Distributed Management of Network Slices in Beyond 5G

Energy Efficient Techniques of MonB5G

Infocom World 2022 – Athens, Greece November 29, 2022

Presenter: Mrs. Vasiliki Vlachodimitropoulou

Hellenic Telecommunications Organization S.A. (OTE)

Fixed Network R&D Programs Section

R&D Department, Fixed & Mobile

This Project has received funding from the European Union´s Horizon 2020 research and innovation programme under Grant Agreement No. 871780

General info

Management of a massive number of network slices across domains that is hierarchical, distributed, scalable, and driven by artificial intelligence, with the end goal of zero-touch management

Grant Agreement No.: 871780

www.monb5g.eu

@monb5g

monb5g

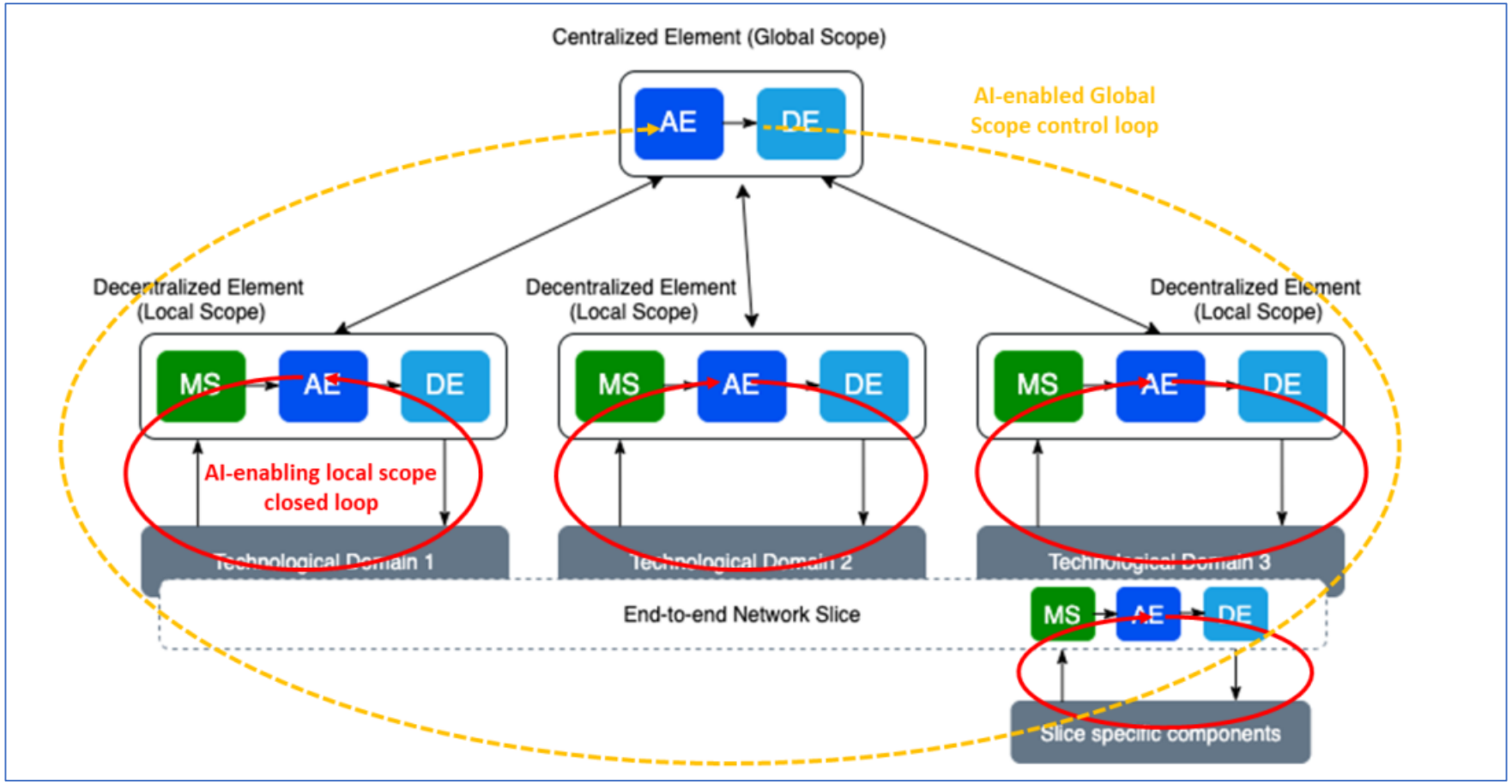
Centre Tecnologic De Telecomunicacions De Catalunya Spain	Eurecom France	Hellenic Telecommunications Organization S.A. - Ote Ae Greece	L.M. Ericsson Limited Ireland
Citrix Ellas Monoprosopi Etairia Periorismenis Evthinis Greece	Orange Sa France	Orange Polska Spolka Akcyjna Poland	Aalto Korkeakoulusaatio Sr Finland
B-Com France	Iquadrat Informatica Sl Spain	NEC Laboratories Europe Gmbh Germany	eBOS Technologies Limited Cyprus

