“AI based Network Resource Management (1)”

Choong Seon Hong

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

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
- Use case 4: Joint Communication, Computation, and Control for Computational Task Offloading in Vehicle-Assisted Multi-Access Edge Computing
- Use case 5: Collaboration in the Sky: A Distributed Framework for Task Offloading and Resource Allocation in Multi-Access Edge Computing

- Concluding Remarks

Introduction

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

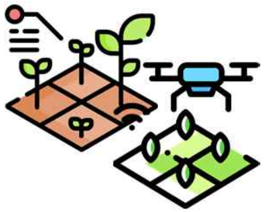
Introduction : Diverse Requirements

Smart Factory

Smart Entertainment

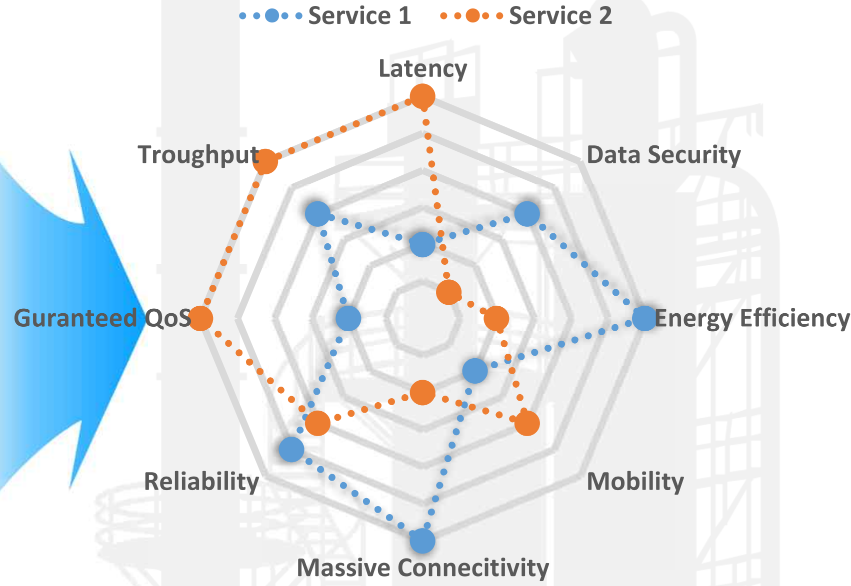
INDUSTRY

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

Introduction : Requirements in Transportation Industry

- 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

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

Introduction : Requirements in Health Sector

- 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.

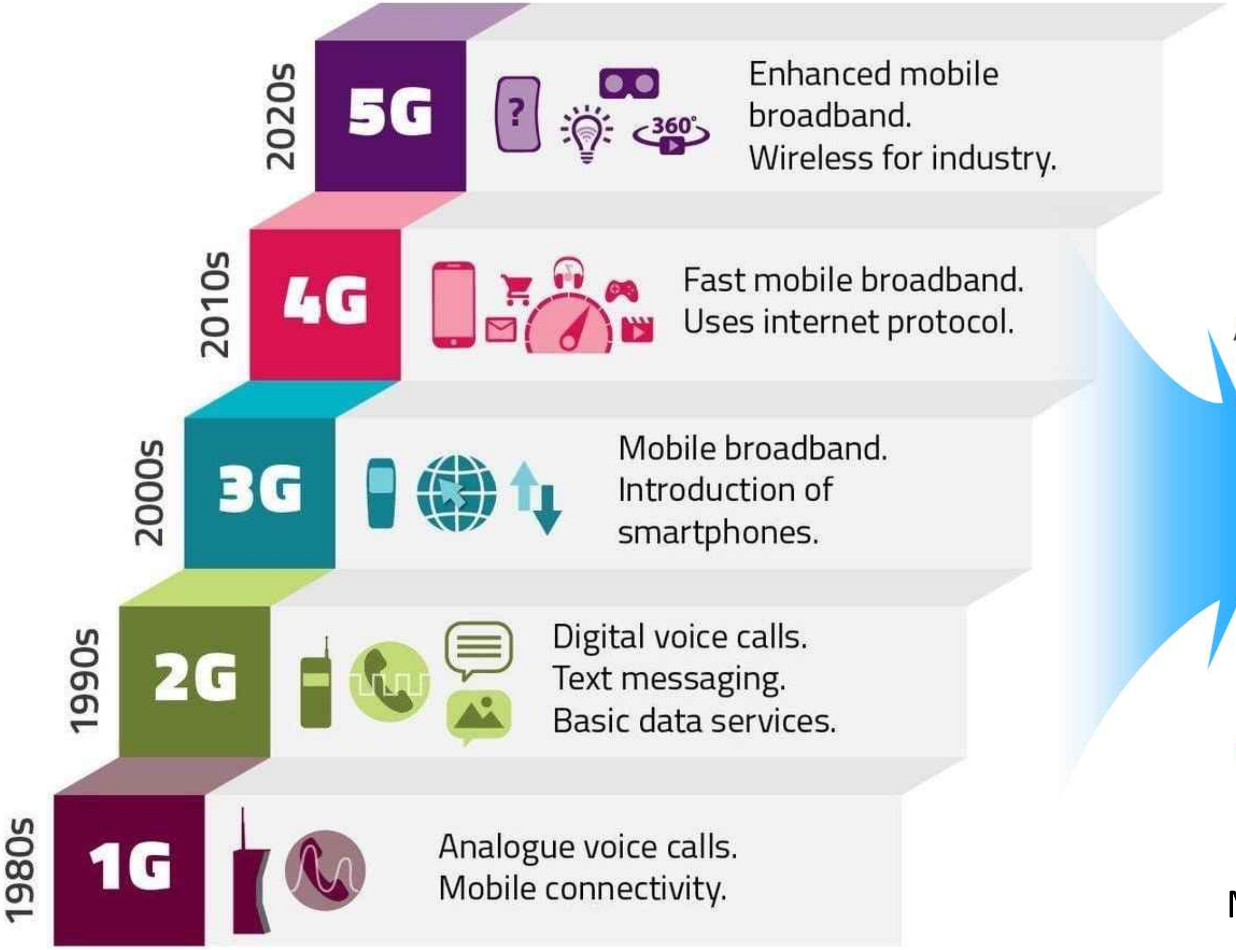
Example applications

- 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

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

Introduction : Evolution of Cellular Systems



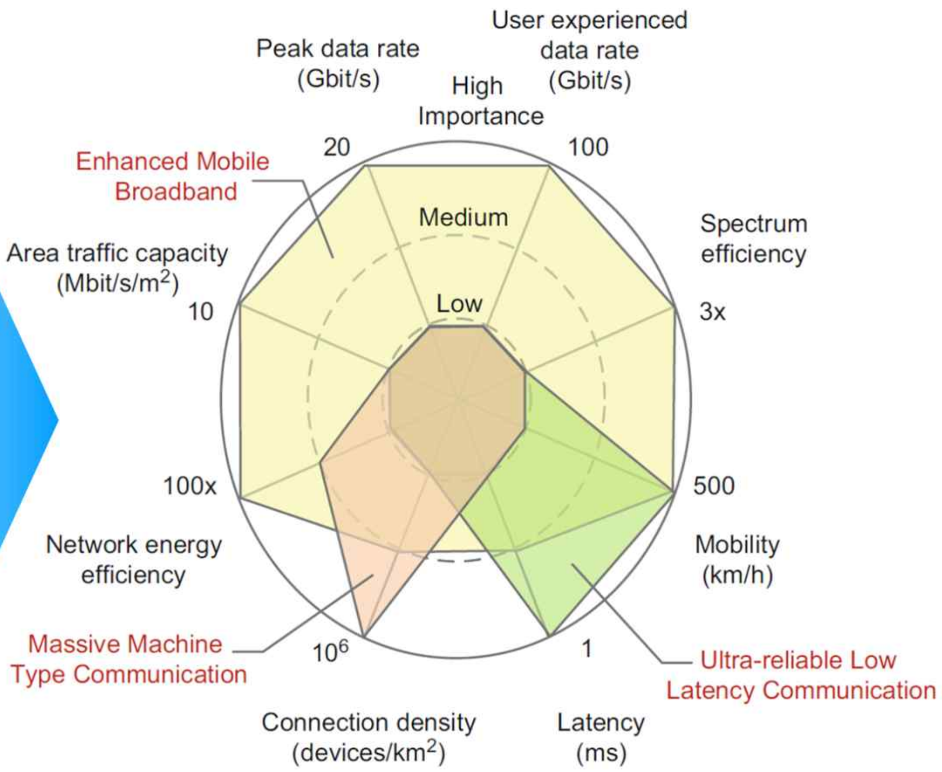
2020s **5G** Enhanced mobile broadband. Wireless for industry.

2010s **4G** Fast mobile broadband. Uses internet protocol.

2000s **3G** Mobile broadband. Introduction of smartphones.

1990s **2G** Digital voice calls. Text messaging. Basic data services.

1980s **1G** Analogue voice calls. Mobile connectivity.



Maximum 5G capabilities for different use cases

https://prc.chapters.comsoc.org/2019/04/01/5g-evolution-wireless-communications/

Introduction : Challenges to realize 5G Networks

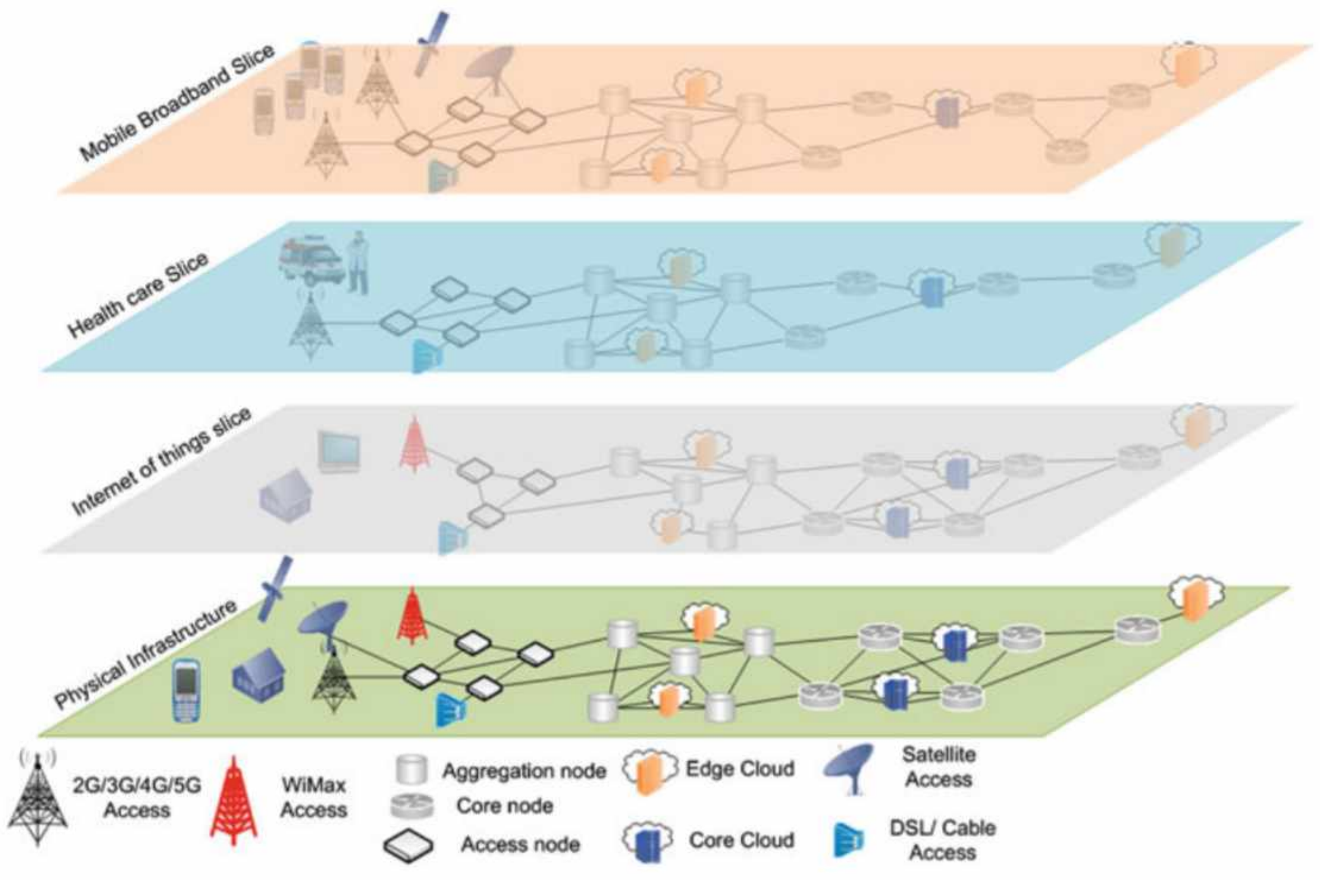
- Generally, to realize diverse 5G use cases there is need to resolve the given challenges:
  - Scalability and Reliability
  - Interoperability
  - Sustainability
  - Network Slicing
  - Security
  - Integration of AI

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.

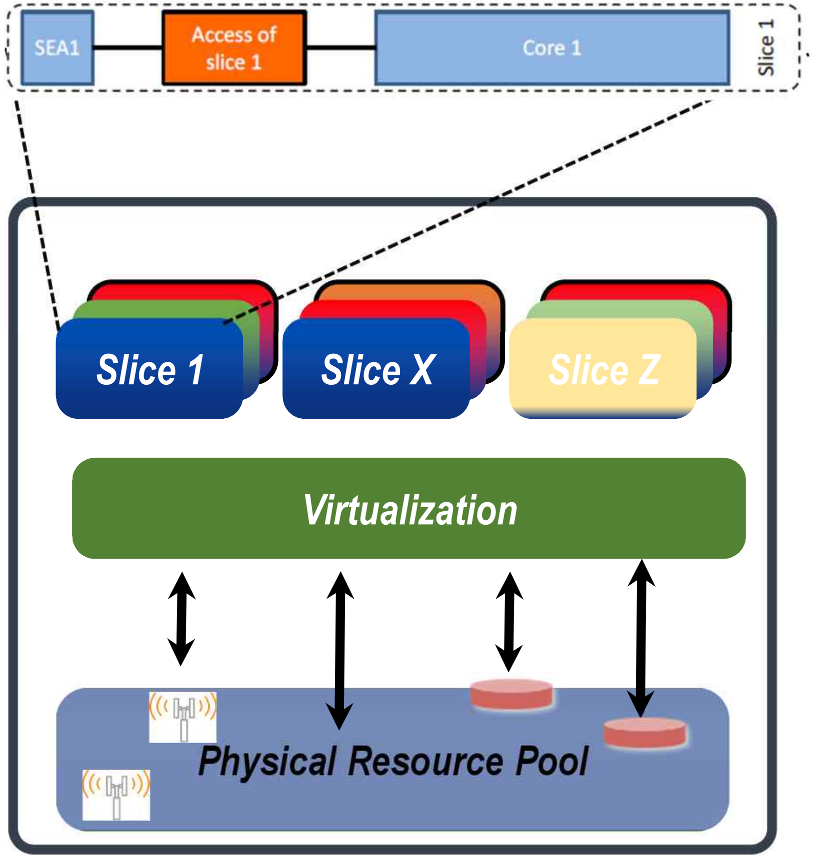


S.M. Ahsan Kazmi, Latif U. Khan, Nguyen H. Tran, Choong Seon Hong, "Network Slicing for 5G and Beyond Networks," ISBN 978-3-030-16169-9, Springer

Network Slicing

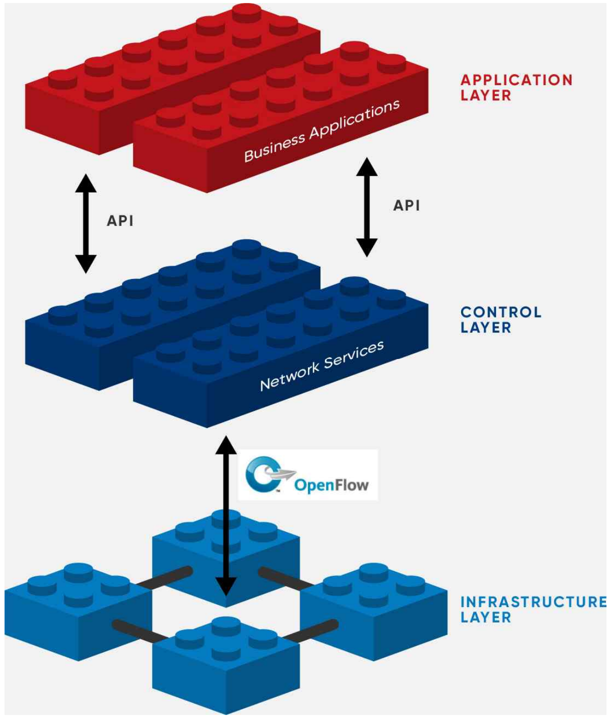
- Network slicing can fulfil the diverse requirements of these novel network services
- Network slicing enables one physical network into multiple, virtual, end-to-end (E2E) networks, each logically isolated including device, access, transport and core network
- A slice is dedicated for different types 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

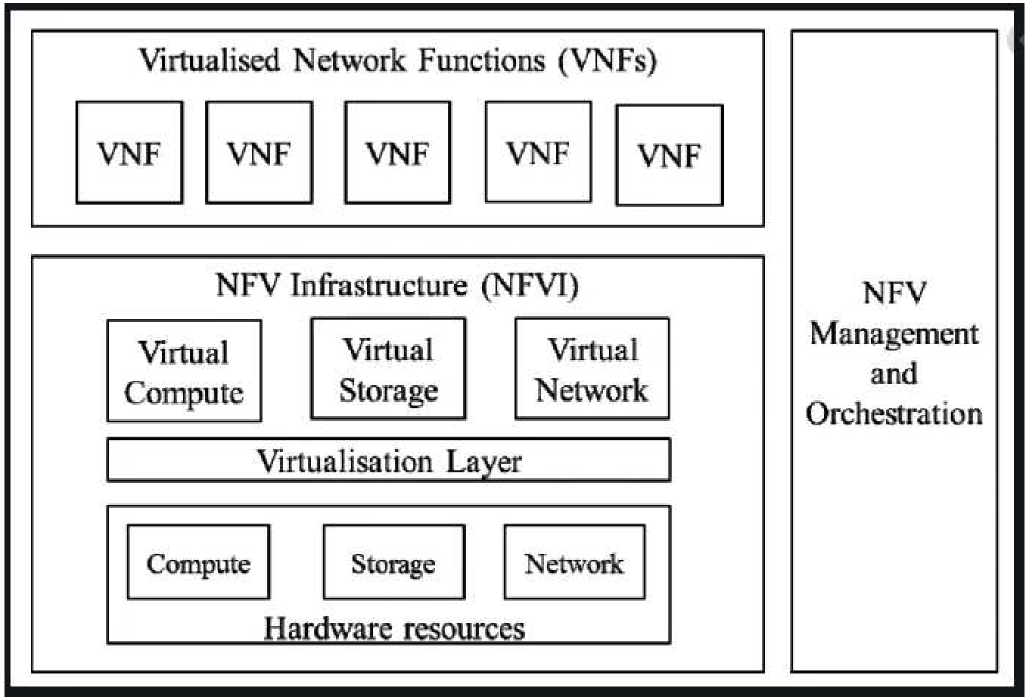


Network Slicing

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

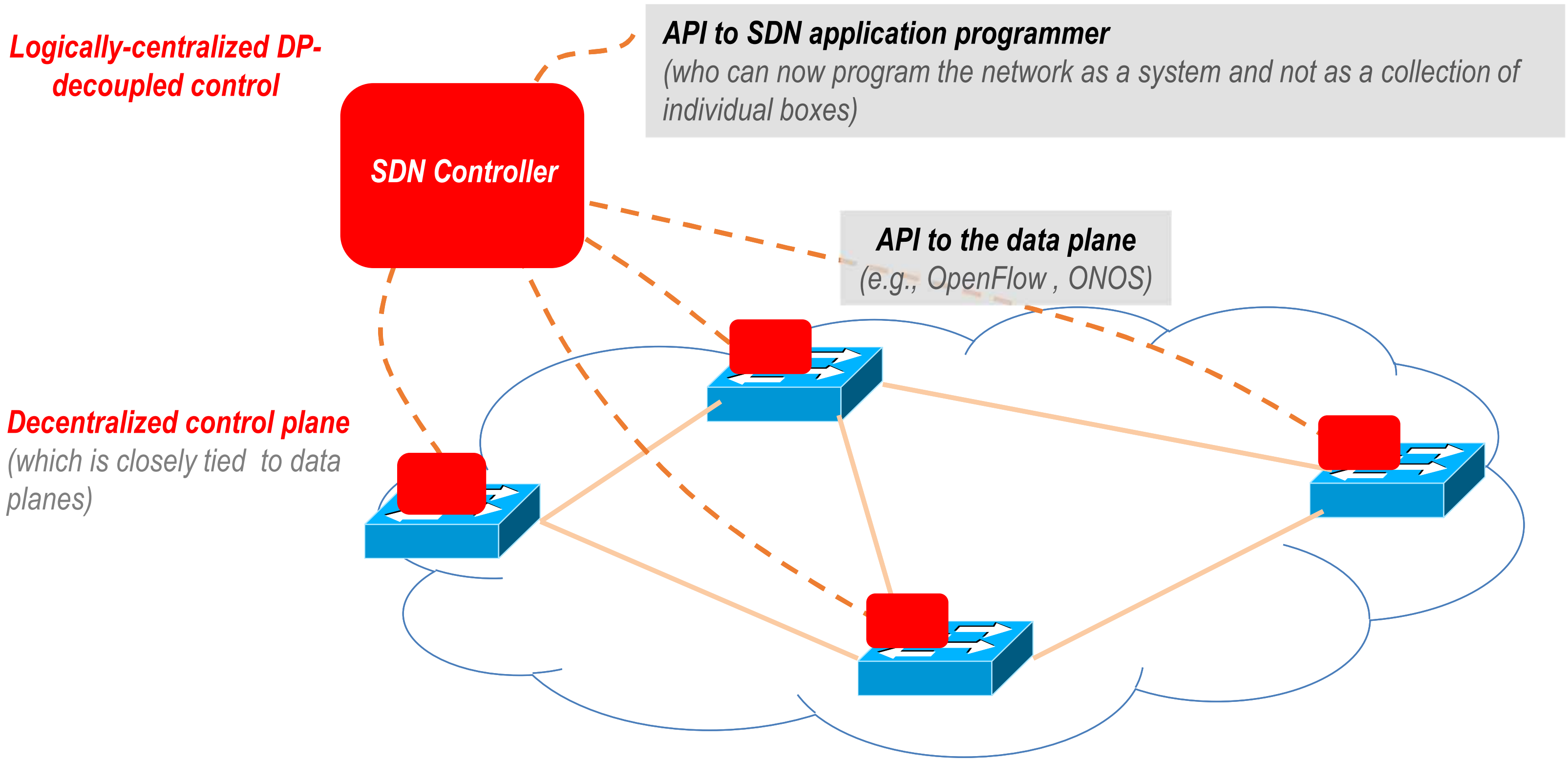


https://www.opennetworking.org/sdn-definition/



High-level NFV framework. Source: ETSI

Network slicing enablers: Software defined network (SDN)



