Energy-Efficient Orchestration of Metro-Scale 5G Radio Access Networks

Rajkarn Singh, Cengis Hasan[*], Xenofon Foukas[┼], Marco Fiore[†], Mahesh Marina, Yue Wang[❁]

The Univ. of Edinburgh, [*]Univ. of Luxembourg, [┼] Microsoft, [†]IMDEA, [❁]Samsung UK

Coseners MSN'21

1st July 2021

**Presented at IEEE INFOCOM 2021*

THE UNIVERSITY of EDINBURGH informatics

EPSRC Centre for Doctoral Training in Pervasive Parallelism

UNIVERSITÄT LUXEMBURG

Microsoft Research

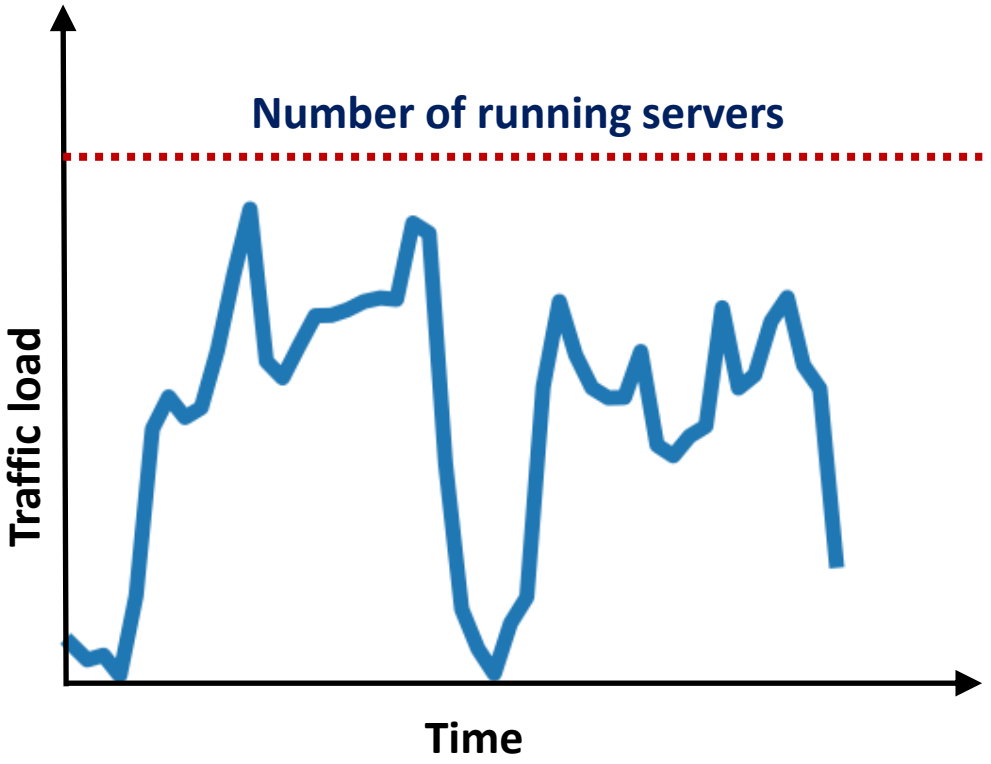
institute IMDEA

SAMSUNG

RAN Energy Consumption Problem

Overprovisioning

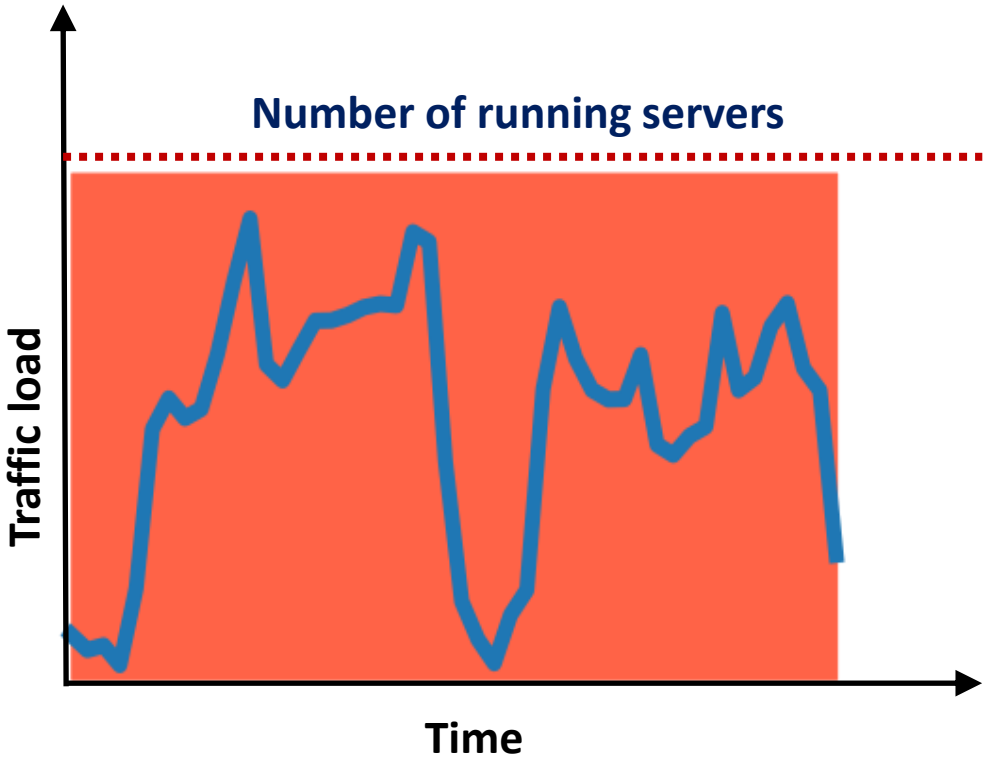
1. Traffic load varies with time
2. Server capacity is greater than demand



RAN Energy Consumption Problem

Overprovisioning

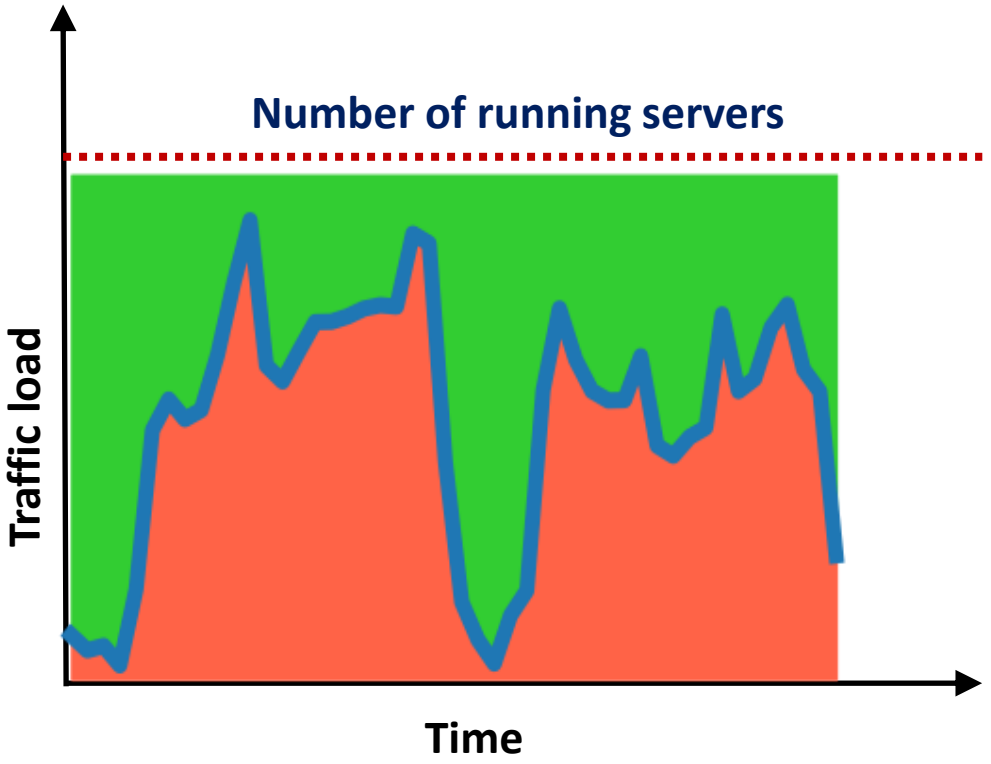
1. Traffic load varies with time
2. Server capacity is greater than demand
3. At lower traffic, static energy incurs huge costs
4. Leads to inefficient resource utilization



RAN Energy Consumption Problem

Overprovisioning

1. Traffic load varies with time
2. Server capacity is greater than demand
3. At lower traffic, static energy incurs huge costs
4. Leads to inefficient resource utilization