MonB5G concept for Energy Efficiency

- MonB5G architecture
- Main energy management and orchestration architectural building blocks
- MonB5G reference architecture for energy management
- Main energy-efficient algorithmic innovations of MonB5G
- Conclusions
- MonB5G Proof-of-Concepts

High level Architecture Concept

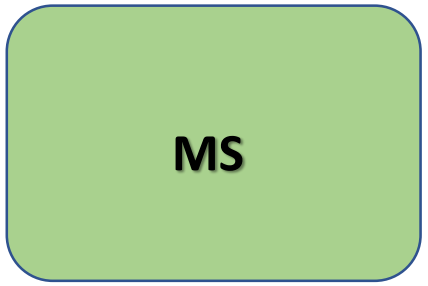
MonB5G framework has been designed for AI-driven management and orchestration of massive number of NSIs



Building blocks of the Architecture

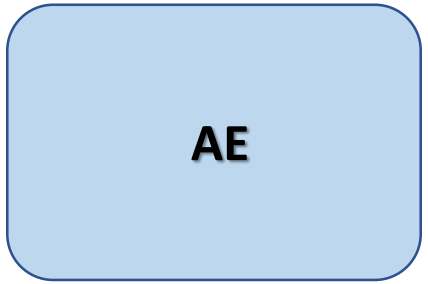
Closed control loop with security and energy management AI-driven mechanisms

MS: Monitoring System



- ✓ Distributed Monitoring
- ✓ Graph-based representation
- ✓ Low monitoring load

AE: Analytics Engine

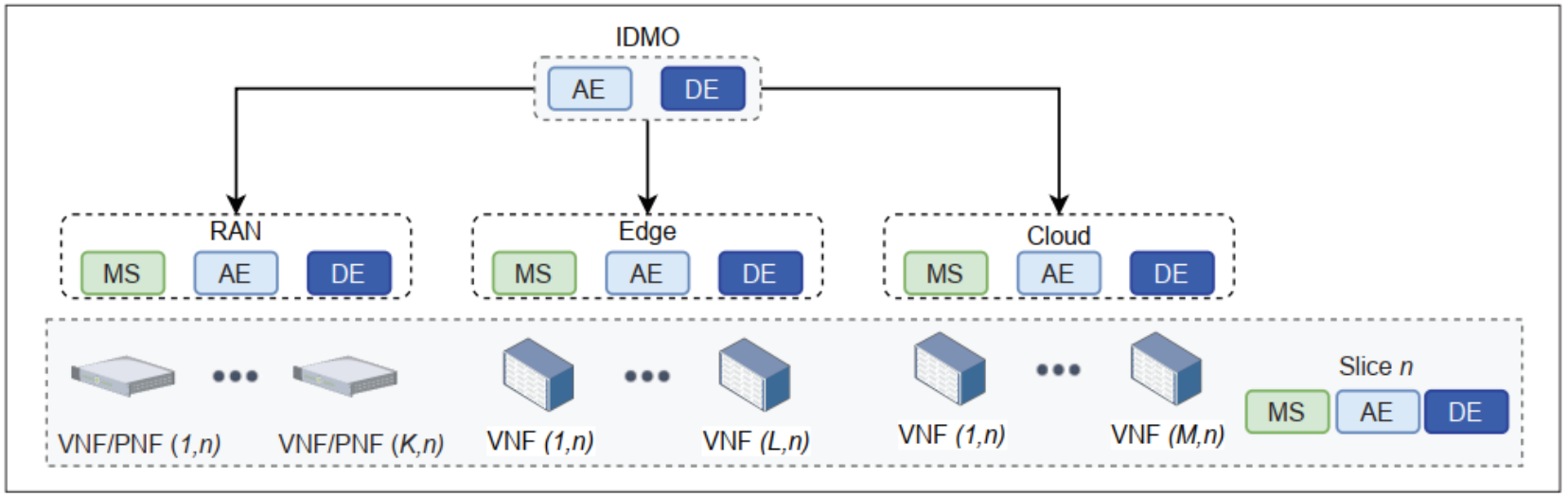


- ✓ Distributed ML (Federated Learning)
- ✓ Slice-level KPI prediction
- ✓ Auto-encoder compress layer management

DE: Decision Engine

- ✓ Distributed Reinforcement Learning for slice orchestration
- ✓ Data-driven inter-slice management
- ✓ Energy-driven reward functions for DRL-based DE

