DISTRIBUTED NETWORK SLICING

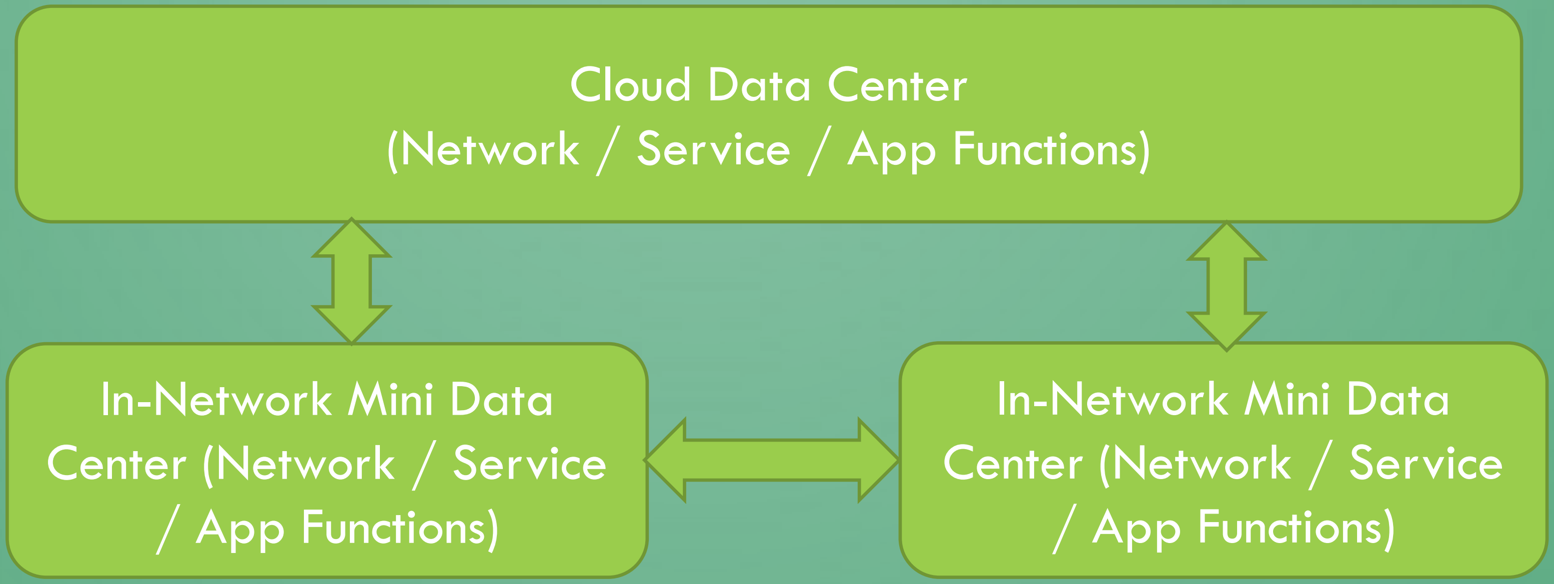
DILIP KRISHNASWAMY (DILIP@IEEE.ORG)

RELIANCE JIO INFOCOMM

AUGUST 11, 2018

DISTRIBUTED FUNCTIONS VIRTUALIZATION (DFV)

With the availability of compute & storage in-networks, explore the placement of network / service / application functions across DCs

