Change Request Form

OPG.01 CR1001 Operator Platform Telco Edge Proposal baseline

Document Summary

Official Document Number, Document Title and Version Number	OPG.01 Operator Platform Telco Edge Proposal v1.0 (Current)
Official Document Type	White Paper
Change Request Security Classification	Non-confidential
Is this a new document or a Major or Minor Change?	New Document
Will this Change Request result in a Major or Minor version update?	Major Version
This document is for	Approval
Input Editor and Organisation	Faisal Zia (GSMA)
Additional Contributors	Carolina Cadenas (GSMA), Divya Khoker (GSMA), Emma Wood (GSMA), Ian Pannell (GSMA), Roger Brown (GSMA), Tom Van Pelt (GSMA)
Issuing Group/Project	TG
Approving Group/Project	TG
Change Request Creation Date	02/10/2020
What are the reasons for and benefits of creating this new document or Change Request?	This documents provides a proposal requirements and a high-level architecture for an Operator Platform. This Operator Platform will expose capabilities in the operator networks to 3rd party developers and to enable a federation between operators to enable every entry point for developers to reach the combined operators' footprint. In the first phase, the focus is on exposing Edge Compute capabilities. The document is meant to provide a baseline to discuss with other SDOs and Open Source organisations on how they can support the development of this Operator Platform through the development of specifications and implementations that align with these

Change Request Form	
	requirements.

Review Log (to be completed by GSMA Support Staff)			
Workflow Step	***Document Review Comments***	***GSMA Support Staff Name***	***Comments Date***
Step 1: Change Request Creation (no comments required)			
Step 2: Document Quality and/or Legal Review			
Document Quality Team	Minor edits suggested for consideration.	Ian Pannell	02/10/2020
Legal Review	INSERT COMMENTS HERE Please enter details for the Legal Review Confirm Legal feedback Record any issues, actions and key decisions	GSMA Support Staff Name	DD/MM/YY
Step 3: Formal Review			
Group(s)/Project(s) Review(s) Comments and Feedback	Document was developed in the OPG with continuous availability of baseline documents including the already approved content. Final content approvals had to be done by 17th September 2020 with further submissions accepted until 25th September that focussed on improving the quality and consistency of the document. Those have all been handled and included in the CR that went for final approval of the group on 5th October.	Faisal Zia	9 Sep 2020 05 Oct 2020
Step 4: Formal Approval(s)			
Group(s)/Project(s) Approval(s) Comments and Feedback	OPG approved first version of the document with changes required by TG	Faisal Zia	20/10/2020

1 Introduction

1.1 Overview

The Operator Platform Concept whitepaper [1] introduced the opportunity that the Enterprise segment offers for operators in the 5G era. It explained the need for a generic platform to package and expose network capabilities and to monetise them. It also described the leverage that the Operator Platform can bring to bear to realize this opportunity. These "levers" are, in summary, the existing relationship of operators with enterprises, the vast local footprint of operators, their excellent position to deliver on the digital sovereignty principles and the competence to provide high-reliability services. Finally, the document introduced a first high-level view of the Operator Platform (OP) architecture identifying main functional blocks and interfaces.

Note: The first version of this document describes initial requirements to run cloud services from the edge of an operator's network; it is not definitive, improvements and recommendations are invited from SDOs, Open Source Projects, industry fora, and market participants from across the cloud services value chain, please send these to futurenetworks@gsma.com.

After the concept described in [1], this document is proposed to act as a next step by suggesting suitable technical requirements, functional blocks and interface characteristics. It also maps these to specifications produced in different Standards Developing Organisations (SDOs) and identifies the gaps where these SDOs would be requested to fill in such specifications. To facilitate the development of interoperable Operator Platform products it also provides a reference of the Open Source communities where source code implementing the Operator Platform functionality and interfaces is available.

This document is written as a proposal to the wider industry with the intention of providing a guide to future specifications for an Operator Platform. The target for this document is all impacted organisations, including platform developers, edge cloud providers, SDOs, Open Source Projects, industry fora, and market participants.

1.2 Scope

The intention for this document is to guide the entire industry ecosystem; operators, vendors, OEMs and service providers to define a common solution for network capabilities exposure. As a first phase, the document will provide an end-to-end definition of the Operator Platform for support in edge computing environments.

This document covers the following areas:

- Operator Platform requirements

    - **Focus on Edge Computing**: The first phase will define edge computing exposure, integrating the network services to the Application Providers and enabling a simple and universal way of interacting towards edge computing platforms.
    - **Open to new services**: The definition should allow the evolution of the platform to expose additional services going forward, such as IP Communications and networking slicing, among others.

- Architecture and functional modules

  - **Reference architecture for edge computing**: Definition of modular architecture suitable for implementation at the network edge.
  - **Reference interfaces**: Definition of interconnection for the end-to-end service, between service providers to end users, network elements and federated platforms. As a first approach, this document will focus on Northbound, Operator Platform Federation and User to Network interfaces.
  - **Mobility**: Network and terminal integration should allow service continuity against end user mobility in home and visited networks.

- Standardisation and Open Source communities

  - **Gap evaluation in the standards**: This document analyses gaps in the current standards and identifies which SDO should be addressed to request further definition of each part of the architecture and functionalities through detailed specifications, protocols and Application Programming Interfaces (API).
  - The **Detailed specification** of architecture and APIs **will be defined by SDOs**, using the baseline in this document.

    The GSMA shall review progress to ensure that the end-to-end system is defined consistently across SDOs.

  - **Open Source communities**: The document identifies which existing Open Source organisation is the most suitable to host a community that can handle the development of the Operator platform.

- Speed to market and resonance

  - **Fit with established ecosystems**: The OP defines the Mobile Operator staging of a broader cloud ecosystem. To meet tight market timing and minimise heavy lifting, it has to fit into existing structures and staging, enabling Application Providers to spin their existing capabilities into the Mobile Edge space. Wherever possible, the OP will reuse existing and established structures and processes.

The first version of this document focusses on the use of the Operator Platform to provide services to devices attached to their home network. It does include high-level requirements that go beyond this scenario because they may influence future architecture choices, but further versions of this document are planned to cover, for example, the following areas in greater depth:

- Service access by devices that are attached to networks other than their home network (e.g. roaming, Wi-Fi, etc.),
- Access to OP services in a network different from the one to which the device is attached (e.g. those provided on another operator's network),
- Low latency interaction between OP applications in different networks,
- Local interfaces on an end-user device,
- Serverless models,
- Management plane,
- Resource reservation,

- Device mobility,
- Call flows.

Note: The document is a proposal and is non-binding.

1.3 Reference Architecture

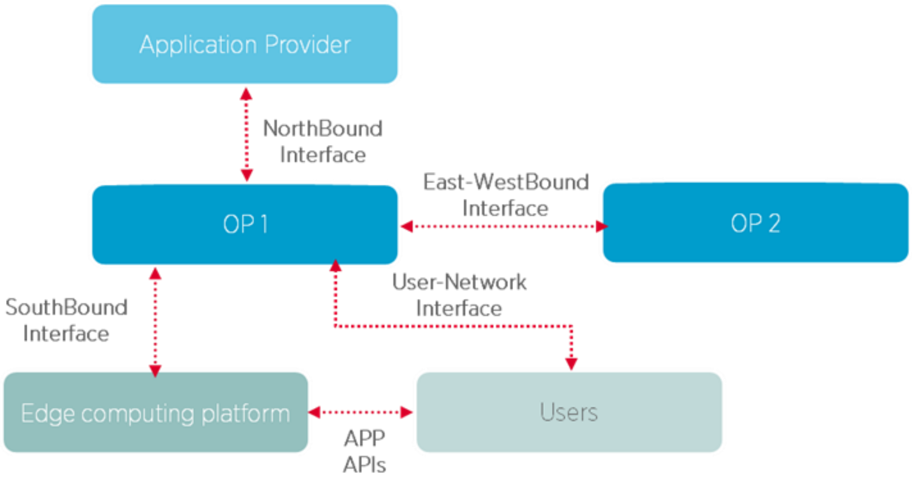

Figure 1: High-Level Reference Architecture

The proposed OP architecture consists of a common exposure & capability framework and is based on a four sided approach:

- **Northbound Interface (NBI)**: enables service management and fulfilment of enterprise and Application Providers' use case requirements.
- **East/Westbound Interface (E/WBI) APIs**: the interface between instances of the OP that extend an operator's reach beyond their footprint.
- **Southbound Interface (SBI)**: connecting the OP with the specific operator infrastructure that will deliver the network services and capabilities to the user.
- **User-Network Interface (UNI)**: enables the User Client (UC) hosted in the user equipment to communicate with the OP.

1.4 Definitions

Term	Description
Application Client	The application functionality deployed on the User Equipment. It works with the User Client to use the Edge Cloud service provided by the Operator Platform
Application Instance	A single deployment of an Edge Application.
Application Provider	The provider of the application, which accesses the OP to deploy its application on the Edge Cloud, and thereby uses the Edge Cloud Resources and Network Resources.
Availability Zone	An OP Availability Zone is the equivalent of an Availability Zone on Public Cloud. An Availability Zone is the lowest level of abstraction exposed to a developer who wants to deploy an Application on Edge Cloud. Availability Zones exist within a Region. Availability Zones in the same Region have anti-

Term	Description
	affinity between them in terms of their underlying resources - this ensures that in general terms when a developer is given a choice of Availability Zones in a Region, they are not coupled, ensuring separation and resilience.
Capability Exposure Role	The OP role in charge of the relationship with the Application Providers. It unifies the use of multiple Edge Clouds, which may be operated by different Operators and accessed through different Operator Platforms.
Certificate Authority	An entity that issues digital certificates.
Cloudlet	A point of presence for the Edge Cloud. It is the point where Edge Applications are deployed. A Cloudlet offers a set of resources at a particular location (either geographically or within a network) that would all provide a similar set of network performance.
Data Protection	Legal control over access to and use of data stored in computers.
East/Westbound Interface	Interface between instances of the OP that extend an operator's reach beyond their footprint.
Edge Application	The application functionality deployed on the Cloudlet
Edge Cloud	Cloud-like capabilities located at the infrastructure edge including, from the Application Provider's perspective, access to elastically allocated compute, data storage and network resources. Edge Clouds are targeted mainly at Edge-Enhanced Applications and Edge-Native Applications. In the context of this document, the Edge Cloud is managed by an Operator Platform. The phrase "located at the infrastructure edge" is not intended to define where an Operator deploys its Edge Cloud. The Edge Cloud is expected to be closer (for example latency, geolocation, etc.,) to the Application Clients than today's centralised data centres, but not on the User Equipment, and could be in the last mile network, for example. (Note 1)
Edge Cloud Resources	In the context of this document, resources of the Edge Cloud Service that are managed by the Service Resource Manager Role.
Edge-Enhanced Application	An application that is capable of operating in a centralised data centre, but which gains performance, typically in terms of latency, or functionality advantages when provided using an Edge Cloud. These applications may be adapted from existing applications that operate in a centralised data centre or may require no changes. (Note 1)
Edge-Native Application	An application that is impractical or undesirable to operate in a centralised data centre. This can be due to a range of factors from a requirement for low latency and the movement of large volumes of data, the local creation and consumption of data, regulatory constraints, and other factors. These applications are typically developed for, and operate on, an Edge Cloud. They may use the Edge Cloud to provide large-scale data ingest, data reduction, real-time decision support, or to solve data sovereignty issues. (Note 1)
Federation Broker Role	The OP role in charge of easing the relationship between federated OPs. For example, it allows an OP to access many other OPs through a single point of contact and to simplify its contractual relationships. The Federation Broker Role is optional since federation can be performed