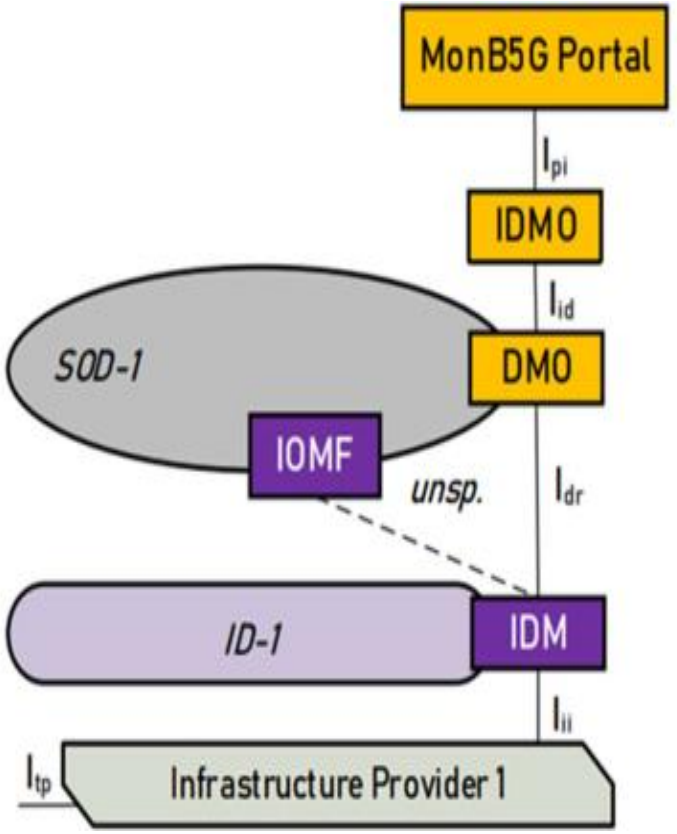
MonB5G architecture overview

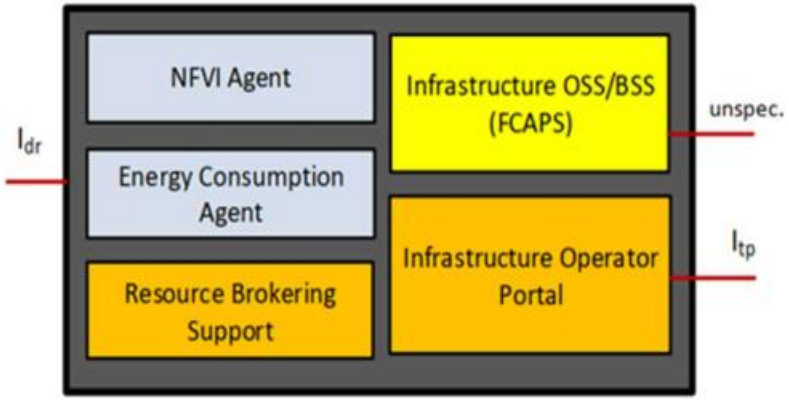


MS, AE, and DE instances at

- Virtual network function (VNF)/physical network function (PNF) level
- Slice-level
- Domain-level (RAN, edge, cloud)
- **IDMO** -Inter-domain manager and orchestrator (IDMO) level, which, like Network Slice Management Function (NSMF), manages the life cycle of E2E network slices by making global E2E analyses or decisions at both the cross-slice and cross-domain levels

MonB5G reference architecture for energy management





Infrastructure Domain Manager (**IDM**)

Domain specific **DMO**(s) management and orchestration

IOMFs Infrastructure orchestrated management functions can optimize infrastructure utilization efficiency and achieve effective infrastructure management

The framework is **aware of the energy costs** associated with infrastructure resources

Energy Efficiency techniques

Minimization of the MS Measurement Load by Adding an Internal Memory like a time-series database (TSDB). This memory block enables avoidance of implementing energy-demanding synchronization among the MS, DE, and AE

Energy efficiency at RAN and Edge

Energy-efficient Statistical FL-based decentralized AEs

Stochastic FL-based policy for scalable AE

AE constrained federated learning is considered to reduce the amount of raw data exchanged between local AEs and the end-to-end AE

Decentralized Cross-Domain Energy Efficient DE

DEs use Decentralized Deep Reinforcement Learning strategies (Multi-Agent DRL, Federated DRL) to perform cross-domain energy-aware VNF and SFC placement in 5G service-customized network slices

DEs choose the ideal compromise solution to achieve a balance between energy efficiency and SLAs