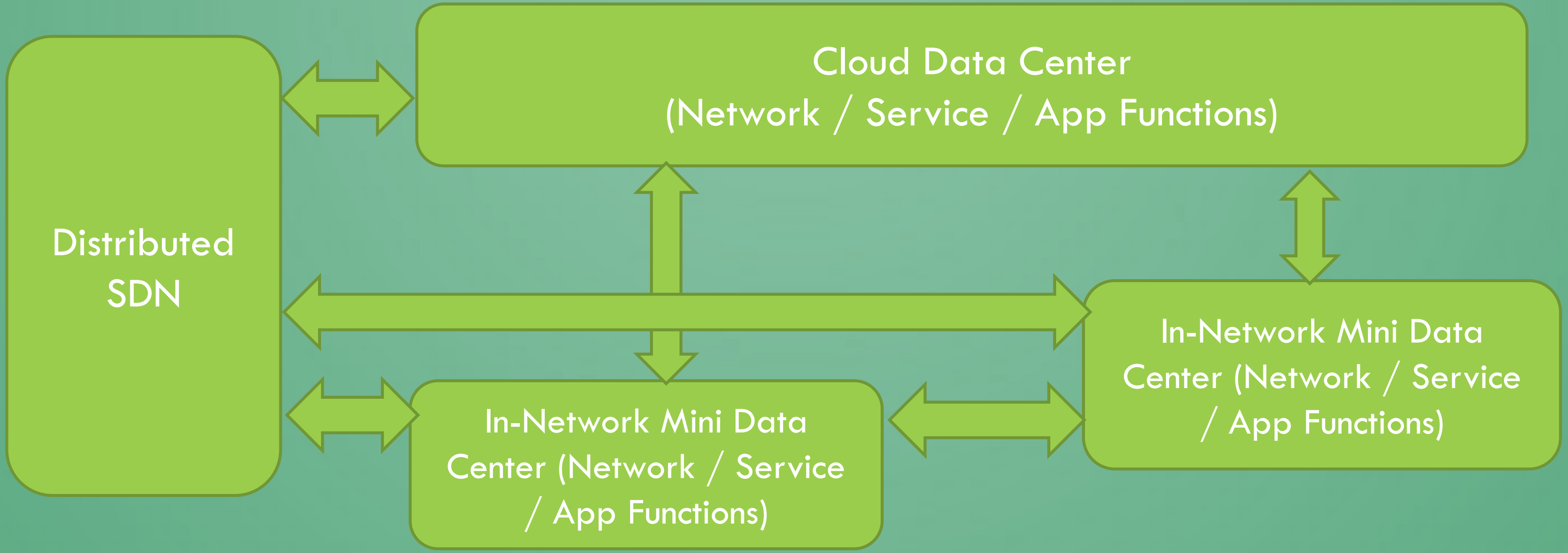


→ How can we best place VNFs in hierarchical data centers taking care of latency constraints associated with VNFs, user mobility, energy cost of utilization, resource availability? How can we leverage distributed predictive machine learning / intelligence for dynamic distributed resource management?

→ How can we best utilize the dynamic availability of such distributed compute/network/storage/energy resources?

DISTRIBUTED SDN FOR DISTRIBUTED NFV

Distributed SDN and NFV will enable smart distributed processing of functions across data centers



→ *Partition / Collapse / Replicate functions across data centers*

→ *Address latency constraints, user mobility, dynamic resource availability, security (compute/network/storage/energy)*

→ *Dynamic Monitoring, Analytics, Optimization, Orchestration, Scaling*

INTER-DC VIRTUAL FUNCTION PLACEMENT & RESOURCE MGMT

Map users to VNFs on DCs based on latency, availability, DC load, energy, mobility

Dynamically direct new user flows to utilize VNFs at the most appropriate DCs

Dynamically divert new VNF resource requirements away from a DC if heavily loaded

Example Depiction – Serving users in Santa Clara and Sunnyvale

Phoenix DC (Solar Powered)

Low Energy cost, Higher Latency

San Francisco DC

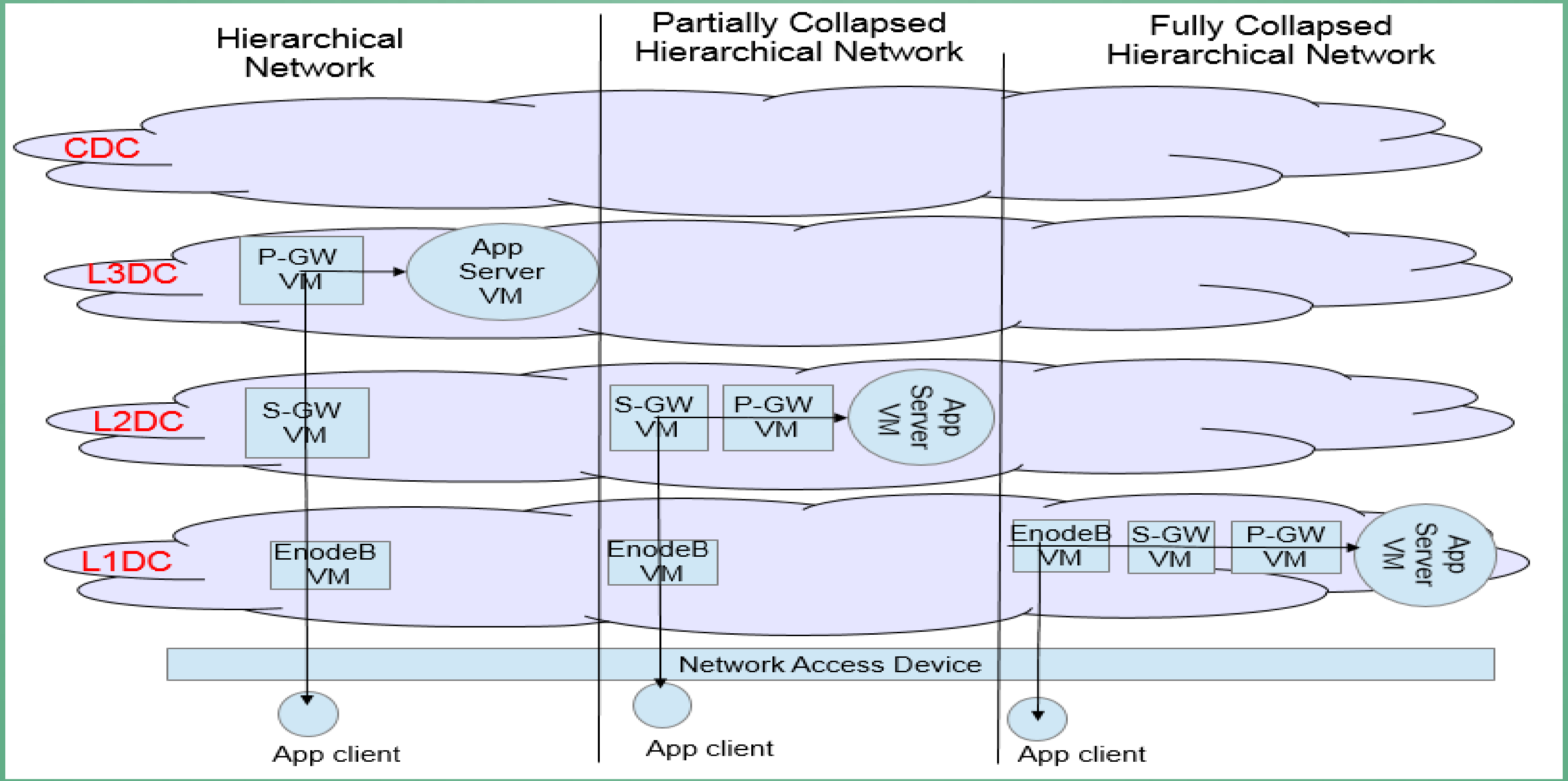
High Energy cost, Medium Latency,
Higher Capacity compared to Small DC