Term	Description
	directly between two Federation Managers (in a one-to-one relationship).
Federation Manager Role	The OP role that publishes and provides access to the resources and capabilities of another OP, including its Capability Exposure Role and Service Resource Manager Role.
Flavour	A set of characteristics for compute instances that define the sizing of the virtualised resources (compute, memory, and storage) required to run a workload. Flavours can vary between operator networks.
Leading OP	The Operator Platform instance that is connected to the Application Provider and receives the onboarding requests, sharing them to the selected federated platforms/operators.
Network Resources	In the context of this document, the network services and capabilities provided by the Operator that are managed by the Service Resource Manager Role.
Northbound Interface	Interface that exposes the Operator Platform to Application Providers
Operator	In the context of GSMA OP, an Operator is a network operator that deploys an Edge Cloud, provides connectivity to User Equipment and is an Operator Platform.
Operator Platform	An Operator Platform facilitates access to the Edge Cloud capability of an Operator or federation of Operators. It follows the architectural and technical principles defined in this document. NOTE: Future versions of this document may extend the capabilities of the Operator Platform.
Region	An OP Region is equivalent to a Region on Public Cloud. It is the higher construct in the hierarchy that is exposed to a developer who wishes to deploy an Application on the Edge Cloud and broadly represents a geography. A Region will typically contain one or multiple Availability Zones. A Region exists within an Edge Cloud.
Regional Controller	The Regional Controller functions at the geographic region level wherein it manages Cloudlets within that geography. The size of Cloudlets and scope of geography under a regional controller is up to the operator to define.
Routing Domain	Operator's inbuilt networking structure in a region which dictates how data path is routed to an anchor like PGW/UPF to internet and back to the device.
Service Resource Manager Role	The OP role in charge of orchestrating Edge Cloud Resources and Network Resources for use by Application Providers and end users, including application load management over the Edge Cloud, the configuration of network capabilities and the management of the User Client relationship.
Southbound Interface	Connects the OP with the specific operator infrastructure that delivers the network services and capabilities.
Tenant	A Tenant is the commercial owner of the applications and the associated data. Note: It is for further study how to align this concept with the commercial track.
Tenant Space	A Tenant Space is a subset of resources from a Cloudlet that are dedicated to a particular tenant. A Tenant Space has one or more VMs running native or containerised workloads.

Term	Description
User Client	Functionality that manages on the user's side the interaction with the OP. The User Client represents an endpoint of the UNI and is a component of the User Equipment. NOTE: Different implementations are possible, for example, OS component, separate application software component, software library, SDK toolkit and so on.
User Equipment	Any device used directly by an end-user to communicate. The term includes an IoT device (Internet of Things). User Clients and Application Clients are deployed on the User Equipment.
User-Network Interface	Enables the User Client (UC) hosted in the user equipment to communicate with the OP.

Note 1: This definition is based on that in “Open glossary of edge computing”, v2.0 [3].

1.5 Abbreviations

Term	Description
5G	5th Generation Mobile Network
5GC	5G Core
AAA	Authentication, Authorisation and Accounting
AAF	Application Authorisation Framework
AF	Application Function
AMF	Access and Mobility Management Function
API	Application Programming Interface
AR	Augmented Reality
B2B	Business to Business
B2B2C	Business to Business to Consumer
B2C	Business to Consumer
CDM	Common Data Model
CISM	Container Infrastructure Service Manager
CPU	Central Processing Unit
CRUD	Create, Read, Update and Delete
DBaaS	DataBase as a Service
DC	Data Centre
DNAI	Data Network Access Identifier
EA	Edge Attribute
ETSI	European Telecommunications Standards Institute
E/WBI	East/Westbound Interface
eMBB	Enhanced Mobile Broadband
FPGA	Field Programmable Gate Array
FQDN	Fully Qualified Domain Name

Term	Description
GDPR	General Data Protection Regulation
GMLC	Gateway Mobile Location Centre
GPS	Global Positioning System
GPSI	Generic Public Subscription Identifier
GPU	Graphic Processing Unit
GW	GateWay
HPLMN	Home Public Land Mobile Network
HTTP	HyperText Transfer Protocol
IaaS	Infrastructure as a service
ID	IDentifier
IMSI	International Mobile Subscriber Identity
I/O	Input/Output
IoT	Internet of Things
IP	Internet Protocol
IPsec	Internet Protocol Security
ISG	Industry Specification Group
ITU	International Telecommunication Union
KPI	Key Performance Indicator
L4	Layer 4
LADN	Local Area Data Network
LAI	Location Area Identification
LBO	Local BreakOut
LCM	Life-Cycle Management
MCC	Mobile Country Code
MEC	Multiaccess Edge Computing
MNC	Mobile Network Code
MR	Mixed Reality
MSISDN	Mobile Subscriber Integrated Services Digital Network Number
NBI	Northbound Interface
NDS	Network Domain Security
NEF	Network Exposure Function
NPU	Neural Processing Units
NUMA	Non-Uniform Memory Access
NWDAF	Network Data Analytics Function
OP	Operator Platform
OS	Operating System
OTT	Over the Top
PaaS	Platform as a service

Term	Description
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RBAC	Role-Based Access Control
RNIS	Radio Network Information Service
RRS	Resource Requirements Specification
SAAS	Software as a service
SBI	Southbound Interface
SBI-CR	Southbound Interface – Cloud Resources
SBI-NR	Southbound Interface – Network Resources
SCEF	Service Capability Exposure Function
SDK	Software Development Kit
SDO	Standards Developing Organisation
SLA	Service Level Agreement
SPR	Subscriber Profile Repository
SR/IOV	Single Root I/O Virtualisation
SRM	Service Resource Manager
SUPI	SUbscription Permanent Identifier
TAI	Tracking Area Identification
TLS	Transport Layer Security
UC	User Client
UE	User Equipment
UNI	User to Network Interface
UPF	User Plane Function
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VIM	Virtualised Infrastructure Manager
VM	Virtual Machine
VNF	Virtualised Network Function
VPLMN	Visited Public Land Mobile Network
VPU	Vision Processing Unit
VR	Virtual Reality
WAC	Wholesale Application Community
Wi-Fi	Wireless network protocols, based on the 802.11 standards family published by the IEEE.

1.6 References

Ref	Doc Number	Title
[1]		Operator Platform Concept – Phase 1: Edge Cloud Computing https://www.gsma.com/futurenetworks/resources/operator-platform-concept-whitepaper/
[2]		void
[3]		Open Glossary of Edge Computing, Linux Foundation Edge, https://github.com/State-of-the-Edge/glossary/blob/master/edge-glossary.md
[4]	3GPP TS29.522	5G System; Network Exposure Function Northbound APIs https://www.3gpp.org/DynaReport/29522.htm
[5]	3GPP TS 29.122	T8 reference point for Northbound APIs https://www.3gpp.org/DynaReport/29122.htm

2 Proposed Architectural Requirements

2.1 High-Level Requirements

2.1.1 General

The OP and its architecture shall comply with the following requirements:

1. The OP shall expose operator network functions and resources to 3rd party applications.
2. For each operator supporting the OP, there shall be an OP instance that has the sole responsibility for managing the OP resources and services that the OP exposes in that operator's network.

    a) This instance may be operated by the operator themselves or be outsourced to a 3rd party.

3. The OP shall be able to effectively isolate each Tenant's applications from all other tenants' applications.
4. The interfaces that an OP instance offers to other parties (e.g. 3rd party providers, other OP instances, clients, etc.) shall be provided using common definitions based on the requirements in this document.

2.1.2 Functionality offered to 3rd party Application Providers

The OP and its architecture shall fulfil the following requirements related to the functionality offered to 3rd party Application Providers:

1. The OP architecture shall allow a 3rd party to use a single interface to manage OP applications deployed towards the subscribers of multiple operators subject to there being an agreement with the operators involved.

    Note: such an agreement could result in the federation of OPs between involved operators.

2. The interfaces that an OP provides to 3rd parties for the development and deployment of OP applications shall allow for easy deployment of applications developed for public clouds.
3. The OP shall hide the complexity of the OP architecture, the involved operator networks and client access to those networks from 3rd party OP Application Providers.
4. There is a “separation of concerns” of the OP and Application Providers, meaning that the Application Providers and OP do not require knowledge of each other’s internal workings and implementation details, for instance:

    a) the OP does not expose its internal topology and configuration, Cloudlets’ physical locations (see note), internal IP addressing, and real-time knowledge about detailed resource availability (Resources are provided as a virtualised service to an Application Provider);
    b) the OP does not know how the application works (for instance, it does not know about the application’s identifiers and credentials).

    Note: The OP provides information on the geographical region(s) where the edge cloud service is available. The Application Provider provides information sufficient for the OP to process the request and (if accepted) fulfil it.

5. The OP architecture shall allow a 3rd party deploying an OP application to monitor the application’s usage across the networks on which it is deployed.
6. The OP architecture shall allow an OP application deployed within an operator network to interface securely with backend infrastructure of the application that is outside of the operator network.
7. The OP architecture shall allow a 3rd party application deployed in the OP to store data in a manner that is secure and compliant with applicable local regulations.

2.1.3 Functionality offered to End Users/Devices

The OP and its architecture shall fulfil the following requirements related to the functionality offered to end users and their devices:

8. The OP shall allow end user devices to access services provided through OP applications.
9. Services provided as OP applications to end user devices shall remain available while that device moves within the operator network and when it moves to another operator's network subject to their being an agreement between the involved operators (i.e. home and visited).

    Note: Because it applies only to visiting subscribers, such an agreement may be different from a federation agreement to expose deploy OP applications on another operator's infrastructure to their subscribers.

10. The OP shall allow the end user to access services deployed on the OP seamlessly and securely.

2.1.4 Functionality offered to Operators

The OP and its architecture shall fulfil the following requirements related to the functionality offered to operators:

11. The OP architecture shall allow an operator to monitor and track OP resource and OP-related network resource usage in their network.
12. The OP architecture shall enable an operator to monitor the OP and network resource usage of their subscribers in a visited network.
13. The OP architecture shall allow the OP to control the quality of service delivered by the network for the interaction between an end user device and an OP application.
14. If the Operator Platform is part of the operator's security domain [3], then it can access through the SBI (and any other operating interface) the network and cloud resources.

    Note: An operator may choose to outsource some of its functionality to another party. For example, an operator could devolve the management of its edge cloud service to an external OP. Then the external OP would know some details about the operator's internal workings, such as its Cloudlets' physical locations. This approach would require an agreement covering commercial, data protection, security, legal issues, etc.

    Note: Security Domains administer and determine the classification of an enclave of network equipment/servers/computers. Networks with a different security domain are isolated from other networks. Security Domains are managed by a single administrative authority. Within a security domain, the same level of security and usage of security services is typical. A network operated by a single operator or a single transit provider typically constitutes one security domain, although an operator may subsection their network into separate sub-networks. 3GPP TS 33.210 Network Domain Security (NDS); IP network layer security.