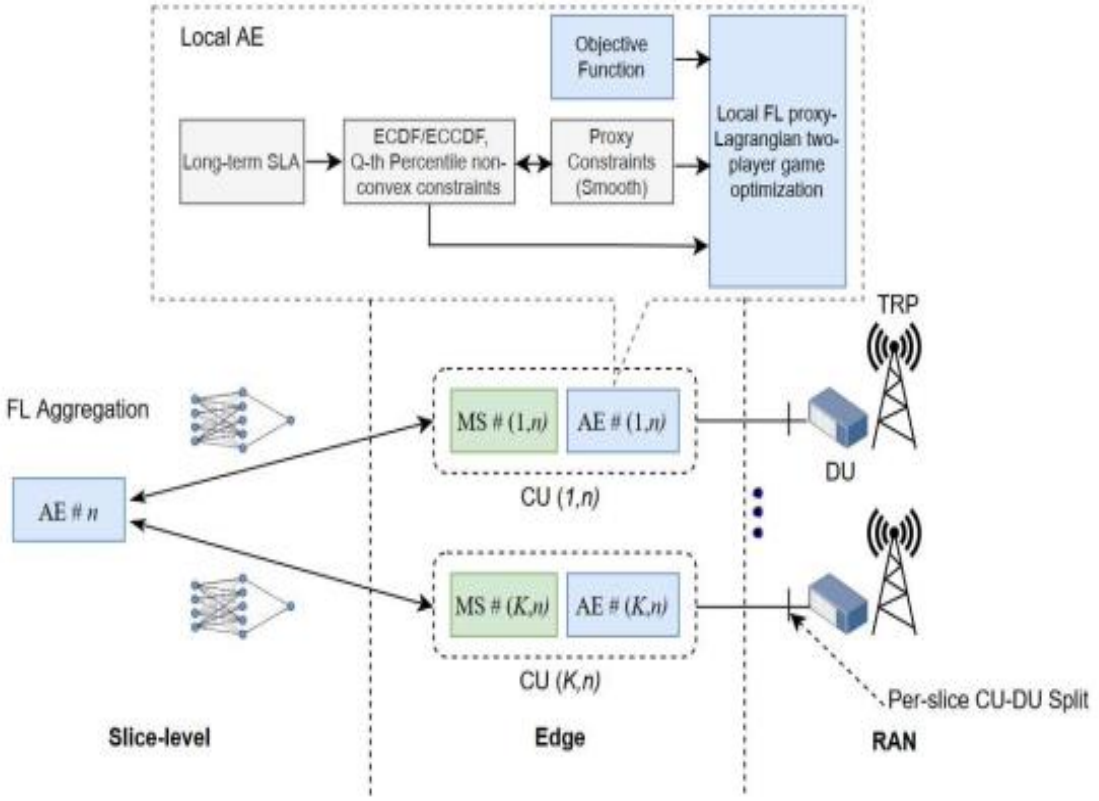
Distributed AI enables the local processing of management information, consequently decreasing the exchange of management information between entities

Energy Efficient Statistical FL-Based Analytics Engines

To address the FL resource provisioning task at the local analytic engines (AE), we define the related SLA-constrained optimization problem within a proxy-Lagrangian framework and solve it with a non-zero sum two-player game strategy

An innovative SLA-driven stochastic FL policy is designed to ensure scalability under massive slicing

Implement the proposed solution in a cloud-native containerized environment



Energy Efficient Statistical FL-Based Analytics Engines

To demonstrate the general framework of Statistical FL, three main slices are considered:

- **eMBB:** *NetFlix, Youtube and Facebook Video*,
- **Social Media:** *Facebook, Facebook Messages, Whatsapp and Instagram,*
- **Browsing:** *Apple, HTTP and QUIC*

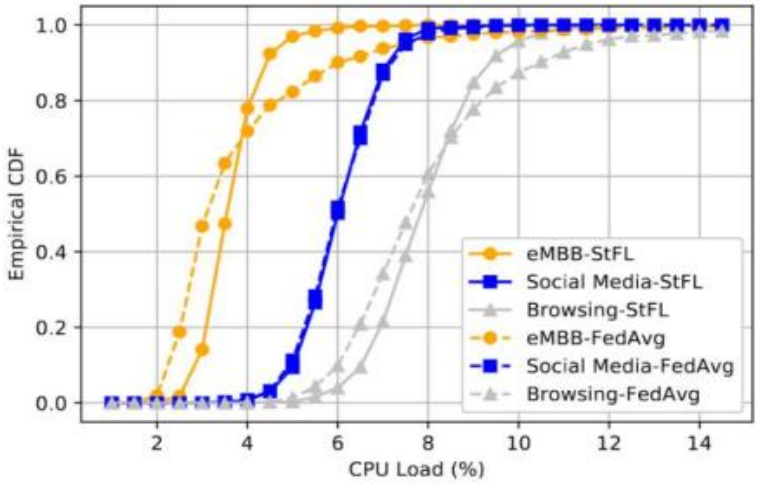
SLA: any assigned resource to the tenant should not exceed a range with a probability higher than an agreed threshold

This translates into learning the CPU resource allocation model under **empirical cumulative density function (CDF)** constraints

The proposed Statistical FL enables controlling of the long-term statistical behaviour of the SLA compared to FedAvg baseline