Need advanced measures to perform optimal resource orchestration

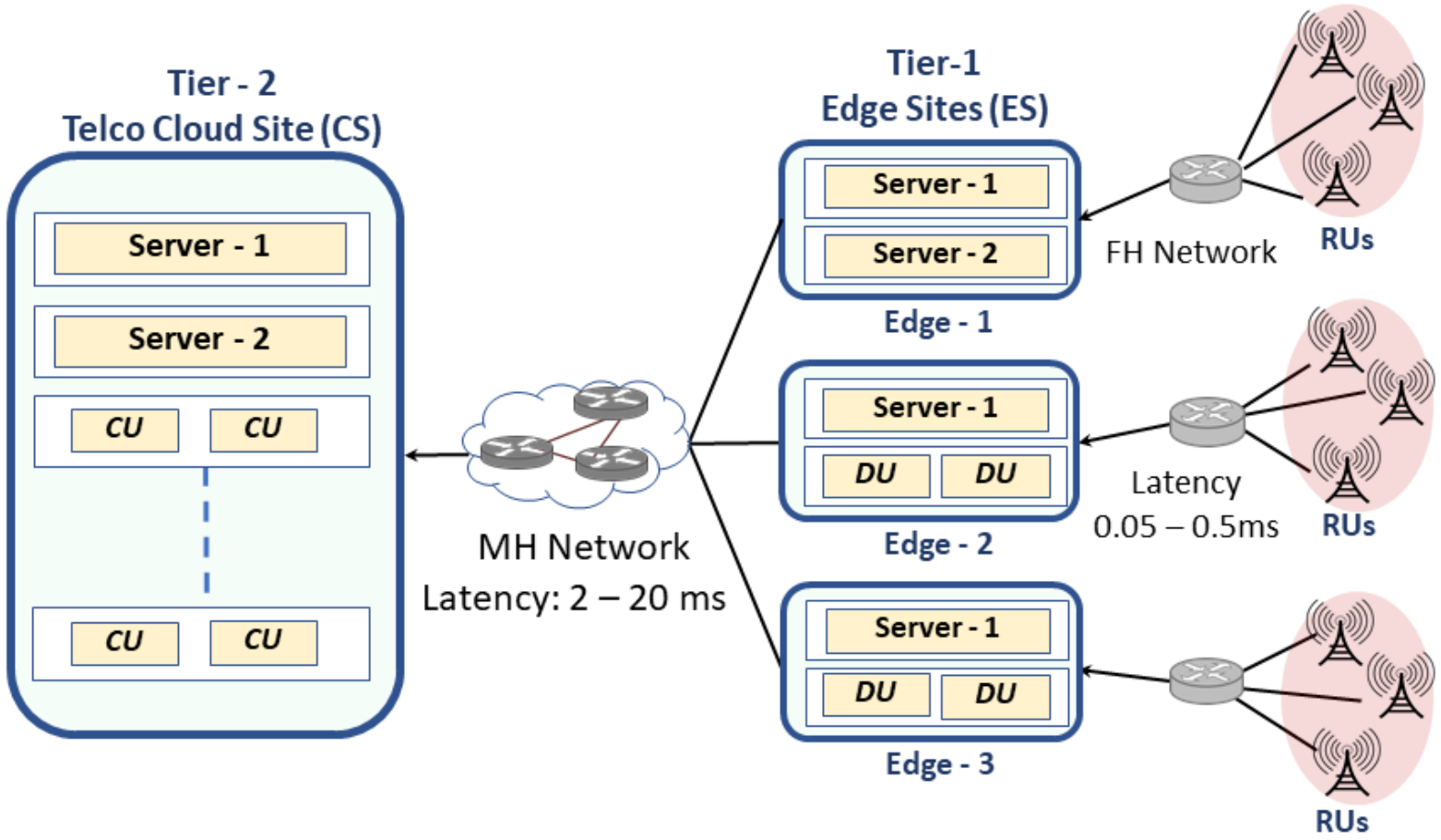
Our Contributions

- Realistic multi-tier cloudified RAN model
- Comprehensive energy model and problem formulation
- A distributed and scalable solution: GreenRAN
- Evaluation with metro-scale real world dataset

Our Contributions

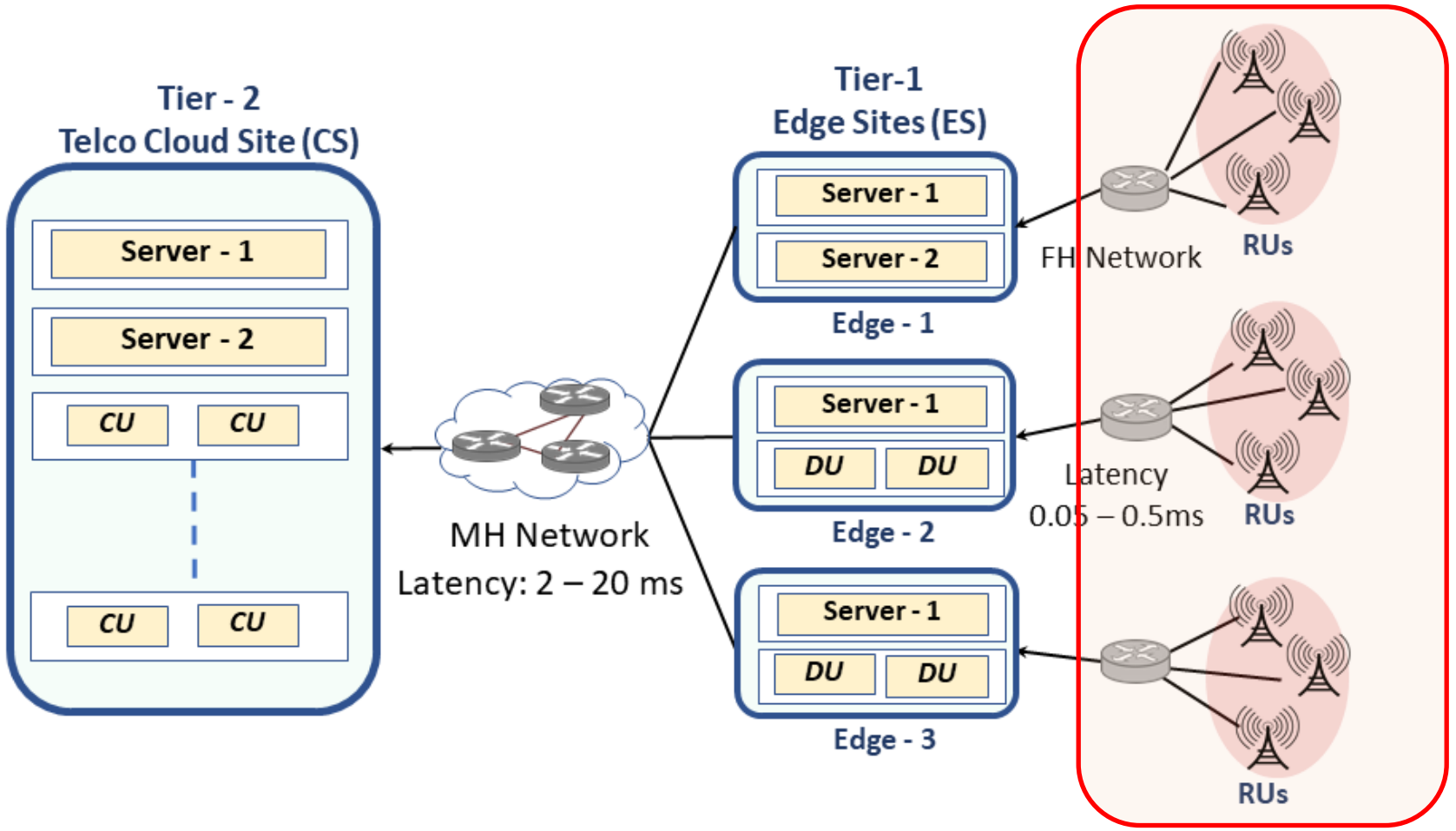
- **Realistic multi-tier cloudified RAN model**
- Comprehensive energy model and problem formulation
- A distributed and scalable solution: GreenRAN
- Evaluation with metro-scale real world dataset