Logically-centralized DP-decoupled control

API to SDN application programmer
(who can now program the network as a system and not as a collection of individual boxes)

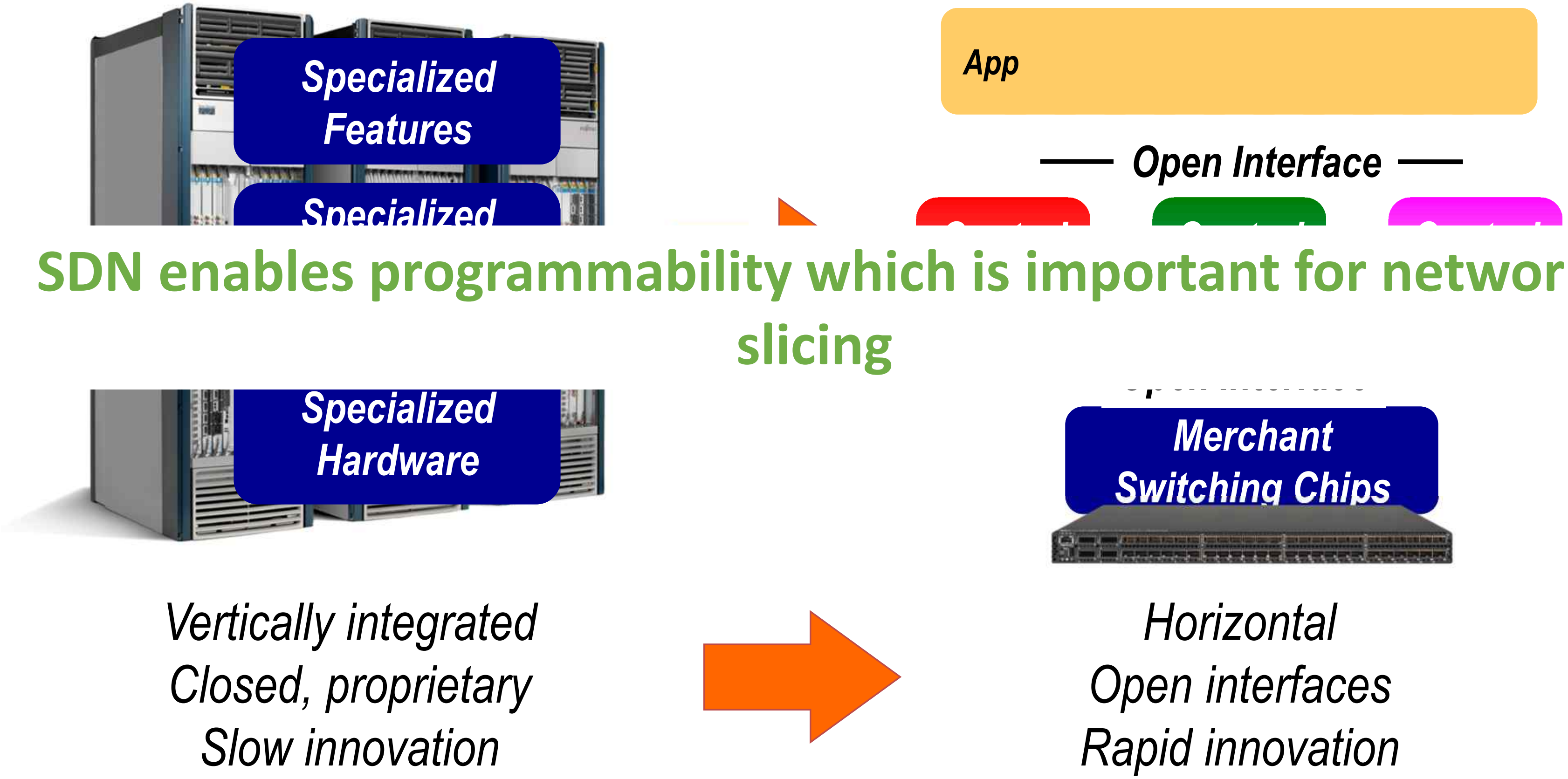
SDN Controller

API to the data plane
(e.g., OpenFlow , ONOS)

Decentralized control plane
(which is closely tied to data planes)

`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.`

SDN enables programmability



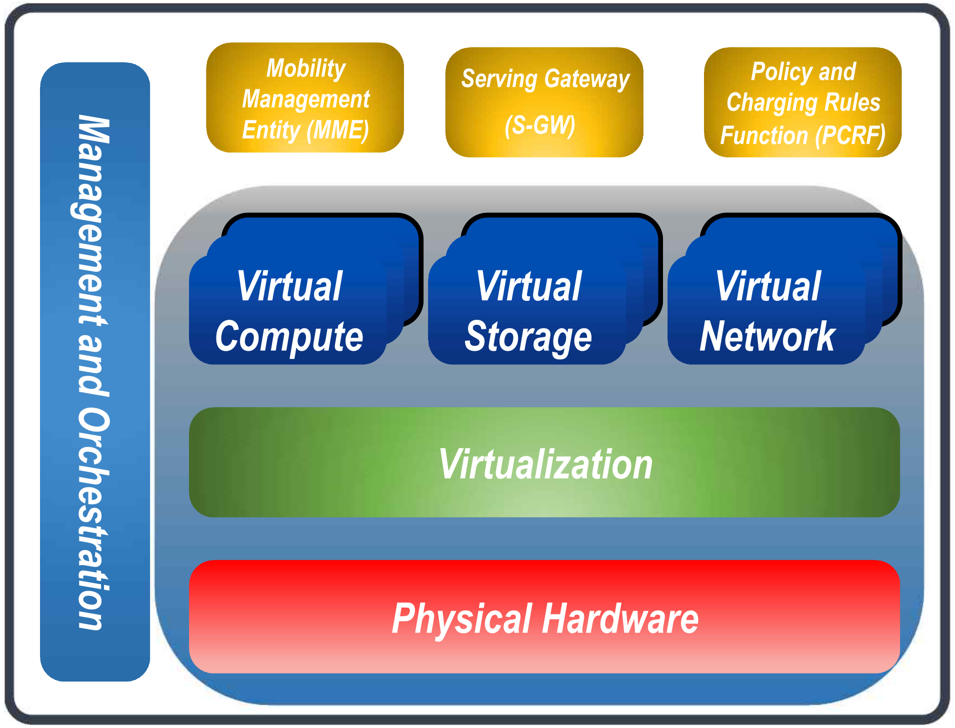
SDN enables programmability which is important for network slicing

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]

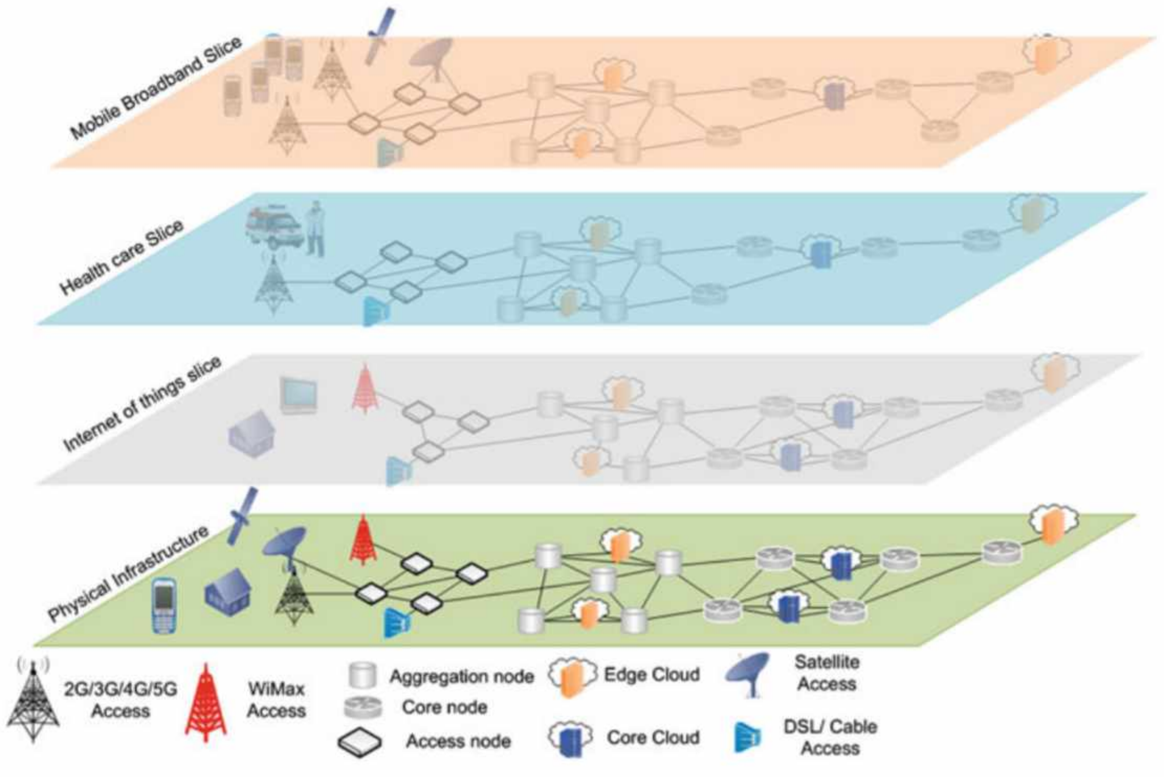
- Wireless network virtualization (WNV) is a novel concept for *virtualizing the RANs* of future cellular networks

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

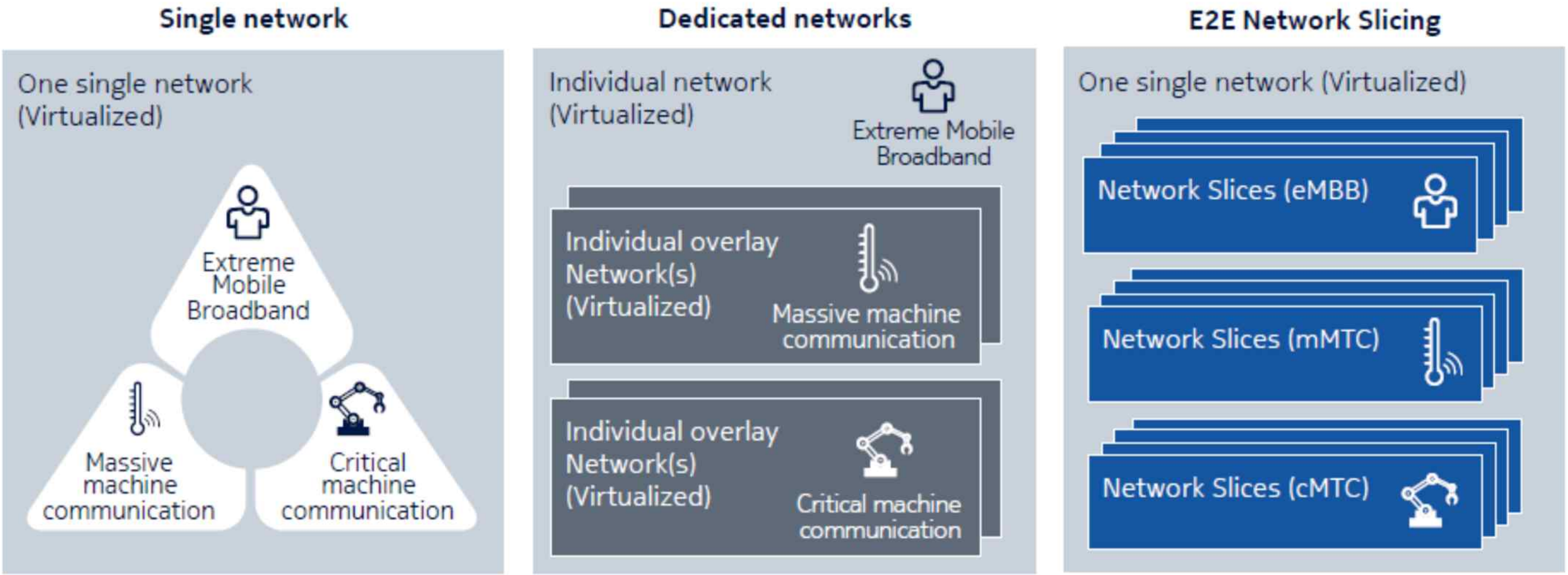
[1] https://opencord.org/
[2] http://www.radisys.com/radisys-m-cord-open-platform-emerging-5g-applications

Network Slicing (4)

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



Three network scenarios provide comparative costs



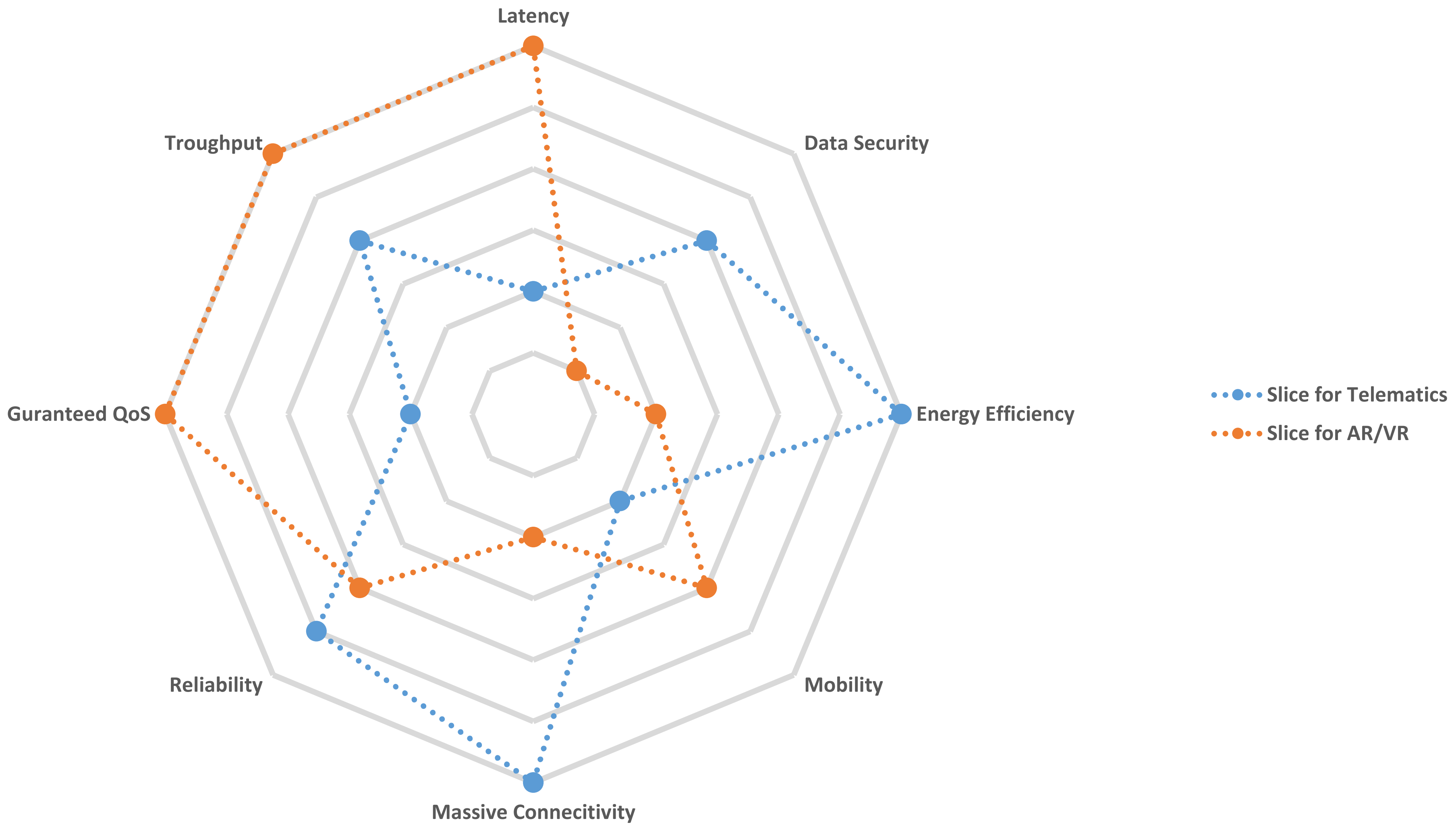
Best-effort support of eMBB, cMTC and mMTC

