

Edge Architecture & Work Items

- General Background
 - The need
 - Service Deployment Goal
- Edge Application Profiles
- Edge Infrastructure Profiles
 - Edge Infrastructure Profile Summary
- Key Challenges w/Centralized ONAP Architecture for Network Function Edge App/Infra Profiles
- Landscape for addressing aforementioned Challenges
- Hierarchical (ONAP Central, Edge) Architecture
 - Single Provider - Hierarchical Architecture
 - Single Provider - Edge Functional Decomposition
 - Single Provider - Sequence Diagram

⚠ Re-arranging content...and, cleaning up....

General Background

A broad set of transformations are taking place:

- *Business transformation: OTT services, faster TTM, Monetization*
- *Technical transformation: QoE, ULL, SDN/NFV/OMEC integration, Edge Analytics, Big data, Virtualization, Automation, C->E, R->E*
- *Architectural transformation: 4 views "NORMA-like" Cloud, ECOMP, Flexible architecture (RAN, Core, CDN, Application delivery, Automation, IoT, fog...)*
- *Industrial transformation: ICT&E*

To efficiently and effectively deploy 5G network supporting ultra low latency and high bandwidth mobile network, we need to deploy variety of applications and workload at the edge and close to the mobile end user devices (UE or IoT). That would include various virtualized RAN and core network elements, content (video), various applications (AR / VR, industrial automation, connected cars, etc.). We might deploy near-real time network optimization, customer experience / UE performance enhancement applications at edge. Edge cloud must support deployment of third party application (e.g. Value added optional services, Marketing, Advertising, etc.). We must deploy mechanisms to collect real time radio network information, process them in real-time (e.g. Geo Location data), summarize, anonymize, etc. and make them available to third party applications deployed at the edge or central location or outside service provider environment. Edge data collection could also be used for training machine learning models and fully trained models can be deployed at the edge to support network optimization.

The need

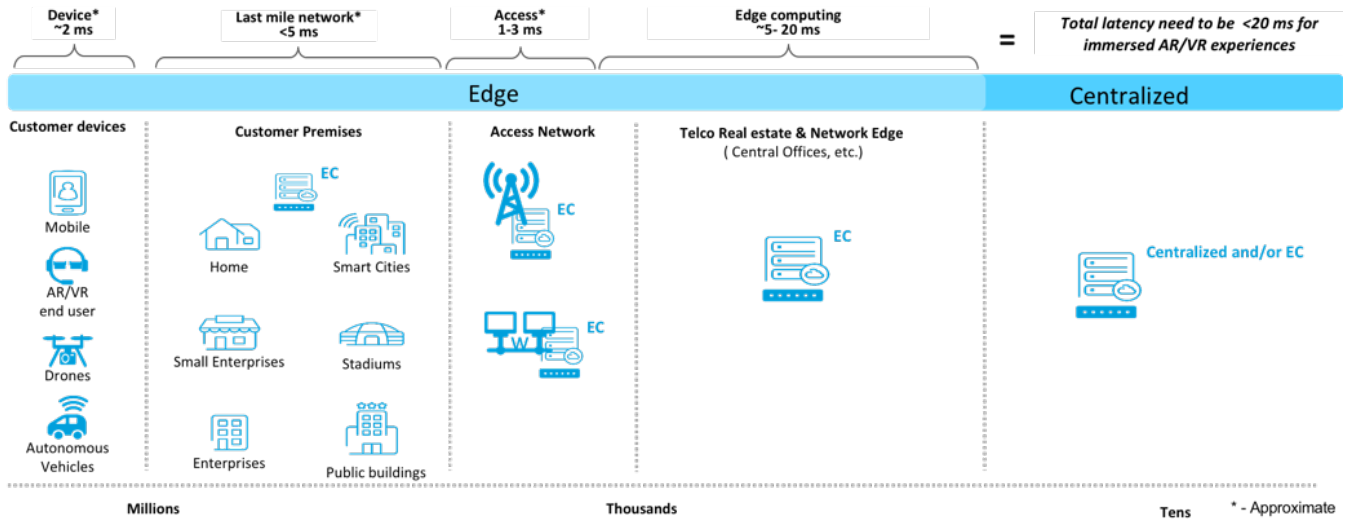
End users and other devices, cyber-physical systems will benefit from a broad set of context information that can enhance and enrich the delivery of a broad set of applications.

Service Deployment Goal

Deliver Application SLAs while minimizing TCO.