Multi-tier vRAN Architecture



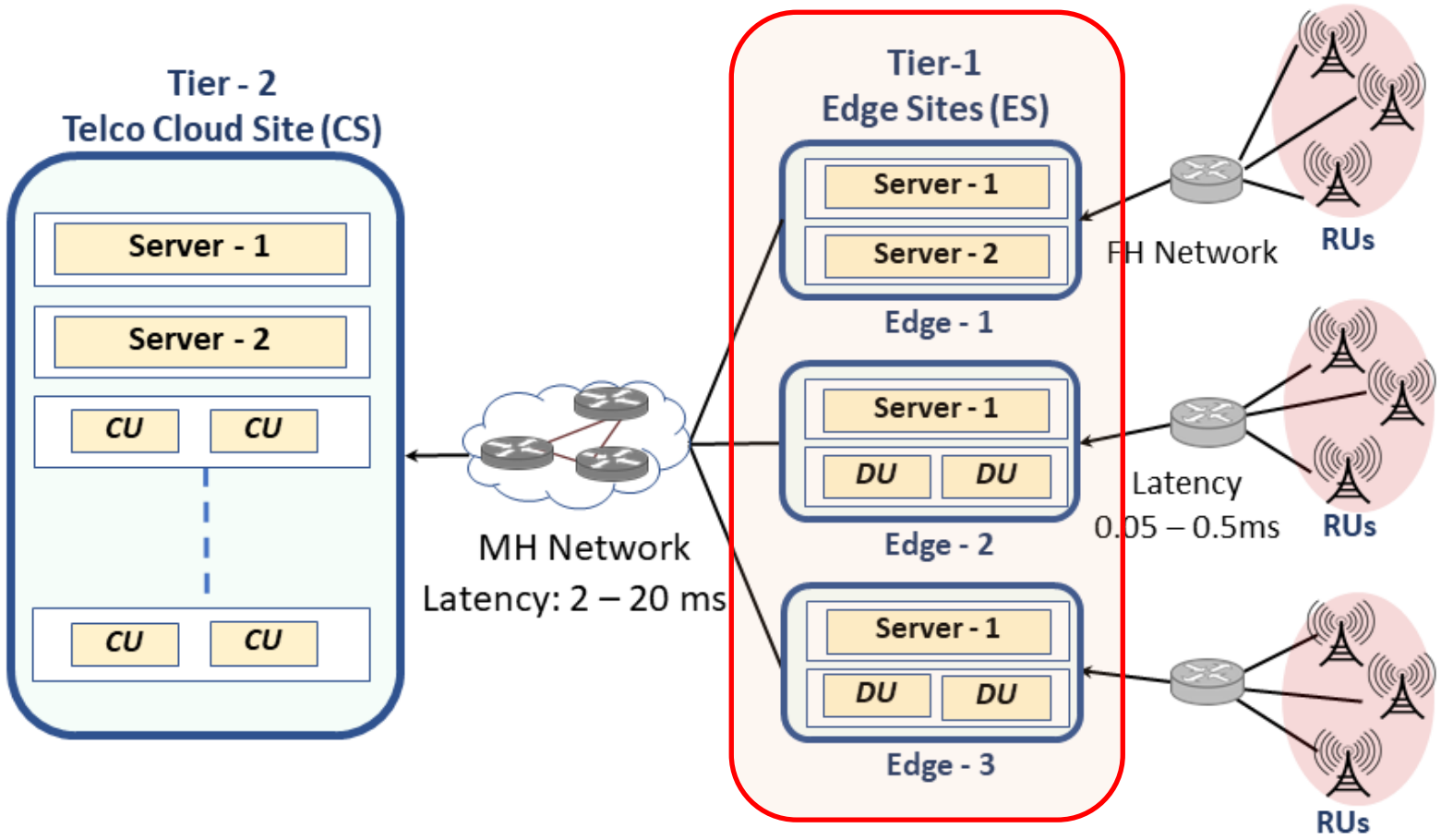
Multi-tier vRAN Architecture

- lowPHY using ASIC & FPGA



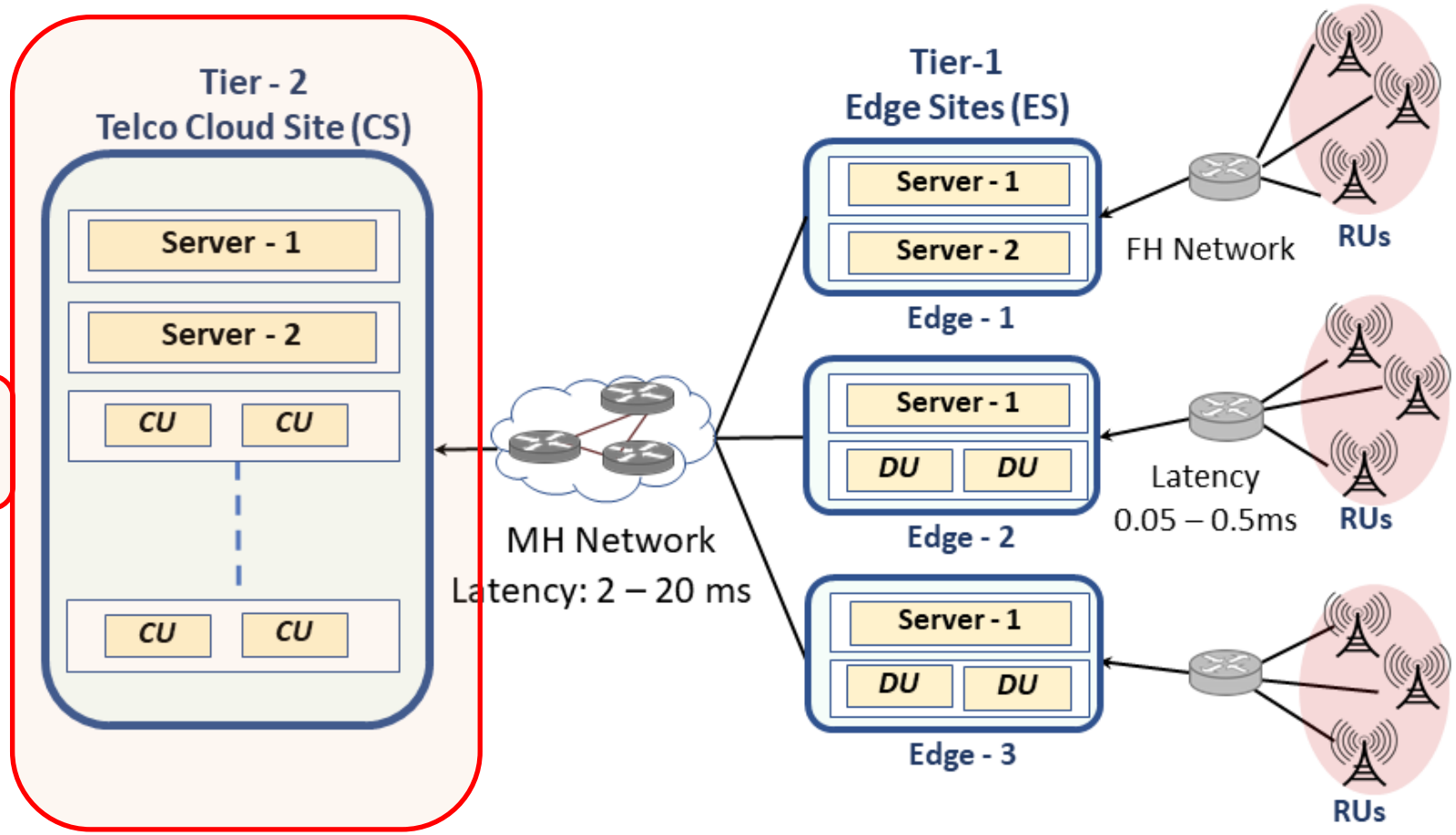
Multi-tier vRAN Architecture

- lowPHY using ASIC & FPGA
- General purpose compute
- VNF consolidation



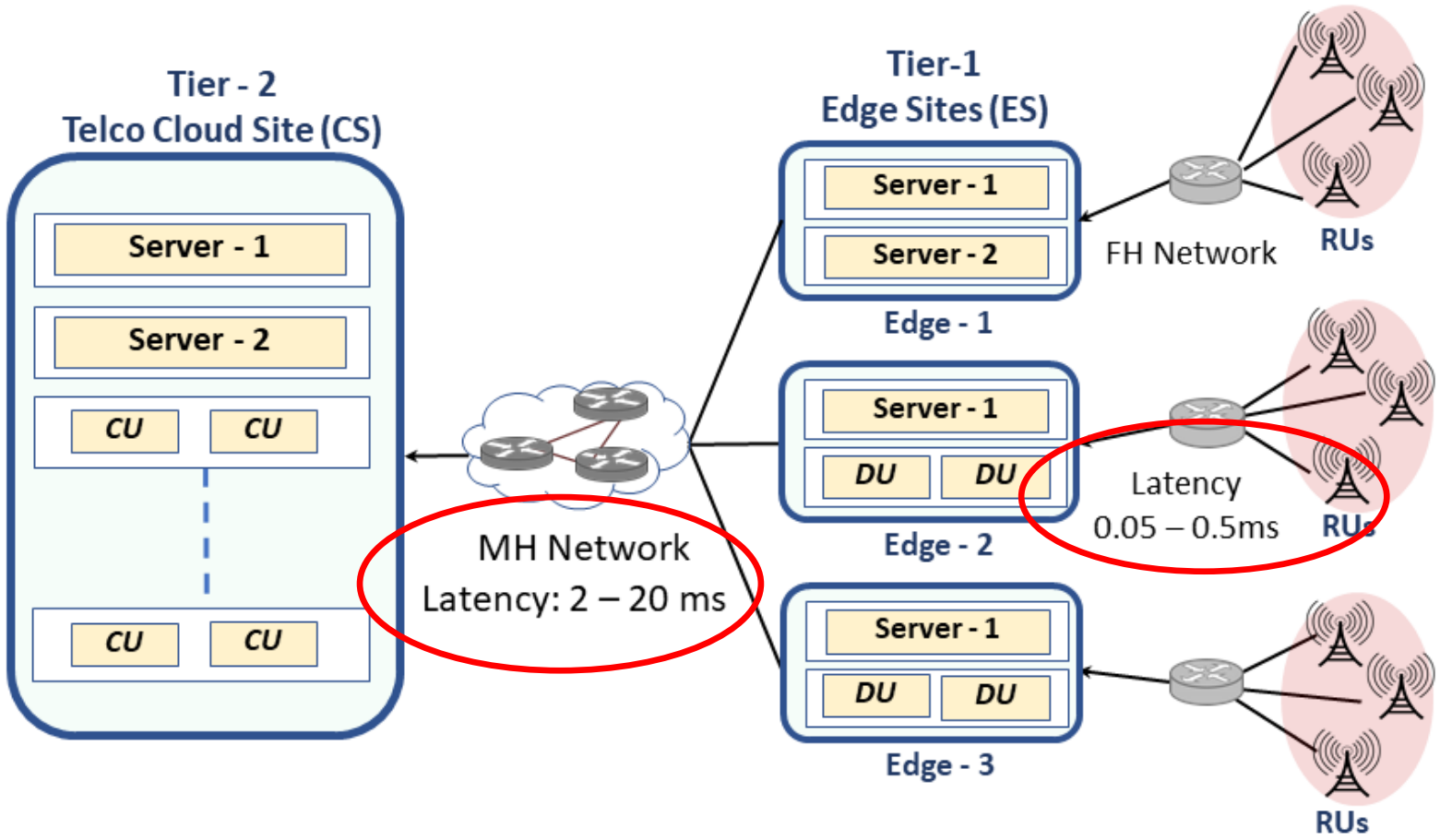
Multi-tier vRAN Architecture

- lowPHY using ASIC & FPGA
- General purpose compute
- VNF consolidation
- More energy efficient