Customized dedicated networks for mMTC and cMTC

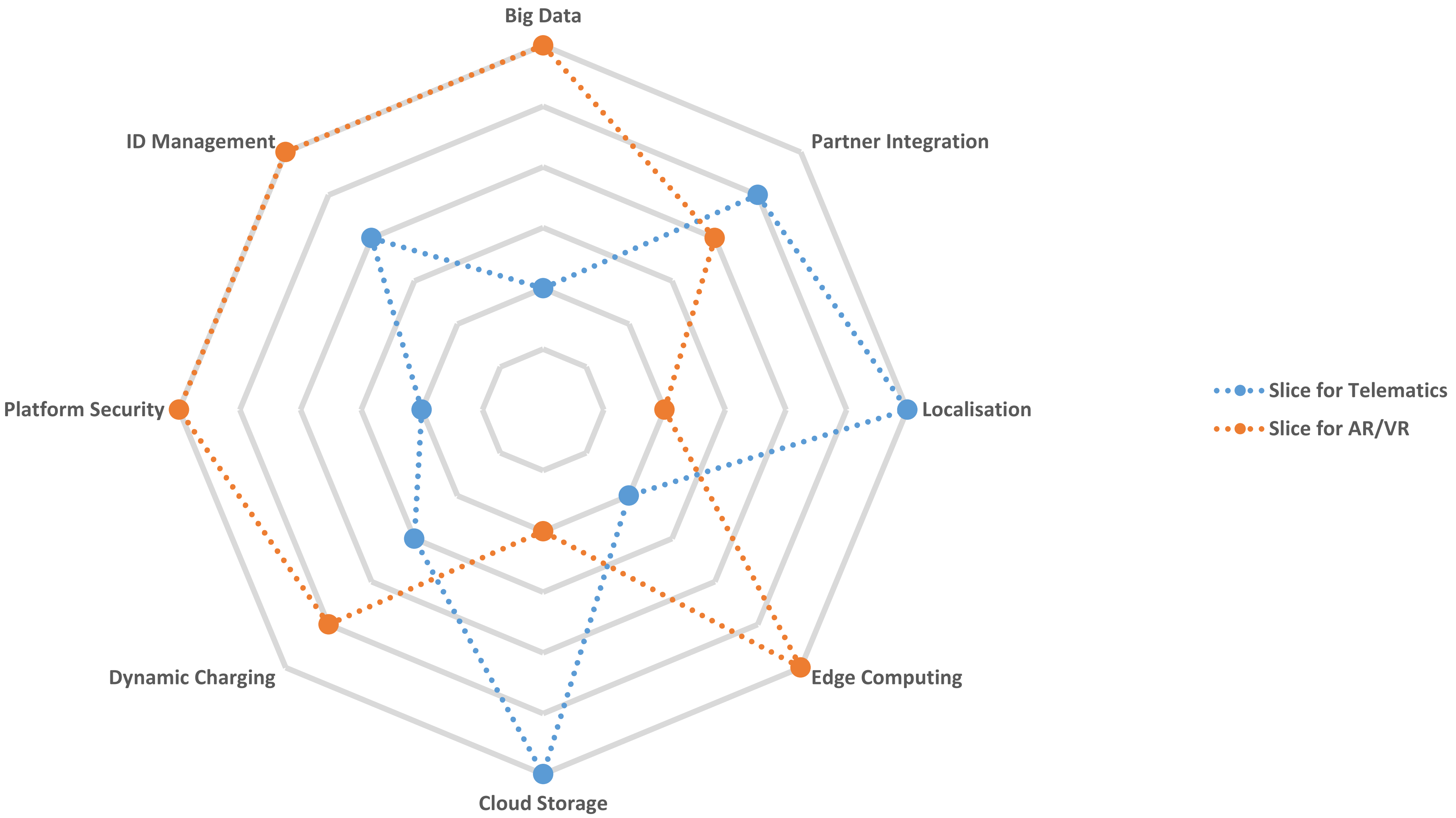
Customized network slices for eMBB, mMTC and cMTC

Nokia: Unleashing the economic potential of network slicing

Network Slicing based on Network Capability



Network Slicing based on Network Services



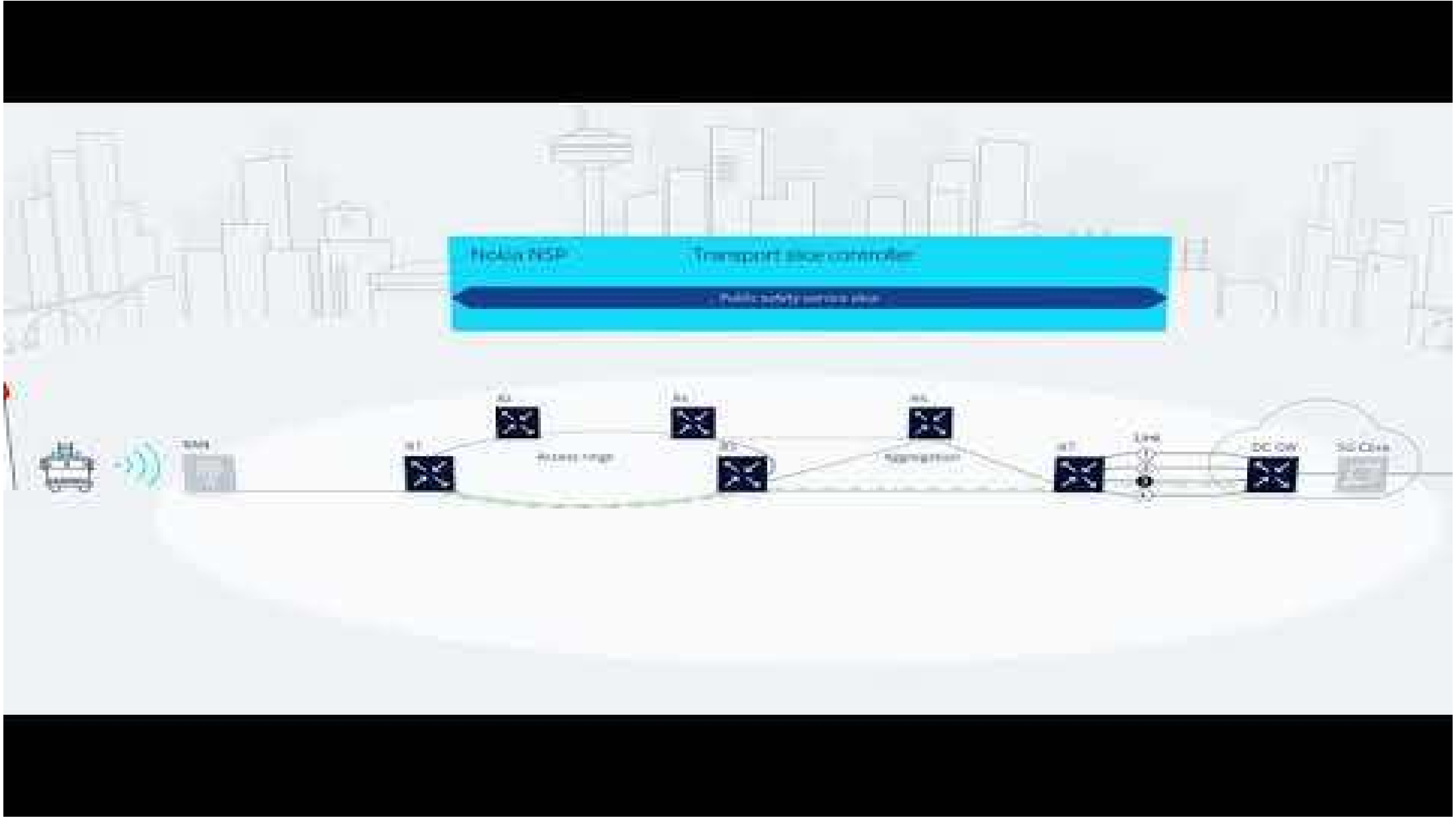
Network Slicing: Industrial Efforts (1)

- **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)

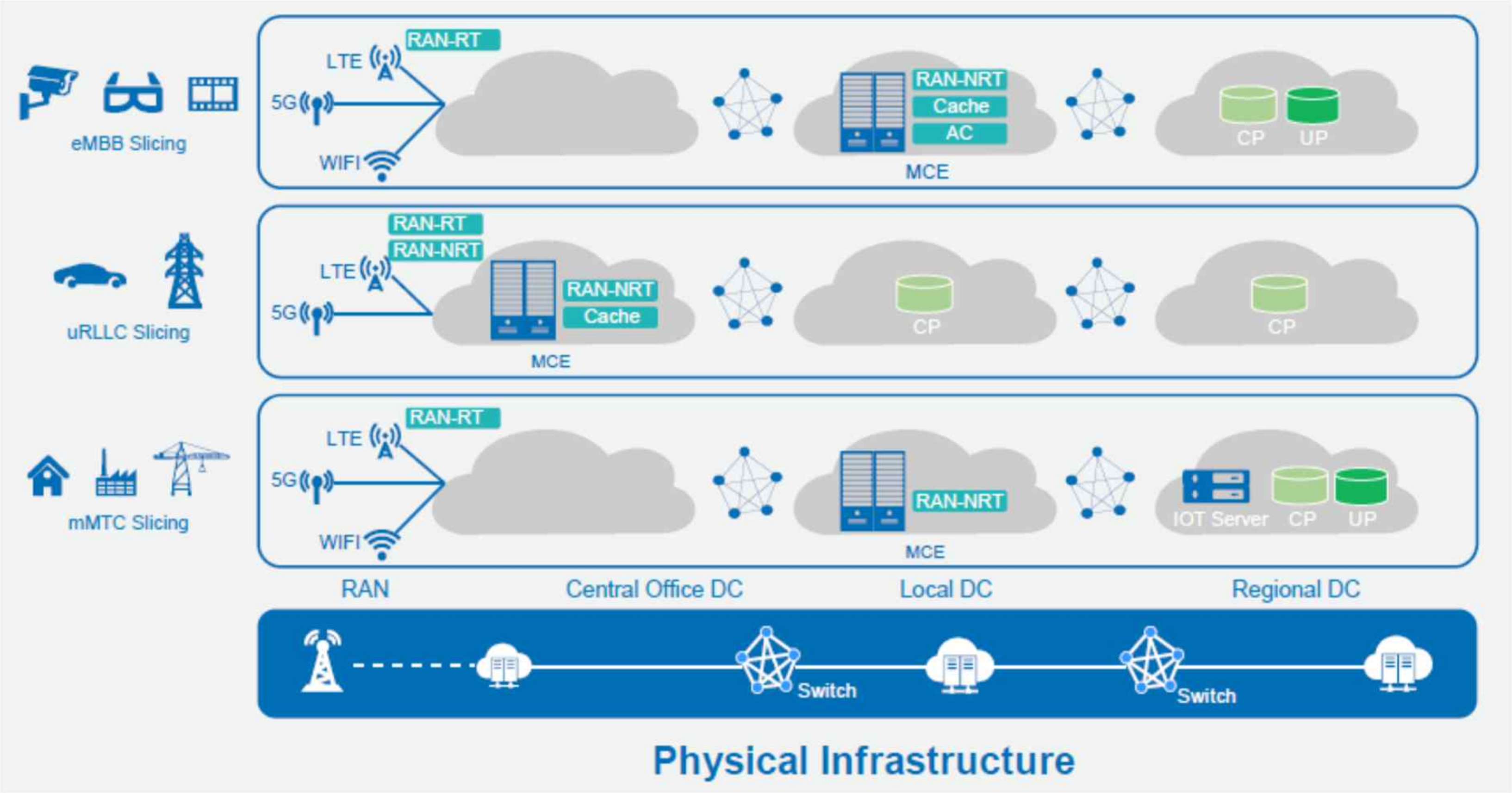
https://youtu.be/nd6LrACl_qM

Network Slicing: Industrial Efforts (3)



https://youtu.be/lyMUFksnXwI

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

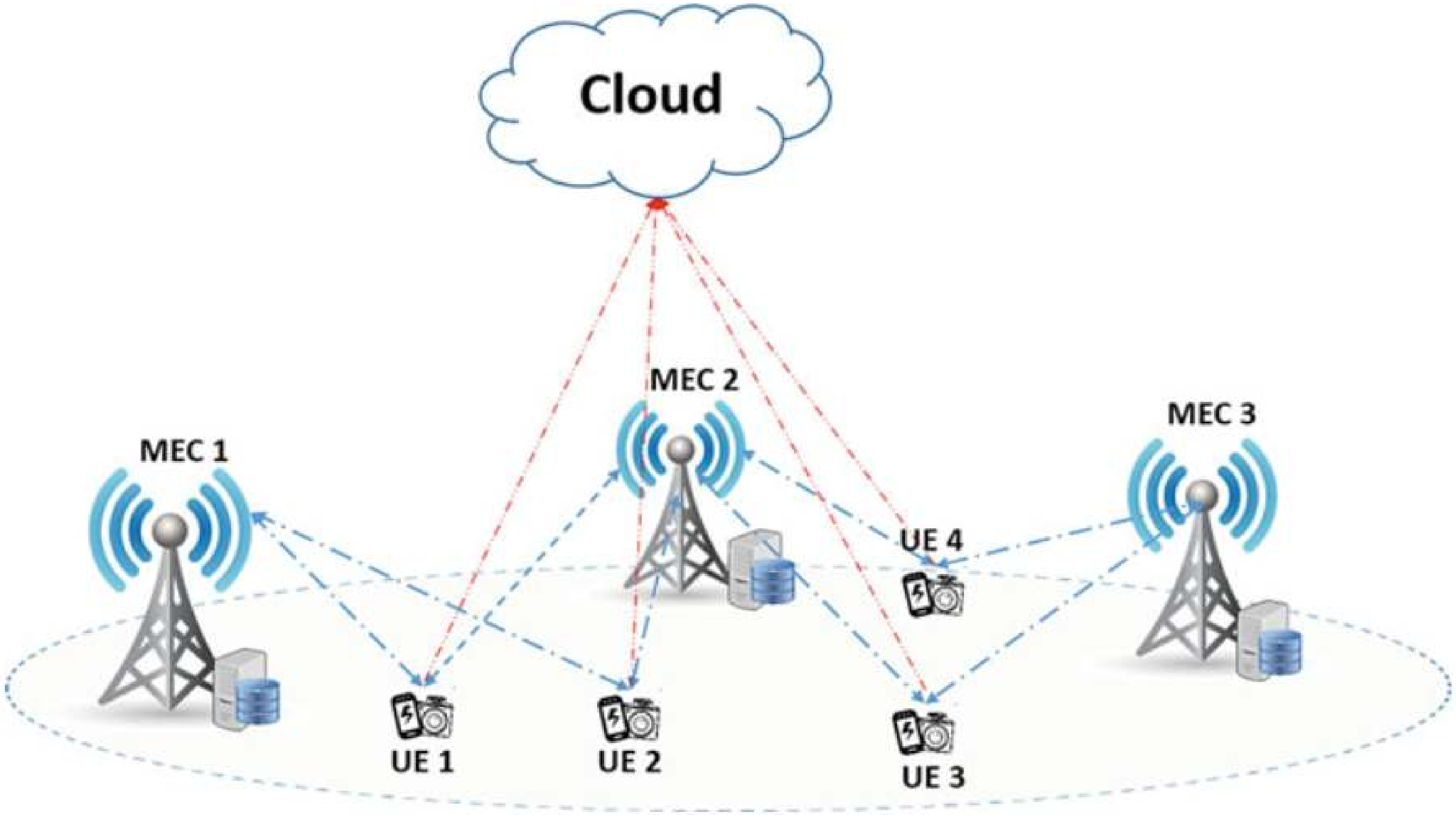
- Introduction
- System Model
- Problem Formulation
- Solution Approach
- Simulation Results

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)



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

$\mathcal{M} = \{1, \ldots, M\}$ 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)

Budget

Computational task to offload for each mobile user

$$\tau_i = \{S_i, T_i, C_i\}$$

Video size

Worst case execution time of the task

Vertical pixels

Video length

Horizontal pixels

Frames per second

Color depth

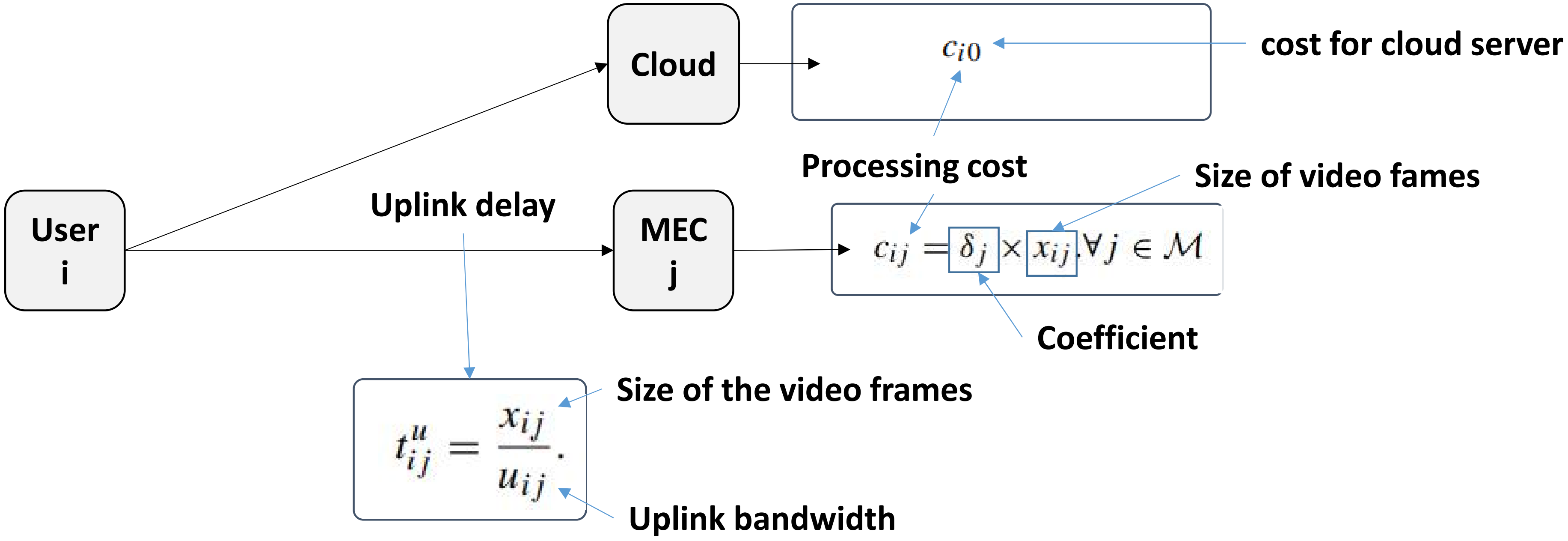
Video size

$$S_i = \frac{H \times V \times F \times L \times 3}{\boxed{1024 \times 1024}}$$

Bit to bytes division

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)



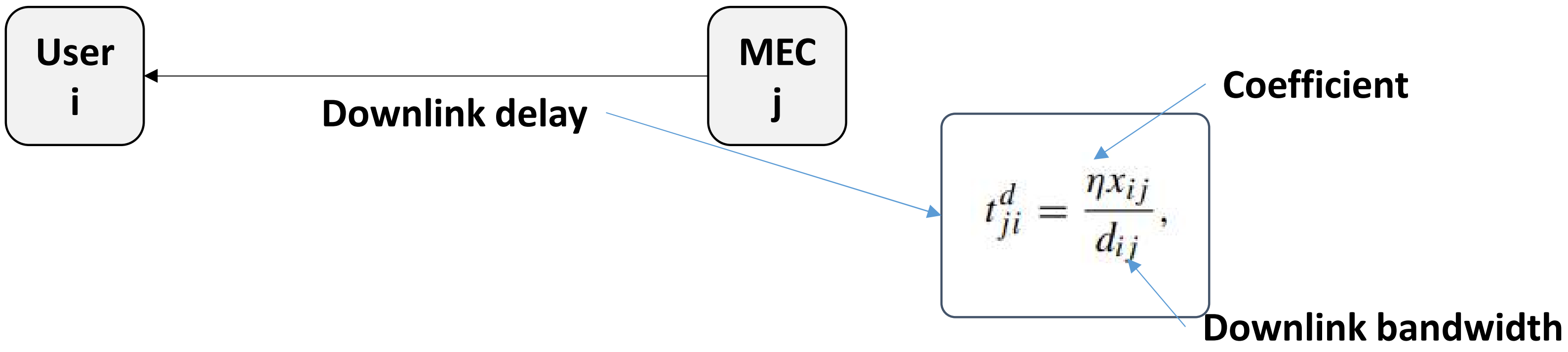
c_{i0}

$c_{ij} = \boxed{\delta_j} \times \boxed{x_{ij}} . \forall j \in \mathcal{M}$

$$t^u_{ij} = \frac{x_{ij}}{u_{ij}}.$$

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

Virtual Reality: System Model (2)



User i

MEC j

Downlink delay

Coefficient

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

Downlink bandwidth

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: Problem Formulation

Total transmission and processing cost minimization problem.

$$\underset{x}{minimize} : \sum_{i \in N} \alpha \left(\overbrace{\boxed{\frac{\sum_{j \in M} t_{ij}^{u}}{t_{i0}^{u}}}}^{\text{Uplink delay}} + \overbrace{\boxed{\frac{\sum_{j \in M} t_{ji}^{d}}{t_{i0}^{d}}}}^{\text{Downlink delay}} \right) + (1-\alpha) \overbrace{\boxed{\frac{\sum_{j \in M} c_{ij}}{c_{i0}}}}^{\text{Processing cost}} \tag{3.10}$$

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

Guarantees that all users are served by network.