Statistical Federated Learning Results

Dramatic overhead reduction at convergence



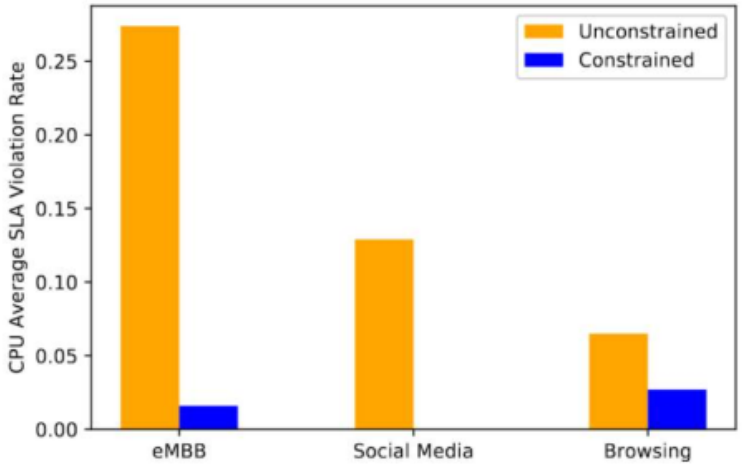
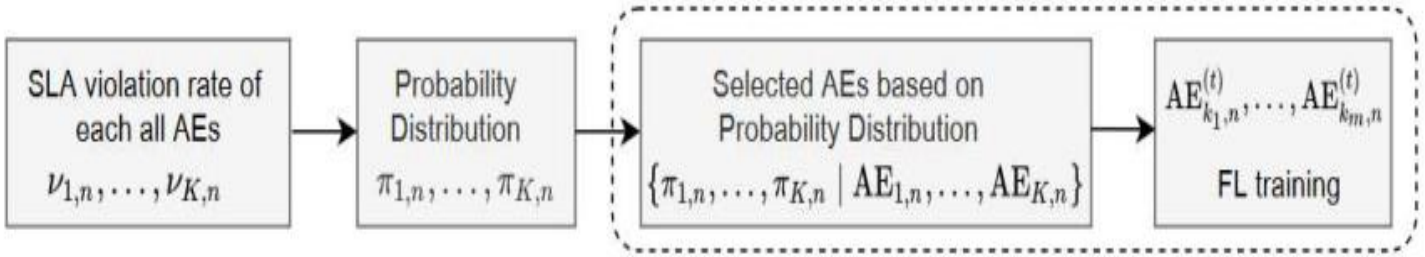


Table I: Overhead and energy comparison

Rounds	50	60	70	80
Overhead CCL (KB)			18750	
Overhead StFL (KB)	1055	1266	1477	1688
Energy CCL (mJ)			118.3	
Energy StFL (mJ)	6.7	8	9.3	10.7
Energy Gain	×**17.8**	×**14.8**	×**12.7**	×**11.1**

SLA-driven stochastic AE Selection Policy



In each FL round, only a subset of active AEs can be chosen

We suggest a stochastic AE selection policy driven by SLA, with the aim to

- ✓ reduce the network data overhead
- ✓ optimize the FL computation time
- ✓ increase the system's energy efficiency

AEs with a low SLA violation have a higher probability to participate in the FL round (softmin function)

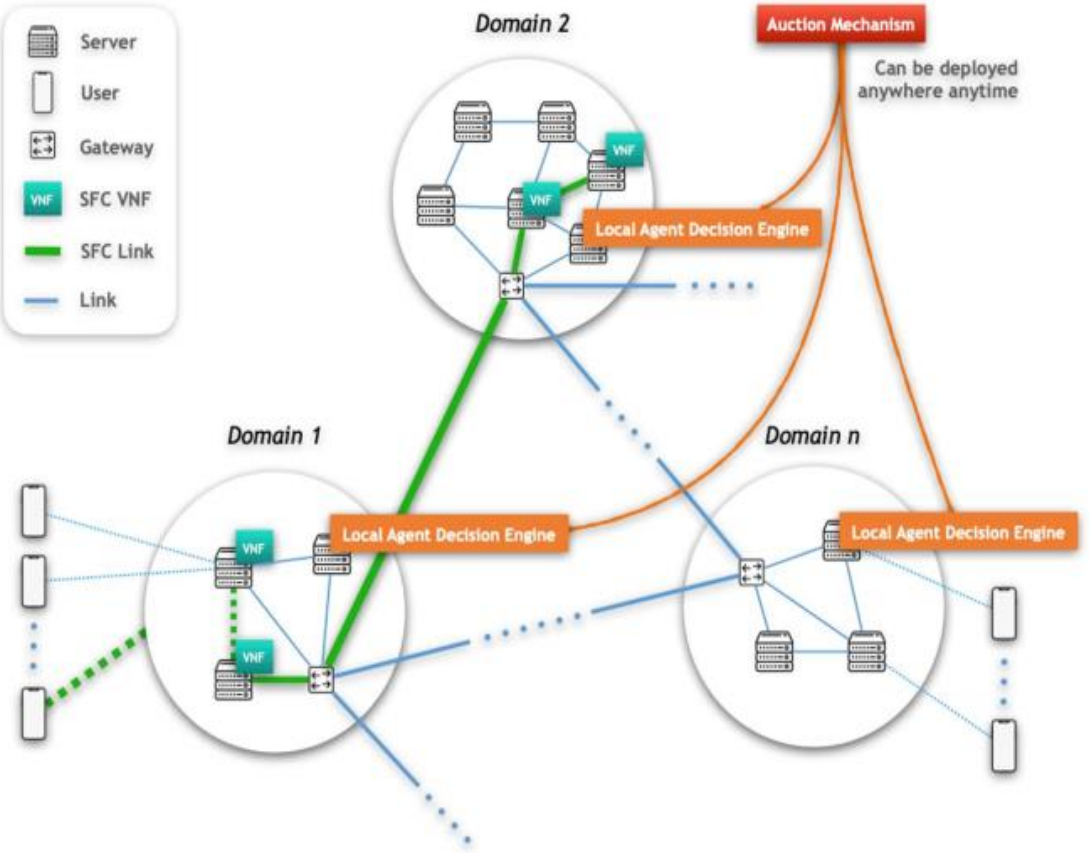
Scalable, Energy-Aware, Multidomain Orchestration for Beyond-5G URLLC Services (SCHEMA)