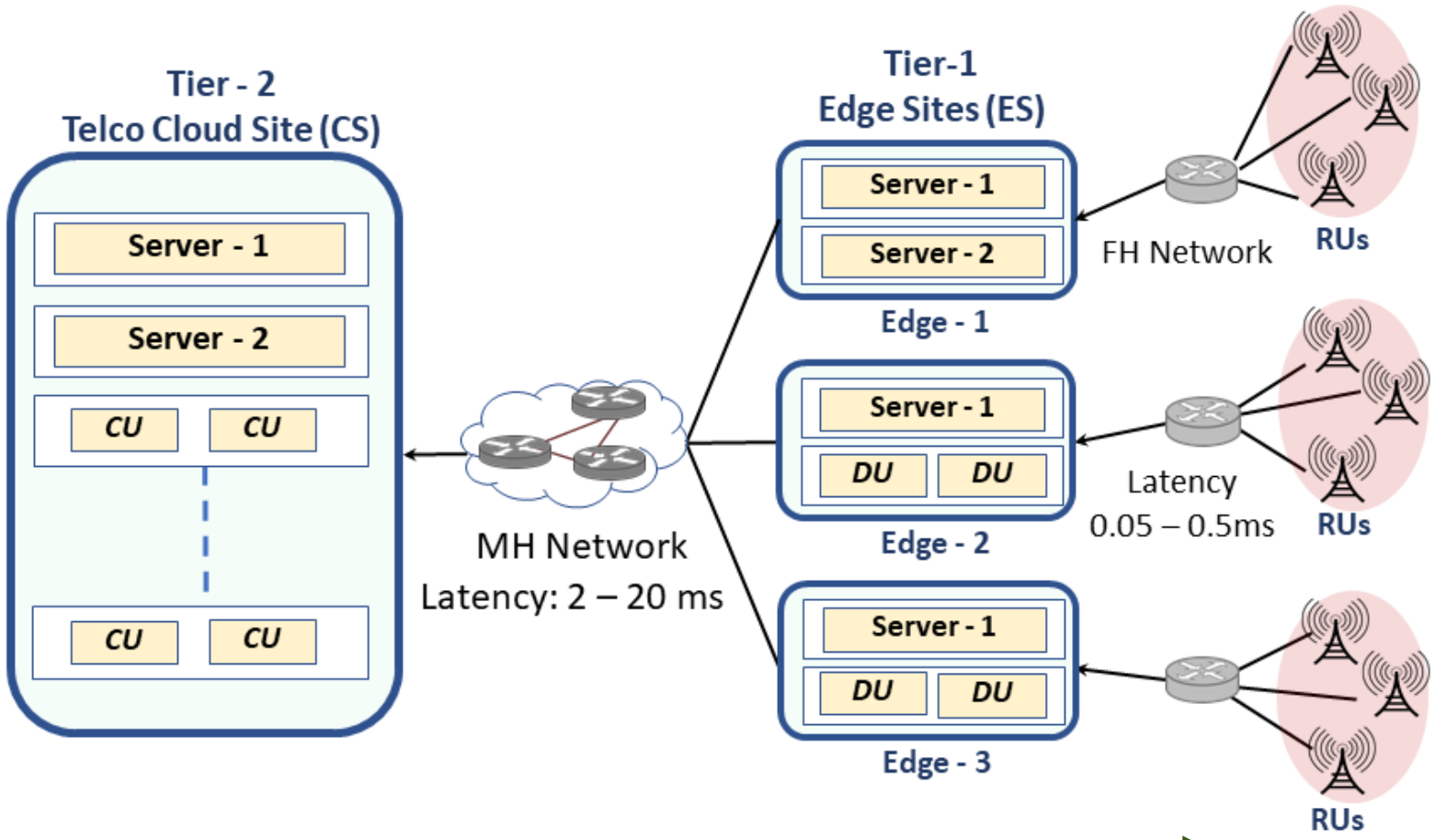
Multi-tier vRAN Architecture

- lowPHY using ASIC & FPGA
- General purpose compute
- VNF consolidation
- More energy efficient
- However, latency issues
- MH link capacity issues

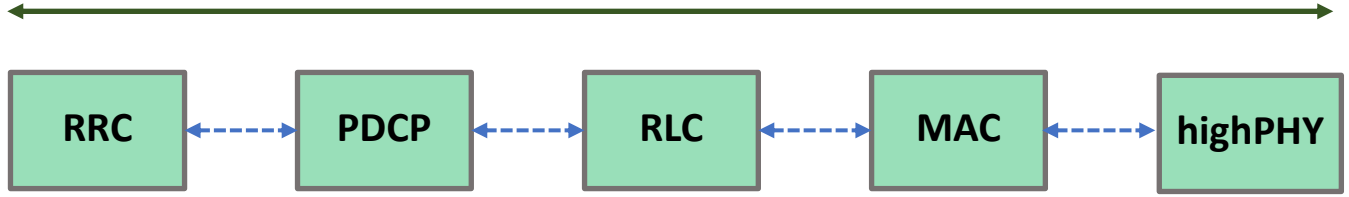


Multi-tier vRAN Architecture

- lowPHY using ASIC & FPGA
- General purpose compute
- VNF consolidation
- More energy efficient
- However, latency issues
- MH link capacity issues

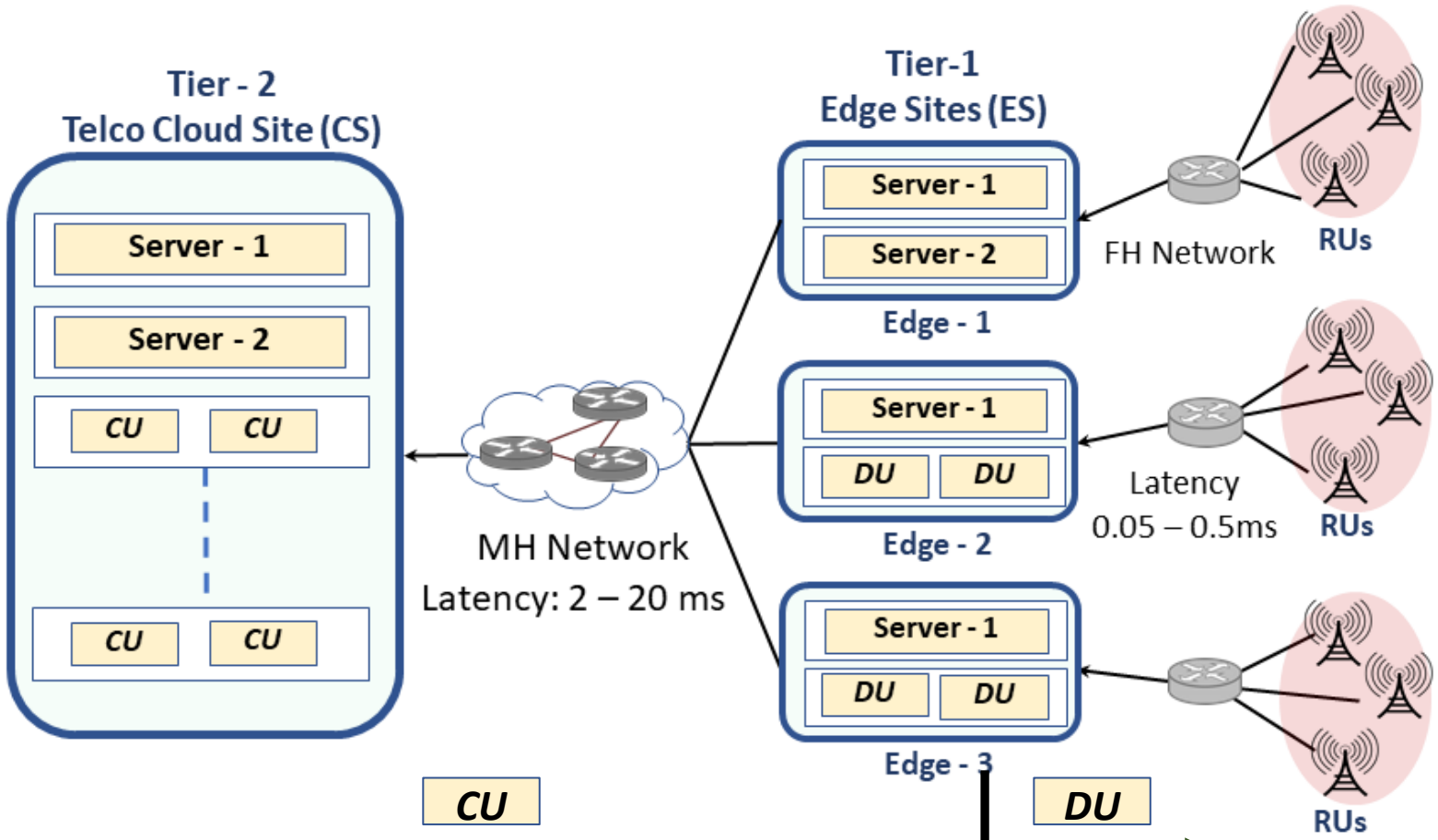


RAN Functional Splits

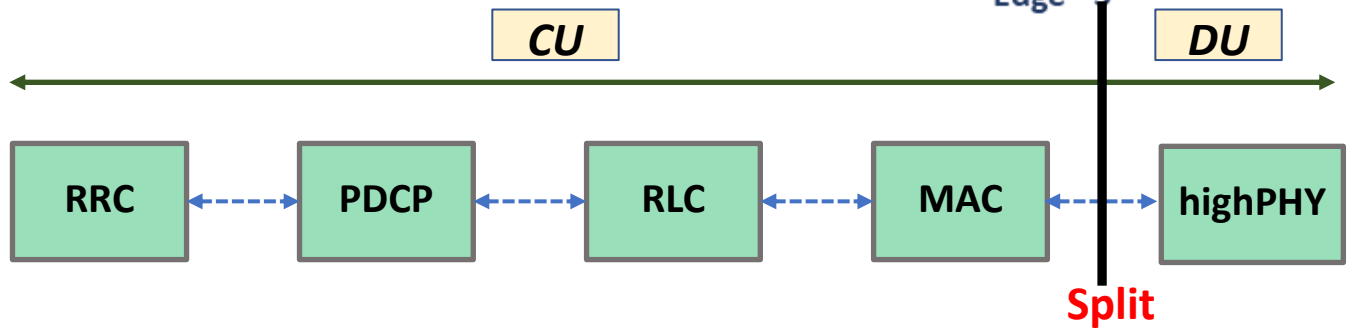


Multi-tier vRAN Architecture

- lowPHY using ASIC & FPGA
- General purpose compute
- VNF consolidation
- More energy efficient
- However, latency issues
- MH link capacity issues