$$\sum_{i \in N} x_{ij} \leq \Gamma_j, \forall j \in M, \tag{3.12}$$

Maximum MEC server processing capacity

$$\sum_{j \in M} t_i \leq T_i, \forall i \in N, \tag{3.13}$$

Latency limit constraint

$$\sum_{j \in M} c_i \leq C_i, \forall i \in N, \tag{3.14}$$

Budge cost must not be greater than processing cost

$$t_i' < 1, \forall i \in N, \tag{3.15}$$

Total time constraint variable

$$c_i' < 1, \forall i \in N, \tag{3.16}$$

Total cost constraint variable

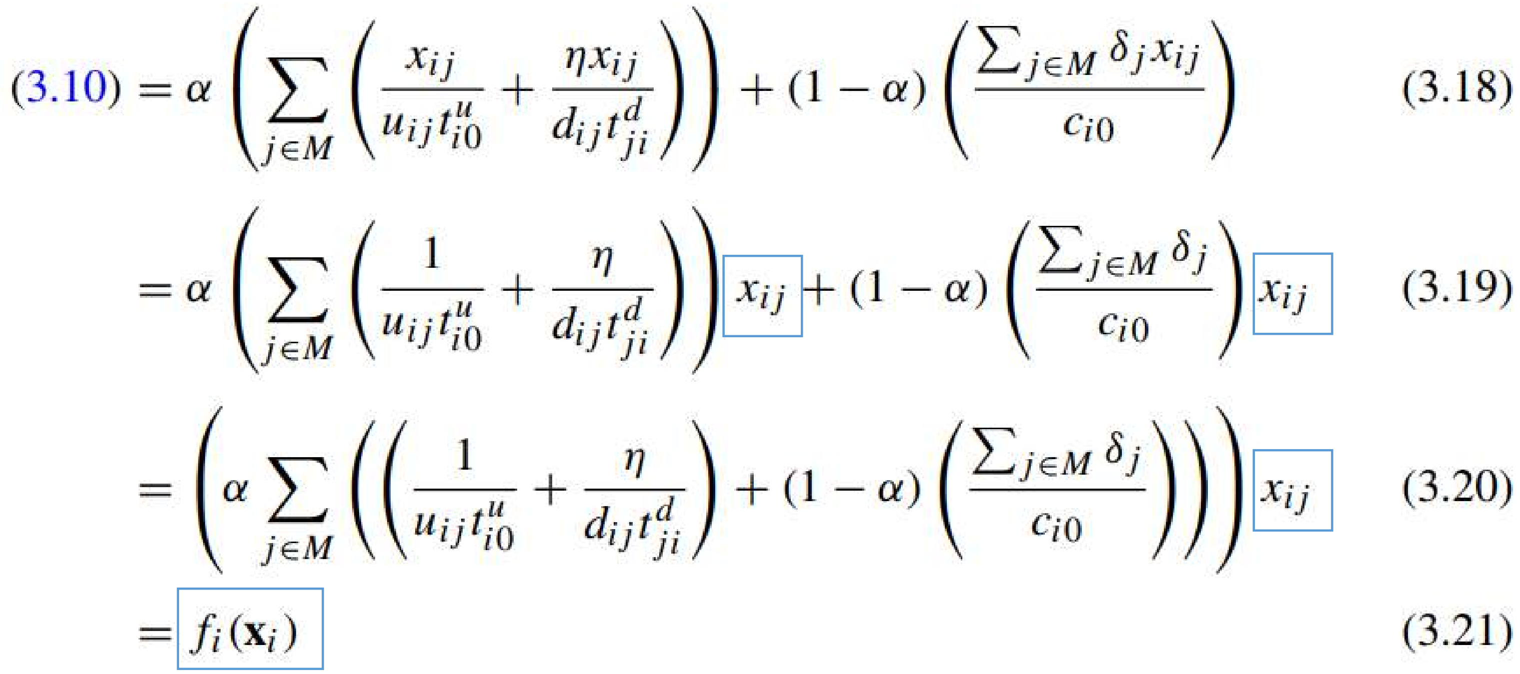
$$x_{ij} \geq 0, \forall i \in N, \forall j \in M. \tag{3.17}$$

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 (1)

Re-write the objective function.



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

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

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

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

ADMM: Alternating Direction Method of Multipliers

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

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

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

Guarantees that all users are served by network.

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

Maximum MEC server capacity constraint

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

Latency limit constraint

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

Budge cost must not be greater than processing cost

$$t_i' < 1, \forall i \in N \tag{3.27}$$

Total time constraint variable

$$c_i' < 1, \forall i \in N \tag{3.28}$$

Total cost constraint variable

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

Video frames size variable

KYUNG HEE UNIVERSITY

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

NETWORKING INTELLIGENT LAB

ADMM-Based Solution (3)

For ADMM-based solution new variable z is introduced.

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

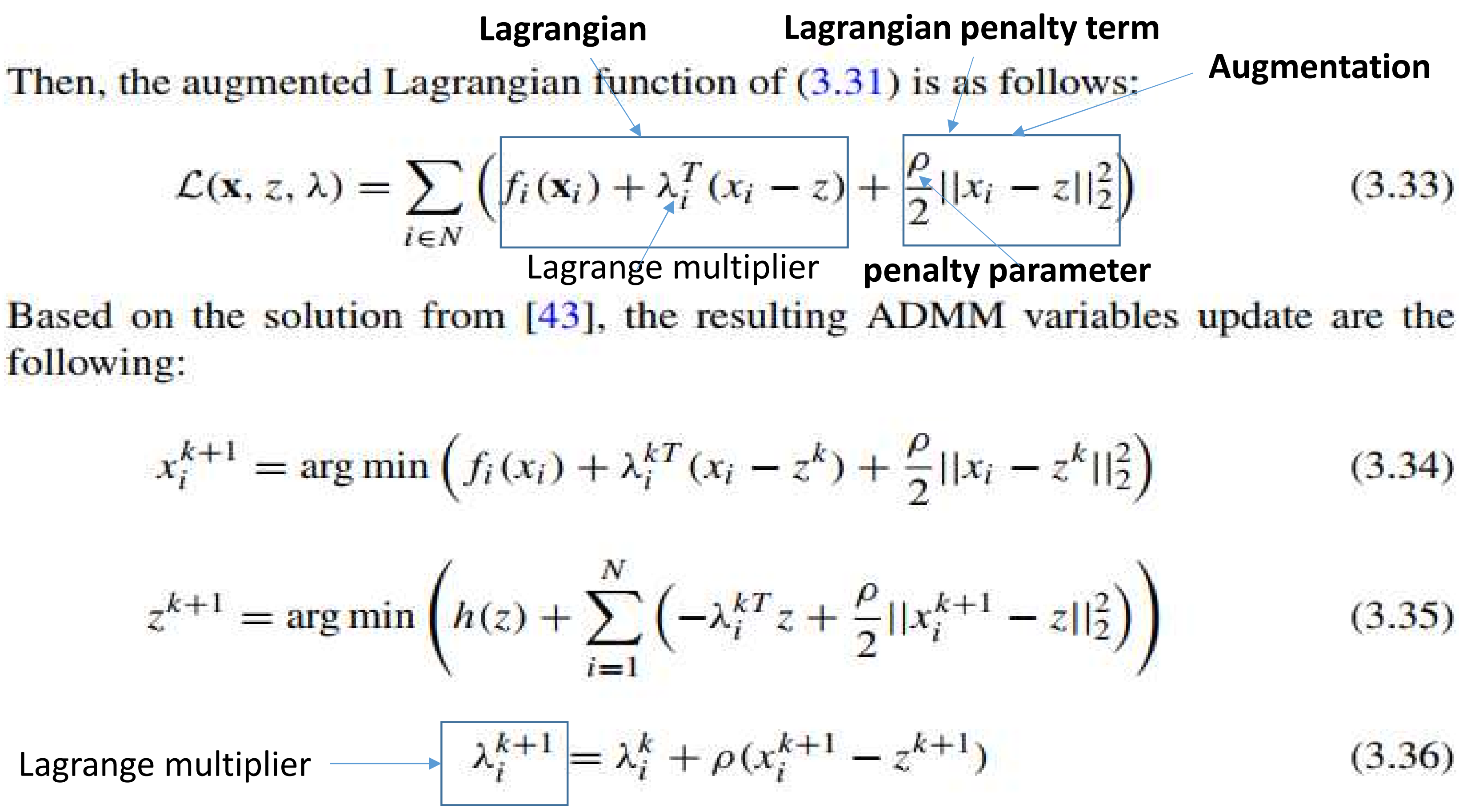
where $h(z) = 0$ when $\mathbf{x}_i \in \mathcal{X}$.

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

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.



Lagrangian **Lagrangian penalty term**

Augmentation

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

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

Lagrange multiplier **penalty parameter**

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

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

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

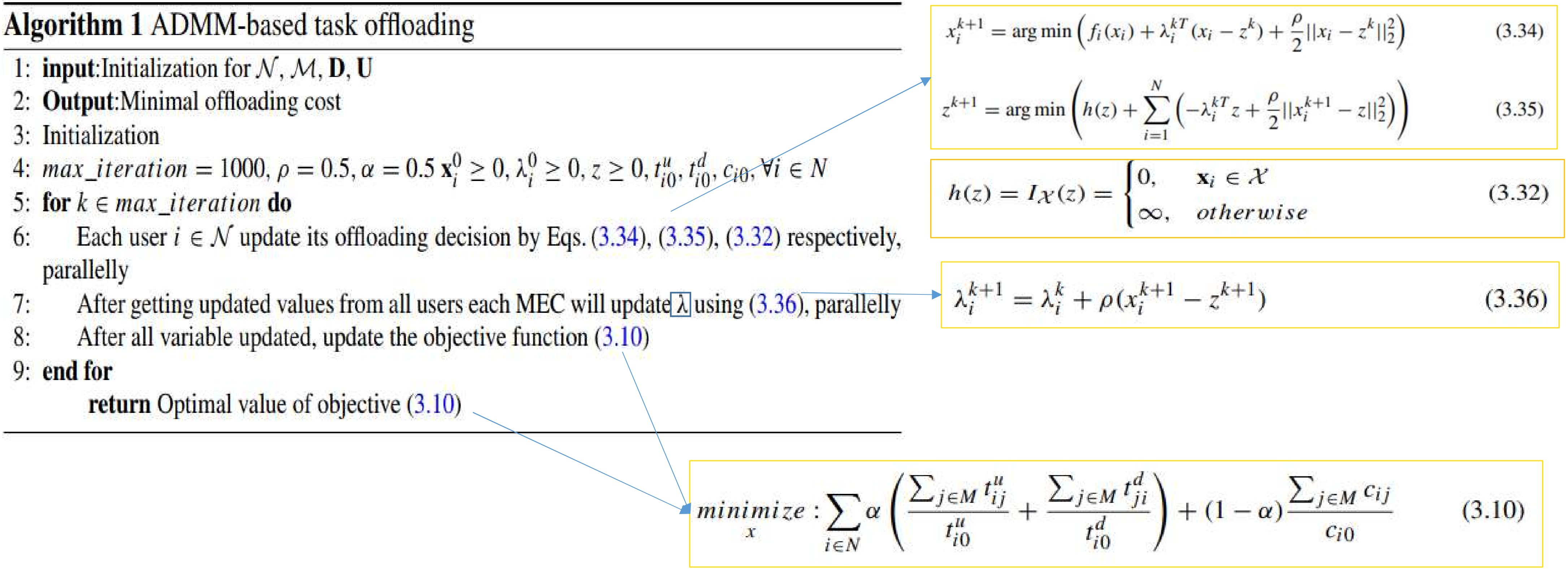
Lagrange multiplier

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

[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 Solution (4)