Edge Application Profiles

Placement varies depending upon the use case, latency, space availability, etc.,



No	Application Classification (based on required RTT)	Application Examples	Network / Service Behavior Type	Deployment Component/ APIs	ONAP Managed	Edge Deployment Hard /Soft Constraint (Based on RTT)	Potential Application Provider	Casablanca Candidate	Additional Information
1	Real-time (20ms -100ms)	In service path optimization applications which run in open CU-CP platform (also known as RAN Intelligent Controller, or SD-RAN controller).	Real-Time Network State Control	Open 5G CU-CP (CU - Control Plane) – VNFC.	Yes	Hard	NF Vendor /Service Provider/3rd Party	Yes	These applications include load balancing, link set-up, policies for L1-3 functions, admission control and leverage standard interface defined by oRAN / xRAN between network information base (or context database) and third party applications. Data collection through is B1 and implemented using x technology.
2	Near-real-time (500ms and above)	Slice monitoring, performance analysis, fault analysis, root cause analysis, SON applications, Optimization (SON Drive Test Minimization etc.), ML methodologies for various apps.	Network Analytics & Optimization	DCAE	Yes	Soft	NF Vendor /Service Provider	Yes	
3	Near-real-time (500ms and above)	Video Analytics, Video Optimization, Customer geoLocation information, Anonymized customer data etc.	Workload Analytics, Optimization & Context processing	Cloud Edge or Cloud Central	No	Soft	3rd Party	NA. Out of scope for ONAP	The apps are OTT and the service provider is offering their infrastructure as a service to OTT providers.

4	Real-time (10-20 ms)	Third party applications that directly interacts with the UEs, like AR/VR, factory automation, drone control, etc.	Workload Automation / AR-VR / Content, etc.	UE or Cloud Edge	No	Hard	3rd Party	NA. Out of scope for ONAP.	These are third party applications, developed by enterprise customers (e.g. factory automation) or content creators (AR /VR applications). In this case, messages or requests or measurements directly go from UE (via UPF or GWs) to the applications and applications respond back.
5	same as 3)	same as 3)	Value Added Services + same as 3)	same as 3) + MEC/Cloud APIs (Note 1)	Yes	same as 3)	same as 3)	Stretch	Service Provider could be offering video surveillance (video analytics /optimization apps etc.) as a service to enterprises.
6	same as 4)	same as 4)	Value Added Services + same as 4)	same as 4) + MEC/Cloud APIs (Note 1)	Yes	same as 4)	same as 4)	Stretch	Service Provider could be offering factory automation as a service to enterprises.

Note 1: API Details

- e.g., MEC APIs - Location info, Radio control info etc.
- e.g., Cloud APIs - IaaS/PaaS + Context Awareness (time, places, activity, weather etc.)

Edge Infrastructure Profiles

(⚠ example based on Akraino Edge Stack..but, need to generalize)

Profiles	Workloads	Compute	Networking	Storage	Control	Security	Edge Application Infrastructure
----------	-----------	---------	------------	---------	---------	----------	---------------------------------

Large	Support for VMs and containers. Commentary: • VNFs from Operators and Edge applications from customers of Operators. • Number of tenants to be supported???	>50 Compute Servers Accelerators: SRIOV based QAT for Crypto and Compression acceleration. ML/DL Accelerators Compute profiles: Fixed number of profiles are expected to be supported. (Will add profiles)	SRIOV Networking for High performance Data plane VNFs. vSwitch (OVS-DPDK) based networking for all other workloads Multiple leaf switches and two spine switches WAN - Underlay : • L3VPN Support (BGPVPN) • L2VPN support (E-VPN, PBB VPN, VPLS?) Underlay realization options • PE at the Edge (MPLS/BGP start at the Edge) as physical appliance • PE at the edge as virtual appliance • CE- Physical at the edge • CE - Virtual at the edge Overlay realization options • GENEVE based networks (for workload migration, redundancy and scalability) IPv4 and IPv6 support NAT44 with LSN (Large Scale NAT) support by providers. Support for dedicated public IP addresses Commentary: Network sharing among container and VM workloads will need to be supported. DVR (Distributed Virtual Routing) for forwarding packets locally among vSwitch based networks. Leaf/Spine switches for forwarding traffic among SRIOV based networks and for networks between vswitch and SRIOV based networks. Few fixed profiles for following: • Local network profiles • Fabric topology profiles • WAN connectivity profiles	Block device support using Ceph Dedicated nodes for storage (3 nodes) Storage profiles representing whether the nodes are dedicated for storage, use compute nodes for storage, Number of nodes for storage etc... Is support for Object storage required in Edges?	Dedicated nodes for control stack Automation Offload Platform (Offloading ONAP) at the Edge. Few control profiles • Profile 1: ◦ Openstack for VM workloads ◦ K8S for Container workloads ◦ Dedicated nodes for VMs and containers. • Profile 2: ◦ K8S control for both VMs and containers. No need to dedicate the computes. Automation Offload Platform profiles consists of following: ◦ VNF Life Cycle management ◦ Fabric Control ◦ WAN Control ◦ Analytics Offload	Transport : TLS 1.2 and above between ONAP and Edge Services Infra Security: TPM 2.0/SGX for private key security and secret/password protection, Remote attestation to detect any software tampering of compute, storage and control nodes.	MEC Platform as a VNF to provide contextual information to Edge applications.
Medium	Same as above	Same as above. Number of compute nodes are >10 and < 50	Same as above	Same as above, except that there is no dedication of nodes to Ceph cluster	Same as above with respect to control, but Automation Offload Platform is not part of the Edge. No dedicated control nodes. Control functionality is shared with compute nodes. Support for K8S profile as it can support both VMs and containers	Same as above	Same as above
Small	Same as above, but may support very less number of tenants	Same as above. Number of compute nodes are < 10	Same as above, but no PE and CE at the Edge. Fabric itself acts as CE.	Same as above, no dedication of nodes to Ceph cluster	No control at the Edge No Automation offload platform at the Edge Regional sites are expected to provide control and AOP services. Support for K8S based control.	Same as above	Same as above

