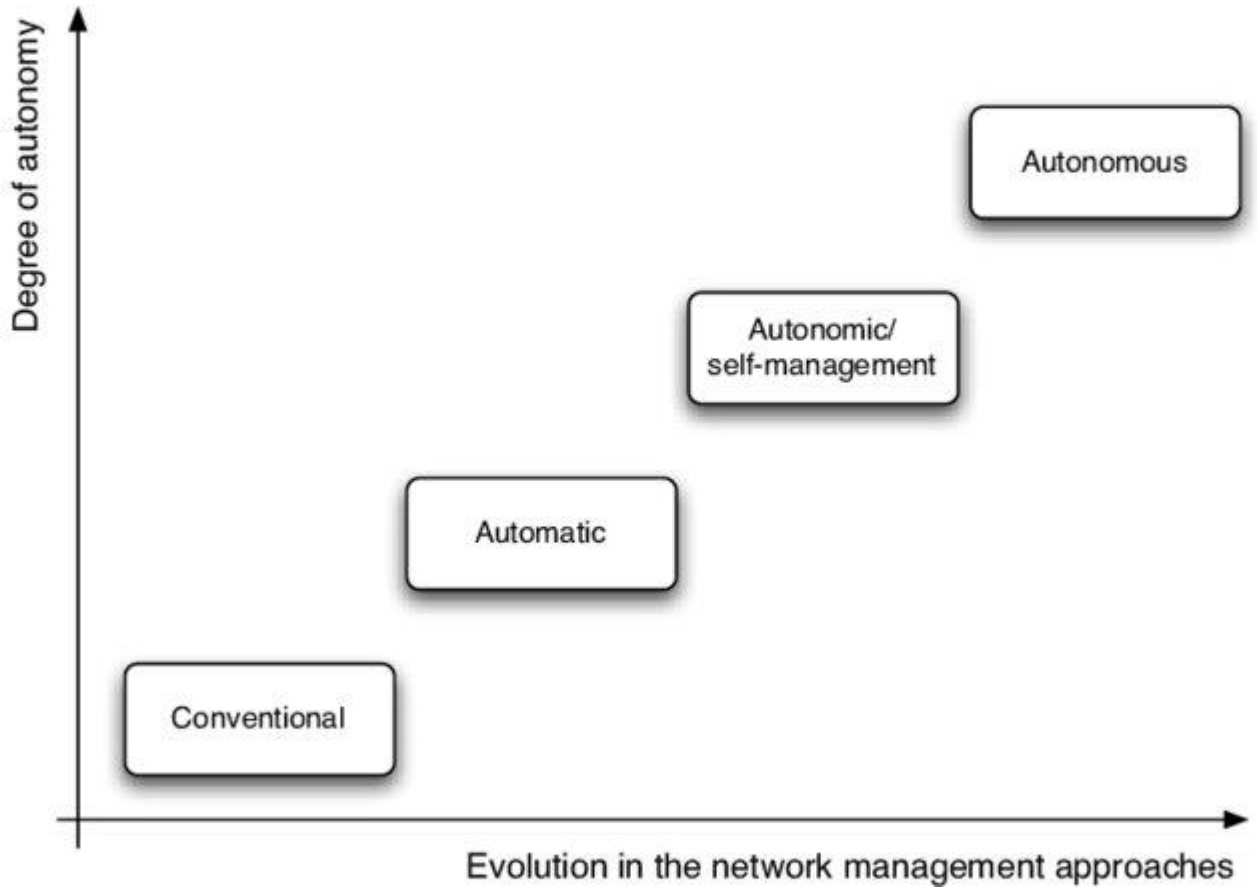
- This lecture is mainly focus to understand a full view of the resource management problem in 5G networks.

- We learned
  - The requirements and enabling technologies of 5G networks.

What is *missing*?

  deliverables.
  - The recent research works' motivation, issues, challenges, and solutions.
  - Some open issues for future research and their potentials.

Evolution in the Network Management Approaches



Next Lecture

- The use of artificial intelligence will play a vital role for enabling a variety of applications in 5G and beyond wireless networks.

- AI definitively provides precious opportunities to analyze trends and recognize patterns. However, it is difficult to perfectly predict the desired results by using traditional simple models such as shallow ANNs

- Deep Neural Networks are envisioned to fill this gap and serve as key predicting enabler to support the 5G networks

- Network slicing coupled with AI will be defining the future of wireless networks

Thanks !!!

Q&A