Problem:

- Zero-touch Service Function Chain (SFC) orchestration for multi-domain networks, targeting the latency reduction of URLLC services while improving energy efficiency for Beyond-5G networks.

Solution:

- DEs use Decentralized Deep Reinforcement Learning strategies to perform cross-domain energy-aware VNF and SFC placement in 5G service-customized network slices.
- We split the network in an inter-domain level graph and multiple intra-domain level graphs.

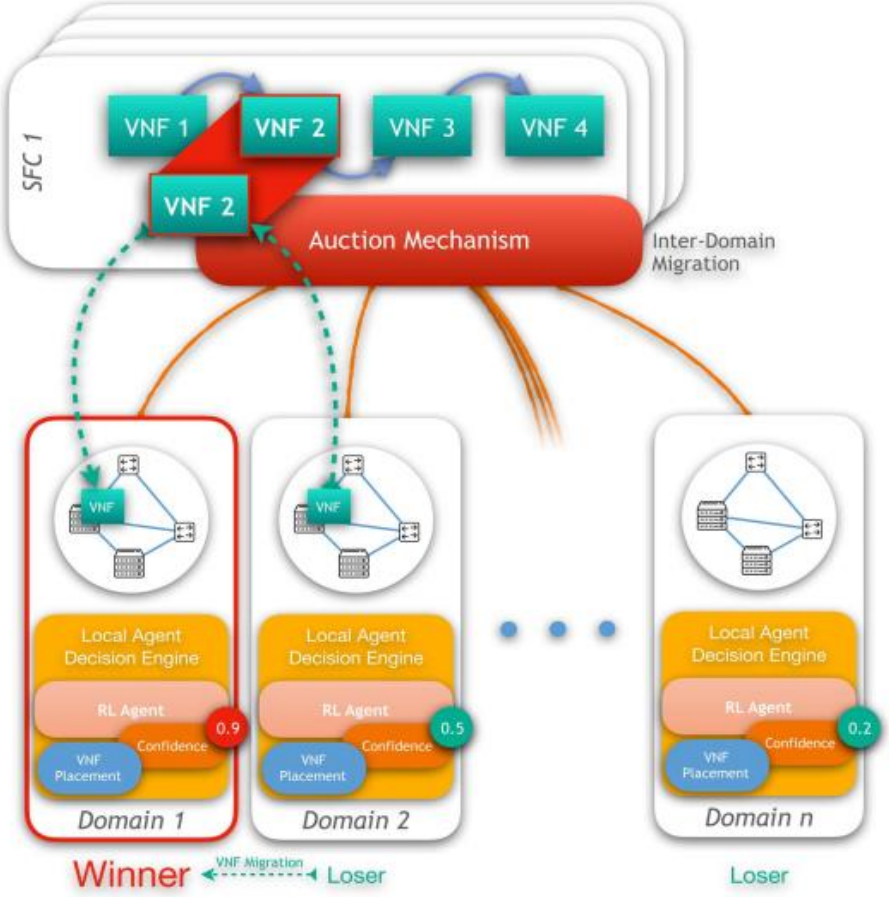


Key SCHE2MA algorithm steps:

1. For every VNF in the network, we perform an auction, letting multiple local domain RL agents bid to receive a specific VNF
2. Local RL agents perform intra-domain orchestration to avoid global network slice reconfiguration

Auction mechanism steps:

1. Auction Initiation
2. Distributed Operation
3. Global Operation
4. Orchestration
5. Iteration



Scalable, Energy-Aware, Multidomain Orchestration for Beyond-5G URLLC Services (SCHE2MA)

Baseline Scenarios:

Two reference cases from the literature are used to evaluate the proposed SCHE2MA solution