15. Similarly, there is a "separation of concerns" of the operators from each other, and between OPs. Where the Operator Platform is not part of the operator's security domain, then there is also a "separation of concerns" of the operators from the OP. "Separation of concerns" again means that they do not require knowledge of each other's internal workings and implementation details, for instance: the operators do not expose their internal topology and configuration, Cloudlets' (exact) physical locations, internal IP addressing, and real-time knowledge about detailed resource availability from one operator to other.

2.1.5 Functionality offered to other OPs

The OP and its architecture shall fulfil the following requirements related to the functionality offered to other OPs:

16. The OP architecture shall allow an OP to deploy applications provided by 3rd parties on another OP (e.g. to enable a federation).
17. A federation of independent operators in an Operator Platform enables additional capabilities, such as:

a) the User Equipment (UE) can continue to use the Edge Cloud service if it moves into a “visited network”.

18. The OP architecture shall allow an OP to receive workloads/applications from other OPs to serve the operator's subscribers supported by the OP, to serve roaming subscribers from the operator supported by the requesting OP, or both.
19. The OP architecture shall allow such an intermediate OP to monitor and track resource usage of an application in the OP on which it has been deployed.

2.2 Edge Enabling Requirements

2.2.1 High-Level Requirements

The following requirements apply for the OP related to enabling access to the edge:

20. The OP shall allow the operator to expose compute and storage resources at the edge of the operator network on which services can be deployed for use by specialised and regular end user devices.
21. The OP architecture shall allow an OP application deployed at the edge of the operator network to interact with low latency with OP applications deployed at nearby operator network edges, including those of other operator networks in the same area.

2.2.2 Resource management requirements

2.2.2.1 General principles

“Resource” refers to edge compute resources (processing and storage) and associated networking.

As general principles:

- The OP provides edge compute resources as a virtualised service to an Application Provider or another party in the OP ecosystem (for example, an aggregator or another operator).
- This Application Provider or other party – and only this one - is responsible for the management of the Edge Applications on the virtualised resource that they have been provided.

Note: Having exactly one entity managing a virtualised resource avoids the technical complexity of multiple controllers, which would require capabilities such as grants and reservations, as well as more complex commercial considerations.

2.2.2.2 Resource management

The OP manages edge compute resources (processing and storage) and associated networking:

22. An OP shall provide edge compute resources on a virtualised basis to another party in the OP ecosystem (for example, an Application Provider, an aggregator or another operator).

23. An OP or Application Provider is responsible for managing the virtualised resources with which it has been provided. For example, in the case of an Application Provider, this includes the allocation, de-allocation and potentially in-life management (such as scaling) of virtualised resource to a specific application client.
24. If one OP or Application Provider overloads the virtualised resource it has been allocated, this should not degrade the performance of others.
25. An OP or Application Provider does not have visibility of the resources that another is allocated or is using.
26. All parties in the OP ecosystem use the same data model for the virtualised resources.
27. It is optional for resource management to provide telemetry or other metrics from the edge node.

2.2.3 Cloud application development

The OP shall retain the generic benefits of cloud application development, hosting and staging native to public cloud deployments. This includes:

28. Support for Continuous Development, through Code development pipelines similar to those provided in a public cloud.
29. Support for Continuous Integration, through staging in edge test sites.

2.2.4 Edge deployment enhancements

The OP shall enhance the edge deployment of workloads and applications to make it easy to integrate workloads and applications coming from the public cloud.

Note: Details are part of the next version of this document.

2.2.5 Data Protection Management

The OP shall offer Data Protection management. Specifically:

30. Data protection regulations differ between countries and regions (such as the EU). The Application Provider shall be able to restrict where the Edge Application is deployed (country, region) to meet Data Protection requirements.
31. The OP shall be able to serve the Data Protection needs of Application Providers and enterprises by protecting data beyond regulatory requirements.

2.2.6 Lifecycle management of Edge Applications

The process lifecycle management of Edge Applications should be based on the following suggested workflow for deployment:

32. Create Tenant Space: a tenancy model which allows auto-scaling and deploying microservices as a set of containers or Virtual Machines (VMs).
33. Create the application manifest, specifying the workload information, defining an application mobility strategy that includes QoE, geographical store and privacy policies;
34. Create the application backend instance, including autoscaling.

Note: The management plane and regional controllers to support the above workflow shall be part of the next version of this document.

3 Proposed Target Architecture

3.1 Introduction

The Operator Platform's primary goal is providing a global and common way of exposing certain services to external Application Providers, whether through a direct connection from the resource owner towards the final consumer, or employing intermediate integration platforms.

The OP environment will host multiple actors who may need to interwork to complete end-to-end service delivery, resource sharing and footprint expansion; which implies defining a common way of enabling actors to interact with each other.

To solve both requirements above, the OP shall enforce a multi-layer architecture with multi-role separation of the complete functionalities and requirements presented in Chapter 2.

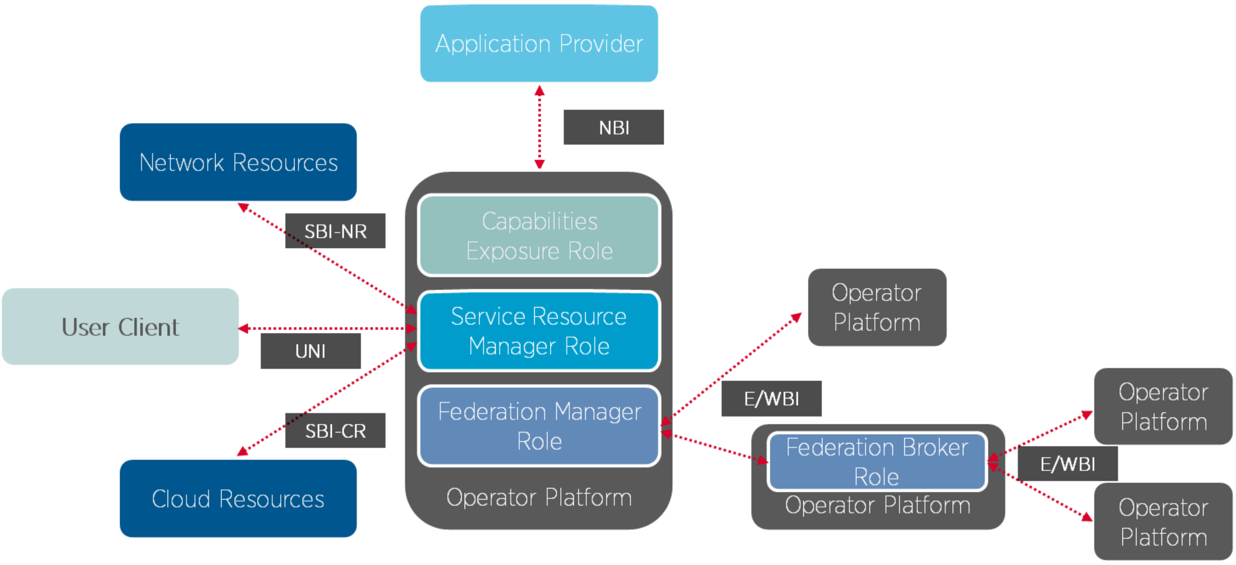

Figure 2: OP Roles and Interfaces Reference Architecture

The following sections will cover the functionalities and role separation, as well as the relationship between each player, or role, via the different interfaces.

3.2 Roles and Functional Definitions

3.2.1 General

The OP functionality is realised via multiple roles. These roles together enable an OP instance to interact with and to execute scenarios from/towards other actors in the OP ecosystem, namely Application Providers, other OP instances, the Cloud Resources and the Network Resources.

The functions of these roles would be implemented via modules that are discussed in the following sections. The modules can be inside or outside the OP instance.

This section lists these roles along with their key functions.

3.2.2 Capabilities Exposure Role

The Capabilities Exposure Role in the OP is responsible for exposing the capabilities of the OP towards the Application Providers via the NBI.

Typical scenarios enabled by the Capabilities Exposure role are:

- Application Onboarding;
- Application Metadata/Manifest Submission;
- Application Lifecycle Management;
- Application Resource Consumption Monitoring;
- Edge Cloud Resource Catalogue exposure;
- The geographical footprint reachable via the OP (either via own resources or partner OP resources).

3.2.3 Service Resource Manager Role

The Service Resource Manager role in the OP is responsible for Management of Cloud and Network resources from the Edge Cloud(s) via the SBI and UNI interfaces.

Typical scenarios enabled by the Service Resource Manager role towards the different interfaces are:

- **SBI:**
  - Inventory, Allocation and Monitoring of Compute resources from Edge Cloud Infrastructure via the Southbound Interface – Cloud Resources (SBI-CR);
  - Orchestration of Application workloads on the Edge Cloud Infrastructure via the SBI-CR interface;
  - Interacting with the Mobile Network via the Southbound Interface – Network Resources (SBI-NR) for:
    - Fetching Cloudlet locations based on the mobile network data-plane breakout location;
    - UE Mobility notifications to assist application mobility on UE mobility.
    - Configuring traffic steering in the Mobile Network towards MEC applications orchestrated in Edge Clouds;
    - Receiving statistics/analytics via Network Data Analytics Function (NWDAF)/ Network Exposure Function (NEF), e.g. to influence Application placement or mobility decisions.
- **UNI:**
  - Application Instantiation/Termination, e.g. based on triggers from the UNI;
  - Application Endpoint exposure towards User Clients via the UNI;
  - Application Placement decisions, e.g. based on measurements/triggers from the UNI.

3.2.4 Federation Broker

The Federation Broker and Manager roles in the OP are responsible for interfacing with other OPs via the East-West Bound Interface.

Typical scenarios enabled by the Federation Manager role are:

- Federation Interconnection Management;
- Edge Cloud Resource Exposure and Monitoring towards partner OPs;
- Application Images and Application metadata transfer towards partner OPs;
- Application Instantiation/Termination towards partner OPs;
- Application Monitoring towards partner OPs;
- Service Availability in visited networks.

The Federation Broker is an optional role. It acts as a broker to simplify interconnection between multiple OPs.

3.3 Federation Management

The Federation Management functionality within the OP enables it to interact with other OP instances, often in different geographies, thereby providing access to a larger footprint of Edge Clouds and Operator's capabilities for the Application Providers.

The following are prerequisites to enable the federation model:

- Operators need to have an agreement to share Edge Cloud resources;
- Operators would need to agree on an Edge Cloud resource sharing policy;
- Operators would need to enable connectivity between the OP instances over which East/West Bound Interface signalling would flow.

The Federation Management functionality is realised via the multiple functional blocks within an OP instance as listed below.

Note: There may be legal constraints restricting the distribution of specific applications to certain regions that would need to be considered in the agreement when Federation is being planned among multiple operators. The technical impact of such legal constraints on OP is for further study.

3.3.1 Federation Interconnect Management

The Federation Interconnect management functional block in the OP deals with the establishment and sustenance of the Federation Interconnect (E/WBI) between the OP instances. The Federation Interconnect uses secure transport, plus capabilities such as integrity protection for the E/WBI messaging between OP instances.

During the Federation Interconnect establishment, the Federation Managers of the participating OPs need to verify each other's identities through mutual authentication.