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 $\boxed{\lambda}$ using (3.36), parallelly
    After all variable updated, update the objective function (3.10)
end for
    return Optimal value of objective (3.10)
```

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

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

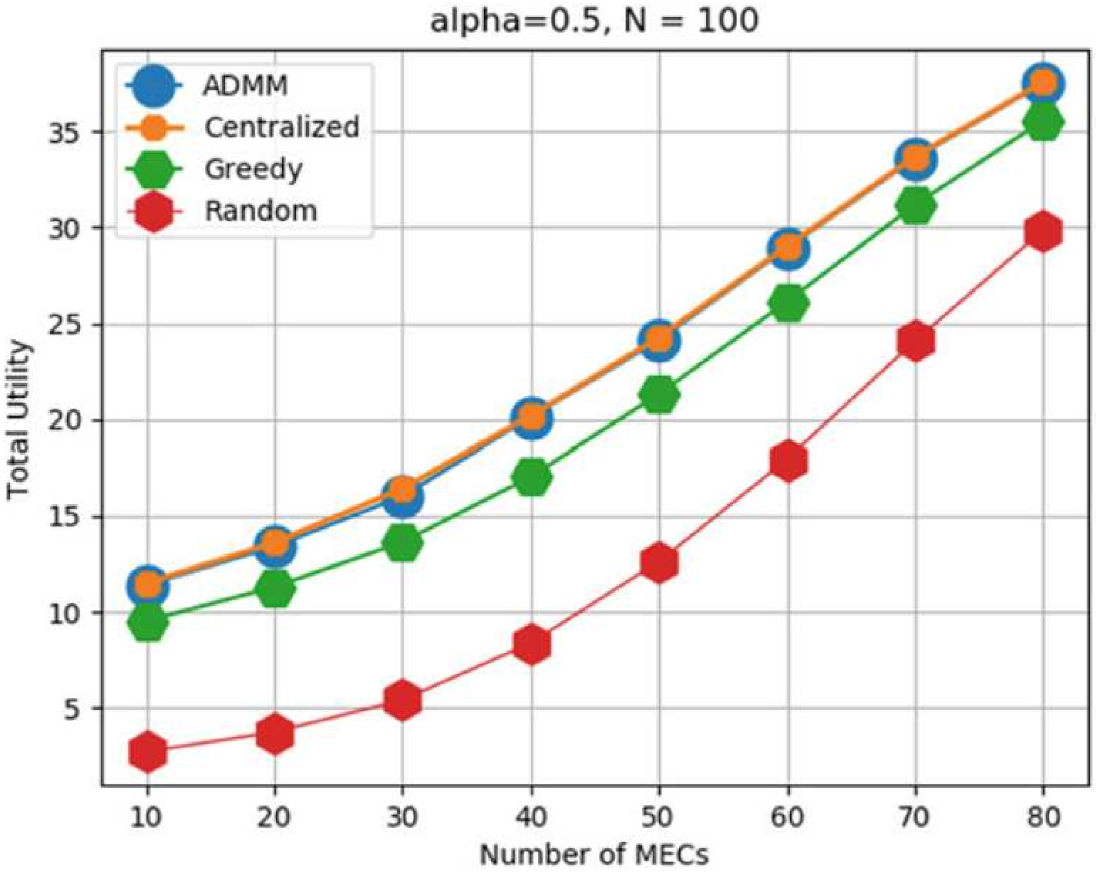
$$h(z) = I_{\mathcal{X}}(z) = \begin{cases} 0, & \mathbf{x}_i \in \mathcal{X} \\ \infty, & otherwise \end{cases} \tag{3.32}$$

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

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

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

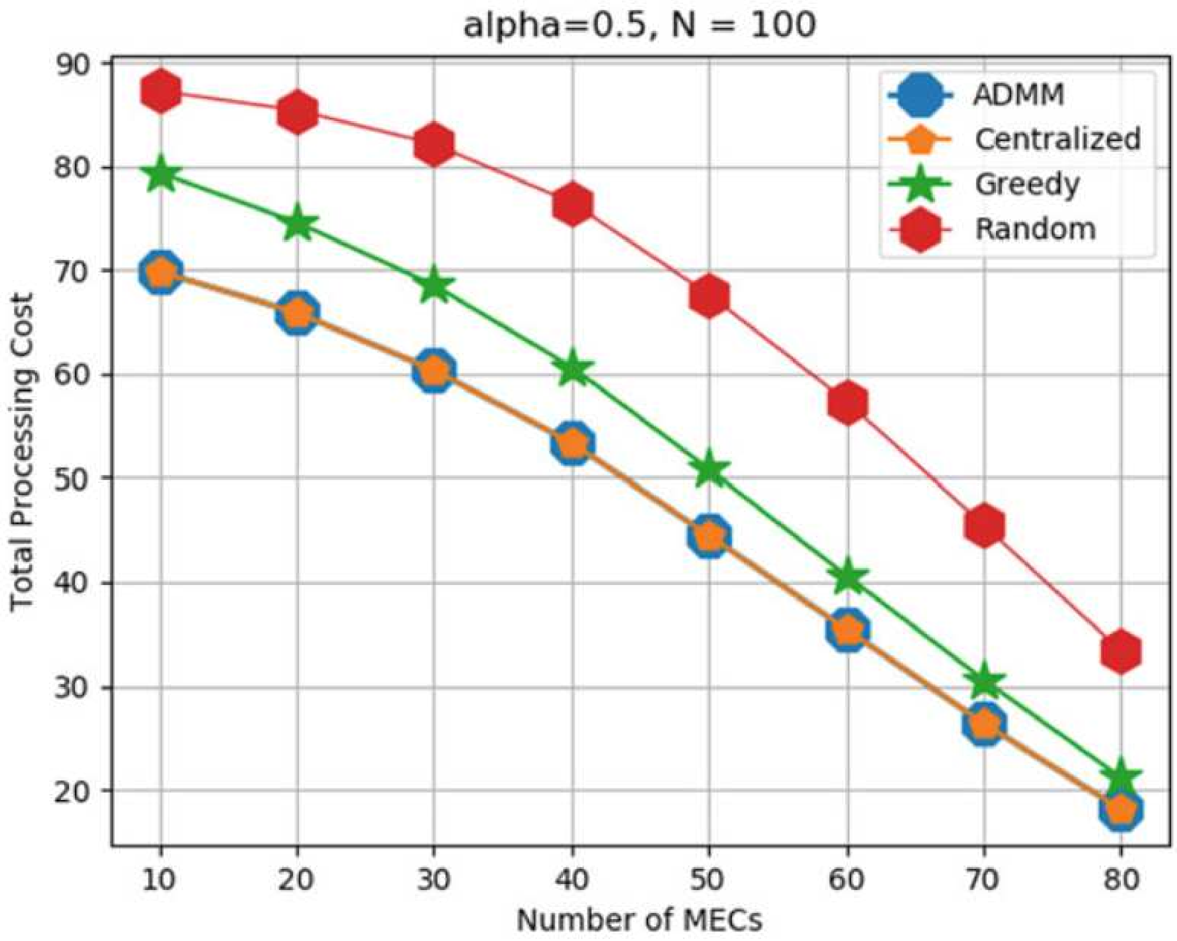
Performance Evaluation (1)



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

Performance Evaluation (2)



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 chapter.
- Numerous key resources for network slicing are discussed.
- A use case of virtual reality is along with its ADMM-based solution is presented.

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 2: Chunk-Based Resource Allocation

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

Introduction

- 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

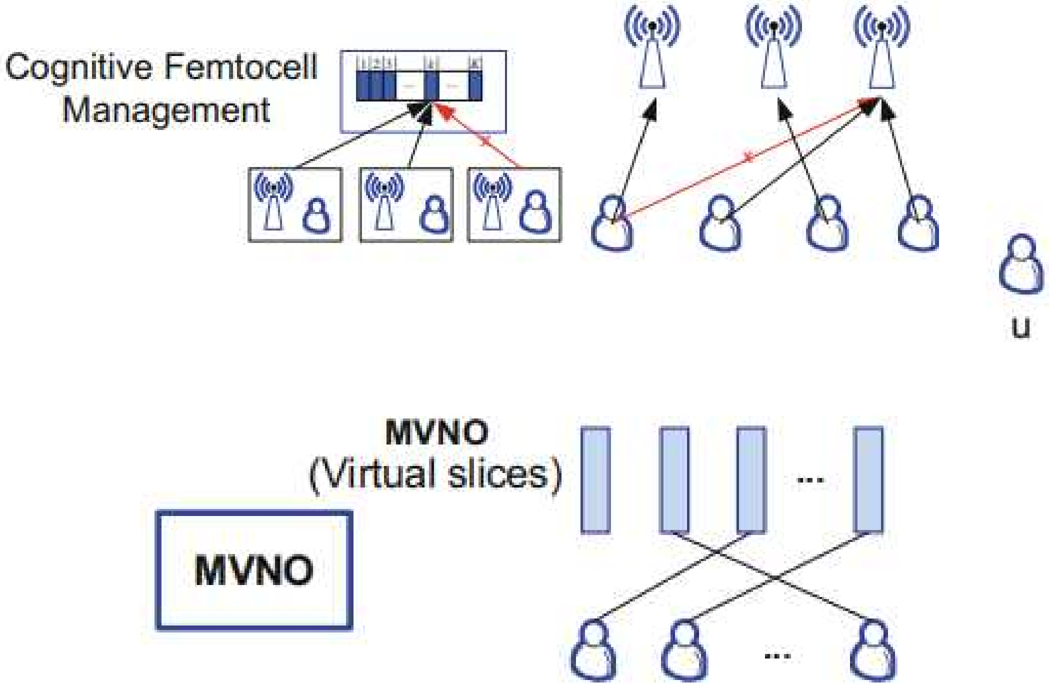
S.M. Ahsan Kazmi, Latif U. Khan, Nguyen H. Tran, Choong Seon Hong, "Network Slicing for 5G and Beyond Networks," ISBN 978-3-030-16169-9, Springer

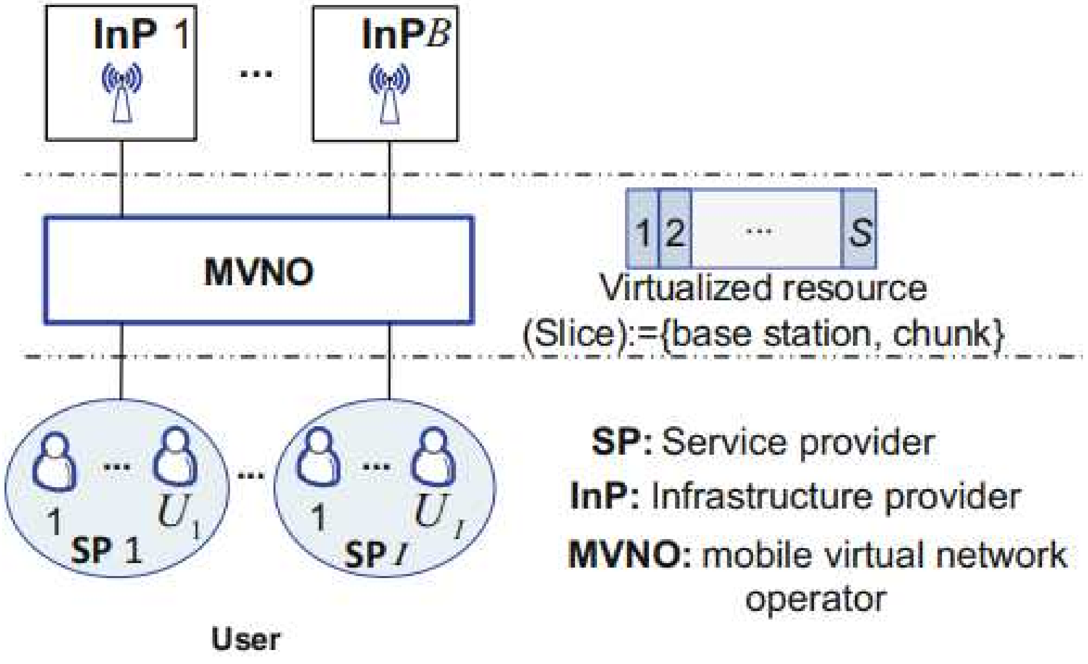
System Model

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

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

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





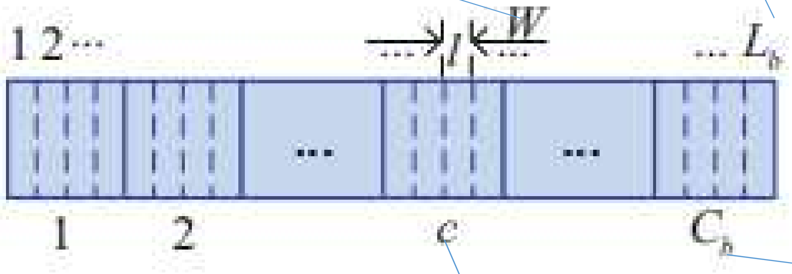
Virtualized cellular network hierarchical model

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

System Model

The bandwidth of narrowband orthogonal subcarriers is W

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



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

Chunk-based radio resources of InP b

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

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

Problem Formulation

$\mathcal{B}$	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

Data rate of a user u associated with service provider i

Binary indicator matrix

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

instantaneous channel gain

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

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

transmit power

$$\boxed{U_{\mathrm{MVNO}}(\boldsymbol{\alpha}, \boldsymbol{P}) = U^{\mathrm{rev}}(\boldsymbol{\alpha}, \boldsymbol{P}) - U^{\mathrm{cost}}(\boldsymbol{\alpha}, \boldsymbol{P}), \qquad (6.4)}$$

Problem Formulation

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

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

(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)

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

Problem Formulation

The Lagrangian function of (6.5) is as follows

$$L(P,\lambda,\mu,\beta) = U_{MVNO}(\alpha,P) +$$

$$\lambda_i \left(R_u^i(\lambda,\beta) - R_{u,min}^i\right) +$$

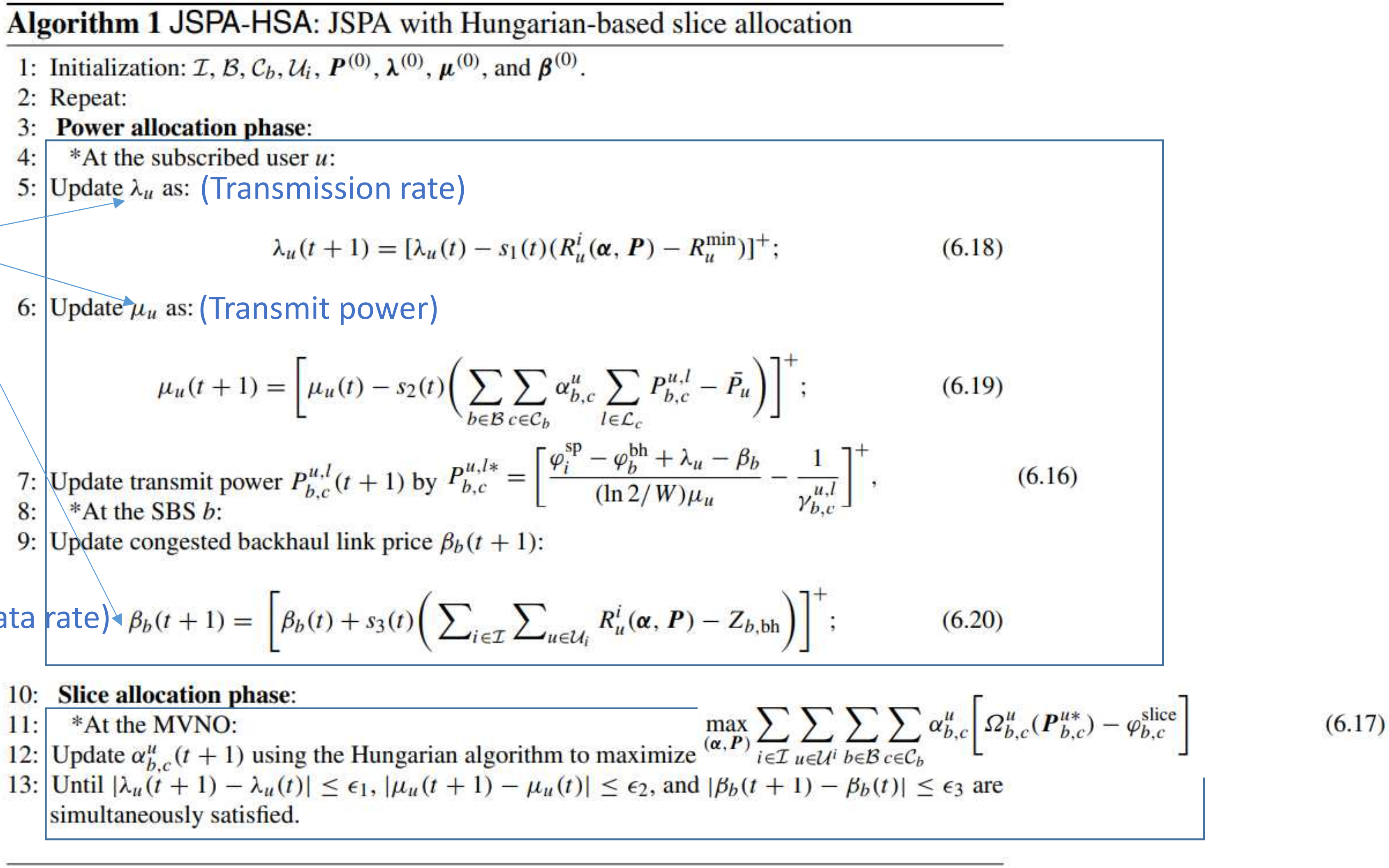
$$\beta_i \left(Z_{b,bh} - \sum_{i \in I} \sum_{u \in U} R_u^i(\alpha,\beta) \right) +$$

$$\mu_i(\bar{P}_u - \sum_{b \in B} \sum_{c \in C_b} \alpha_{b,c}^u \sum_{l \in L_c} P_{b,c}^{u,l})$$

KYUNG HEE
UNIVERSITY

NETWORKING
INTELLIGENT LAB

Solution Approach-1



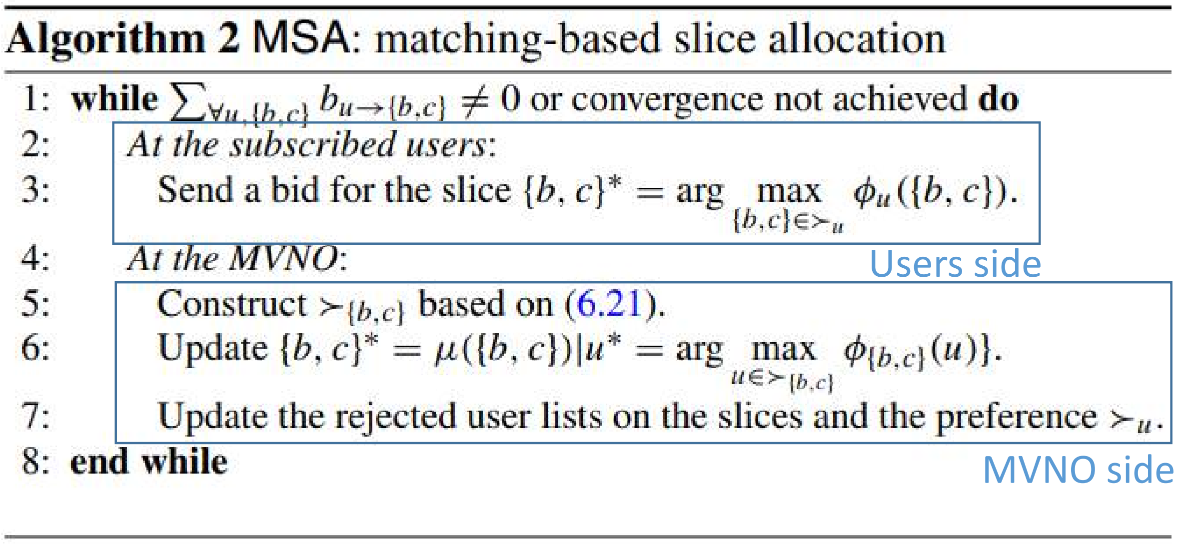
Algorithm 1 JSPA-HSA: JSPA with Hungarian-based slice allocation

```
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 $\lambda_u$ as: (Transmission rate)
      $\lambda_u(t+1) = [\lambda_u(t) - s_1(t)(R_u^i(\boldsymbol{\alpha}, \boldsymbol{P}) - R_u^{\min})]^+;$   (6.18)
    Update $\mu_u$ as: (Transmit power)
      $\mu_u(t+1) = \left[\mu_u(t) - s_2(t)\left(\sum_{b\in\mathcal{B}}\sum_{c\in\mathcal{C}_b} \alpha_{b,c}^u \sum_{l\in\mathcal{L}_c} P_{b,c}^{u,l} - \bar{P}_u\right)\right]^+;$   (6.19)
    Update transmit power $P_{b,c}^{u,l}(t+1)$ by $P_{b,c}^{u,l*} = \left[\frac{\varphi_i^{\mathrm{sp}} - \varphi_b^{\mathrm{bh}} + \lambda_u - \beta_b}{(\ln 2/W)\mu_u} - \frac{1}{\gamma_{b,c}^{u,l}}\right]^+,$   (6.16)
    *At the SBS $b$:
    Update congested backhaul link price $\beta_b(t+1)$:
      (Backhaul data rate) $\beta_b(t+1) = \left[\beta_b(t) + s_3(t)\left(\sum_{i\in\mathcal{I}}\sum_{u\in\mathcal{U}_i} R_u^i(\boldsymbol{\alpha}, \boldsymbol{P}) - Z_{b,\mathrm{bh}}\right)\right]^+;$   (6.20)
  Slice allocation phase:
    *At the MVNO:
    Update $\alpha_{b,c}^u(t+1)$ using the Hungarian algorithm to maximize $\max_{(\boldsymbol{\alpha},\boldsymbol{P})} \sum_{i\in\mathcal{I}}\sum_{u\in\mathcal{U}^i}\sum_{b\in\mathcal{B}}\sum_{c\in\mathcal{C}_b} \alpha_{b,c}^u \left[\Omega_{b,c}^u(\boldsymbol{P}_{b,c}^{u*}) - \varphi_{b,c}^{\mathrm{slice}}\right]$   (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