Edge Infrastructure Profile Summary

- Distributed
 - 1000's of edge locations of varying capacity
 - Casablanca - Implementation
 - 10-100 edge locations (simple starting point)
- Performance-awareness
 - GPU, FPGAs, SR-IOV etc.
 - Casablanca - Implementation
 - SR-IOV desired for Data Plane (5G CU-UP)
 - NIC offload desired for tunnel encap/decap e.g. 5G CU-UP GTP tunnel
- Resource Isolation through fine-grained QoS
 - Support both Latency-sensitive and General purpose applications
 - Support ONAP Management plane components in the same cloud with Workloads
 - Casablanca - Implementation
 - Min/Max resource reservation model desired
- Security
 - Workloads are often deployed in external (non-dc-type) locations and need HW security (TPM etc.)
 - 3rd party applications which need additional HW security (VM, Containers in VM etc.) and SW security (Inter-component TLS etc.)
 - Casablanca - Implementation
 - Edge Clouds with private IP addresses, i.e. reachable via private connections
 - For example, edge cloud in a public cloud provider reachable via AWS direct connect or Azure express route or Google partner interconnect
- Capacity constraints

- Very small footprint (few nodes per physical location), Medium footprint (10's of nodes per physical location), Large footprint (100's of nodes per physical location)
- Casablanca - Deployment
 - Need number of cores per servers; Need storage capacity/pool
- Cloud Diversity
 - Private and Public Cloud Providers
 - Casablanca - Implementation
 - Note: ONAP currently supports private edge clouds based on VMware VIO, Wind River Titanium Cloud, Upstream OpenStack
 - Desire to have at least one Public Cloud Provider (Azure, AWS, GCE etc.) as an Edge Cloud Provider
 - ONAP central instantiates an Edge Cloud instance (blue cloud provider in gliffy) via a IaaS API to cloud provider
 - ONAP central instantiates one or more ONAP edge components as need, e.g. DCAE
 - ONAP central instantiates one or more NFs, e.g. 5G CU-UP/CP
- Configuration Diversity
 - 5G Factory Automation, 5G General Mobility Services etc. – User Plane components (DU, CU-UP, UPF etc.)

Key Challenges w/Centralized ONAP Architecture for Network Function Edge App/Infra Profiles

WAN network bandwidth & latency issues for the following key functions

- Conveying Application & Infrastructure metrics, faults in real-time from Edge Cloud to Centralized ONAP for closed-loop fault management for a single edge site with a single Cloud control plane
- Conveying Application & Infrastructure metrics, faults, alerts in real-time from Edge Cloud to Centralized ONAP for dynamic network function closed-loop fault management/optimization across multiple edge sites with a single Cloud control plane
- Conveying Application & Infrastructure metrics, faults, alerts in real-time from Edge Cloud to Centralized ONAP for dynamic network function closed-loop fault management/optimization across multiple edge sites with multiple Cloud control planes

Exemplary Network Function Placement/Service Assurance Policy for demonstrating the aforementioned challenges