Santa Clara
Small DC

Sunnyvale
Small DC

Lower Latency, Lower Capacity
High Energy Cost

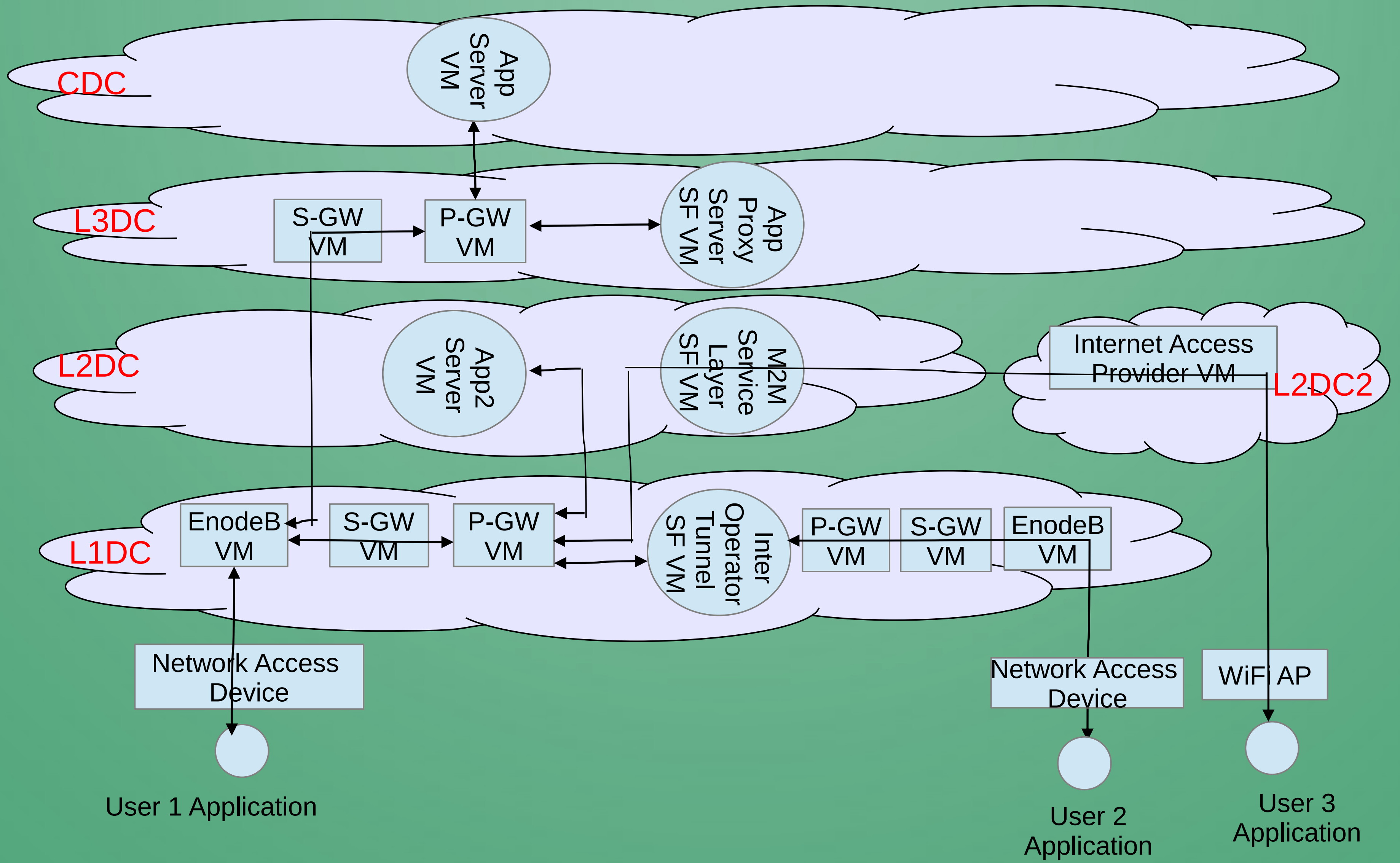
HIERARCHICAL COLLAPSED FUNCTIONS (IEEE NFV-SDN'15)