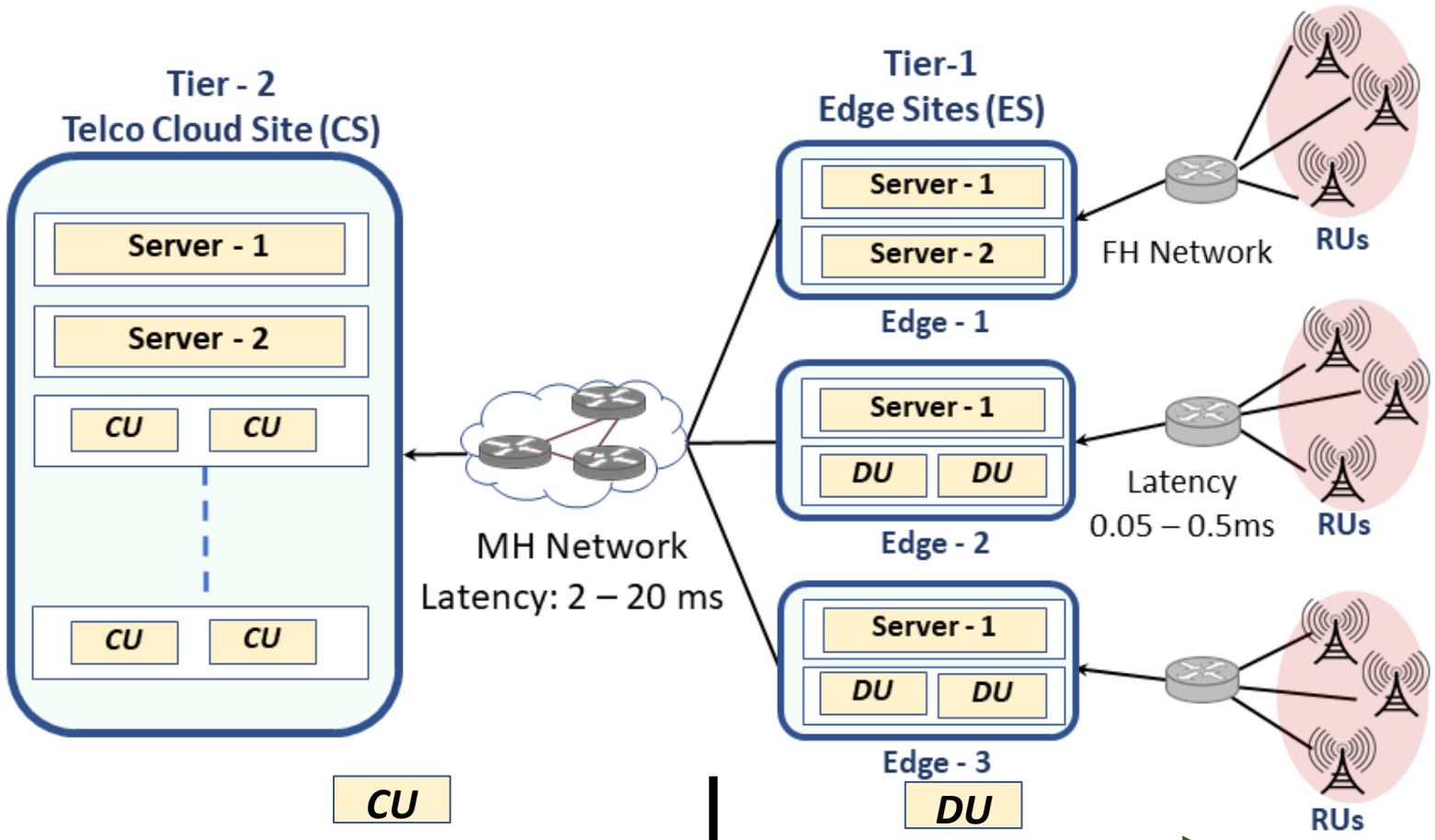


RAN Functional Splits

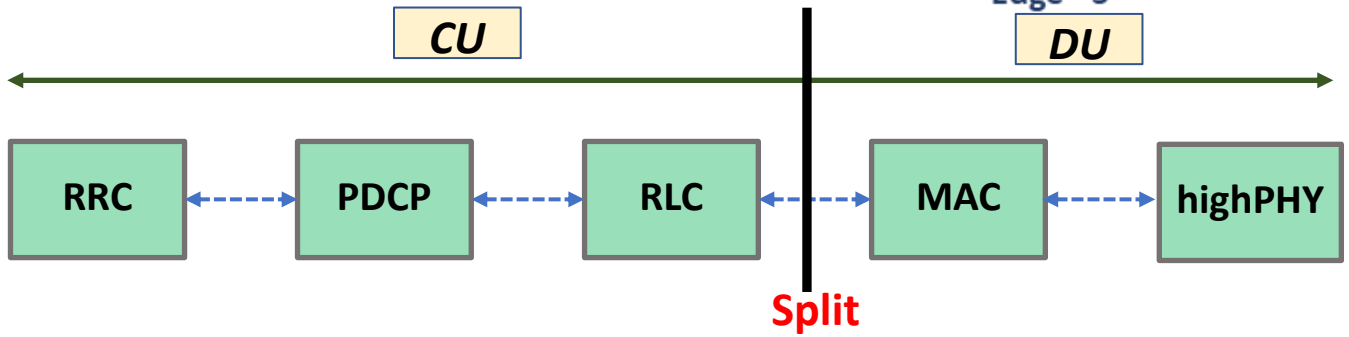


Multi-tier vRAN Architecture

- lowPHY using ASIC & FPGA
- General purpose compute
- VNF consolidation
- More energy efficient
- However, latency issues
- MH link capacity issues

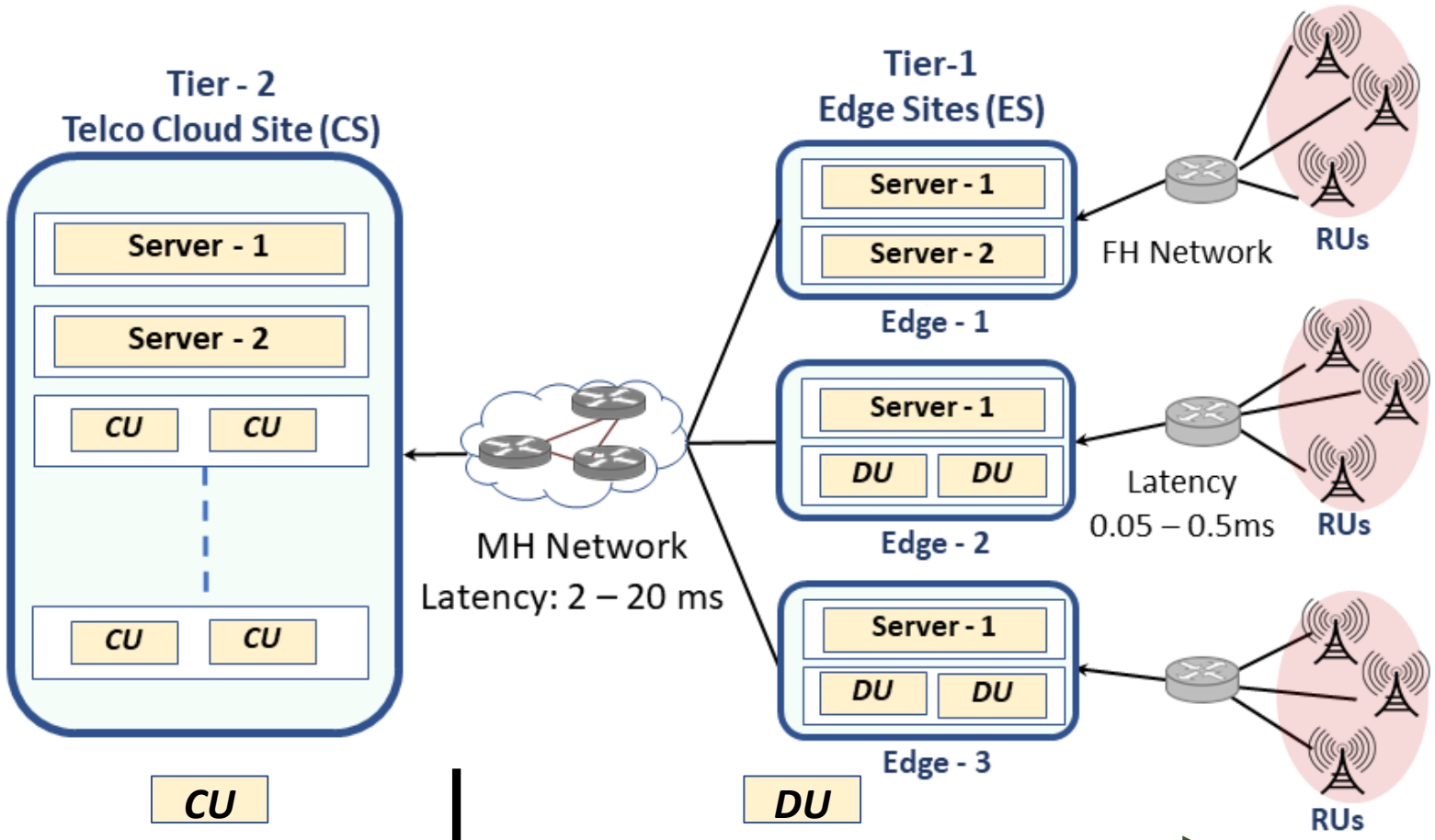


RAN Functional Splits

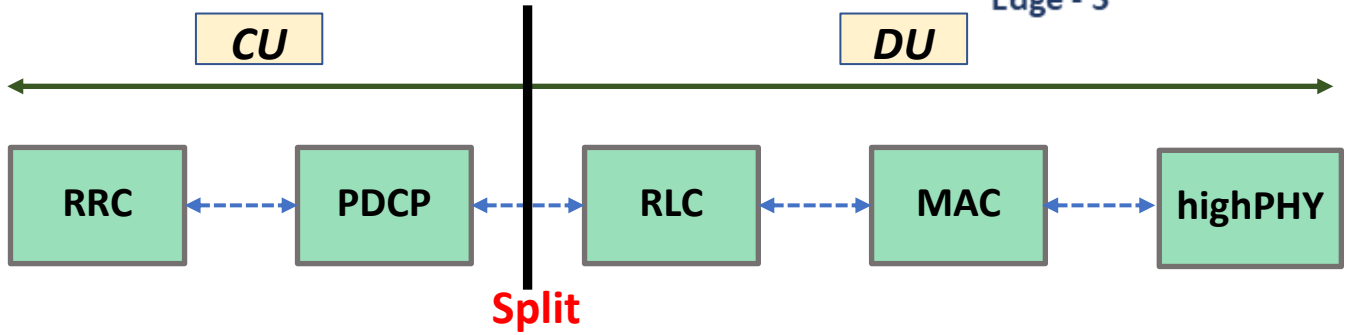


Multi-tier vRAN Architecture

- lowPHY using ASIC & FPGA
- General purpose compute
- VNF consolidation
- More energy efficient
- However, latency issues
- MH link capacity issues