Solution Approach-2

Matching-Based Low-Complexity Algorithm



Algorithm 2 MSA: matching-based slice allocation

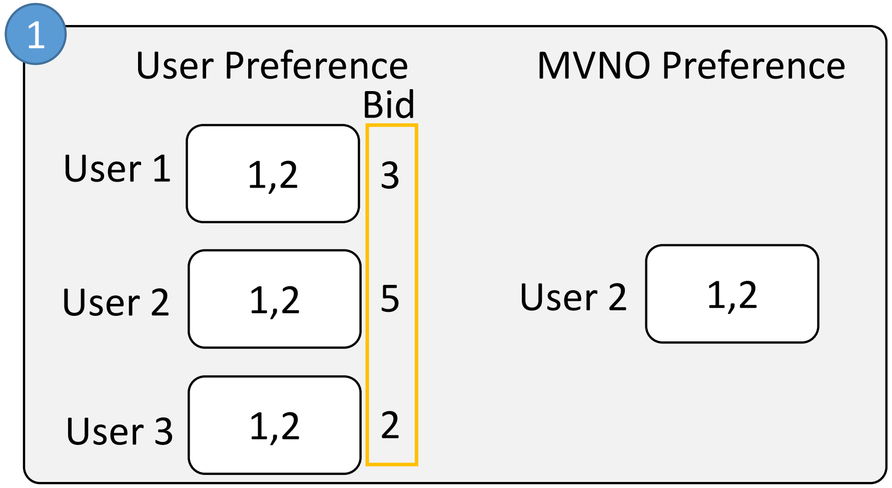
```
while $\sum_{\forall u,\{b,c\}} b_{u\to\{b,c\}} \neq 0$ or convergence not achieved do
    At the subscribed users:
        Send a bid for the slice $\{b,c\}^* = \arg\max_{\{b,c\}\in\succ_u} \phi_u(\{b,c\})$.
    At the MVNO:
        Construct $\succ_{\{b,c\}}$ based on (6.21).
        Update $\{b,c\}^*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

Users side

MVNO side

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

①



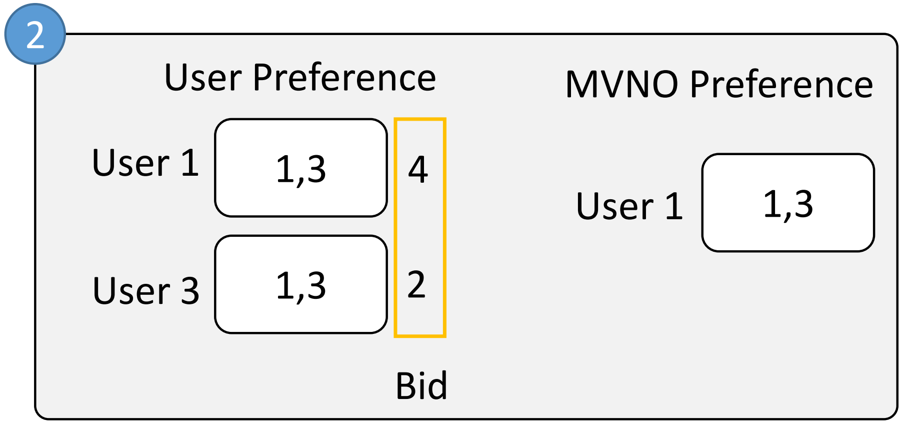
User Preference

		Bid
User 1	1,2	3
User 2	1,2	5
User 3	1,2	2

MVNO Preference

User 2	1,2

②



User Preference

		Bid
User 1	1,3	4
User 3	1,3	2

MVNO Preference

User 1	1,3

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

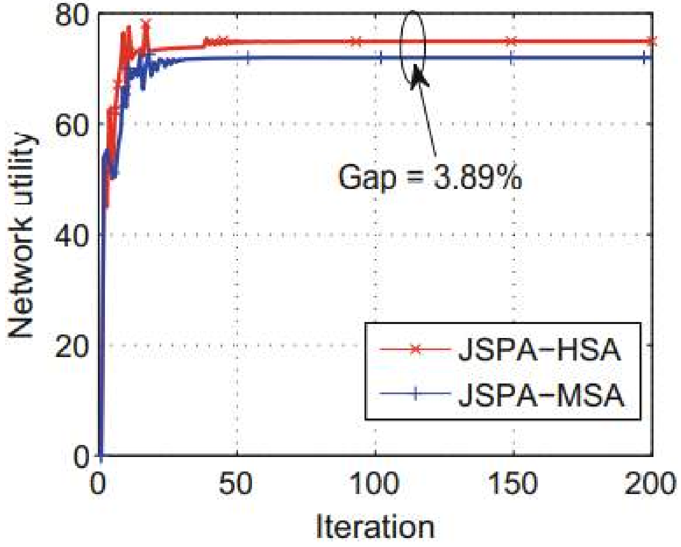


Fig. 6.3 Evaluation results of Network utility with $Z_{b,\mathrm{bh}} = 10\,\mathrm{Mbps}$

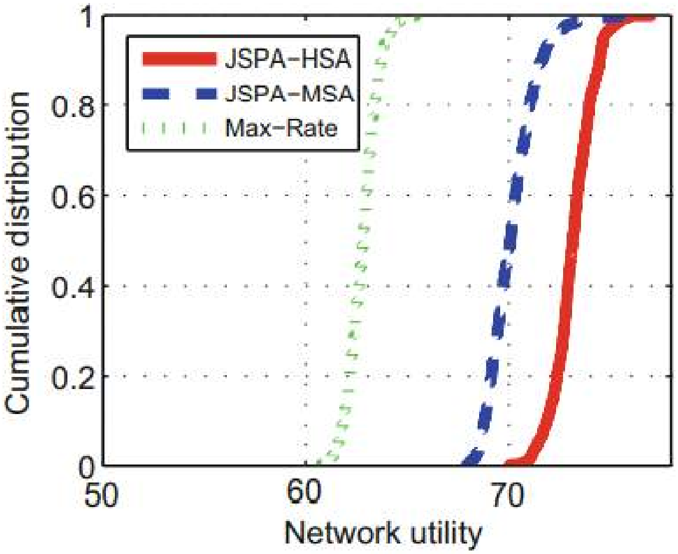


Fig. 6.4 CDF of network utilities

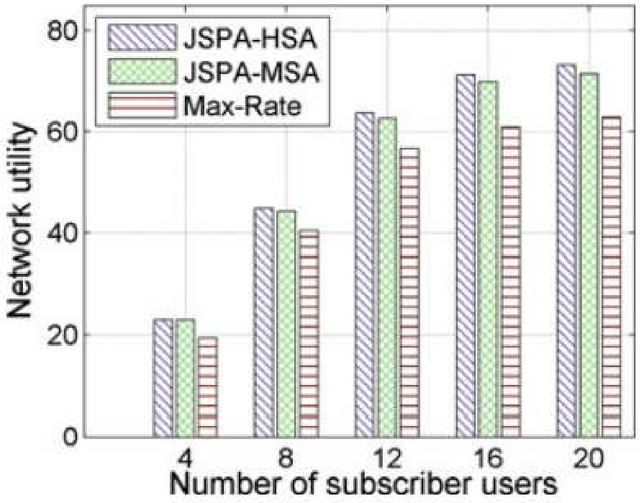


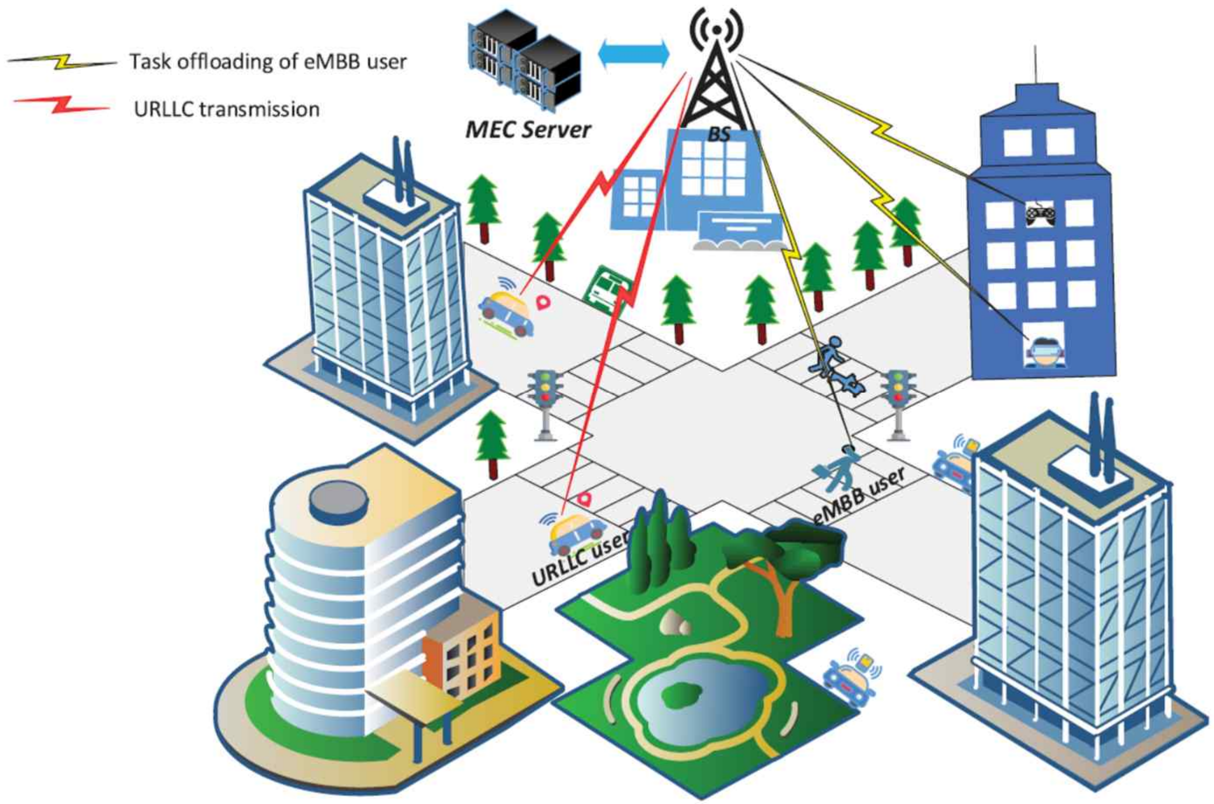
Fig. 6.5 Network utility versus number of subscriber users

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

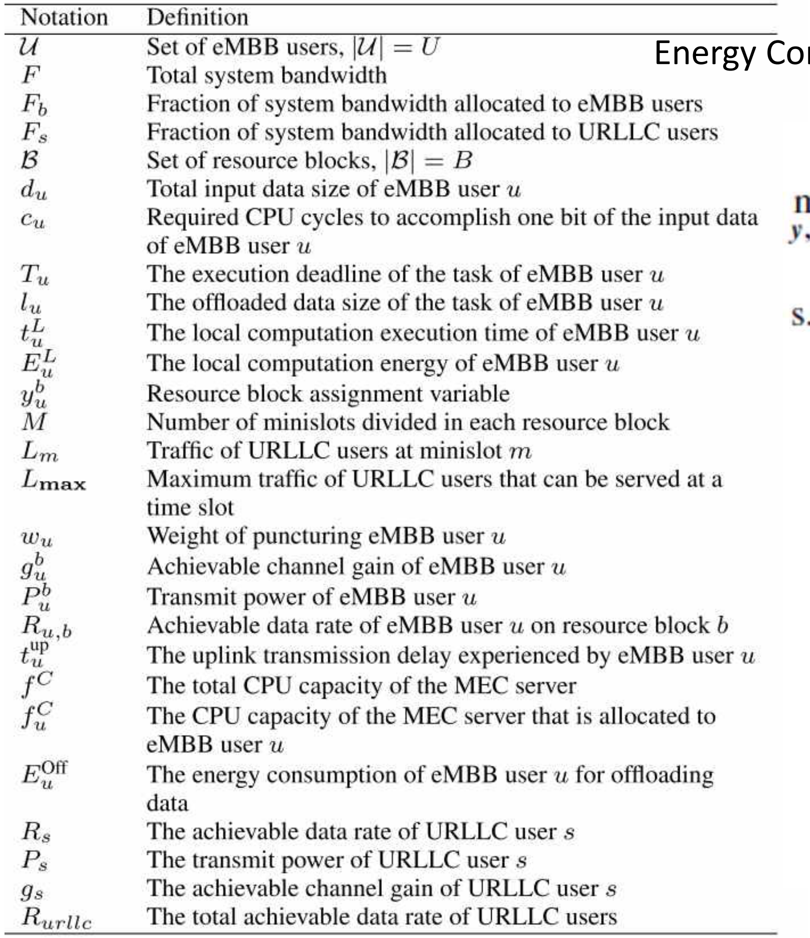
Use Case 3 – System Model



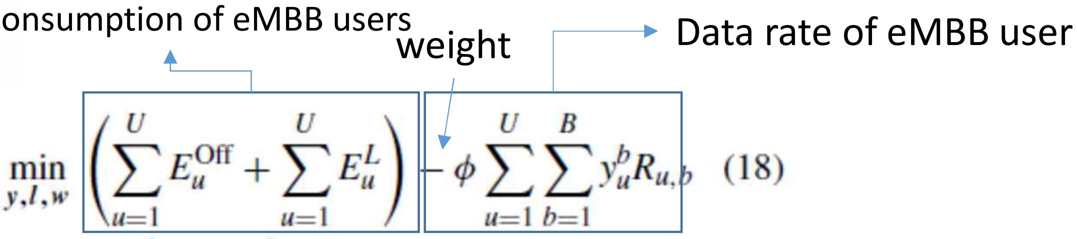
Yan Kyaw Tun, Do Hyon Kim, Madyan Alsenwi, Nguyen H. Tran, Zhu Han, Choong Seon Hong, "Energy Efficient Communication and Computation Resource Slicing for eMBB and URLLC Coexistence in 5G and Beyond," IEEE Access, Vol.8, pp.136024-136035, Jul. 2020

Problem Formulation

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
$L_{\mathbf{max}}$	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

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

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

execution latency constraint of the eMBB users

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

execution latency constraint of the eMBB users

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

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

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

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

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

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

reliability constraints of URLLC users

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

reliability constraints of eMBB users

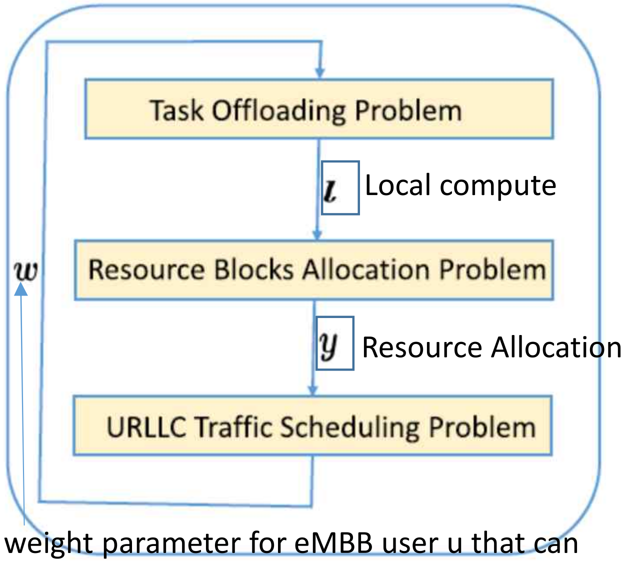
$$\text{C8}: y_u^b \in \{0,1\}, \quad \forall u \in \mathcal{U}, \forall b \in \mathcal{B}, (26)$$

one resource block can be allocated to only one eMBB user

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

Solution Approach

BLOCK COORDINATE DESCENT BASED SOLUTION



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