Metric	Fully Hierarchical	Partially Collapsed	Fully Collapsed
Round trip time	179 ms (median)	64 ms (median)	22 ms (median)
Connection setup	3.7 sec (median)	1.3 sec (median)	0.7 sec (median)
TCP bandwidth	3.19 Mbps	3.45 Mbps	3.72 Mbps
Playout Stalls	12	2	0

GENERALIZED DFV (IEEE NFV-SDN 2015)

NETWORK, SERVICE, AND APPLICATION FUNCTION VM PARTITIONING



MOBILITY AWARE VNF-PLACEMENT (IEEE NFV-SDN 2017)

(COLLABORATION WITH AKANKSHA PATEL, PROF MYTHILI VUTUKURU)

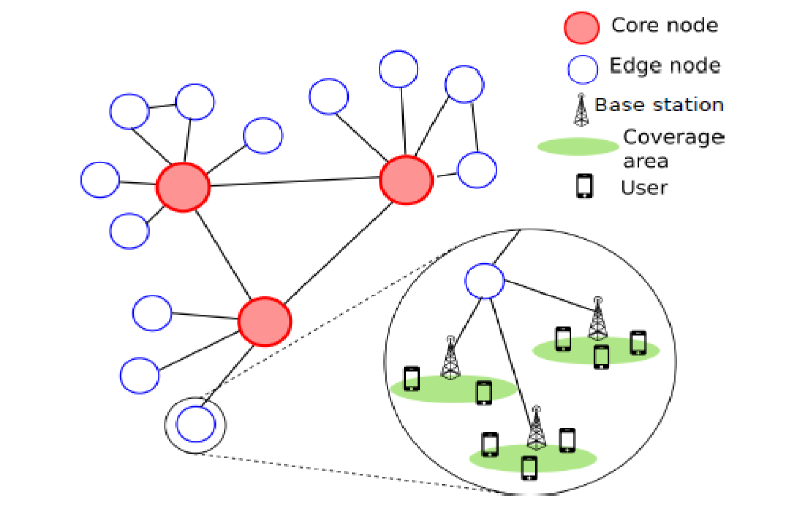


Fig. 1. Physical substrate network of a telecom operator

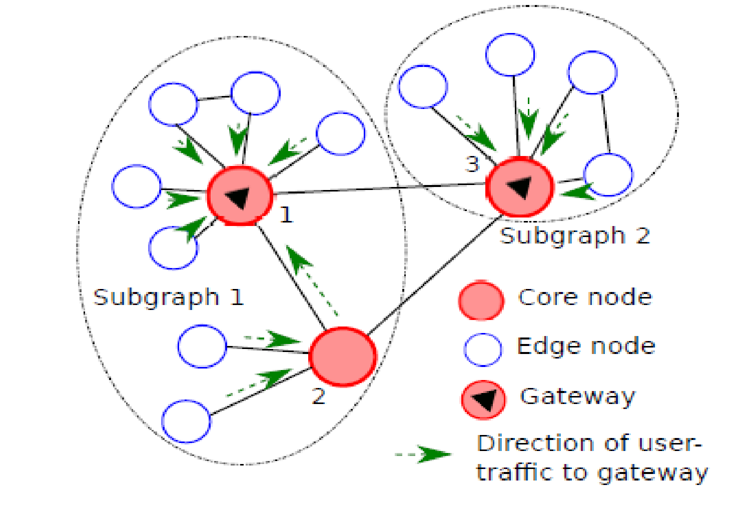


Fig. 4. Division of graph into subgraphs

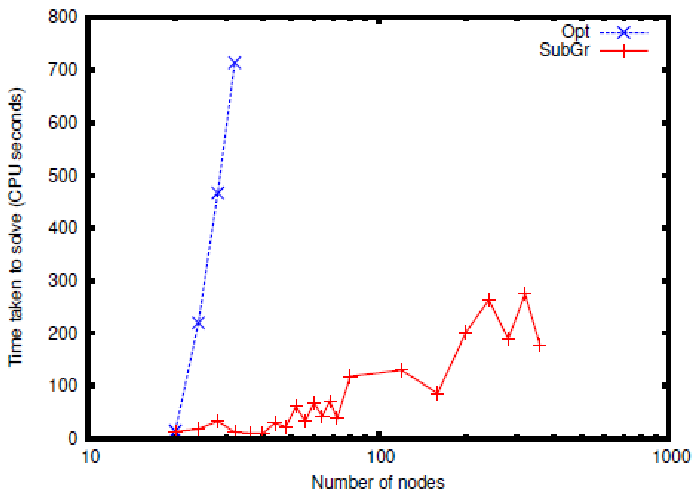


Fig. 8. Comparison of time taken to obtain the optimal placement by running model on complete graph vs subgraphs