Federation interconnect management functionality also ensures that the partner OP is authorised to establish and maintain the interconnect according to the terms of the federation agreement between the partnering OPs/Operators.

3.3.2 Resource Synchronisation and Discovery

The OPs shall exchange and maintain the types of resources offered to each other.

This exchange includes information about Availability Zones:

- A Region identifier (e.g. geographical area);
- Compute Resources Offered: e.g. a catalogue of resources offered (CPUs, Memory, Storage);
- Specialised Compute Offered: catalogue of add-on resources, e.g. Graphic Processing Units (GPU), Vision Processing Units (VPU), Neural Processing Units (NPU), Field Programmable Gate Arrays (FPGA).

This information may change and can be updated via the E/WBI whenever the geographical area or the types of resources offered to an OP by a partner changes due to Operational or Administrative events (e.g. due to scheduled maintenance).

A subscription/notification mechanism will be supported over the E/WBI to achieve the above.

3.3.3 Application Management

This procedure corresponds to the forward of a northbound request from one operator to accommodate an Edge Application in another operator's Cloudlets. Operators authorise the deployment based on available resources and federation agreement.

In the Federated model, one OP can coordinate with partner OPs to assist application onboarding, deployment and monitoring in the partner OP Edge Clouds. The E/WBI interface must provide capabilities to support application onboarding, deployment and monitoring in partner OP Edge Clouds.

The Application Providers interact with one OP instance and provide their requests over the NBI along with the intended geographical regions that they want to target. The OP instance translates the NBI interactions to the E/WBI.

There may be multiple models possible for how application orchestration is performed via the E/WBI.

On a federation relationship, the decision of which Edge Cloud(s) to deploy the applications is decided by the partner/target OP based on the Zone/Region preferences indicated by the Application Provider. In doing so, the Application Provider criteria are used by the partner OP as provided to it via the E/WBI.

The E/WBI, therefore, enables the partner OP to be informed about the Application Provider's requirements - information which the home OP has learnt from the Application Provider, through the NBI.

3.3.4 Service Availability on Visited Networks Management

When a User Client (UC) requires accessing the Edge Cloud service from a visited network, the federation model facilitates service availability for this UC. The service could be (preferably) provided via local Edge Cloud resources of the visited OP if local breakout is available for roaming UEs. If local breakout is not possible, the UC may be served via the home OP. For that reason, and considering the credential and authoritative ownership of the users to the home operator, the first register request shall always be driven to the home operator's OP.

Note: Home Public Land Mobile Network (HPLMN) identifiers or pre-provisioned IDs can be used to form the home Service Resource Manager (SRM) URL. e.g. http://register.opg.mnc.mcc.pub.3gppnetwork.org.

During UC registration, to support the Edge service discovery procedure for the UC in the visited OP, the home OP shall identify that the UC is in a visited network and provide the UC with the discovery URL of the visited OP to redirect the UC registration. The home OP shall be aware of the discovery URL of the visited OP either:

- via E/WBI communication, or
- by deriving it, when the UC performs the home OP registration procedure, from the visited operator’s identity, i.e. the Mobile Network Code (MNC) and Mobile Country Code (MCC).

  Note: NEF/SCEF event and information retrieval may be used to identify the Visited Public Land Mobile Network (VPLMN) ID and the visited OP URL where the user is connected.

To facilitate service availability in a visited OP, the E/WBI shall allow the visited OP to access the UC profile within the home OP to perform authentication and authorisation and grant the service access. When the UC tries to access a service when on visited networks, the visited OP authenticates the UC using the authentication information received via the E/WBI from the home OP of the UC as part of the secured federation interconnection.

This procedure is network-driven, which means that it shall only be triggered after a network change or a token expiration. Once a UC is registered on a visited OP, that platform shall remain responsible for providing applications to the UE until any network change, not per application request.

3.3.5 Edge Node Sharing

Two operators may decide to share edge nodes to maximise their edge coverage. Using the figure below as an example, Operator A may deploy edge sites on the north region of a country and operator B on the south Region of that country. In such a scenario, Operator B may have an application deployed in Operator A edge node and allowing its subscriber to access it.

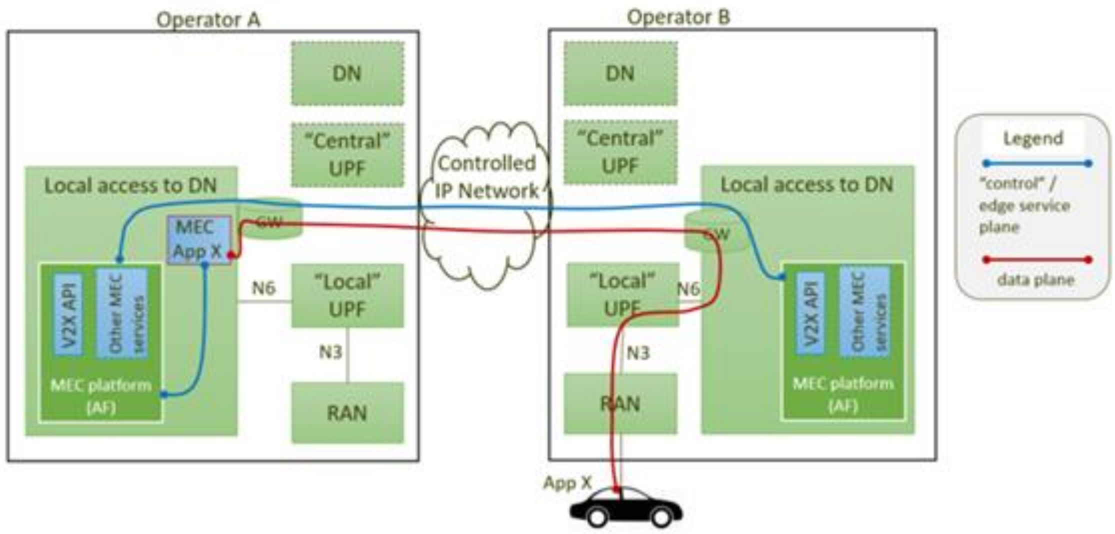

Figure 3: Edge Node Sharing

Figure 3 above shows a car as a subscriber of Operator B accessing a MEC service that is offered by Operator A in the current location of the subscriber. The blue connectivity between the two OPs refers to the E/WBI interface.

The East/West interface will apply as in a similar fashion as in the user in the visited network scenario by allowing the serving Operator platform to access the UC profile information from the home operator

Operator B subscriber will access its home network/operator platform and ask for the closest edge node running the required Edge Application. Operator B OP will identify that the nearest edge node is in operator A and will redirect the subscriber to operator A OP.

Operator A OP will authenticate the UC by interrogating Operator B's Subscriber Profile Repository (SPR) and will allocate the closest edge node allowing the Edge Application access. The same procedure as for service availability can be followed here for the credential exchange among OPs employing the federation interface.

3.3.6 Configurations

An OP shall provide various configuration capabilities to establish and manage the Federation Interconnect.

3.3.6.1 Partner OP Provisioning

An OP shall allow mechanisms to provision partner OP information that would be used for Federation Interconnect establishment and management. This information would include:

- Partner Name;
- Partner's geographical area (e.g. Country of operation);
- Partner identifiers;
- Partner's federation interconnect E/WBI endpoint;
- The federation agreement validity duration.

3.3.6.2 Authentication and Authorisation

When an OP connects to a partner OP via the federation interconnect, it needs to authenticate itself to the partner OP. This authentication requires that authentication information (e.g. digital certificate or passphrase) be provisioned in the OP. This mechanism can be mutually agreed between operators as a first step. A more generic solution based on a Certificate Authority could be considered going forward within the GSMA.

An OP may authorise a partner OP for a limited duration (based on federation agreement) or for specific Availability Zone(s) where it has Edge Cloud resources. This information would need to be provisioned during partner provisioning.

3.3.6.3 Resource sharing policies

An OP shall provide controls to the Operator to specify Availability Zones to be made available to a partner OP. These controls shall allow all or part of the resources of an Availability Zone to be shared. Availability Zone sharing is dependent on the Federation agreement that exists between the OPs.

3.3.7 Operational visibility.

The OPs shall have an operational view of each other, allowing Fault Management and Performance management within the limits of their agreements in the federation contracts.

A subscription/notification mechanism should be available to allow the above.

3.3.8 Edge Cloud resource consumption

An OP monitors Edge Cloud resource consumption by the Edge Applications. These applications would also include applications from the partner OPs. The OP informs the resource consumption statistics of the partner OP applications to the partner OPs via the E/WBI.

The amount of resources shall be identified per Operator and Edge Application and may be reported per Availability Zone.

An OP would use this information as an input for billing, audit and settlement purposes.

3.4 Common Data Model

The Common Data Model (CDM) introduces a set of standardised data schemas for describing characteristics of the elements of an OP system. The common data model presented here covers elements of an operator platform, including applications, OP roles, and edge clouds, as well as functional aspects, such as security.

The data model should define the information elements required to deploy and manage an OP system.

It should define a minimum set of mandatory information elements and should allow for reasonable default values for these elements to be inferred where they make sense.

The model should accommodate optional information elements following a common syntax, to allow OP systems to evolve. Examples of optional information elements are:

- Infrastructure configuration deemed necessary by an application for proper operations, such as Non-Uniform Memory Access (NUMA) node affinity or core sequestration.
- Optional QoS attributes that not all networks may support, e.g., Packet Error Loss Rate (from 3GPP 23.203).

Optional information attributes default to “not specified” if not expressed in a data object.