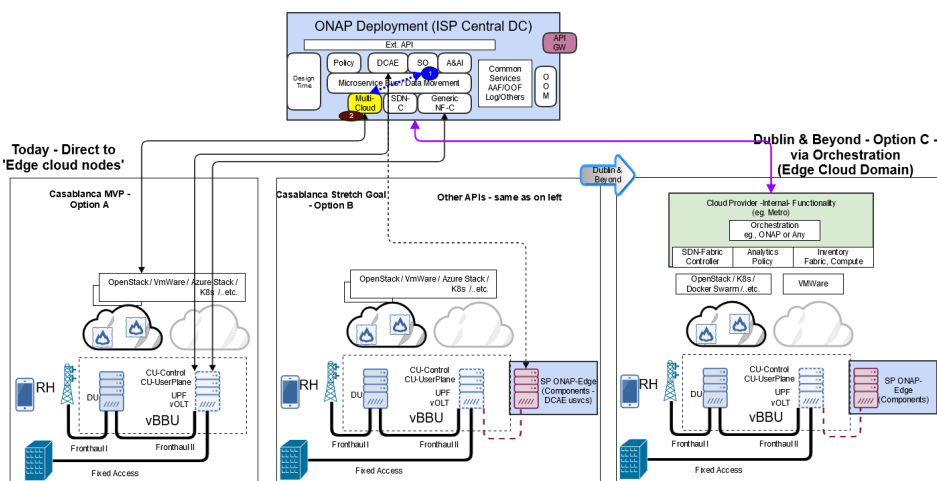
- "5G CU-UP VNF location to be proximal to a specific physical DC based on 5G DU, bounded by a max wire latency from physical DC"

Landscape for addressing aforementioned Challenges

- Cloud Providers
 - Public Clouds such as Azure, Amazon, GCP do not support dynamic network function closed-loop fault management/optimization across multiple edge sites with single/multiple Cloud control planes
- Akraio Open Source
 - Defines Blueprints for Edge Cloud using OpenStacks (VMs) and K8S (Containers)
 - Plans to use ONAP for VNF/Service Orchestration for addressing aforementioned challenges

Hierarchical (ONAP Central, Edge) Architecture

Single Provider - Hierarchical Architecture



- Cloud Provider Business Unit:
 - Provides hosting of Workloads, ie., IaaS/PaaS

- SP installs and manages ONAP in separate 'Management Cloud' instances
 - SP installs and manages Network Services + 3rd Party Apps in separate 'Services/Apps Cloud' instances
- Cloud Provider Business Unit:
 - Provides SaaS, eg., Analytics/Closed Loop as a Service, LCM of Apps, etc.
 - ONAP Edge may not be needed
- Cloud Provider Business Unit:
 - Types of virtualized cloud resource tenant and their characteristics
 - Virtualized Network Workload Cloud Resource Tenant Category
 - Network Management Cloud Resource Tenant Category
 - Virtualized Application Workload Cloud Resource Tenant Category
 - Application Management Cloud Resource Tenant Category
 - Physical Network Function and their characteristics
 - Part of Edge Cloud Orchestrator
- Immediate interest to ONAP for Network Function use cases
 - Virtualized Network Workload Cloud Resource Tenant Category
 - Guaranteed
 - Burstable (with minimum guarantee)
 - Best Effort
 - Network Management Cloud Resource Tenant Category
 - Burstable (with minimum guarantee)

Single Provider - Edge Functional Decomposition

Function	Statefulness	ONAP Project Mapping	Details
Inventory	yes	A&AI	
IP Address Management (IPAM)	yes	SDN-C	
Multi-Cloud Support	no	Multi-VIM	
Initial Placement	no	OOF	
Closed Loop Controller	no	APP-C	
Closed Loop Policy	no	Policy	
Infra Closed Loop Analytics	no	Multi-VIM, DCAE	
App Closed Loop Analytics	no	DCAE	
Logging	no		
Infra/App Monitoring events, statistics	no		

Single Provider - Sequence Diagram

ONAP Activity Goal #1: ONAP requires IaaS/PaaS attributes (see ongoing work – Edge Cloud Infrastructure Enablement in ONAP, 5G Items for Casablanca) from Cloud providers for Infrastructure profiles that allow Distributed, Highly-secure, Config/Cloud-diverse, Capacity-constrained and Performance/Isolation-aware

ONAP Activity Goal #2: Define hierarchical ONAP Central, Edge Architecture/functional interactions (API reference points) to support aforementioned Application/Infrastructure profile in Any "Cloud" (internal Business Unit or external Partner) at Any "Location" edge, regional or central.

Sequence Diagram