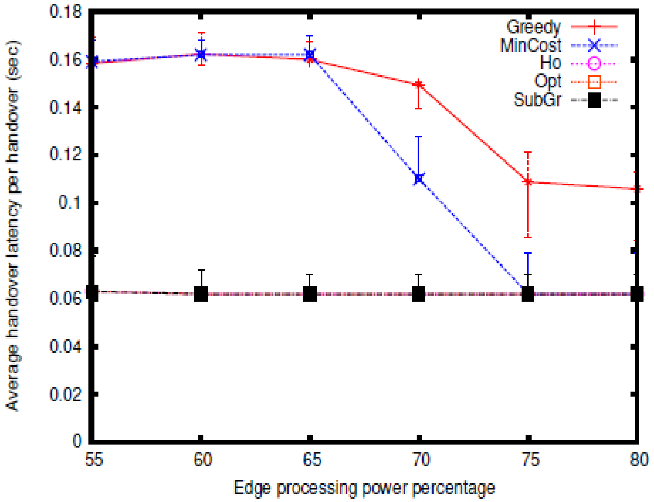
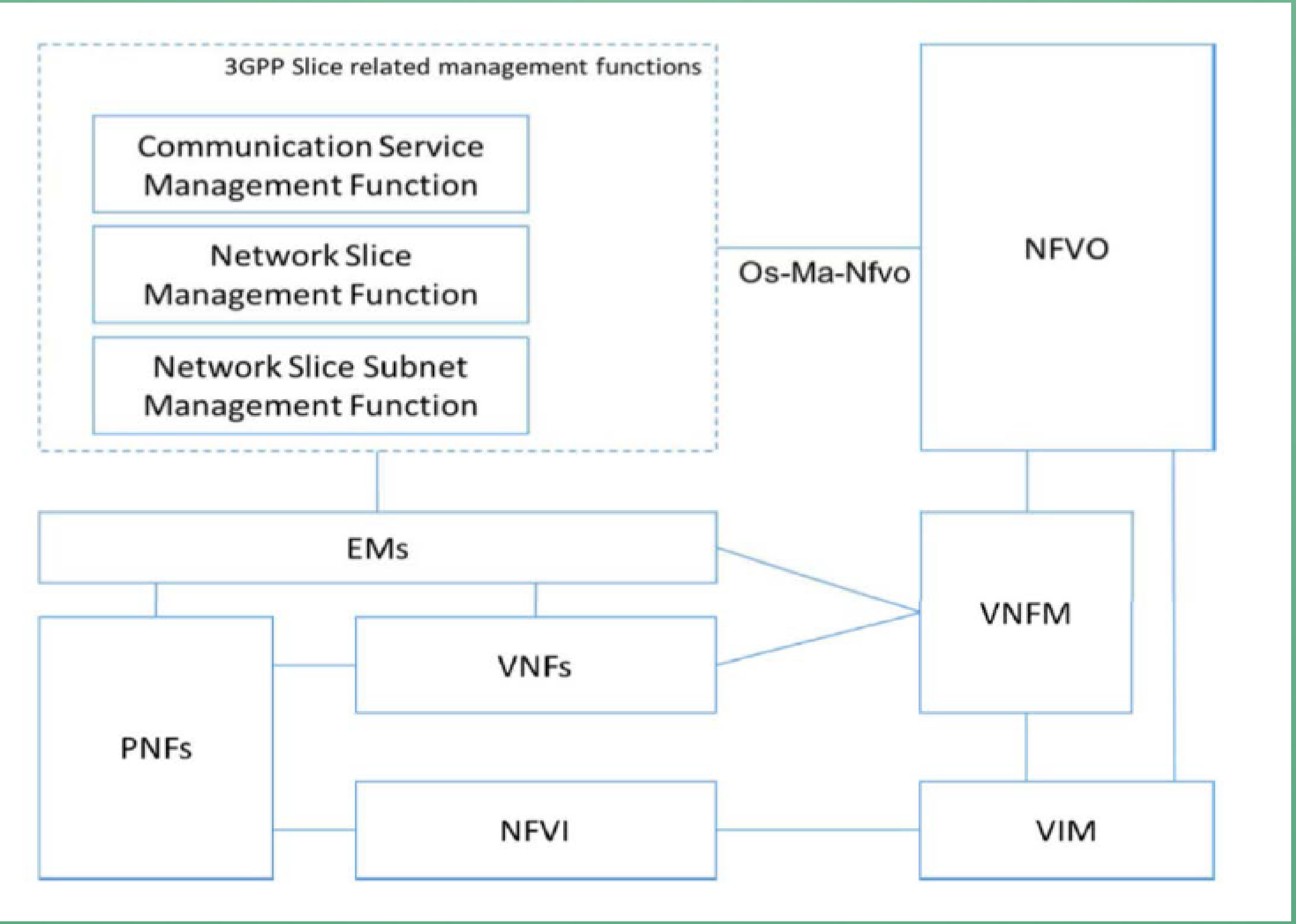


Fig. 6. Comparison of the average handover latency for different approaches (min, median, and max values have been plotted)