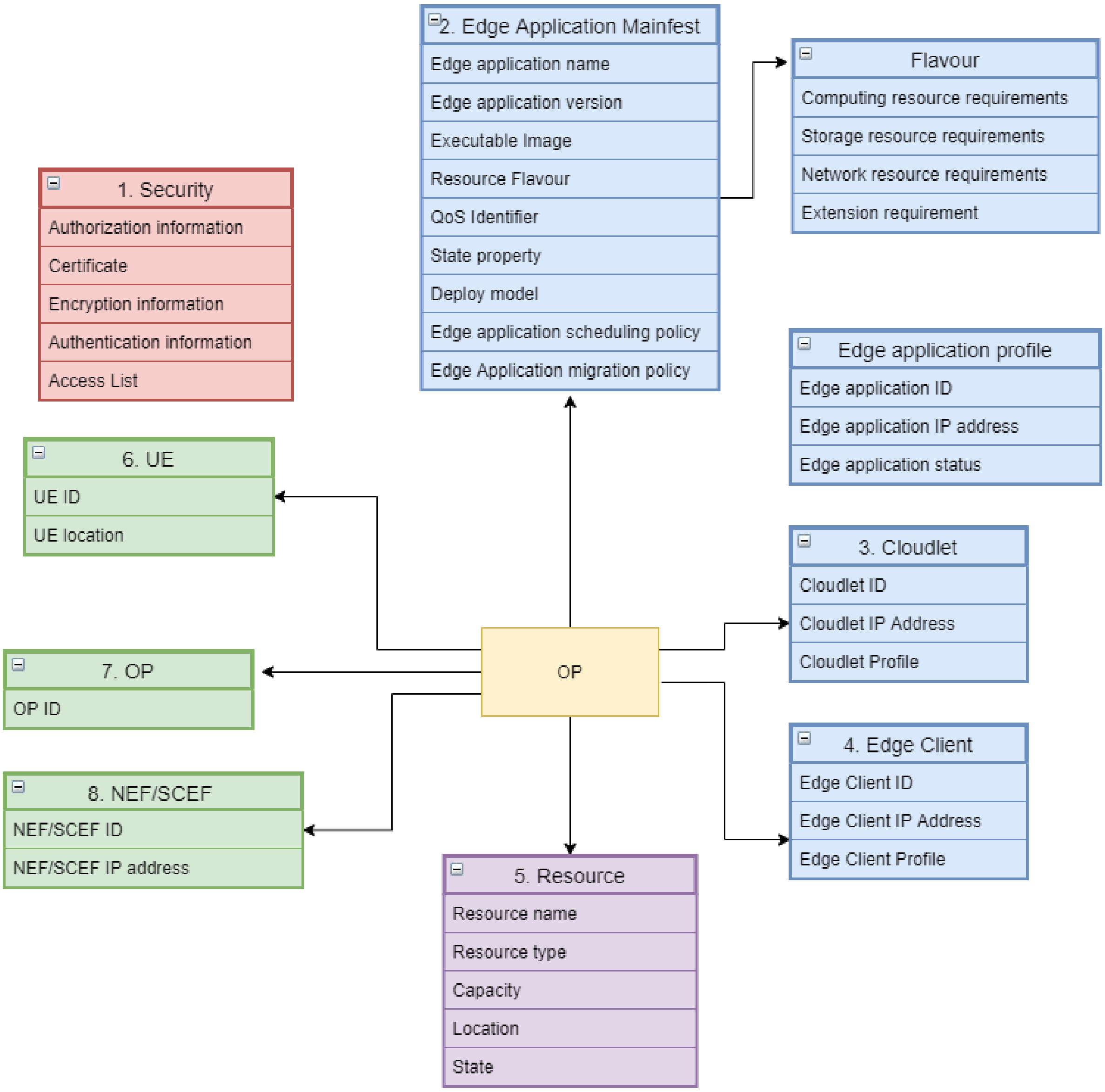

Figure 4: Common Data Model

3.4.1 Security

The security element of the data model provides information elements to allow trust domains, entities, credentials, and other information required to support secure processing among the roles of an OP platform. The following table should be interpreted as the information elements maintained by a role (e.g., OP, Application Provider) about other trusted domains.

Data type	Description	Applicability
Authorisation information	Authorisation information of the Application Provider	UNI/East/West/NBI
Certificate	The certificate of the Application Provider	UNI/East/West/NBI
Encryption information	To encrypt data transmission and data streams, or cryptographic credentials (e.g., TLS certificates) used for information exchange among trust domains	UNI/East/West/NBI
Authentication information	Certified identities of other trusted domains	UNI/East/West/NBI
Access List	For information elements that may be requested by an API between trust domains, the list of identities authorised to make a request	UNI/East/West/NBI

Table 1: Common Data Model – Security

3.4.2 Edge Application

The data model of the Edge Application contains the information about the application object required to instantiate it (the Edge Application manifest), and the information about the instantiated application required to manage it (the Edge Application profile).

An application instantiation is created by an OP. More precisely, an edge cloud instantiates it in response to an OP's request. As such, it is in the OP’s trust domain. The input to this operation is an application manifest, and the output, besides an application instantiation, is an application profile.

An application manifest is created and should be owned by an Application Provider. An OP that instantiates an application from the application manifest should request the manifest from the Application Provider. This requirement implies that Partner OPs should be able to request the application manifest from the Leading OP for the Application Provider.

The application manifest shall contain mandatory data elements and may include optional data elements. A data element may be described by a separate sub-model below (e.g., the QoS specification for an application is a sub-model).

The application profile is a data object created and owned by an OP. It describes an application instantiation on an OP. It shall contain any data elements specified in the application manifest used to create it, together with the values used in its instantiation.

The following table describes the information elements in the Application Manifest data model. In addition to the elements listed, the model should allow the definition of additional attributes, at the Application Provider' or OP's discretion. A possible realisation of optional elements is key-value pairs, as is used in various data models.

Data type	Description	Applicability	Optionality
Edge Application name	Name of the Edge Application. The name is an artefact created by the Application Provider. The name is namespaced to the Application Provider. There is no default value; this must be supplied.	East/West/NBI	Mandatory
Edge Application version	The version of the Edge Application. The default value is 1.0.	East/West/NBI	Mandatory
Executable Image	A URI (or another similar name) of the executable image (or container) to be installed and executed by the OP.	East/West/NBI	Mandatory
Resource Flavour	The “name” or identifier of the Flavour that should be used to instantiate the application, as selected by the Application Provider. The default value is “Default Flavour”. “Flavour” is defined below.	East/West/NBI	Mandatory
QoS Identifier	A “name” or identifier of the QoS description for network traffic, as selected by the Application Provider. The default value is “Default QoS”, which is described below.	East/West/NBI	Optional
State property	Indicates whether the application has state (e.g., persistent file systems, database, and location-dependent associations with other elements that must be migrated in a coordinated manner when an application is relocated). The default value is “stateless”.	East/West/NBI	Optional
Deploy model	Indicates whether an application may be located freely by the OP, or whether the Application Provider specifies the edge cloud, on which it is to be deployed. The default value is “free”.	East/West/NBI	Optional
Edge Application scheduling policy	Indicates whether a backend application can be scaled up or down based on offered traffic. The default value is “not scalable”.	East/West/NBI	Optional
Edge Application migration policy	Indicates whether a backend application may be moved from its current operator network or current geographic region (i.e., without violating the General Data	East/West/NBI	Optional

Data type	Description	Applicability	Optionality
	Protection Regulation (GDPR)).		

Table 2: Information elements in the Application Manifest data model

The following table is the data module of the Edge Application profile

Data type	Description	Applicability
Edge Application ID	The ID of the Edge Application running on the edge node	East/West/NBI
Edge Application IP address	The IP address of the Edge Application running on the edge node	East/West/NBI
Edge Application status	The status of the Edge Application running on the edge node	East/West/NBI

Table 3: Edge Application profile

A Flavour is a description of a set of resource requirements used by an application instantiation. It should have a name that can be used to identify the description uniquely and that should be global across OPs in an OP system.

A resource description should be consistent with those appearing in Flavours available in public clouds. This means that a Flavour should specify CPU, memory, storage, I/O bandwidth, CPU architecture, special hardware (e.g., accelerators).

A Flavour definition ensures that, if an Application Provider selects a Flavour for a manifest, the application will run successfully if instantiated into a cluster containing at least the resources specified.

Flavours are not standardised (at this time) in this document. The Federation should undertake to produce and maintain a Flavour catalogue.

Data type	Description	Applicability	Optionality
Computing resource requirements	The computing resource requirements of the Edge Application	East/West/NBI	Optional
Storage resource requirements	The storage resource requirements of the Edge Application	East/West/NBI	Optional
Network resource requirements	The network resource requirements of the Edge Application	East/West/NBI	Optional
Extension requirement	The extension requirements of the Edge Application	East/West/NBI	Optional

Table 4: Flavour profile parameters

A QoS description is a characterisation of traffic between an Application Client and an Edge Application and carried by a flow between the client and backend. A QoS description allows an Application Provider to describe the physical constraints in an edge network that should be met for the application to run successfully and provide a correct Quality of Experience (QoE) for the end user at the UE.

Various standards organisations have investigated QoS and have specified definitions of QoS classes. For example, research in the 5G community has led to a description of QoS traffic classes that are common (or are expected to be common) in 5G networks. The reader is directed to 3GPP 23.502, Table 5.7.4-1. In this table, the traffic classes are defined via a collection of metrics, including:

- “resource type” (i.e., whether a flow is guaranteed the service requested, or only gets best effort);
- Packet Delay Budget;
- Packet Error Rate;
- Maximum Data Burst Volume.

These are aggregate statistics, collected over a time window, and the length of that window is specified by the operator. These statistics apply to the path from the UE to the User Plane Function (UPF).

For edge computing, QoS on this path is necessary, but not complete. It does not cover the segment from the UPF to the backend application. Including this path in a QoS latency budget is important because the details of scheduling an application on a host concerning core assignment or NUMA node assignment have a material effect on the latency and particularly the jitter experienced by an application.

Based on this discussion:

- The QoS spec may contain, as optional attributes, latency, bandwidth, and jitter.
- The attributes shall be measured from UE to backend application, over a time window consistent with the duration of a gaming session.
- Optional attributes shall be permitted, following the requirements of the data model as a whole.
- Considerations of QoS from UE to UPF, and definition of QoS classes from UPF to backend application, are noted as requiring further investigation.

Data type	Description	Applicability	Optionality
Bandwidth	Bidirectional data rate between UE and Edge Application measured end-to-end with a “loopback” application	East/West/NBI	Optional
Latency	The round trip delay between UE and Edge Application measured end-to-end with a “loopback” application	East/West/NBI	Optional
Jitter	The variance of round-trip delay between UE and Edge Application measured end-to-end with a “loopback” application	East/West/NBI	Optional

Table 5: QoS profile

3.4.3 Cloudlet

The application functionality deployed on the Cloudlet, the Cloudlet data model provides various resources (including storage, GPU/NPU support, etc.) for applications. The common data model of Cloudlet involves Cloudlet ID, Cloudlet IP address and Cloudlet Profile.

Data type	Description	Applicability
Cloudlet ID	The FQDN defining the Cloudlet of where the Edge Client shall connect. This ID shall be unique per OP domain.	UNI/East/West
Cloudlet IP address	The IP address of the Cloudlet of where the Edge Client shall connect	UNI/East/West
Cloudlet Profile	Gathers the Cloudlet information (e.g. ID, IP address) and characteristics (e.g. storage, GPU support, etc.)	UNI/East/West

Table 6: Common Data Model of Cloudlet

3.4.4 Edge Client

The Edge Client represents an endpoint of the UNI and is a component of the User Equipment. Different implementations are possible, for example, OS component, separate application software component, software library, Software Development Kit (SDK) and so on. The data module of the edge application includes Edge Client ID, Edge Client IP address and Edge Client Profile.

Data type	Description	Applicability
Edge Client ID	A unique value that defines the client ID accessing the OP	UNI/East/West
Edge Client IP address	The IP address of the Edge client	UNI/East/West
Edge Client Profile	Reflects the profile of the edge client and the level of Authorisation to access the edge nodes.	UNI/East/West

Table 7: Common Data Model of Edge Client

3.4.5 Resource

The resource can be provided by cloud and edge. The common data model of resource properties includes the type, capacity, location and state of the resource.

Data type	Description	Applicability
Resource name	The name of the resource	East/West/NBI
Resource type	The type of resource	East/West/NBI
Capacity	The capacity of the resource	East/West/NBI
Location	The location of the resource	East/West/NBI
State	The state of the resource	East/West/NBI

Table 8: Common data model of resource properties

3.4.6 UE

UE is the User Equipment. The common data model of UE includes the UE ID, UE location. There is a need to preserve the UE ID in multiple scenarios such as roaming, authentication and charging.