1. *Centralized RL*

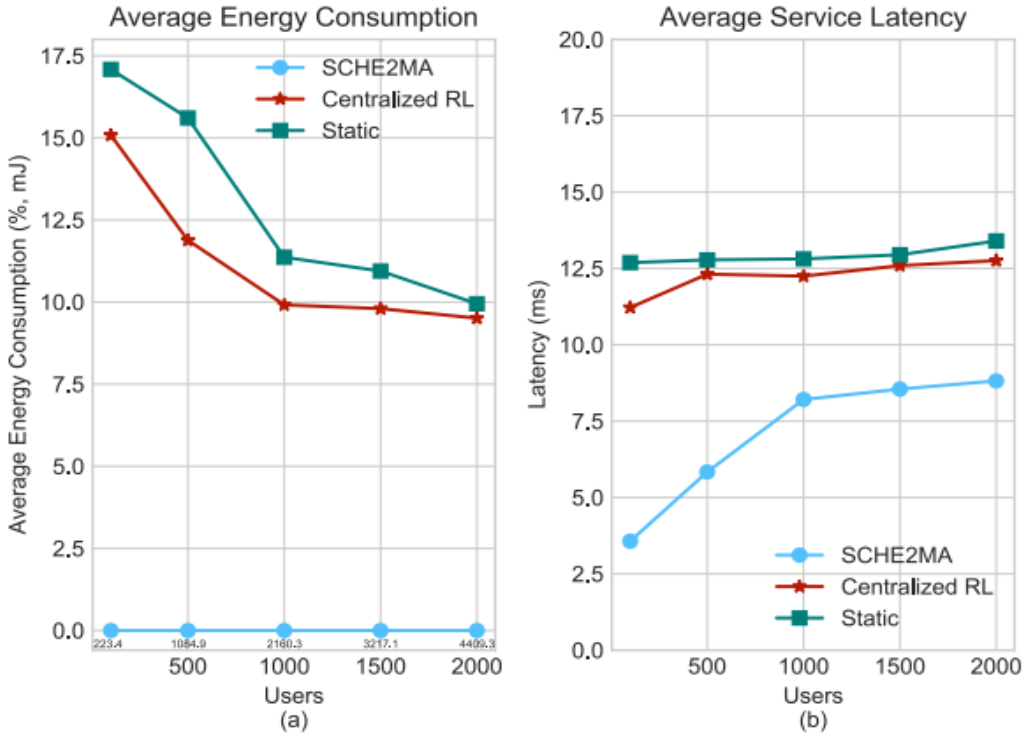
   The central orchestration algorithm where in every VNF in the network we perform an auction, letting multiple local domain RL agents bid to receive a specific VNF.

2. *Static Placement*

   In this strategy the VNF placement is static and the VNFs remain hosted in the initial node throughout the experiment.

Results:

- The performance of SCHE2MA outperforms the two baselines SCHE2MA demonstrates a clear indication of its ability to conceive better VNF placements that satisfy the latency and energy consumption trade-off.
- This behavior is a result of SCHE2MA's ability to cluster VNFs within servers, minimizing the number of transmissions in physical media therefore decreasing costly communication between servers.