ETSI NFV + SLICING ARCHITECTURE



NETWORK SLICING – KEY ASPECTS

- Scalable, flexible, programmable, on-demand resource allocation for flows in 5G networks
- Slice allocation for different use-cases / applications / services
- Can involve both physical and virtual network functions
- Can combine RAN and Core Network Slices
  - 1-1 or one-to-Many
  - Multiple RAN slices may connect to a single Core Network Slice
  - One RAN slice can connect to multiple Core Network Slices
- Slice paths can be stitched dynamically based on requirements
- Slice allocation can be done for multiple user plane flows with admission control

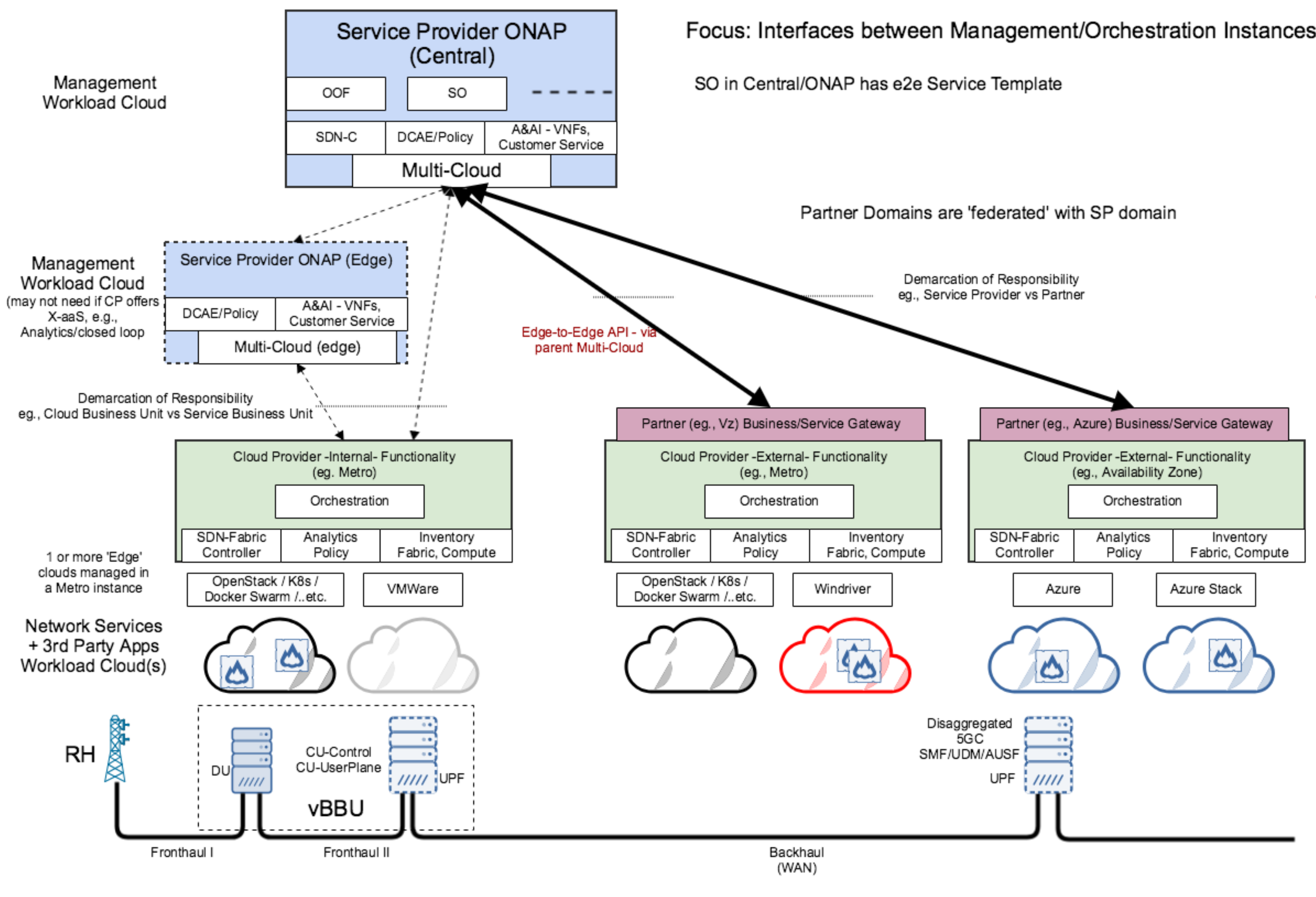
NETWORK SLICING – KEY ASPECTS

- Intra-Slice separation for greater reliability using multiple physical servers
- Latency, Bandwidth, Memory, Compute, Energy, Redundancy, Service Availability are all critical parameters in selecting resources to allocate for a network slice
    - Delay Tolerant flows (typical mMTC), Delay Sensitive Flows, Ultra-delay-sensitive flows (uRLLC), low bandwidth flows, high bandwidth flows, very high bandwidth flows (eMBB), …..
- While this may seem like the network is going to go crazy, this is not necessarily true – most flows have well defined needs and can be mapped to an appropriate slice
    - Slices can be predictively allocated, and scaled up or down dynamically
    - New Crazy slices can be created however for those new crazy apps that we can think of