```
Algorithm 1 Iterative Algorithm for the Relaxed Problem
Initialization: Set $k = 0$, $\epsilon_1, \epsilon_2, \epsilon_3 > 0$, and initial solutions $(\boldsymbol{l}^{(0)}, \boldsymbol{y}^{(0)}, \boldsymbol{w}^{(0)})$;
repeat
    Compute $\boldsymbol{l}^{(k+1)}$ from (P1) at given $\boldsymbol{y}^{k}$, and $\boldsymbol{w}^{k}$;
    Compute $\boldsymbol{y}^{(k+1)}$ from (P2) at given $\boldsymbol{l}^{(k+1)}$, and $\boldsymbol{w}^{k}$;
    Compute $\boldsymbol{w}^{(k+1)}$ from (P3) at given $\boldsymbol{l}^{(k+1)}$, and $\boldsymbol{y}^{(k+1)}$;
    $k = k + 1$;
until $\| \boldsymbol{l}^{(k+1)} - \boldsymbol{l}^{(k)} \| \leq \epsilon_1$, $\| \boldsymbol{y}^{(k+1)} - \boldsymbol{y}^{(k)} \| \leq \epsilon_2$, and $\| \boldsymbol{w}^{(k+1)} - \boldsymbol{w}^{(k)} \| \leq \epsilon_3$;
Then, set $\left(\boldsymbol{l}^{(k+1)}, \boldsymbol{y}^{(k+1)}, \boldsymbol{w}^{(k+1)}\right)$ as the desired solution.
```

Optimization framework

$$\text{(P1)}: \min_{l} \sum_{u=1}^{U} E_u^{\text{Off}} + \sum_{u=1}^{U} E_u^{L} \qquad (30)$$

$$\text{s.t. (C1)-(C3),} \qquad (31)$$

$$\text{(P2)}: \min_{y} \sum_{u=1}^{U} E_u^{\text{Off}} - \phi \sum_{u=1}^{U} \sum_{b=1}^{B} y_u^b R_{u,b} \qquad (34)$$

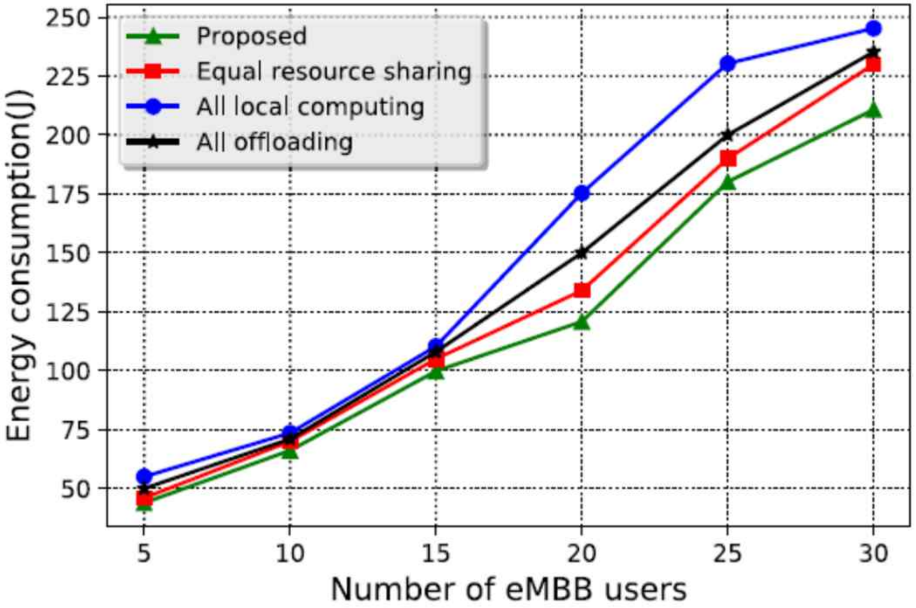
$$\text{s.t. (C1), (C7), and (C8),} \qquad (35)$$

$$\text{(P3)}: \min_{w} \sum_{u=1}^{U} E_u^{\text{Off}} - \phi \sum_{u=1}^{U} \sum_{b=1}^{B} y_u^b R_{u,b} \qquad (41)$$

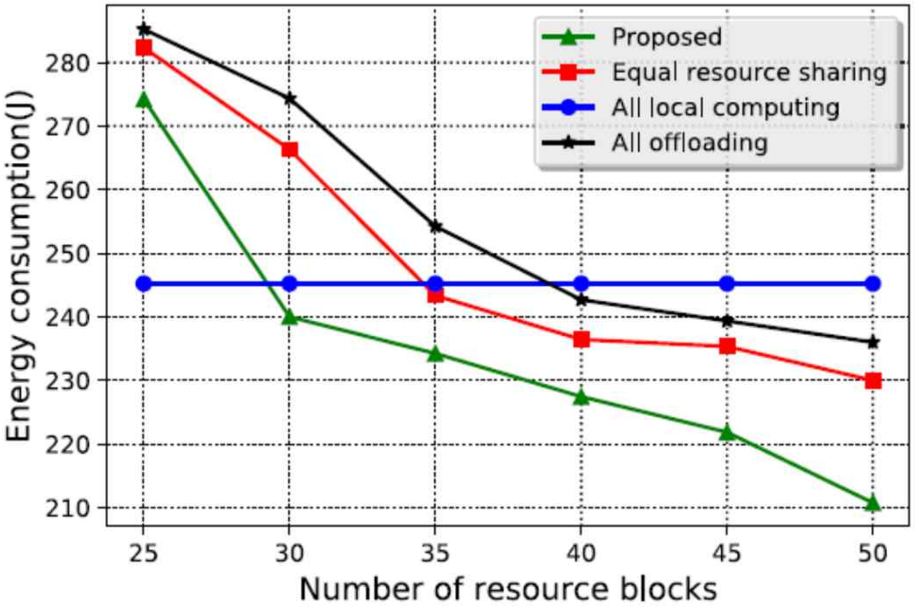
$$\text{s.t. (C4), (C5), and (C6),} \qquad (42)$$

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

Simulation Results



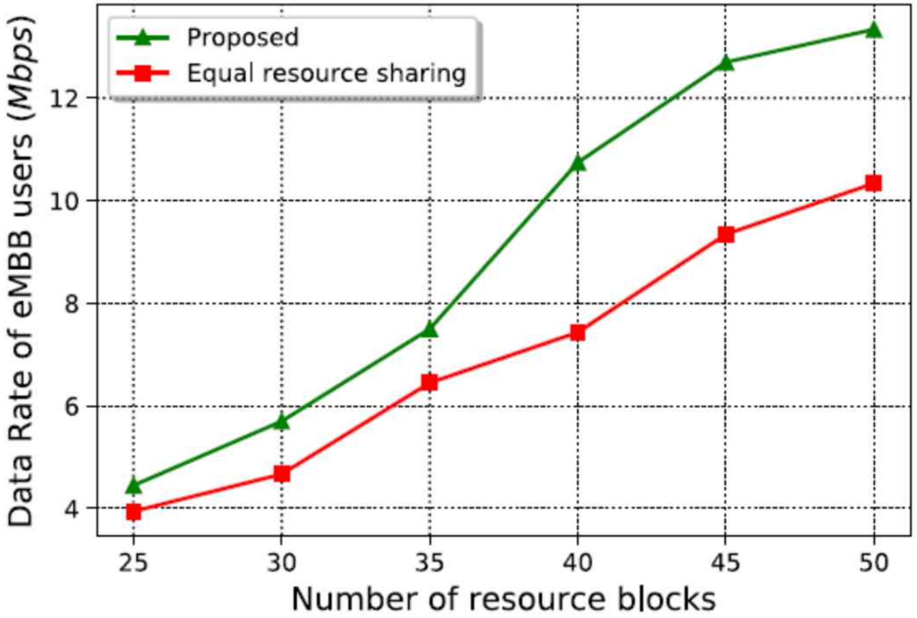
Energy consumption under different number of users.



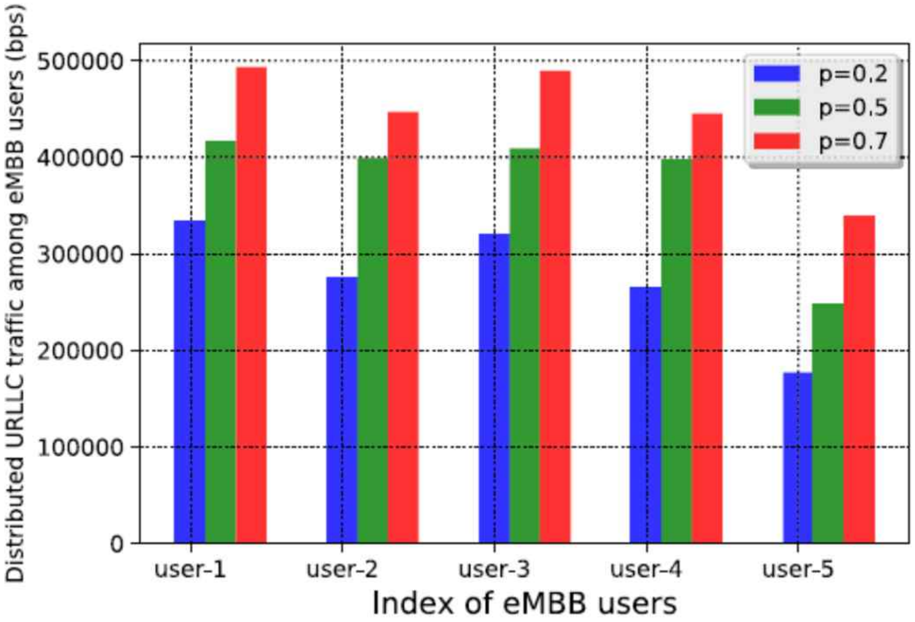
Energy consumption under different number of resource blocks

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

Simulation Results



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



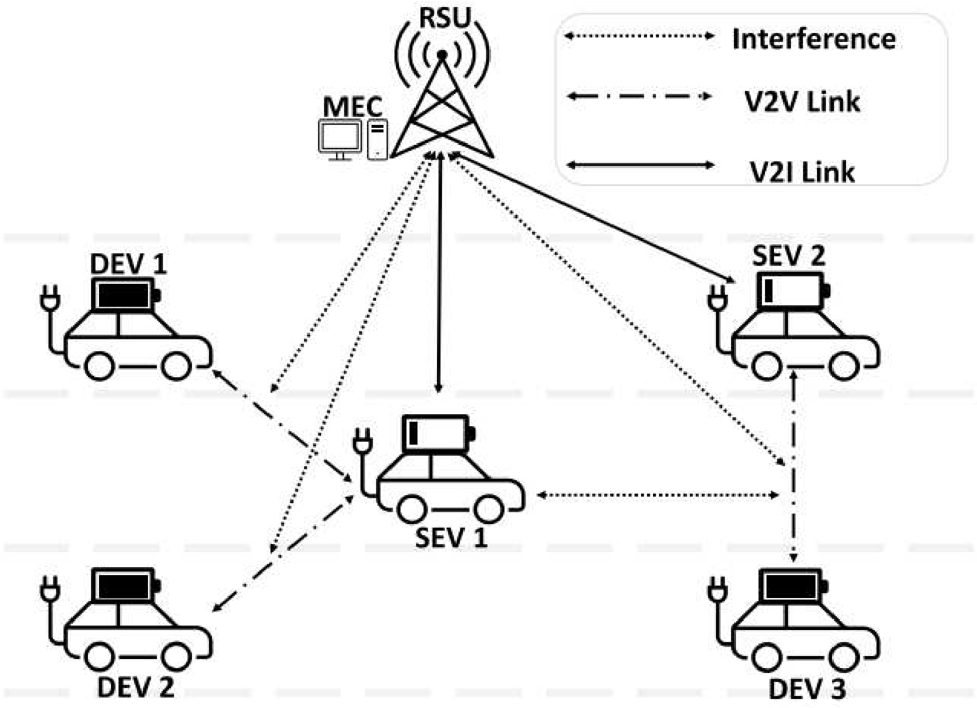
Data rate of URLLC users on the puncturing resource of eMBB users.

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

Use case 4: Joint Communication, Computation, and Control for Computational Task Offloading in Vehicle-Assisted Multi-Access Edge Computing

- System model
- Problem formulation
- Solution approach
- Simulation results

Use Cases 4: System model



- ➢ Decision making for task-offloading in multiple location: EVs, MEC, etc.
- ➢ Communication resource allocation for V2V, V2I links
- ➢ Interference Management for V2V, V2I links w.r.t. protection constraint, *i.e., maximum interference for each link*

Goal: Minimize the trade-off between Energy Consumption and Latency in EV-assisted MEC

Tri Nguyen Dang, Choong Seon Hong et al. "Joint Communication, Computation, and Control for Computational Task Offloading in Vehicle-Assisted Multi-Access Edge Computing." *IEEE Access* 10 (2022): 122513-122529.

Use Cases 4: Problem formulation

Trade-off coefficient

Normalization factor

Energy consumption

$$\mathcal{F}(\boldsymbol{\zeta},\boldsymbol{p},\boldsymbol{h},\boldsymbol{x},\boldsymbol{y},z)=\sum_{i\in\mathcal{O}}\Bigg[\psi\,\Phi\Bigg(\underbrace{\kappa_i h_i^2 x_i}_{local\ energy}+\underbrace{\sum_{j\in\mathcal{D}}\Bigg(\kappa_j h_{i,j}^2+p_{i,j}\frac{S_i}{\sum_{b\in\beta}R_{i,j}^b\zeta_{i,j}^b}\Bigg)y_{i,j}}_{V2V\ energy}+\underbrace{\Bigg(\kappa_0 h_{i,0}^2+p_{i,0}\frac{S_i}{\sum_{b\in\beta}R_{i,0}^b\zeta_{i,0}^b}\Bigg)z_{i,0}}_{V2I\ energy}\Bigg)$$

$$+(1-\psi)\Bigg(\underbrace{\frac{S_i}{h_i}x_i}_{local\ latency}+\underbrace{\sum_{j\in\mathcal{D}}\Bigg(\frac{S_i}{h_{i,j}}+\frac{S_i}{\sum_{b\in\beta}R_{i,j}^b\zeta_{i,j}^b}\Bigg)y_{i,j}}_{V2V\ latency}+\underbrace{\Bigg(\frac{S_i}{h_{i,0}}+\frac{S_i}{\sum_{b\in\beta}R_{i,0}^b\zeta_{i,0}^b}\Bigg)z_{i,0}}_{V2I\ latency}\Bigg)\Bigg].$$

Latency

Note that: Energy and latency are not the same unit.
That’s why we need to normalize them in the same scale

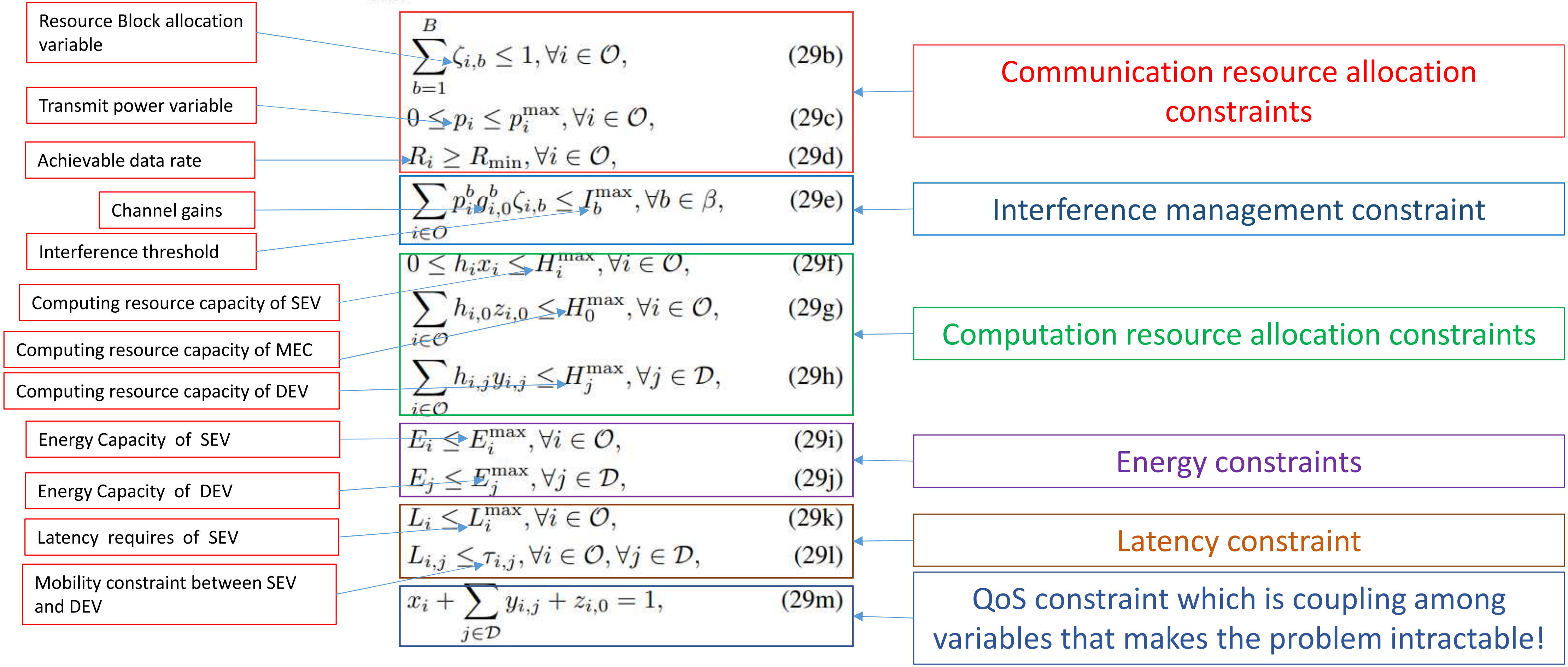
Tri Nguyen Dang, Choong Seon Hong et al. "Joint Communication, Computation, and Control for Computational Task Offloading in Vehicle-Assisted Multi-Access Edge Computing." *IEEE Access* 10 (2022): 122513-122529.

Use Cases 4: Problem formulation

$$\min_{\boldsymbol{\zeta},\boldsymbol{p},\boldsymbol{h},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}} \quad \mathcal{F}(\boldsymbol{\zeta},\boldsymbol{p},\boldsymbol{h},\underline{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}}) \tag{29a}$$

SEV/DEV/MEC decision variables

s.t.:



Resource Block allocation variable

Transmit power variable

Achievable data rate

Channel gains

Interference threshold

Computing resource capacity of SEV

Computing resource capacity of MEC

Computing resource capacity of DEV

Energy Capacity of SEV

Energy Capacity of DEV

Latency requires of SEV

Mobility constraint between SEV and DEV

$$\sum_{b=1}^{B} \zeta_{i,b} \le 1, \forall i \in \mathcal{O}, \tag{29b}$$

$$0 \le p_i \le p_i^{\max}, \forall i \in \mathcal{O}, \tag{29c}$$

$$R_i \ge R_{\min}, \forall i \in \mathcal{O}, \tag{29d}$$

Communication resource allocation constraints

$$\sum_{i\in\mathcal{O}} p_i^b g_{i,0}^b \zeta_{i,b} \le I_b^{\max}, \forall b \in \beta, \tag{29e}$$

Interference management constraint

$$0 \le h_i x_i \le H_i^{\max}, \forall i \in \mathcal{O}, \tag{29f}$$

$$\sum_{i\in\mathcal{O}} h_{i,0} z_{i,0} \le H_0^{\max}, \forall i \in \mathcal{O}, \tag{29g}$$

$$\sum_{i\in\mathcal{O}} h_{i,j} y_{i,j} \le H_j^{\max}, \forall j \in \mathcal{D}, \tag{29h}$$

Computation resource allocation constraints

$$E_i \le E_i^{\max}, \forall i \in \mathcal{O}, \tag{29i}$$

$$E_j \le E_j^{\max}, \forall j \in \mathcal{D}, \tag{29j}$$

Energy constraints

$$L_i \le L_i^{\max}, \forall i \in \mathcal{O}, \tag{29k}$$

$$L_{i,j} \le \tau_{i,j}, \forall i \in \mathcal{O}, \forall j \in \mathcal{D}, \tag{29l}$$

Latency constraint

$$x_i + \sum_{j\in\mathcal{D}} y_{i,j} + z_{i,0} = 1, \tag{29m}$$

QoS constraint which is coupling among variables that makes the problem intractable!

Tri Nguyen Dang, Choong Seon Hong et al. "Joint Communication, Computation, and Control for Computational Task Offloading in Vehicle-Assisted Multi-Access Edge Computing." *IEEE Access* 10 (2022): 122513-122529.

Use Cases 4: Solution Approach

$$\text{RBA}: \max_{\boldsymbol{\zeta}} \mathcal{F}(\boldsymbol{\zeta}) = \sum_{i\in\Omega_S}\sum_{b\in B} R_{i,b}\zeta_{i,b} \tag{30a}$$

s.t.

$$\sum_{b=1}^{B}\zeta_{i,b}\le 1, \forall i\in\Omega_S, \tag{30b}$$

$$\sum_{i\in\Omega_S} p_{i,0,b}\,g_{i,0,b}\,\zeta_{i,b} \le I_b^{\max}, \forall b\in\beta, \tag{30c}$$

$$\zeta_{i,b}\in\{0,1\}, \forall b\in\beta. \tag{30d}$$

Set of EV pairs

RBA: Resource Block Allocation

1. Iteratively solve each subproblem to achieve an sub-optimal solution

$$\text{PCP}: \max_{\boldsymbol{p}} \mathcal{F}(\boldsymbol{p}) = \sum_{i\in\mathcal{O}}\sum_{b\in\beta} F(p_{i,b}) = R_i^b \tag{34a}$$

s.t.:

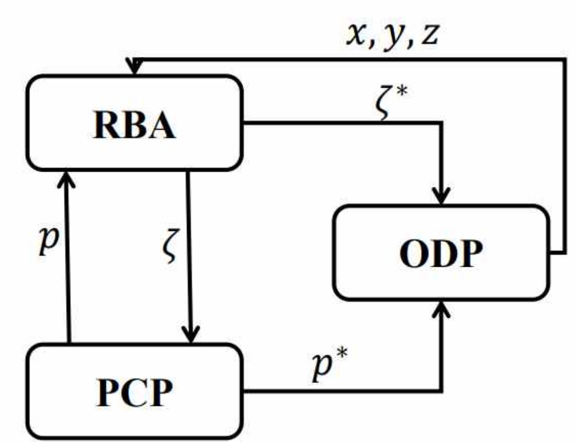
$$0\le p_i^b\le p_i^{\max}, \tag{34b}$$

$$R_i^b\ge R_{\min}, \tag{34c}$$

$$\sum_{i\in\mathcal{O}} p_i^b\,g_{i,0}^b\,\zeta_{i,b}\le I_b^{\max}. \tag{34d}$$

PCP: Power Control Problem

2. Update the solution based on last iteration



3. Always guarantee a stationary solution by the BCD technique - global optimal solution

BCD: Block-Coordination Descent

ODP: Offloading Decision Problem

$$\text{ODP}: \min_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}} \mathcal{F}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) \tag{39a}$$

s.t:

$$0\le h_i\le H_i^{\max}, \tag{39b}$$

$$\sum_{i\in\mathcal{O}} h_{i,0}\le H_0^{\max}, \tag{39c}$$

$$\sum_{i\in\mathcal{O}} h_{i,j}y_{i,j}\le H_j^{\max}, \forall j\in\mathcal{D}, \tag{39d}$$

$$E_i\le E_i^{\max}, \forall i\in\mathcal{O}, \tag{39e}$$

$$E_j\le E_j^{\max}, \forall j\in\mathcal{D}, \tag{39f}$$

$$L_i\le L_i^{\max}, \tag{39g}$$

$$L_{i,j}\le \tau_{i,j}, \forall i\in\mathcal{O}, \forall j\in\mathcal{D}, \tag{39h}$$

$$x_i+\sum_{j\in\mathcal{D}} y_{i,j}+z_{i,0}=1, \tag{39i}$$

$$h_{i,j}\ge 0, \tag{39j}$$

$$h_{i,0}\ge 0, \tag{39k}$$

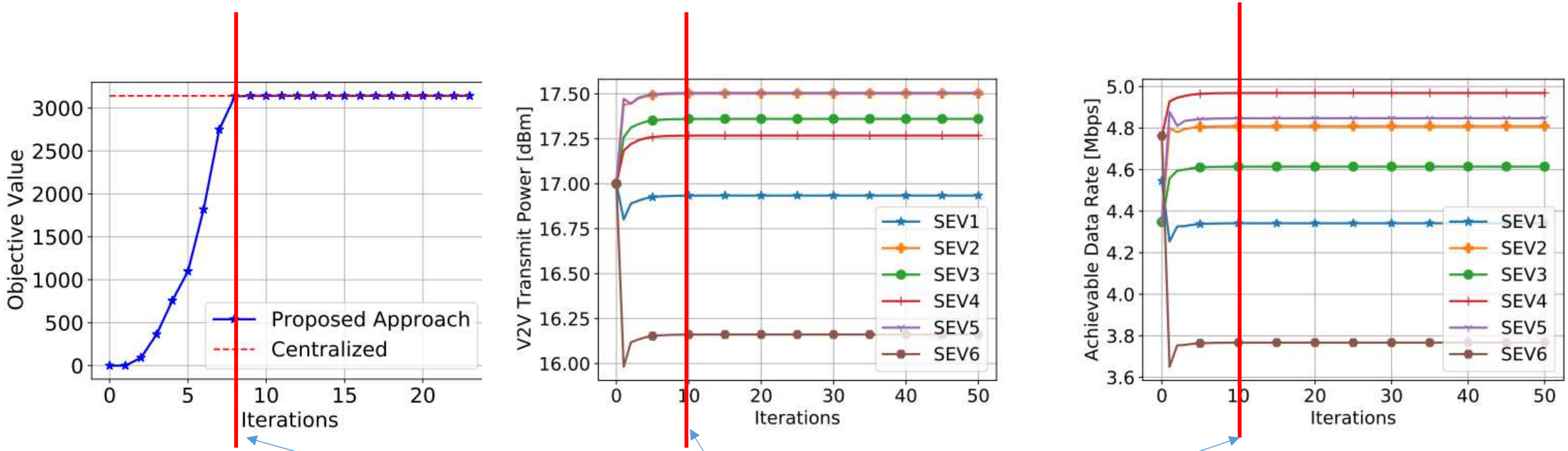
$$x_i\in\{0,1\}, \tag{39l}$$

$$y_{i,j}\in\{0,1\}, \tag{39m}$$

$$z_{i,0}\in\{0,1\},. \tag{39n}$$

Tri Nguyen Dang, Choong Seon Hong et al. "Joint Communication, Computation, and Control for Computational Task Offloading in Vehicle-Assisted Multi-Access Edge Computing." *IEEE Access* 10 (2022): 122513-122529.

Use Cases 4: Numerical Results



Convergence of ODP

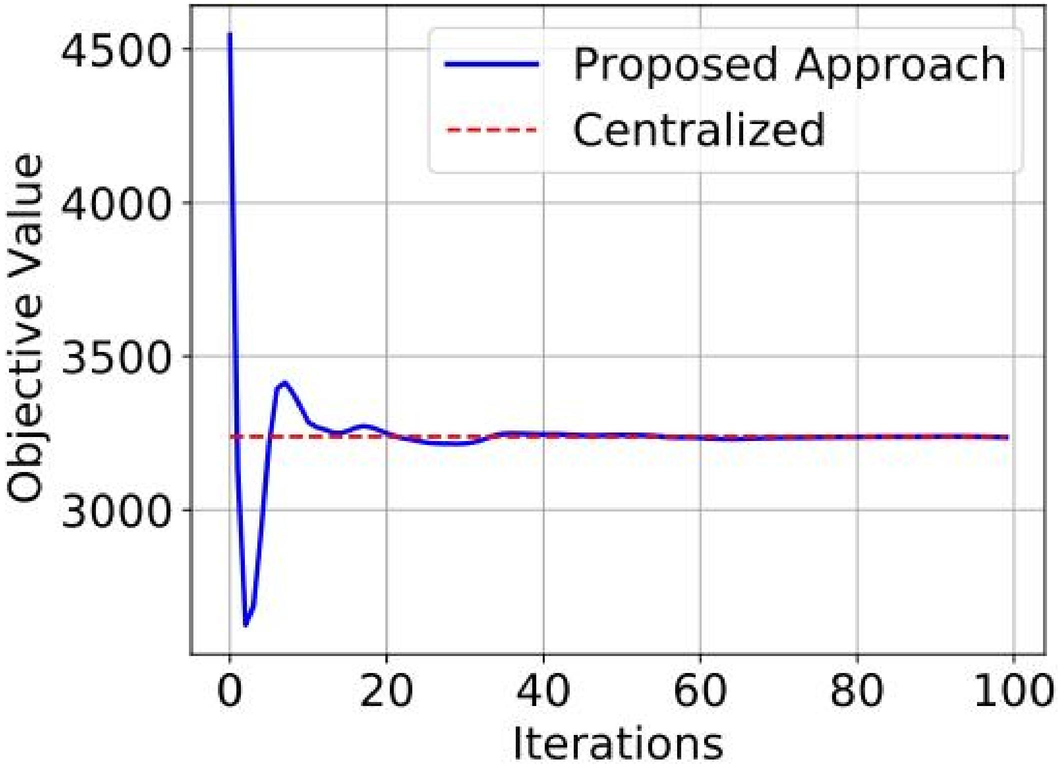
Convergence of PCP

Convergence of RBA

Fast convergence rate within 10 iteration

Tri Nguyen Dang, Choong Seon Hong et al. "Joint Communication, Computation, and Control for Computational Task Offloading in Vehicle-Assisted Multi-Access Edge Computing." *IEEE Access* 10 (2022): 122513-122529.

Use Cases 4: Numerical Results



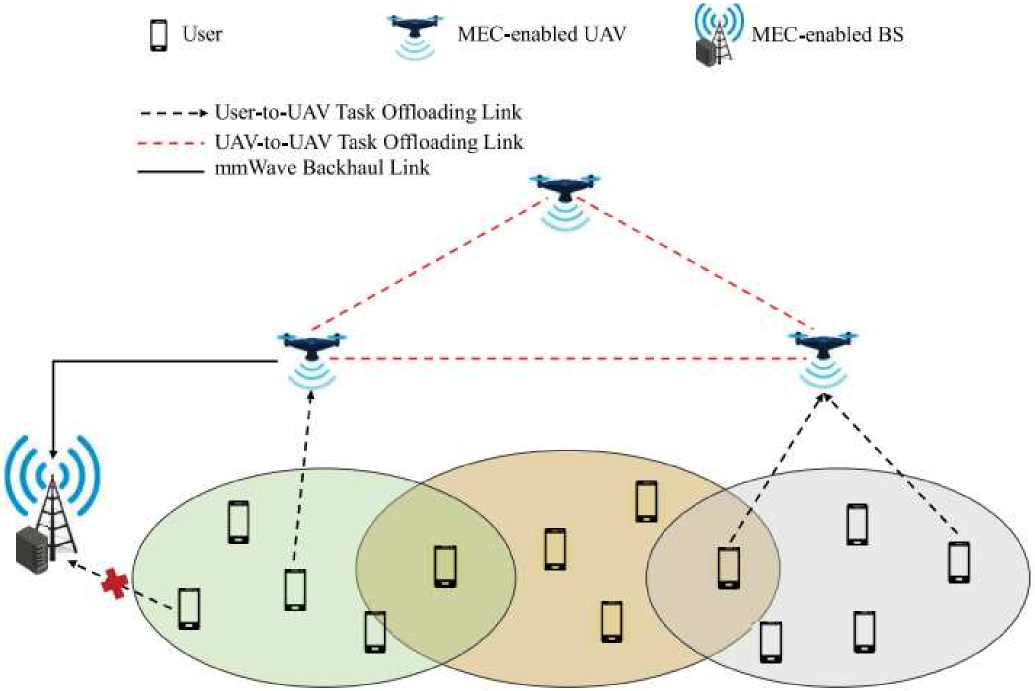