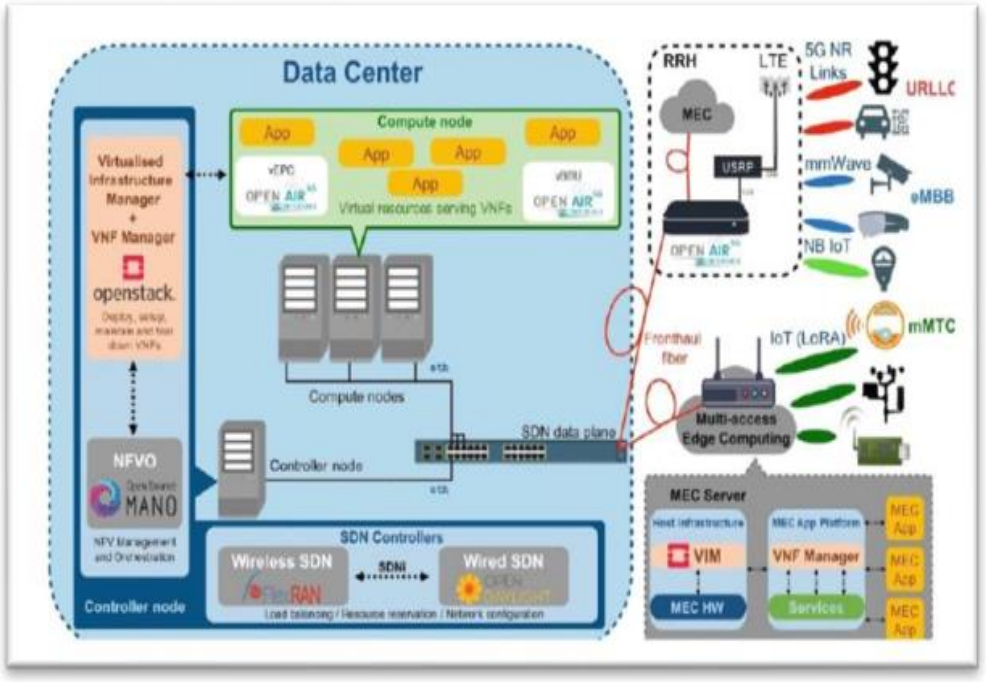


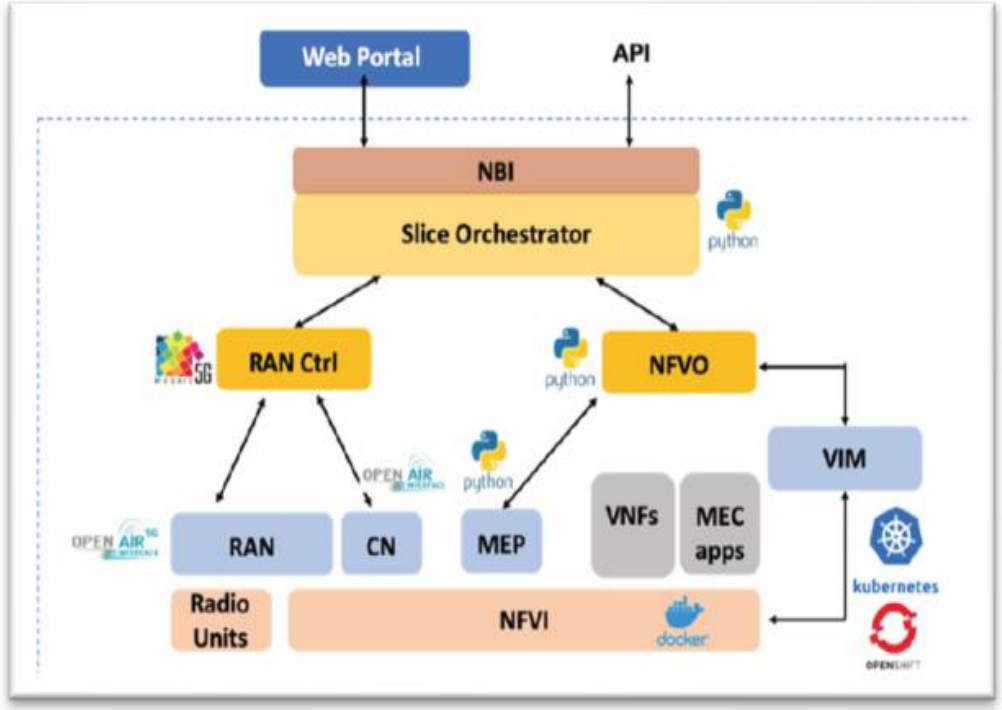
a) Average energy consumption improvement
b) Average service latency per number of users in multiple traffic scenarios

Conclusions

- Distributed management and components with embedded intelligence reduce the amount of the information exchanged for management purposes, the management system response time and the energy cost.
- IOMF functions can improve the effectiveness of infrastructure utilization and contribute to the overall quality of infrastructure management and, as a result, to achieve energy saving goals.
- Scalable cloud-native SLA-driven stochastic FL policy for zero-touch network slicing resource allocation reduces the corresponding computation cost and SLA violation.
- Statistical federated learning (StFL)-based Analytical engine for slice-level KPI prediction achieves greater than an x10 improvement in energy efficiency over its centralized SLA-constrained deep learning equivalent while obtaining an x20 reduction in SLA violations relative to the FedAvg.
- SCHEMA, a Distributed Reinforcement Learning (RL) algorithm through model validation and simulation demonstrated its ability to jointly reduce average service latency and energy consumption compared to a Centralized RL solution and Static Placement.

MonB5G Proof of Concepts





PoC-1: Zero-Touch Network and service management with end-to-end SLAs (CTTC Platform - Barcelona (ES))

PoC-2: AI-assisted policy-driven security monitoring & enforcement (EURECOM Platform - Sophia Antipolis (FR))

References

- Deliverable 5.3 “Final report on AI-driven MonB5G energy efficiency techniques” https://monb5g.eu/deliverables
- H. Chergui, L. Blanco and C. Verikoukis, "Statistical Federated Learning for Beyond 5G SLA-Constrained RAN Slicing," in IEEE Transactions on Wireless Communications, vol. 21, no. 3, pp. 2066-2076, March 2022.
- S. Roy, H. Chergui, L. Sanabria-Russo and C. Verikoukis, "A Cloud Native SLA Driven Stochastic Federated Learning Policy for 6G Zero-Touch Network Slicing,“ IEEE ICC 2022.
- H. Chergui, L. Blanco , L. A. Garrido, K. Ramantas, S. Kuklinski, A. Kasentini, S.Kuklinksli, "Zero-Touch AI-Driven Distributed Management for Energy-Efficient 6G Massive Network Slicing," in IEEE Network, vol. 35, no. 6, pp. 43-49, November/December 2021
- A. Dalgkitsis *et al*.,“SCHE2MA: Scalable, Energy-Aware, Multidomain Orchestration for Beyond-5G URLLC Services” in *IEEE Transactions on Intelligent Transportation Systems*, 2022