DISTRIBUTED NETWORK SLICING

- Create Network Slices with slice resources spanning distributed data centers
- MANO across data centers with the ability to stitch resource chains to create network slices
- Long term slices can be created for standard flows
- Dynamic short term slices can be created on demand
- System needs to be aware of dynamic resource availability / constraints across these data centers
- Distributed optimization required to optimally use these distributed slice resources

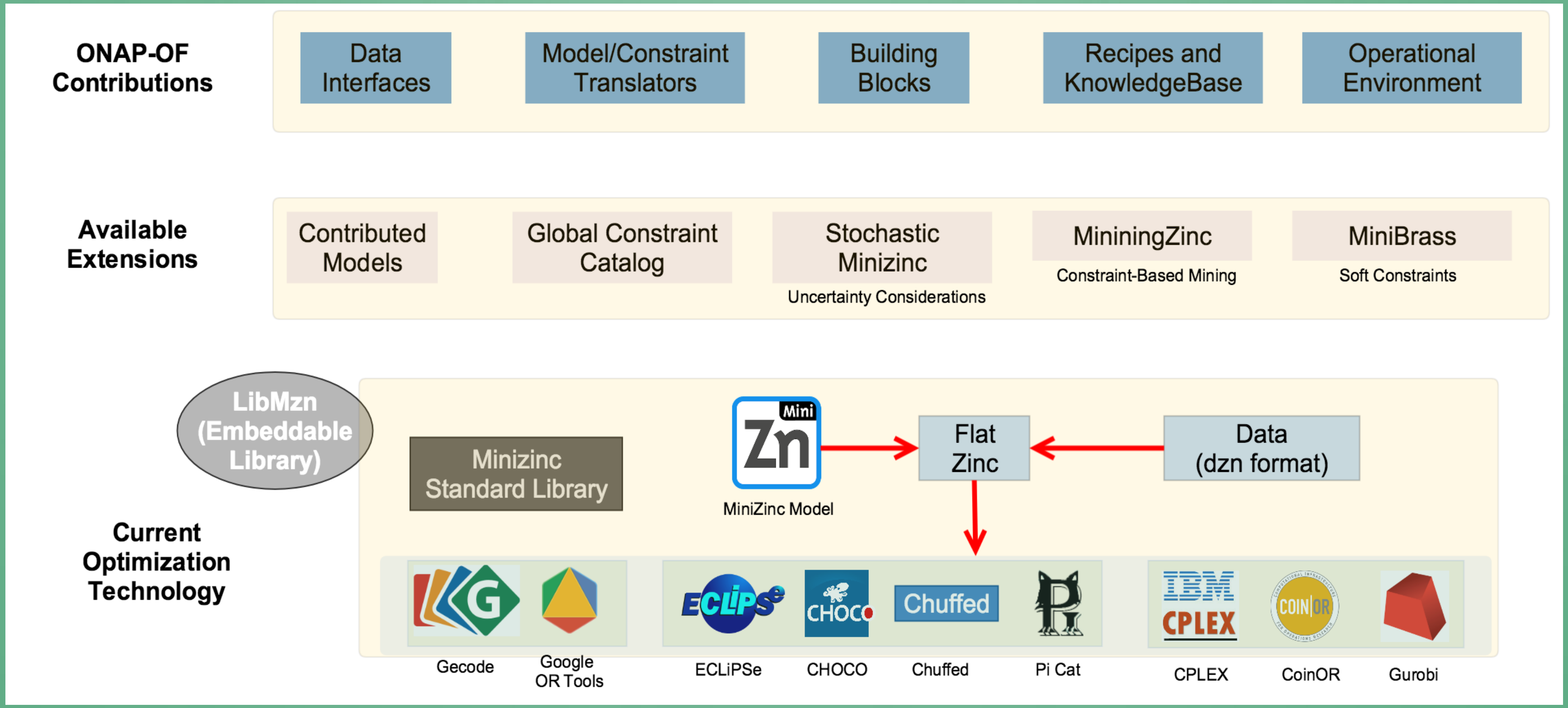
END-TO-END REFERENCE ARCHITECTURE – ONAP PERSPECTIVE



*** This diagram is discussion in progress and not final ***

Discussion in Progress: Edge Automation Through ONAP WG (https://wiki.onap.org/display/DW/Edge+Automation+through+ONAP)

MODEL-DRIVEN OPTIMIZATION FRAMEWORK



- **Designer & Developer friendly** Domain-Specific Modelling Language for **Service Placement/Scheduling Policy**
- Address **Central/Edge** Resource Management/Optimization Requirements
- **Masks** the **Mathematical complexity** of optimization algorithms through **Modelling**
- **Flexibility** to add **Custom optimizers** especially for Edge Resource Management/Optimization
- Drive **Service Creation Agility** for 5G, Edge Computing etc.