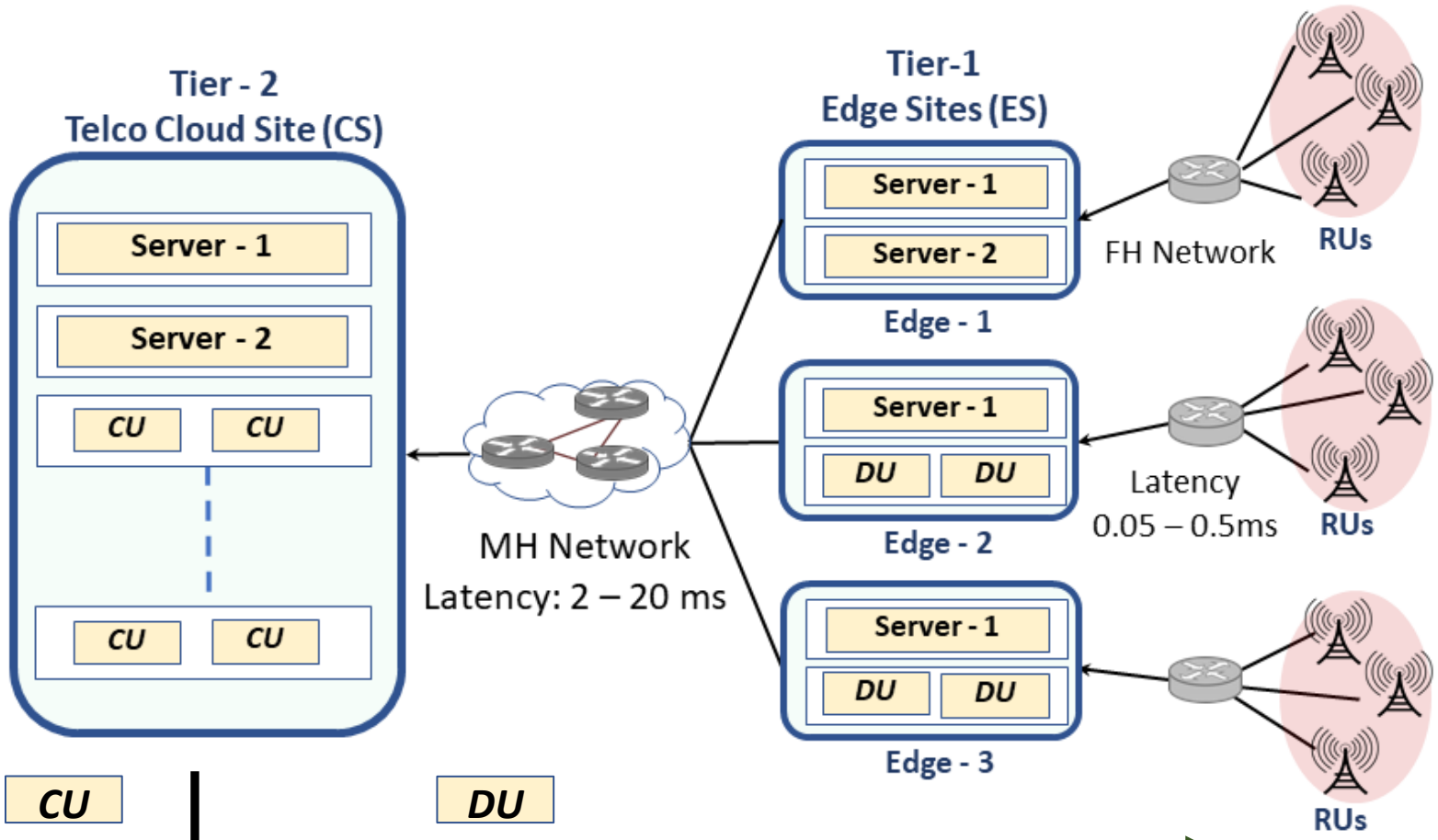


RAN Functional Splits

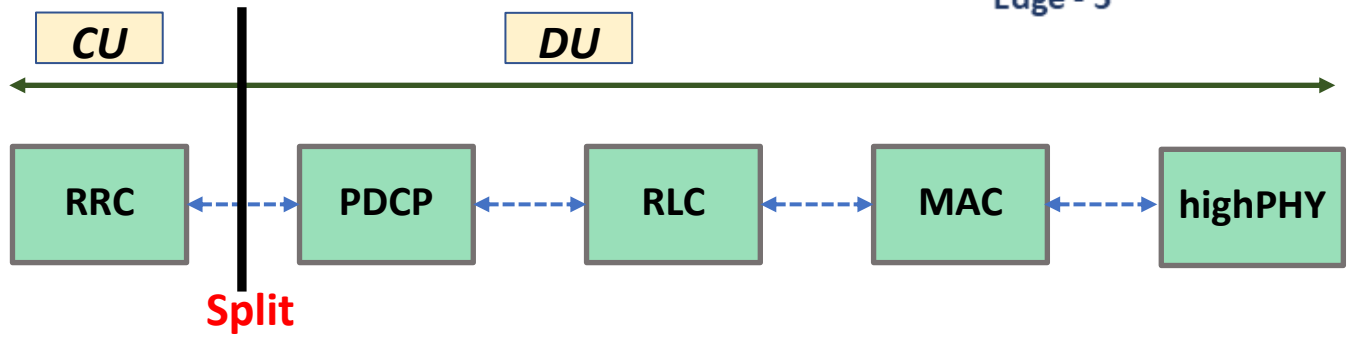


Multi-tier vRAN Architecture

- lowPHY using ASIC & FPGA
- General purpose compute
- VNF consolidation
- More energy efficient
- However, latency issues
- MH link capacity issues



RAN Functional Splits



Our Contributions

- Realistic multi-tier cloudified RAN model
- **Comprehensive energy model and problem formulation**
- A distributed and scalable solution: GreenRAN
- Evaluation with metro-scale real world dataset

Energy Model

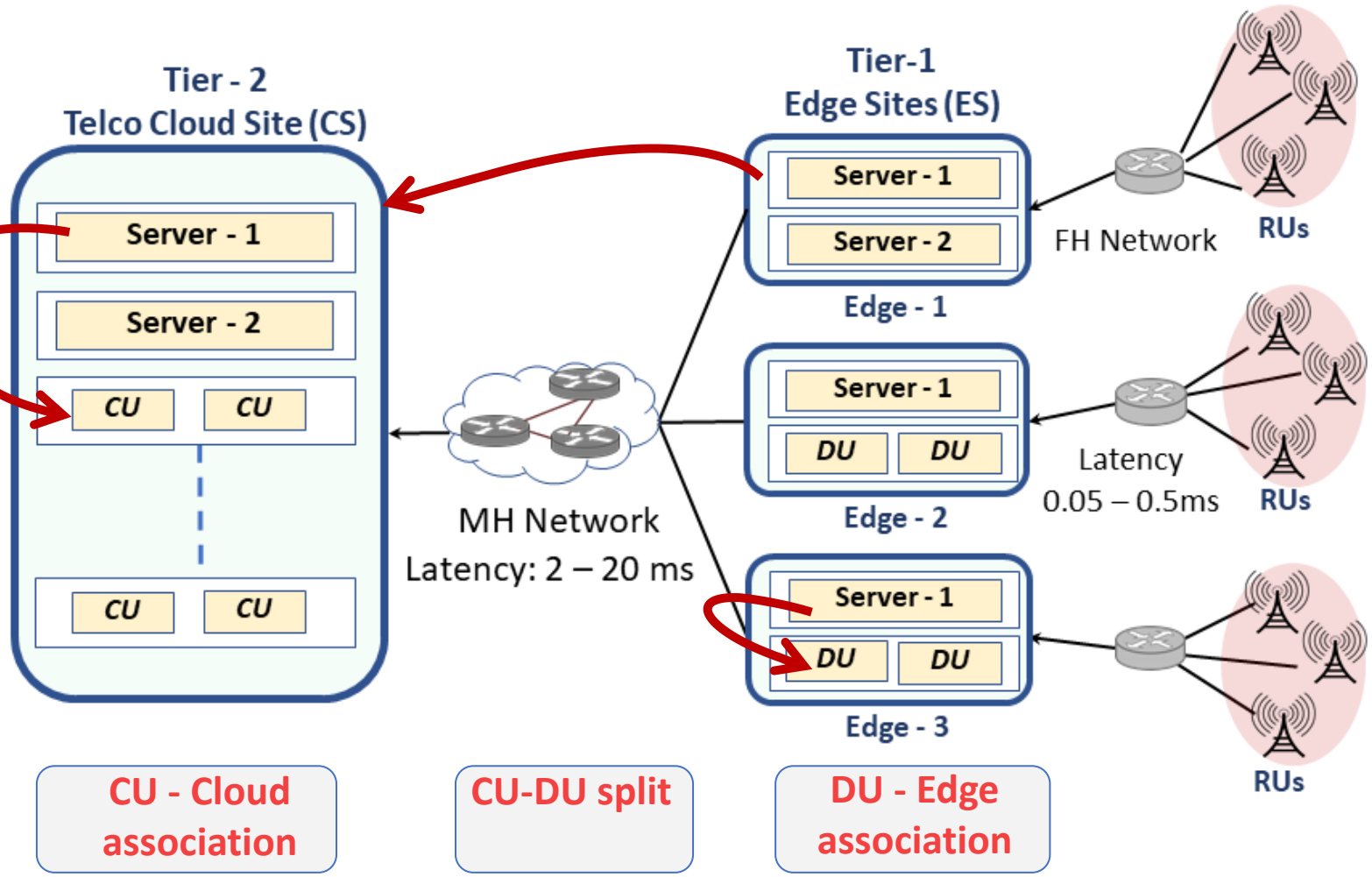
Processing cost (E^P)