Convergence of RBA + PCP + ODP

Tri Nguyen Dang, Choong Seon Hong et al. "Joint Communication, Computation, and Control for Computational Task Offloading in Vehicle-Assisted Multi-Access Edge Computing." *IEEE Access* 10 (2022): 122513-122529.

Use case 5: Collaboration in the Sky: A Distributed Framework for Task Offloading and Resource Allocation in Multi-Access Edge Computing

- System model
- Problem formulation
- Solution approach
- Simulation results

Use Cases 5: System Model



Goal: Minimize total latency of the network

- ➢ UAVs-assisted multi-access edge computing:
  - ✓ UE is able to offload task to UAVs
  - ✓ UAV sharing resource with each other to speedup the task processing
  - ✓ Increase the capability of MEC
- ➢ We propose a joint optimization problem:
  - ✓ Bandwidth allocation
  - ✓ Task offload decision
  - ✓ Collaboration among UAVs

KYUNG HEE UNIVERSITY

Tun, Yan Kyaw, Choong Seon Hong et al. "Collaboration in the sky: A distributed framework for task offloading and resource allocation in multi-access edge computing." *IEEE Internet of Things Journal* 9.23 (2022): 24221-24235.

NETWORKING INTELLIGENT LAB

Use Cases 5: Problem Formulation

Bandwidth allocation

UE-to-UAV

UAV-to-UAV

Total latency of the network

Task partition

UAV-to-MEC

$$\mathbf{Z}_v(\boxed{\boldsymbol{\alpha}}, \boxed{\boldsymbol{\beta}}, \boxed{\boldsymbol{\theta}}, \boxed{\boldsymbol{\gamma}}, \boxed{\boldsymbol{\phi}}) = \left(\boxed{\sum_{u \in U_v} t_u^{v,\text{loc}} + \theta_u^{u \to v} t_u^{v,e} + \sum_{w \in \mathcal{V}, w \neq v} \gamma_u^{v \to w} t_u^{v \to w,e} + \phi_u^{v \to 0} t_u^{v \to 0,e}} \right) \quad \forall v \in \mathcal{V}$$

Local UE latency

$$t_u^{v,\text{loc}} = \frac{C_u(S_u - \alpha_u^v)}{f_u} \quad \forall u \in \mathcal{U} \quad \forall v \in \mathcal{V}$$

Required CPU cycles/Size of task
Local Computing Capacity

UAV-to-UAVs latency

$$t_u^{v \to w,e} = t_u^{v,\text{up}} + t_u^{v \to w,\text{com}} + t_u^{v \to w,\text{comp}} \quad \forall u \in \mathcal{U}_v \quad \forall v, w \in \mathcal{V}$$

UE-to-UAV latency

$$t_u^{v,e} = t_u^{v,\text{up}} + t_u^{v,\text{comp}}$$

UAV-to-MEC latency

$$t_u^{v \to 0,e} = t_u^{v,\text{up}} + t_u^{v \to 0,\text{com}} + t_u^{v \to 0,\text{comp}} \quad \forall u \in \mathcal{U}_v \quad \forall v \in \mathcal{V}.$$

Tun, Yan Kyaw, Choong Seon Hong et al. "Collaboration in the sky: A distributed framework for task offloading and resource allocation in multi-access edge computing." *IEEE Internet of Things Journal* 9.23 (2022): 24221-24235.

Use Cases 5: Problem Formulation

$$\mathbf{P}: \underset{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\phi}}{\text{minimize}} \quad \mathbf{Z}(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\phi}) \qquad (40a)$$

$$\text{subject to} \quad E_u^{\text{loc}} + E_u^{v,\text{up}} \leq E_u^{\mathbf{max}} \quad \forall u \in \mathcal{U}_v \qquad (40b)$$

$$\sum_{u \in \mathcal{U}_v} \theta_u^{u \to v} E_u^{v,\text{comp}} + \sum_{w \in \mathcal{V}, w \neq v} E^{v \to w} + E^{v \to 0} + E^{v,\text{hov}} \leq E_v^{\mathbf{max}} \quad \forall v \in \mathcal{V} \qquad (40c)$$

$$\sum_{u \in \mathcal{U}_v} E_u^{v \to w,\text{comp}} + E^{w,\text{hov}} \leq E_w^{\mathbf{max}} \quad \forall w \in \mathcal{V},\ w \neq v \qquad (40d)$$

$$\sum_{u \in \mathcal{U}_v} \beta_u^v \leq 1 \quad \forall v \in \mathcal{V} \qquad (40e)$$

$$\beta_u^v \in [0,1], \quad \forall u \in \mathcal{U}_v \quad \forall v \in \mathcal{V} \qquad (40f)$$

$$\theta_u^{u \to v} + \sum_{\substack{w \in \mathcal{V},\\ w \neq v}} \gamma_u^{v \to w} + \phi_u^{v \to 0} = 1 \quad \forall u \in \mathcal{U}_v \qquad (40g)$$

$$0 \leq \alpha_u^v \leq S_u \quad \forall u \in \mathcal{U} \quad \forall v \in \mathcal{V}, \qquad (40h)$$

$$\theta_u^{u \to v} \in \{0,1\} \quad \forall u \in \mathcal{U}_u \quad \forall v \in \mathcal{V} \qquad (40i)$$

$$\gamma_u^{v \to w} \in \{0,1\} \quad \forall u \in \mathcal{U}_u \quad \forall v, w \in \mathcal{V} \qquad (40j)$$

$$\phi_u^{v \to 0} \in \{0,1\} \quad \forall u \in \mathcal{U}_u \quad \forall v \in \mathcal{V} \qquad (40k)$$

Energy for offloading and local processing less than or equal to capacity of local user

Base load energy + energy to help other UAVs + energy for offloading to MEC + hovering energy must less than maximum energy of UAV

Bandwidth allocation variable

Wherever to offload, the task must be 100% processed!

Task partitioning: percentage of task to offload

Decision of UAV v on user u's task

Decision of UAV v on request of UAV w

Decision of MEC for request of user u

Energy constraints

Bandwidth constraints

QoS constraint

Decision variable constraints

Tun, Yan Kyaw, Choong Seon Hong et al. "Collaboration in the sky: A distributed framework for task offloading and resource allocation in multi-access edge computing." *IEEE Internet of Things Journal* 9.23 (2022): 24221-24235.

Use Cases 5: Solution Approach

$$\textbf{P1:} \ \underset{\boldsymbol{\alpha}}{\text{minimize}} \ \mathbf{Z}(\boldsymbol{\alpha}) \tag{41a}$$

$$\text{subject to} \ E_u^{\text{loc}} + E_u^{v,\text{up}} \leq E_u^{\mathbf{max}} \quad \forall u \in \mathcal{U}_v \tag{41b}$$

$$0 \leq \alpha_u^v \leq S_u \quad \forall u \in \mathcal{U} \quad \forall v \in \mathcal{V}. \tag{41c}$$

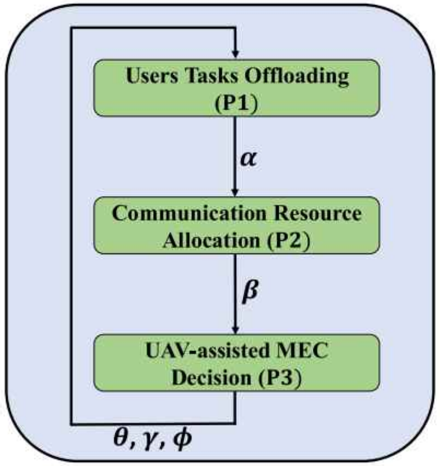
1. Iteratively obtain solution for P1 by fixing P2, and P3
2. Fixed P3, updated P2 based on P1
3. Update P3 based solution of P1 and P2 on last iteration

$$\textbf{P2:} \ \underset{\boldsymbol{\beta}}{\text{minimize}} \ \mathbf{Z}(\boldsymbol{\beta}) \tag{42a}$$

$$\text{subject to} \ E_u^{v,\text{up}} \leq E_u^{\mathbf{max}} \quad \forall u \in \mathcal{U}_v \tag{42b}$$

$$\sum_{u \in \mathcal{U}_v} \beta_u^v \leq 1 \quad \forall v \in \mathcal{V} \tag{42c}$$

$$\beta_u^v \in [0, 1] \quad \forall u \in \mathcal{U}_v \quad \forall v \in \mathcal{V}. \tag{42d}$$



$$\textbf{P3:} \ \underset{\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\phi}}{\text{minimize}} \ \mathbf{Z}(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\phi}) \tag{50a}$$

$$\text{subject to} \ \sum_{u \in \mathcal{U}_v} \theta_u^{u \to v} E_u^{v,\text{comp}} + \sum_{w \in \mathcal{V}, w \neq v} E^{v \to w} + E^{v \to 0} + E^{v,\text{hov}} \leq E_v^{\mathbf{max}} \quad \forall v \in \mathcal{V} \tag{50b}$$

$$\sum_{u \in \mathcal{U}_v} E_u^{v \to w,\text{comp}} + E^{w,\text{hov}} \leq E_w^{\mathbf{max}} \quad \forall w \in \mathcal{V} \quad w \neq v \tag{50c}$$

$$\theta_u^{u \to v} + \sum_{\substack{w \in \mathcal{V}, \\ w \neq v}} \gamma_u^{v \to w} + \phi_u^{v \to 0} = 1 \quad \forall u \in \mathcal{U}_v \tag{50d}$$

$$\theta_u^{u \to v} \in \{0, 1\} \quad \forall u \in \mathcal{U}_u \quad \forall v \in \mathcal{V} \tag{50e}$$

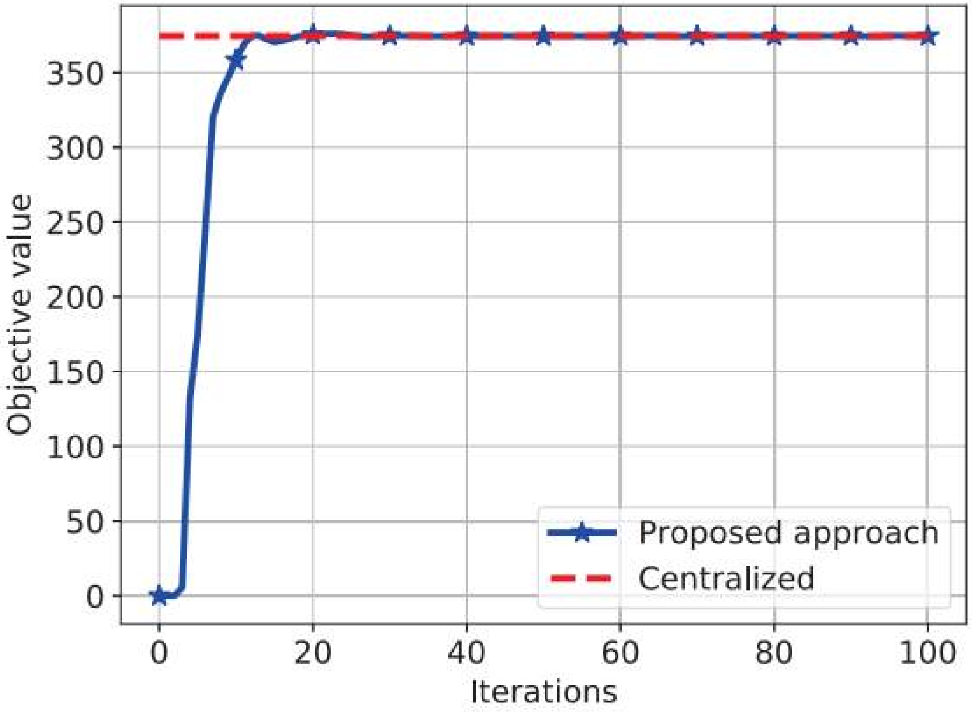
$$\gamma_u^{v \to w} \in \{0, 1\} \quad \forall u \in \mathcal{U}_u \quad \forall v, w \in \mathcal{V} \tag{50f}$$

$$\phi_u^{v \to 0} \in \{0, 1\} \quad \forall u \in \mathcal{U}_u \quad \forall v \in \mathcal{V}. \tag{50g}$$

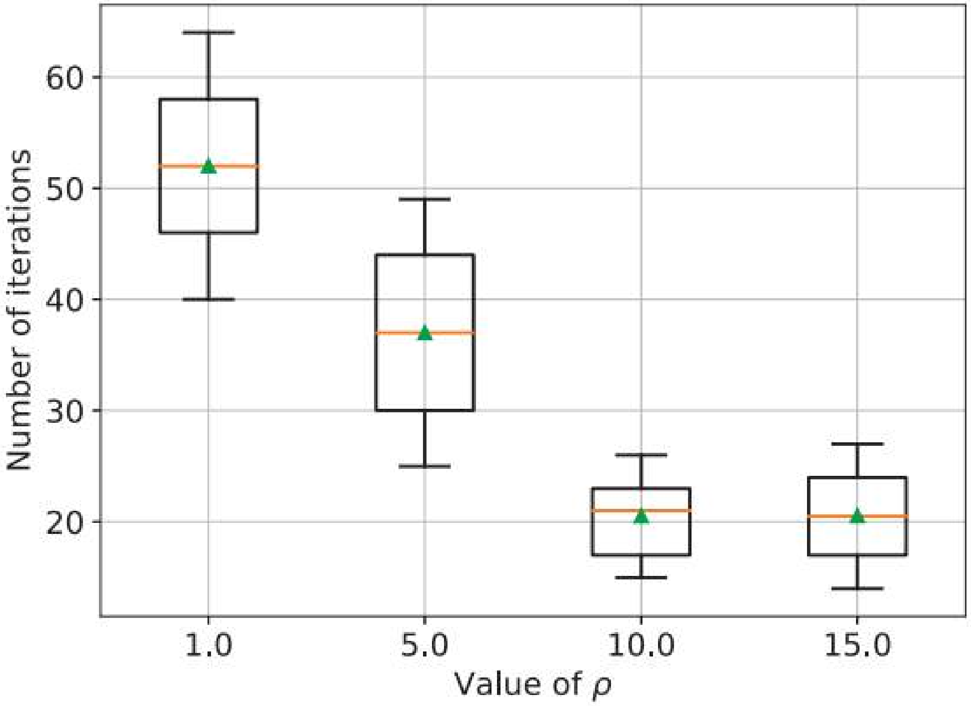
This iterative algorithm always guarantee an optimal solution with in O($\frac{1}{\varepsilon^2}$), where ε is convergence rates which is positive and strictly small, i.e., 1e-4

Tun, Yan Kyaw, Choong Seon Hong et al. "Collaboration in the sky: A distributed framework for task offloading and resource allocation in multi-access edge computing." *IEEE Internet of Things Journal* 9.23 (2022): 24221-24235.

Use Cases 5: Numerical Results



Convergence performance of proposed algorithm

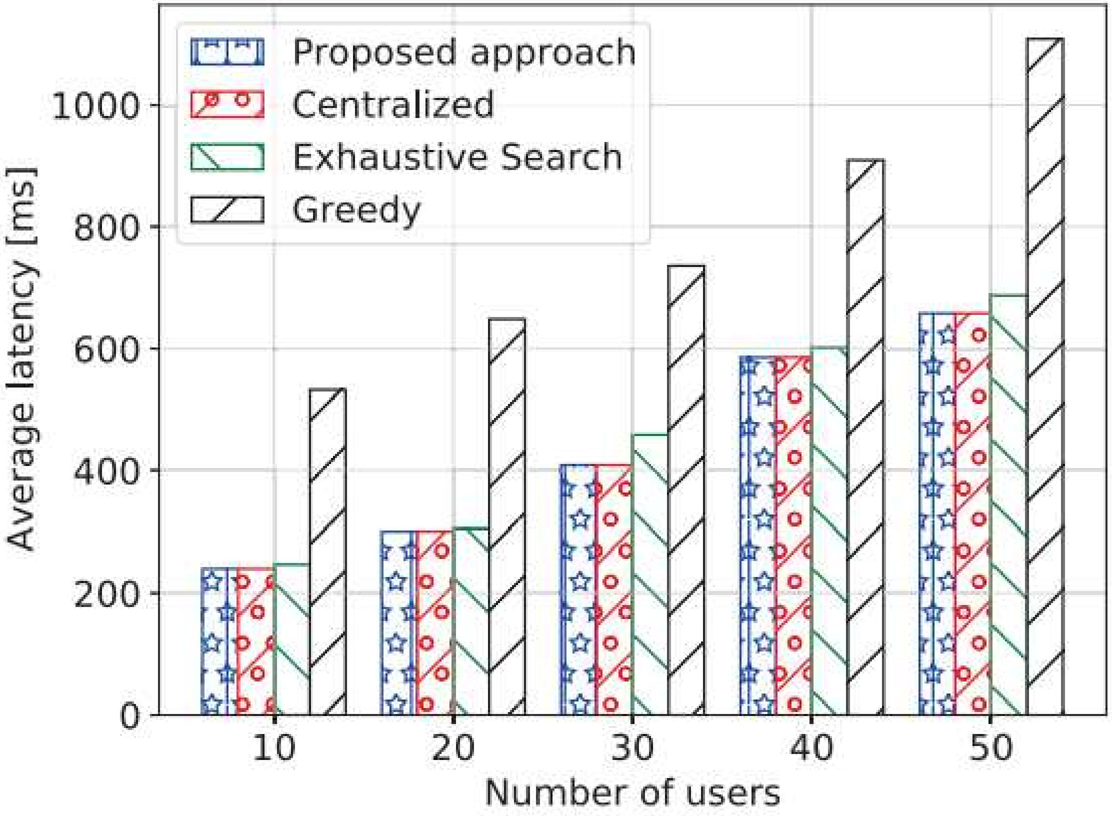


Convergence rate versus penalty parameters

*Rho is penalty parameter for controlling the augmented Lagrangian in P3.

Tun, Yan Kyaw, Choong Seon Hong et al. "Collaboration in the sky: A distributed framework for task offloading and resource allocation in multi-access edge computing." *IEEE Internet of Things Journal* 9.23 (2022): 24221-24235.

Use Cases 5: Numerical Results



Average latency of proposed approach versus based-lines

Centralized Algorithm: the BS serves as a central coordinator
Greedy Algorithm: Considering availability of the neighboring UAVs on the bandwidth between a UAV and it's neighbors and it's neighbors computing resources

KYUNG HEE UNIVERSITY

Tun, Yan Kyaw, Choong Seon Hong et al. "Collaboration in the sky: A distributed framework for task offloading and resource allocation in multi-access edge computing." *IEEE Internet of Things Journal* 9.23 (2022): 24221-24235.

NETWORKING INTELLIGENT LAB

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.
    - Require migration of services 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

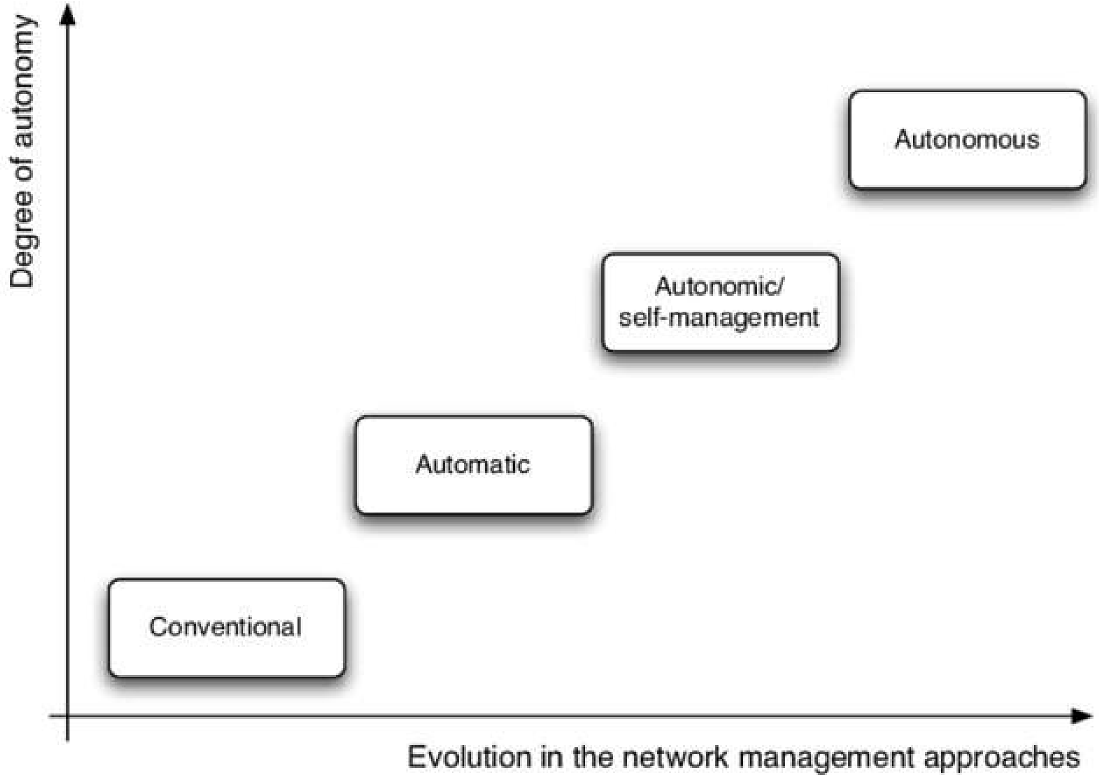
- This lecture is mainly focus to understand a full view of the resource management problem in 5G networks.

- 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