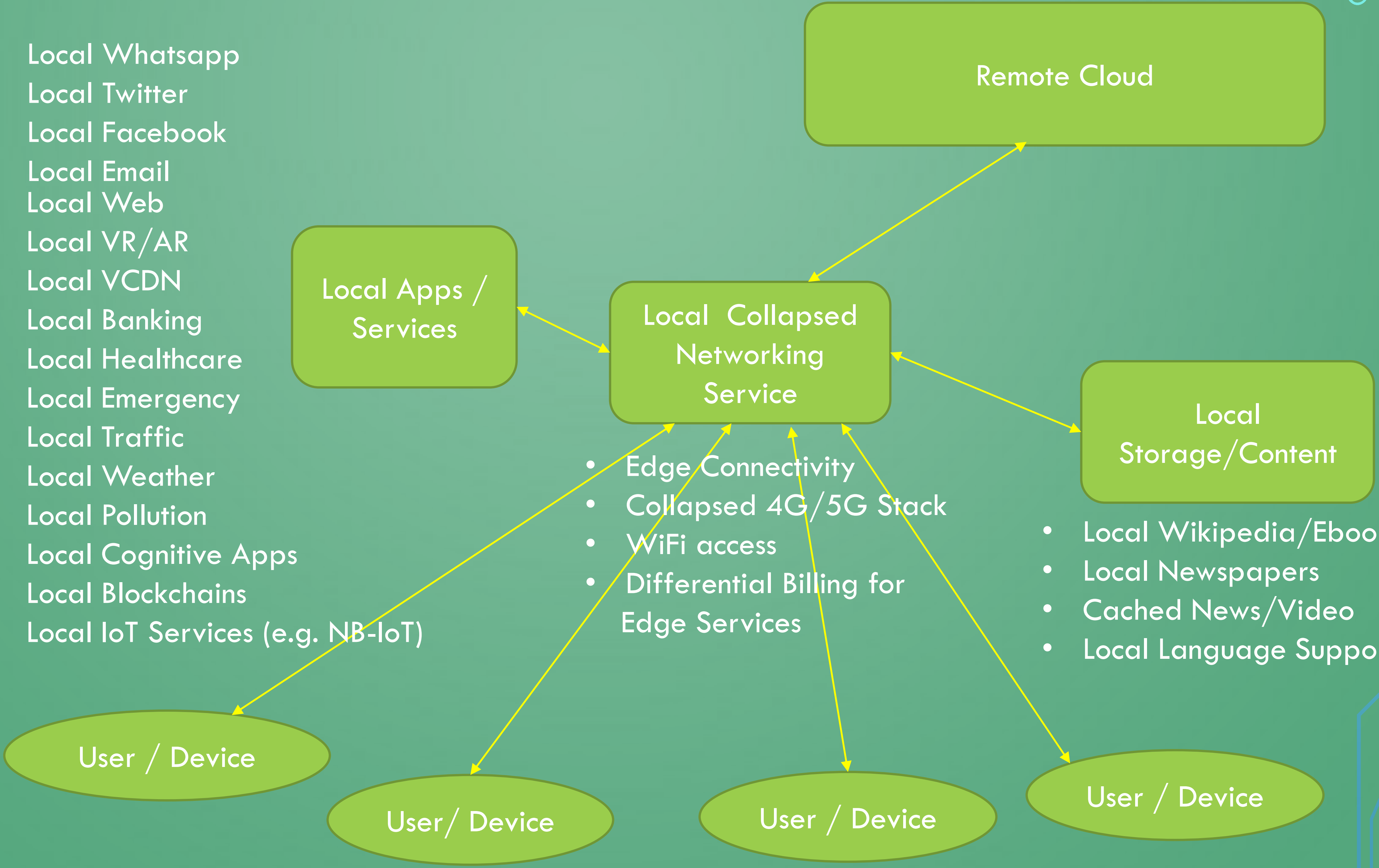
Discussion in Progress: ONAP OOF OSDF
(https://wiki.onap.org/display/DW/Optimization+Service+Design+Framework)

ONAP Optimization Framework Enhancements

TOPIC	ICON	DESCRIPTION
Optimal placement of vNF		Placement of Mobility Virtual Network Elements (CUs) across the highly distributed edge clouds is a fundamental requirement. Service Providers must also optimize the performance of the 5G RAN in real-time.
Slice optimization problem formulation		Ability to model the problem as a constrained optimization problem, closely tied up policies
Slice optimization problem solving		Ability to use appropriate algorithms and solvers to solve the problem in acceptable time frames.
SON - problem formulation		Ability to formulate SON problems as constrained optimization problems, policy driven. Potential use cases - (1) Energy optimization (2) Load balancing
SON - problem solving		Ability to use appropriate algorithms and solvers to solve the problem in acceptable time frames.

SLICING FOR EDGE SERVICES (4G → 5G NSA → 5G SA)

- Local Whatsapp
- Local Twitter
- Local Facebook
- Local Email
- Local Web
- Local VR/AR
- Local VCDN
- Local Banking
- Local Healthcare
- Local Emergency
- Local Traffic
- Local Weather
- Local Pollution
- Local Cognitive Apps
- Local Blockchains
- Local IoT Services (e.g. NB-IoT)



- Edge Connectivity
- Collapsed 4G/5G Stack
- WiFi access
- Differential Billing for Edge Services

- Local Wikipedia/Ebooks
- Local Newspapers
- Cached News/Video
- Local Language Support

THANK YOU

- Questions?

- Contact Email: dilip@ieee.org