- To process VNFs
- Coupled bin-packing

Migration cost (E^M)

- To consolidate VNFs
- Consider previous epoch

Decision variables



Challenges

NP hard

- Integer Quadratic Program

Scale

- Huge parameter space

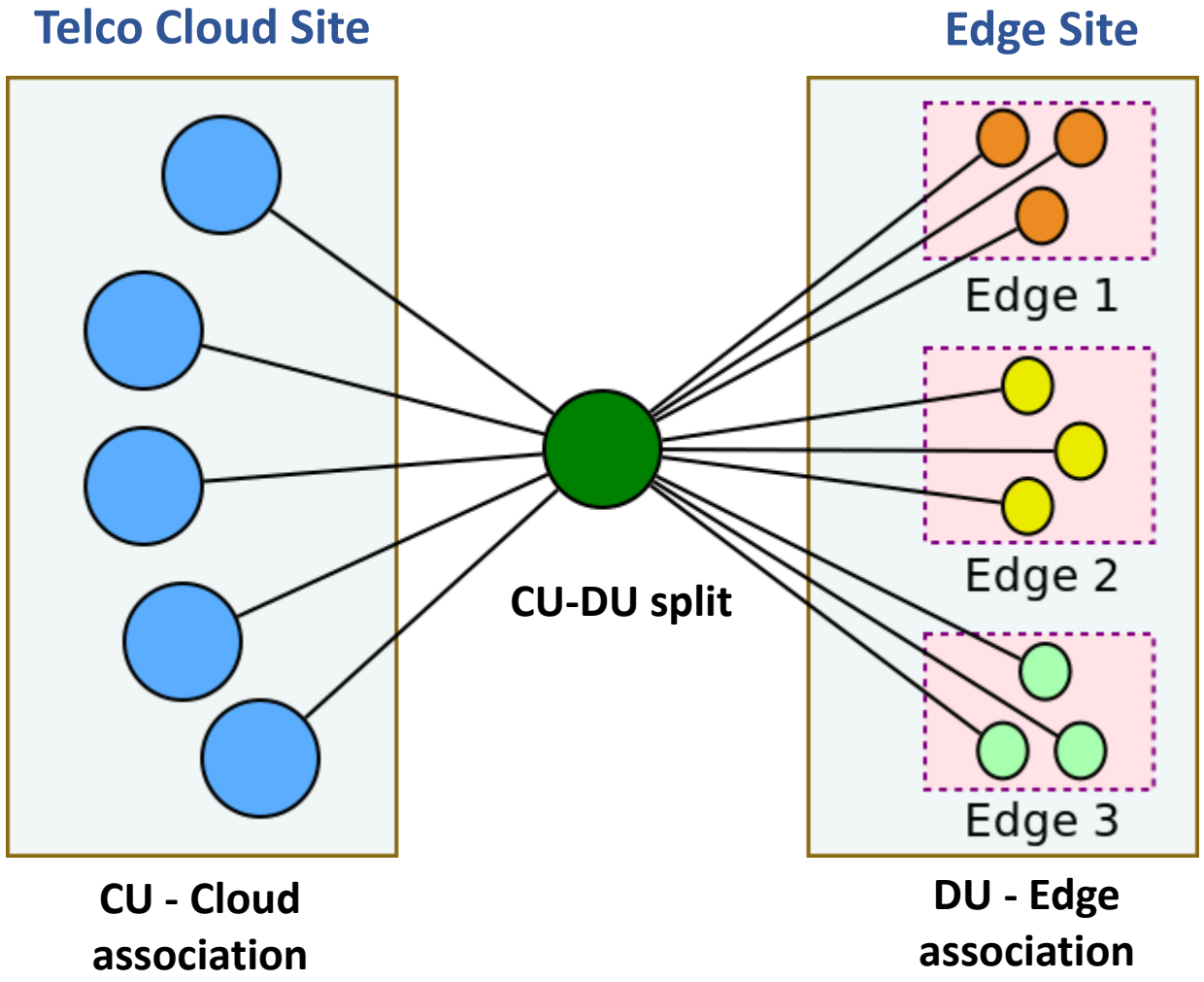
Centralized

- Edge and Cloud variables are inter-dependent

Our Contributions

- Realistic multi-tier cloudified RAN model
- Comprehensive energy model and problem formulation
- **A distributed and scalable solution: GreenRAN**
- Evaluation with metro-scale real world dataset

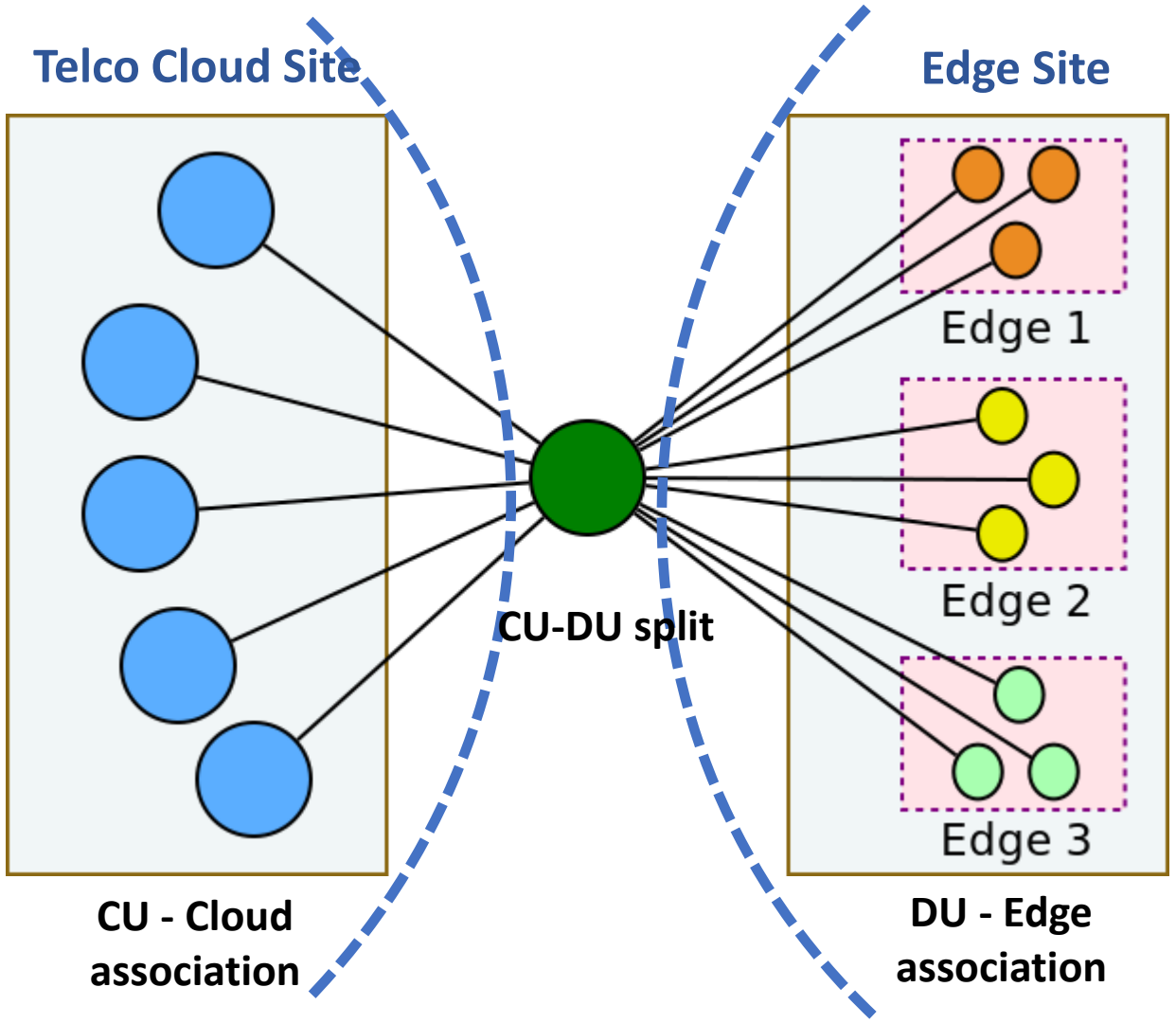
Towards a decentralized solution: GreenRAN