Data type	Description	Applicability
UE ID	The terminal ID. For mobile networks, the ID shall be based on International Mobile Subscriber Identity (IMSI) and Mobile Subscriber Integrated Services Digital Network Number (MSISDN) (in case of 3G-4G access) and General Public Subscription Identifier (GPSI) and Subscription Permanent Identifier (SUPI) in case of 5G access as defined by 3GPP. When presented out of the trusted domain (e.g. NBI exposure), the UE ID may take different format (e.g. a token) bound by the OP to ensure user privacy.	UNI/NBI/East/West/South
UE location	UE location indicates where the UE is connected into the network. In a cellular network, the information can be based on Cell Identity or Tracking Area Identity as defined by 3GPP. When presented out of the trusted domain (e.g. NBI exposure), the UE location may take a different format (e.g. a token) bound by the OP to ensure user privacy.	UNI/NBI/East/West/

Table 9: Common data model of UE

3.4.7 OP

The common data model of Operator Platform includes the OP ID.

Data type	Description	Applicability
OP ID	The ID of the Operator Platform. This ID shall be unique per OP domain	UNI/NBI/East/West/South

Table 10: Common data model of Operator Platform

3.4.8 NEF/SCEF

NEF (Network Exposure Function)/SCEF (Service Capability Exposure Function), as a 5G/4G network capability opening function, provides secure disclosure services and capabilities provided by 3GPP network interfaces.

Data type	Description	Applicability
NEF/SCEF ID	The FQDN of the NEF/SCEF against which the OP shall connect. The ID shall be unique per OP domain	South
NEF/SCEF IP address	The IP address of the SCEF or NEF against which the operator platform shall connect	South

Table 11: Common data model of NEF/SCEF

3.5 Interfaces

3.5.1 Northbound Interface (NBI)

The Edge Cloud is similar to a traditional cloud offering, but with the advantage of better QoS, in particular lower latency in a geographical region or regions which correspond to areas nearby where an operator has deployed Cloudlets. The NBI allows an OP to advertise the above cloud capabilities that it can provide to Application Providers. The NBI allows an Application Provider to request an Edge Cloud service with the resources and features it requires, and for the OP to accept or reject the request (but not to negotiate).

3.5.1.1 General Onboarding Workflow

Application Providers usually have information about their users and the resource requirements of their application. User information may include the number of users and the traffic they generate as a function of time and location, the QoS expectations of the users, and the compute and network resource requirements of the application to function correctly. This information will be referred to as workload information. Application Providers may estimate workload information a priori or use telemetry to collect workload information. Application Providers provide workload information to Orchestration Services to automate and optimise the deployment of Application Instances. Developers may analyse collected workload information to predict changes in users and traffic over time. The deployment of Edge Applications can be independent of network mobility or specific device attachment.

The NBI is the interface between the developers and an OP.

35. To allow a developer to "write once, deploy anywhere", the NBI is a standard, universal interface. In other words, a developer does not need to rewrite its applications to work with another OP.
36. An OP may provide the edge cloud itself directly, or offer it indirectly (that is, using an edge cloud service provided by another party, such as another OP or operator).
37. The capabilities offered through the NBI depend on what is provided (directly or indirectly) by the underlying edge cloud. For example, the geographical regions where the edge cloud is provided, the "granularity" of the edge cloud and network service, the quality of service available, and the type of specialised compute.
38. An Application Provider will not have visibility of the exact geographical locations of the individual Cloudlets, and will not be able to request deployment of its application on a specific Cloudlet. Instead, the OP will offer to Application Providers edge cloud service in Availability Zones. The OP chooses the size of each "Availability Zone and which and how many Cloudlets to use to provide its edge cloud service in each Availability Zone".
39. The NBI will provide a request-response mechanism through which the Application Provider can state a geographical point where a typical user could be, and be informed of the mean latency performance that is expected. As an option, an OP can publish a "heat map" showing expected mean latency performance at different locations; this is not part of the NBI, and the OP could post it on a webpage, for instance.
40. The NBI allows an Application Provider to reserve resources ahead of their usage or to get resources as their applications need them ("reservationless" or "auto-scaling"). An Application Provider can also request that its edge cloud resources are isolated

from those used by other Application Providers. The NBI allows an Application Provider to delete their reservation. A reservation is intended to be relatively long-lasting (for example, not triggered by the activity of one Application Client).

41. These resources include CPU, memory and specialised compute (such as GPU). Since the types of resources are evolving, the NBI must be flexible enough to incorporate future kinds of resources, as they are defined.
42. The NBI allows the OP to advertise the (relatively) static information about the types of resource that it offers (“flavours”) but does not allow the OP to indicate the dynamic information about current availability or usage of the resources.
43. The NBI allows the OP to accept or reject the request, but not to negotiate.
44. The NBI allows an Application Provider to download its application image to the OP. The NBI enables an Application Provider to delete its application image.
45. The NBI allows an Application Provider to request that their application is instantiated. The NBI enables an Application Provider to request that instances of their application be Created, Read, Updated and Deleted (CRUD).
46. The NBI allows an Application Provider to specify that their Edge Applications are restricted to a particular geographical area, corresponding to data privacy (GDPR) restrictions.
47. The NBI allows an Application Provider to specify whether it requires service availability on visited networks (that is, when a UE roams away from its home network operator).
48. The NBI allows the OP to report telemetry information about the performance of the edge cloud service to an Application Provider. Because different Application Providers will require (and different OPs will offer) different degrees of performance information (how fine-grained and how often), the NBI will provide a request-response mechanism to allow an Application Provider to request a particular granularity for the telemetry. Similarly, the NBI will provide an Application Provider with information about faults that (may) affect its edge cloud service.
49. Backend services deployment can be based on several different strategies to enable mobility of Edge Applications, including:

    a) Static, whereby the Application Provider chooses the specific region or edge sites and the particular services for each location.
    b) Dynamic, whereby the Application Provider submits criteria to an orchestration service and the orchestration service makes best-effort decisions about Edge Application placement on behalf of the Application Provider. One implementation of this would have Application Providers choose a region in which they yield control to a system operator’s or cloud operator’s orchestration system. This orchestration system would determine the optimum placement of an Application Instance based on the amount of requested edge compute resources, the number of users and any specialised resource policies. This model assumes the OP is aware of resource needs per Application Instance.

50. The process of Application Instance creation should be based on the following suggested workflow for deployment:

a) Resource reservation and isolation(optional), a tenancy model which allows auto-scaling and deploying microservices as a set of containers or Virtual Machines (VMs);
b) Create the application manifest, specifying the workload information for the Edge Application to Orchestration Services;
c) Create the Application Instance, including autoscaling if required.

51. The other processes of lifecycle management of Edge Applications should follow a similar pattern.
52. For the service provider edge, there are two different views of resource management: orchestration and resource control:

a) Orchestration View: Operators and Application Providers interact to create a running Edge Application. The Application Provider specifies workload requirements, and the Operator uses them (with other information) to orchestrate an Edge Application.
b) Resource Control View: The resource provider manages its Cloudlets in response to Orchestration actions. Resource management includes creating collections of resources as Flavours, which are specified by the Application Provider and used by the Orchestrator.

53. The deletion of Edge Applications should be as follows:

a) Stop the Application Instance;
b) Release the related resources including network, computing and storage;
c) Delete the application in the orchestrator and remove the reserved resource.

54. The NBI shall provide a set of functionalities for Application Providers, including access to Edge Cloud, and image management. Application lifecycle management and operations are also functionalities to be provided through this interface.

3.5.1.2 Resource Requirement Specification

55. The OP shall enable Application Providers to express the resource (e.g., compute, networking, storage, acceleration) requirements of an application running on a Cloudlet.
56. The Resource Requirements Specification (RRS) shall have the following attributes:

a) An application ported from a cloud to a Cloudlet will, in general, have an RRS. The mapping of a cloud RRS to a Cloudlet RRS shall be “natural”, meaning:

i. The attributes that may appear in a Cloudlet RRS should be a superset of those appearing in a cloud RRS. For example, if an attribute set {numcores, memory_size, disk_space, IO_bandwidth} is common across cloud service providers, a Cloudlet RRS should contain these attributes as well.
ii. An “Edge Attribute” (EA) is an attribute that may appear in a Cloudlet RRS, and which describes requirements that an OP deems necessary to perform resource and allocation for an edge app, but which does not appear in cloud RRSs. Edge Attributes should, but need not, be

specified in a Cloudlet RRS. Omitted EAs shall have reasonable default values assigned that are determined by the OP.

iii. One of the RRS formats to be provided shall be that of “flavours”. A flavour is a vector of RRS attribute values that are statically defined and associated with an identifier for the flavour. Selecting a particular flavour identifier is equivalent to specifying the values of each of the attributes that appear in its definition.

b) There shall be no standardised, a priori, definition of flavours. Instead:

   i. The flavours offered by a federation of OPs shall be agreed among the operators in the federation.
   ii. The flavour definitions shall be defined in OP documentation and available to all operators and all Application Providers using the federated platform.
   iii. All OPs in a federation should use the same flavour definitions.
   iv. The protocols and APIs provided by OP should provide consistent "fallback" behaviour for the case that Flavour catalogues between OPs are not consistent.
   v. The protocols and APIs provided by an OP should provide consistent "fallback" behaviour for the case that the app provider requests a flavour that is not available.

c) A Cloudlet RRS should include attributes pertinent to operating an application in an edge location. These attributes may include:

   i. Physical region
   ii. Network delay, jitter, and packet loss rate as measured by an accumulated average of these statistics for traffic originating at an edge zone and terminating in a Cloudlet.
   iii. Variance or confidence interval (e.g., 95% confidence) for network statistics.

d) A Cloudlet RRS shall provide means of specifying technology-related attributes, such as the use of accelerators.

e) A Cloudlet RRS shall provide a means of specifying additional scheduling EAs that relate to modern CPU technology. These attributes could support sequestering on virtual CPUs or taking into account NUMA nodes or high-performance network interface technology like Single Root I/O Virtualisation (SR/IOV).

3.5.1.3 Application Resource Catalogue

57. The NBI shall allow applications providers to access the resource catalogue.
58. The Resource catalogue shall consider local resources.
59. Resources footprint shall be abstracted to Availability Zones, preserving the network topology hiding as stated in sections 2.1.2 and 2.1.4.
60. An Application Provider shall be able to create custom request zones that can be reached by one or more catalogued availability zones, not only at coarse level but also on a private or limited footprint.

3.5.1.4 Application Manifest

An application manifest is created and should be owned by the Application Provider. An OP that instantiates an application from the application manifest should request the manifest from the Application Provider. This requirement implies that other OPs should be able to request the application manifest from the OP.

The application manifest shall contain mandatory data elements and may include optional data elements. A data element may be described by a separate sub-model below (e.g., the QoS specification for an application is a sub-model).

An application manifest describes various properties of the application, including but not limited to the following properties:

61. **Executable Image**

    A URI (or another similar name) of the executable image (or container) to be installed and executed by the OP.