Towards a decentralized solution: GreenRAN

Divide & Conquer Paradigm

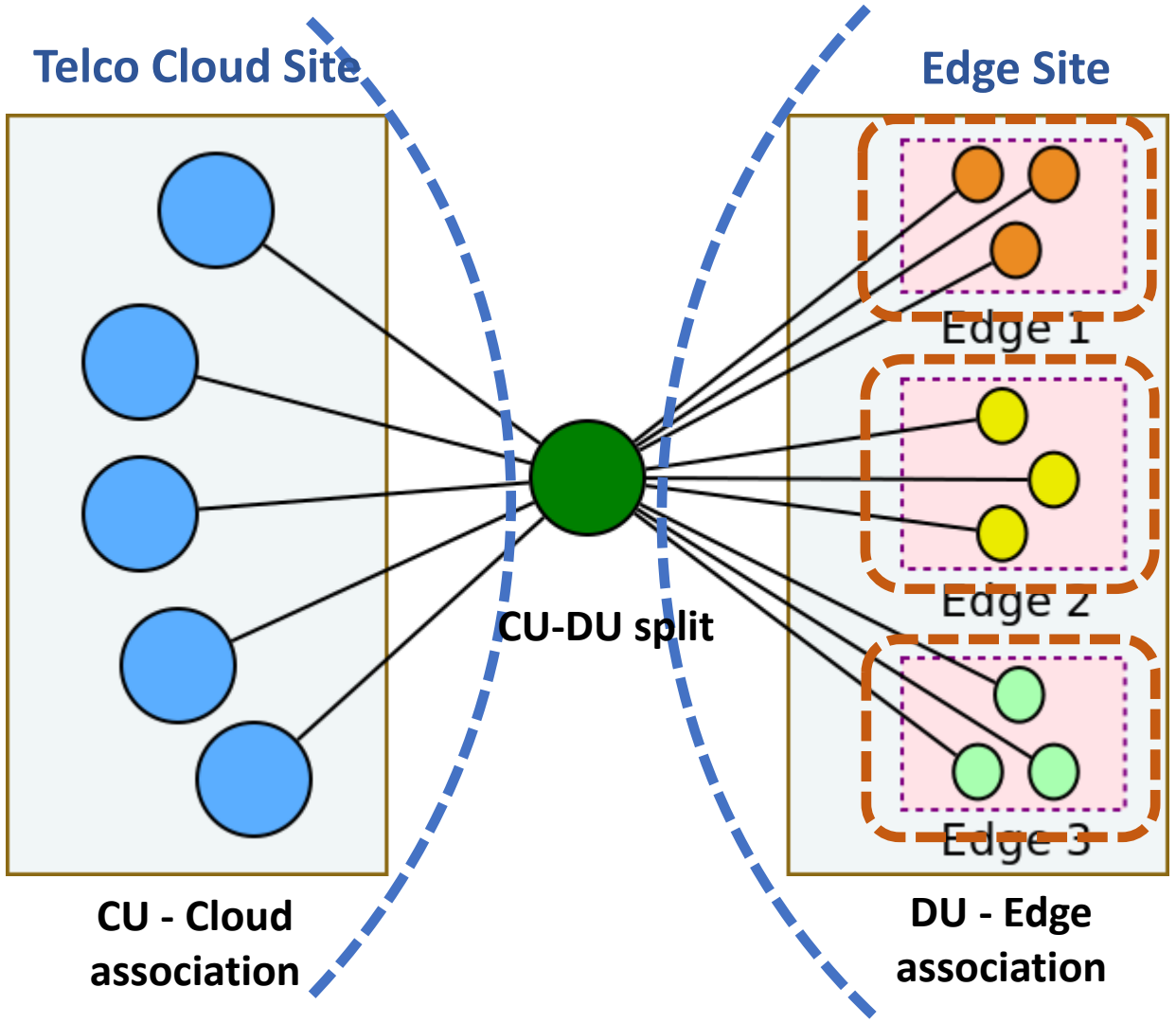
1. Lagrangean Relaxation



Towards a decentralized solution: GreenRAN

Divide & Conquer Paradigm

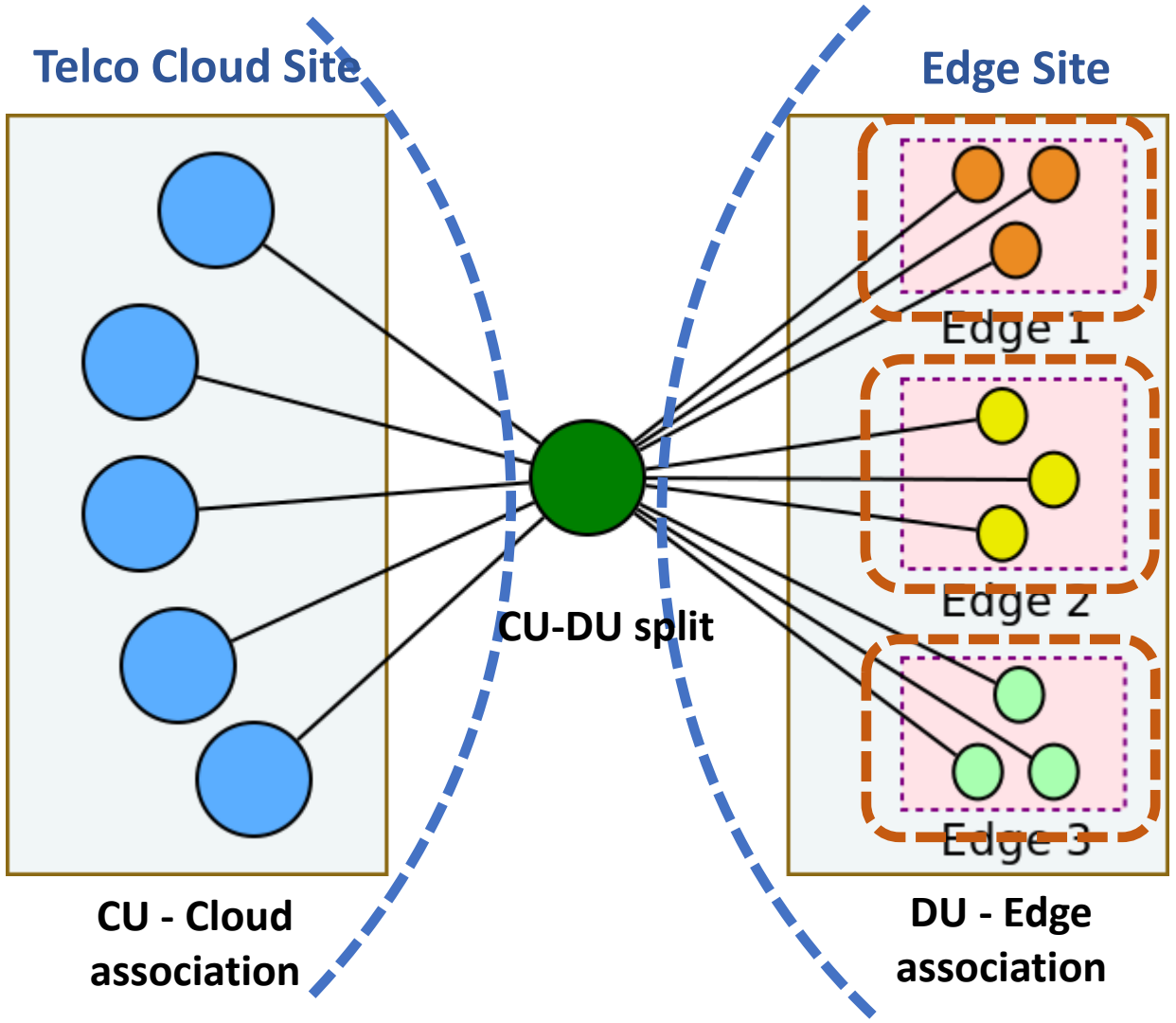
1. Lagrangean Relaxation
2. Per edge formulation



Towards a decentralized solution: GreenRAN

Divide & Conquer Paradigm

1. Lagrangean Relaxation
2. Per edge formulation
3. Simulated Annealing



Our Contributions

- Realistic multi-tier cloudified RAN model
- Comprehensive energy model and problem formulation
- A distributed and scalable solution: GreenRAN
- **Evaluation with metro-scale real world dataset**

Experimental Setup

vRAN Configuration

- Metro-scale **real world traffic** dataset
- **450** RUs
- **18** Far Edge Clouds
  - **15** physical servers per Edge Cloud
- **1** Telco Cloud
  - **30** servers per Telco Cloud

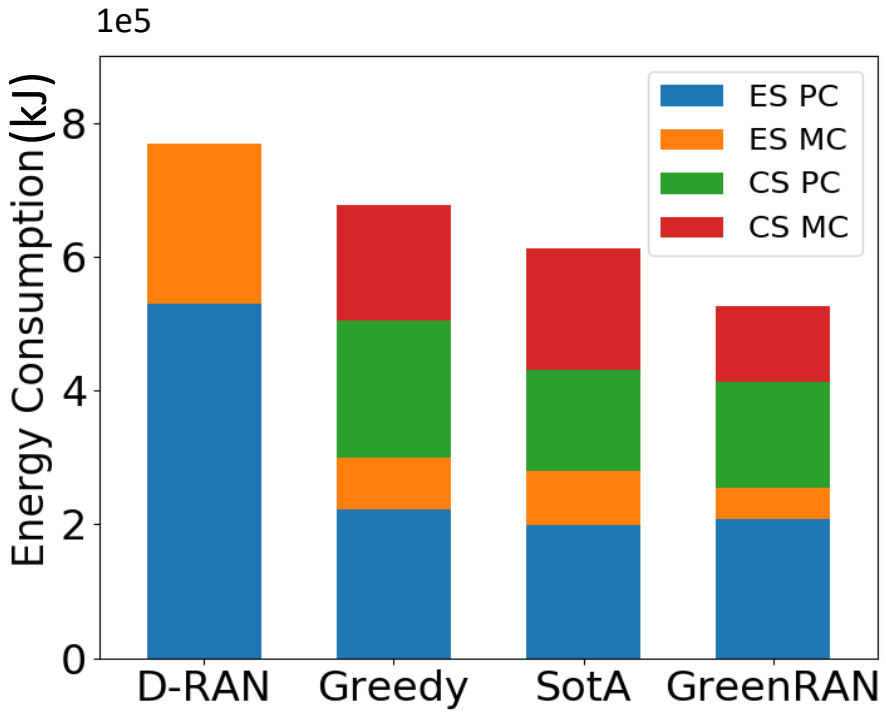
Benchmarks

- **D-RAN**
  - All processing in DUs at Far edge site
- **Greedy**
  - Maximum processing at CU in Telco cloud
- **SotA**
  - Only considers the processing cost

SotA: H. Gupta et al., “Apt-RAN: A Flexible Split-Based 5G RAN to Minimize Energy Consumption and Handovers,” TNSM, 2020

Energy Consumption

- GreenRAN
  - 25% better than SotA
  - 42% better than traditional D-RAN
- D-RAN has fewer multiplexing opportunities
- Greedy offloading to cloud is not optimal
- SotA incurs lesser processing cost
  - But higher migration cost



Conclusion

- Optimizing energy consumption is more challenging with advanced architecture
- We presented a latest vRAN model, realistic energy model considering migration costs
- Our framework, GreenRAN, is a distributed, scalable and highly efficient solution
- Results show 25% improvement over SotA and 42% over D-RAN

Thank You