62. **Resource Flavour**

    A Flavour is a description of a set of resource requirements used by an application instantiation. It should have a name that can be used to identify the description uniquely and that should be global across OPs in an OP system.
    A resource description should be consistent with those appearing in Flavours available in public clouds. This means that a Flavour should specify CPU, memory, storage, I/O bandwidth, CPU architecture, special hardware (e.g., accelerators).
    A Flavour definition ensures that, if an Application Provider selects a Flavour for a manifest, the application will run successfully if provided with at least the resource described in the Flavour.
    Flavours are not standardised (at this time) in this document. The OPs in the federation should collectively undertake to produce and maintain a Flavour catalogue.
    The resource flavour includes the following properties:

    - **Computing Resource**
    - **Storage Resource**
    - **Network Resource**
    - **Extension resource.**

63. **QoS Requirements (optional)**

    A QoS description is a characterisation of traffic between an Application Client and an Edge Application and carried by a flow between the client and backend. A QoS description allows an Application Provider to describe the physical constraints in an edge network that should be met for the application to run successfully and provide a correct Quality of Experience (QoE) for the end user at the UE.
    The QoS requirements include the following properties:

    - **Bandwidth**, bidirectional data rate between UE and backend application, measured end-to-end with “loopback” application;

- **Latency**, round trip delay between UE and backend application, measured end-to-end with “loopback” application;
- **Jitter**, Variance of round-trip delay between UE and backend application, measured end-to-end with “loopback” application.

64. **State Property (optional)**

    The NBI allows an Application Provider to specify their support for a stateful or stateless Edge Application, i.e. whether the Edge Application cannot or can be moved from one edge compute resource to another and this with or without prior notification.

65. **Deploy Model (optional)**

    The NBI allows an Application Provider to specify whether its Edge Application (s) are pre-deployed (based on the Application Provider’s requirements and OP deployment criteria); or whether an Edge Application is deployed, triggered by activity from Application Client(s).

66. **Application Scheduler Policy**

    A scheduler policy is a scheduling strategy representing a mechanism, comprehensively considering factors such as the location of the edge site, the location distribution of the accessed UE, latency requirement, and application resource requirements, assigning application instances to one or more edge nodes. The NBI shall support setting scheduler policy, based on the Application Provider’s criteria, when creating an application instance and the ability to switch to another scheduler policy when it is necessary.

67. **Edge Application Migration Policy**

    Defines a policy when an Edge Application may be moved from its current operator network or current geographic region (i.e., without violating GDPR).

68. **Other Restrictions (optional)**

    There are several further aspects that the Application Provider wants to signal about:

    - Data privacy (GDPR) restriction on the geographical area
    - Service availability on visited networks (roaming): two possibilities: required or not. And maybe: all visited networks; or selected visited networks

3.5.1.5 Application Instances Management

The Northbound interface shall support the management of application instances, including the following abilities:

69. Create application instances;

    The input parameters of an application instance include:

    a) URL for the image for the Application deployment <required>;
    b) Deployment related constraints, e.g. zone, multiplicity, etc. <optional>.

70. Update application instances;
71. Query application instances;
72. Delete application instances.

3.5.1.6 Image Management

An Application Provider deploys the application by making the program software to an image, uploads the image to an image repository and uses the URL of the image for the deployment.

The Northbound Interface shall provide the image repository to manage the image of applications, includes the following abilities:

73. **Upload images;**
74. **Update images;**
75. **Download images;**
76. **Query images;**
77. **Delete images.**

3.5.1.7 Network Event Support

An Application Provider may require to be notified by network events or may request specific information about UE, network status or information.

The NBI shall expose network information towards Application Providers and application instances so that that network capacities can be used alongside the provided edge service.

The capacities, information or services to be provided may be among the following:

- UE location information and events;
- UE network connection events;
- Application to UE connection status.

3.5.2 Southbound Interface

3.5.2.1 SBI-CR

3.5.2.1.1 General

The Southbound Interface of the OP includes all interfaces the OP is consuming from other parts of the service provider’s infrastructure to create the capabilities of the different roles described in section 3.2. The SBI includes interfaces to:

- Infrastructure manager functions of a cloud or edge cloud infrastructure (e.g. resource management for compute and network resources);
- Orchestrator functions facilitating the application and workload lifecycle management and scheduling;
- Service management functions (e.g. platform services, network services, mobility support, etc.);
- Other external modules providing services to the OP.

In many cases, close interworking between resource management, workload lifecycle management, platform services and traffic management services is needed.

The SBI is not defined by the OP, but by the systems consumed.

3.5.2.1.2 SBI Infrastructure manager functions

In most deployments, the OP will make use of cloud infrastructure management. The OP is expected to work over key industry reference infrastructures. There are various options in the industry, most based on OpenStack® or Kubernetes®, but others are also available. OP can also make use of the resource management via an orchestrator function, e.g. as defined by ETSI ISG MEC or ETSI ISG NFV. In these cases, also resource management and workload management are consumed via the orchestrator function.

The SBI is defined here via the interfaces produced by the consumed systems.

In addition to the management of the virtualised resources, also hardware infrastructure needs to be managed via the SBI.

The picture below illustrates some possible SBI-CR integrations between the OP and the cloud resources.

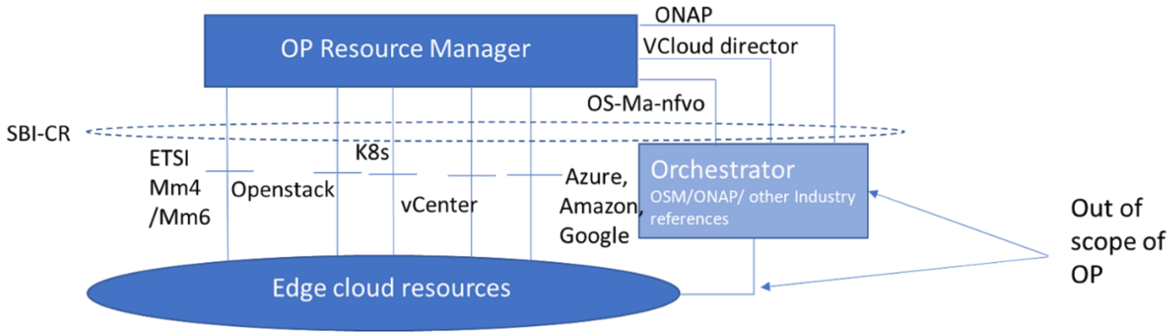

Figure 5: Possible SBI CR integrations

The SBI-CR is expected to reuse current industry standards and connectors. At this stage, no specific required enhancements have been identified.

3.5.2.1.3 SBI Orchestrator functions

Lifecycle management for applications or other workloads can be implemented by internal modules of the OP or externally, e.g. consuming ETSI ISG MEC or ETSI ISG NFV via the SBI.

3.5.2.2 SBI-NR

3.5.2.2.1 Network

The Network Exposure APIs on the SBI-NR, optionally, can help OP to obtain various mobile core network information of a UE and may enable OP to perform some of the tasks as given below:

- UC location information retrieval;
- Requesting specific Quality of Service (QoS);
- Applying local routing and traffic steering rules for Local BreakOut (LBO) of MEC traffic;
- Application relocation on most adequate edge nodes;

- Managing service availability in Local Area Data Network (LADN);
- Influencing Data plane attachment point (re)selection for service continuity;
- Collecting radio network information, e.g. cell change notification, measurement reports etc. for mobility decisions;
- Supporting applications creation in a given network slice.

Some of the functions, namely location info retrieval or requesting specific QoS, can be performed in a 4G network, while others are introduced in 3GPP Release 15. They will be guided by the further developments in the specifications in future revisions.

The functionalities mentioned above are optional, and an OP implementation can choose to use the available interfaces to optimise the platform functionalities.

The above list is not exhaustive but indicates some of the main informational elements and functions which an OP is expected to perform. The SBI-NR interface enables the Service Resource Manager Role in an OP to meet the required Service Level Agreements (SLA) agreed with the external actors like Application Providers, and may help to optimise the utilisation of available network resources in a mobile operator network.

The mobile core network may provide all, or a subset of, the above information via the SBI-NR APIs to the OP. In a 5G mobile core network, the OP, in the role of an Application Function (AF), may communicate with the 5G Core (5GC) network over the standardised interfaces as defined by 3GPP, for example, using the services of the NEF network function.

Additionally, the OP, apart from using the SBI-NR APIs for self-decision, may also provide access to some of the APIs to authorised 3rd party applications. For example, some of the services namely Location Service, Radio Network Information Service (RNIS) as defined by ETSI MEC and available over the ETSI APIs, can be exposed in simplified abstractions to applications for offering location-aware features to end users.

3.5.3 User to Network Interface

3.5.3.1 General Requirements

78. The primary function of the User to Network interface is to enable a User Client to interact with the OP, to enable the matching of an Application Client with an Application Instance on a Cloudlet.
79. The UNI shall allow the communication between the User Client on the user equipment and the Operator Platform.
80. User Client should be capable of being implemented on User Equipment software, e.g. as an SDK or OS add-on.
81. The UNI shall allow the User Client to discover the existence of an Edge Cloud service.
82. The OP's UNI shall allow the user client registration process with the Operator Platform SRM, which entails the following:

    a) It enables the end-user device to establish an encrypted communication channel with the Operator Platform SRM.

b) Authentication and authorisation. In this document we assume that the UE attaches to the 4/5G network so that the OP can rely on AAA done by operator.
c) It enables User client's usage tracking. For example, to support integration with the network operator's billing infrastructure.

83. The OP's UNI shall allow the user client to trigger the selection of a Cloudlet by the OP.
84. The OP's UNI shall allow the user client to trigger the instantiation of an application instance on the selected Cloudlet.
85. The OP shall measure network performance metrics for tracking the average latency characteristics of the edge network.
86. Based on metrics and location information, the User Client may request, though the UNI, that the OP considers a change of Cloudlet.

3.5.3.2 Establishing Chain-of-Trust between architectural elements

The OP shall provide a mechanism to establish a chain-of-trust between:

87. the UE and the OP;
88. the User Client and the OP;
89. the Application Client and the Edge Application;
90. the operator Network and the Edge Application;
91. the end-user and the OP.

The mechanism can use the 4G/5G authentication procedure(s) to establish a chain-of-trust between the UE and the OP.

The mechanism shall use an attestation method to authenticate the UC and therefore establish a chain-of-trust between the UC and the OP.

The procedures for establishing a chain-of-trust between the Application Client and the Edge Application are implementation-dependent.

The procedures for establishing a chain-of-trust between the operator Network and the Edge Application are implementation-dependent.

The mechanism shall use a registration procedure from the UC to the OP Service Resource Manager (SRM) to establish the chain-of-trust between the end user and the OP. The registration procedure assumes that the prerequisite chain-of-trust steps described above have been successfully carried out.

Part of the registration includes authenticating the identity and location of the end user's UE, which must be done via the operator. The SRM is a service trusted by the operator network, allowing it to authenticate identity and location.

In a roaming scenario, the registration may be required to be carried out from the home network SRM.

The mechanism shall ensure security, privacy and commercial confidentiality. To support the privacy of the end user, an obfuscation technique, such as opaque tokens, shall be used.

Additional services may be created to return metadata associated with a User Client. These services may have a chain-of-trust established with the SRM. If they have a chain-of-trust established with the SRM, they may require that an application using them also establish a chain-of-trust.

An example of such a service is “verify location”. The input of "verify location" shall be a nominal physical location and a geographical bound (precision) around that location. The output of the API shall be an indication of "user is in that area" or "user is not in that area". An example usage of this service is to allow an Edge Application at a retail location to verify that a user is close enough to a physical location to be worthwhile pushing a notification to the user’s application client.

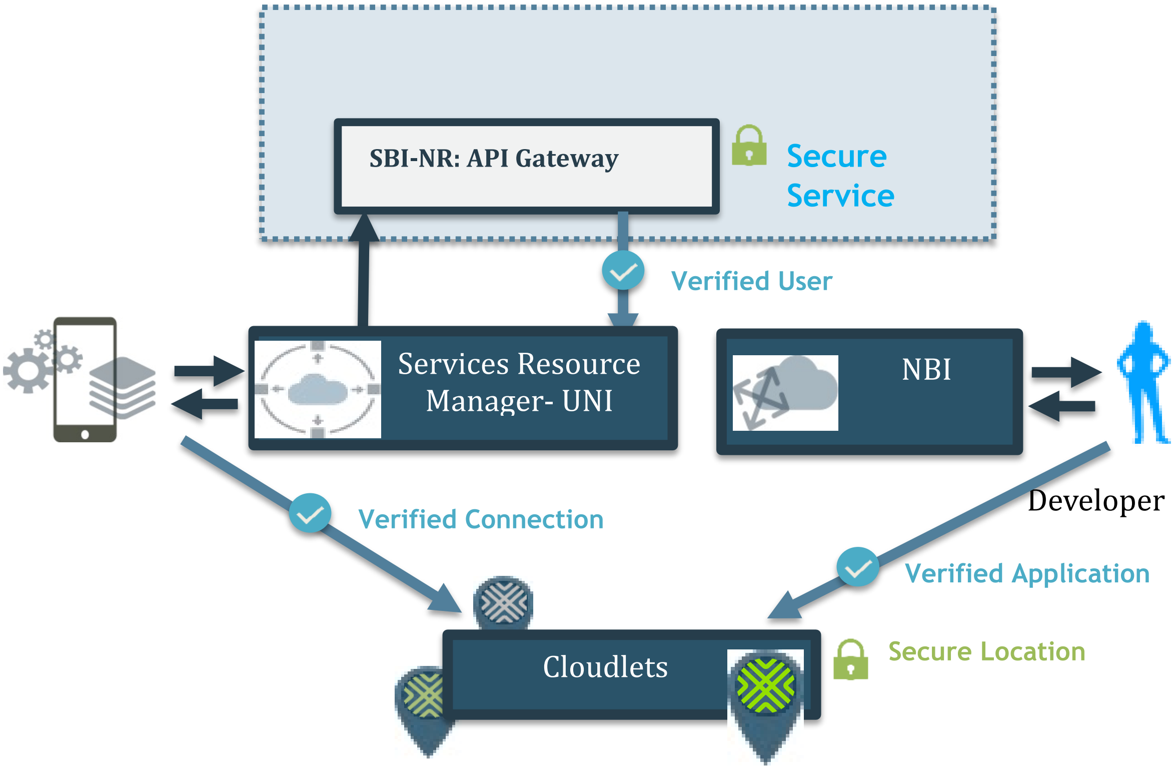

Figure 6: SRM as a trusted service: High-level Diagram

3.5.4 East/Westbound Interface

The E/WBI connects partner OP instances with the primary goal of allowing Application Providers of an OP to utilise the Edge Cloud services of another OP.

The E/WBI is not exposed to the Application Providers and is primarily driven by the Federation Manager functionality within the OP.

The following sections provide a list of services that would be executed on the East/West Bound Interface.

3.5.4.1 East/West Bound Interface Management Service

The East/West Bound Interface Management Service shall be used for setting up and maintaining the East/West Bound interface between OPs.

The service would include APIs for the following:

- Setup of the East/West Bound Interface between OPs;
- Update parameters of the East/West Bound Interface;
- Heartbeat/Keep-Alive of the East/West Bound Interface;
- Termination of the East/West Bound Interface.

3.5.4.2 Availability Zone Information Synchronisation Service

The Availability Zone Information Synchronisation Service shall be used to share and update the Availability Zone specific information corresponding to the Edge Cloud resources of one OP with another.

The Availability Zone information shared over E/WBI shall provide to a partner OP information about which zones are shared with that OP, where do they provide coverage and what amount and type of compute they provide.

The service would include APIs for the following:

- Fetch Availability Zone information of a partner OP via the E/WBI;
- Add Subscription over E/WBI for Availability Zone information update notifications;
- Delete Subscription over E/WBI for Availability Zone information update notifications;
- Update Subscription for Availability Zone information update notifications;
- Notifications for Availability Zone information update (including information about Operational and Administrative states).

3.5.4.3 Application Management

3.5.4.3.1 Application Onboarding Management Service

An OP shall use the Application Onboarding Management Service over E/WBI to onboard applications towards another OP.

The onboarding service shall include the following:

- Transfer of application images and Application Provider criteria towards a partner OP. The procedure may optionally also request the launch of application instance(s) in partner OP edge clouds as a follow-up action after onboarding.
- Publishing of application information to support Edge Node Sharing scenario (as described in Section 3.5.4.3.2).

The Application Onboarding Management Service shall include APIs over E/WBI for the following:

- Submission of applications (application images, application type, Application Provider criteria, target availability zones) towards a partner OP.
- Removal of applications (application images and metadata) from a partner OP.
- Update of application information towards a partner OP (e.g. application versions, Application Provider criteria, target availability zones).

3.5.4.3.2 Edge Node Sharing Enablement Service

Edge node sharing is a scenario wherein an OP, when serving the UNI requests originating from (its own) UCs, decides to provide the application from the Edge nodes of a partner OP (where the application is available). This decision may be due to individual policy controls of the Operator, because of the specific Application Provider restrictions, or due to constraints originating from the federation agreement between the Operators.

An E/WBI service is required to support the publishing of application information, and availability zone information, to enable specific applications to be served from a Leading OP’s Edge Clouds in the following scenarios:

- In a roaming scenario where local breakout (i.e. data plane access to Edge Cloud resources in visited network) is not available, the applications may need to be served from the home OP for consumption by visiting UCs;
- In a non-roaming scenario where an OP needs to allow, for its own UCs, the consumption of the applications published by a partner OP that must be served from that partner’s Edge Clouds.

This service shall include the following API(s) over E/WBI:

- Application publishing API, including application meta information (including information about the policies controlling application distribution restrictions) and availability zone(s);
- Unpublish API, to cancel already published application(s).

3.5.4.3.3 Application Deployment Management Service

The Application Deployment Management Service shall be used by an OP to control launch and termination of applications that have been onboarded on a partner OP.

The Application Deployment Management Service shall include APIs for the following:

- Instantiation of applications based on Application Provider criteria in select Partner OP Zones;
- Termination of running application instances from select partner OP Zones.

3.5.4.4 Events and Notifications Service

The Events and Notifications Service shall be used to set up, send and receive Events and Notifications from one OP to another over the E/WBI.

As indicated under the Availability Zone Information Synchronisation Service, each OP publishes towards its partners the information about the resource levels being provided to each partner. An OP shall send Notifications to partner OPs related to these published resources, for example in the following scenarios:

- The availability state of these resource changes;
- The consumption of resources reaches a pre-defined threshold (e.g. warning notifications when 80% of the agreed threshold value is utilised);
- Federation Agreement expiry imminent.

To enable this, the Events and Notifications Service provides the following APIs over E/WBI:

- Setup Event reporting (e.g. resource threshold levels);
- Update Event reporting parameters;
- Notifications for Events.

3.5.4.5 Service Availability in Visited Network Management Service

This service shall be used to support information exchange between the OPs to enable service availability for UCs in the visited network.

Information elements that need to be shared over E/WBI to support this scenario include:

- Discovery Service URL for a partner OP.
- Authentication information for User Clients.

This service shall include the following APIs over the E/WBI:

- Setup Service Availability in Visited Network related parameters towards partner OPs;
- Update Service Availability in Visited Network related parameters towards partner OPs;
- Fetch UC authentication information for a visiting UC from the home OP.

4 Functional Definition and Service Flows

Service Flows related to the function definition will be provided in a future version of this document.

5 Proposed Requirements on interfaces and functional elements

This section defines the requirements that the interfaces and functional elements that make up the architecture should comply with to cover the different use cases that an OP should provide. They should be fulfilled by solutions developed in SDOs and implementations provided by the open-source community.

5.1 Interfaces

5.1.1 Northbound Interface

5.1.1.1 High-level requirements

92. All Operators and Operator Platforms shall offer the Edge Cloud service through the same NBI.
93. The NBI shall offer the capabilities of the Edge Cloud to Application Providers, in particular:

    a) a low latency service (and perhaps other application QoS metrics) in a geographical region;
    b) Edge Cloud capabilities are offered whatever operator the UE is attached to.

94. In deployment, the NBI shall use profile-based access control to provide appropriate restrictions on the amount of functionality that the NBI offers to a particular system or person, according to the operational profile. Profile-based access control, for

example, RBAC, Role-Based Access Control, restricts the degree of access depending on the person's (or system's) defined privilege and role.

Note: Not all profiles will have access to all the functions listed below - for example, monitoring information would not necessarily be accessible during onboarding; and the detail of monitoring information may depend on the operational profile (for example first-line vs second-line support).

Note: The text below is split into two broad types, but in practice, there is likely to be more granular profiles.

5.1.1.2 Onboarding and Deployment Profile

5.1.1.2.1 General

When an Application Provider accesses the OP portal or uses the OP's NBI APIs to deploy their application, the OP shall get in charge of:

- receiving the request,
- authorising/authenticating the Application Provider, and
- gathering all the necessary data to deploy (onboard and instantiate) the application in the most appropriate edge nodes meeting the Application Provider's request.

The deployment management thus shall allow to onboard and instantiate the application meeting different criteria, sourced by Application Providers as well as the operators owning the OP instance and the underlying resources.

5.1.1.2.2 Application Provider Criteria

The platform shall be able to support the following Application Provider requirements:

95. Footprint/coverage area selection;
96. Customer reach/ operator selection;
97. Infrastructure resources:

    a) CPU;
    b) Memory;
    c) Storage;
    d) Networking definition used by the application.

98. Specific requirements definition:

    a) Use of GPUs.

99. Edge-Cloud requirements:

    a) Latency;
    b) Jitter;
    c) Bandwidth;
    d) The relevant geographical area for data privacy purposes.

100. Type of application instantiation:

a) Static: the application shall be deployed in several edges based on Application Provider's requirements and the operator's deployment criteria. The application shall be deployed upfront (independently of the UC's request).
b) Dynamic: when a UC request the use of an application, the application shall be deployed in the selected edge location (triggered by UNI request(s)).
c) Based on capacity: criteria to define if there will be an instance per user or one instance per specific number of users.

101. Policies that allow the Application Provider to manage circumstances where user conditions do not comply with the deployment criteria.
102. Support for telemetry information from the operator.
103. Policy control concerning support of stateful and stateless applications.

The Application Provider shall be able to indicate that:

a) Its Edge Application cannot be moved from one edge compute resource to another;
b) Its Edge Application can be moved from one edge compute resource to another, without any notification;
c) Its Edge Application can be moved from one edge compute resource to another, with prior notification.

104. Service availability in visited networks required/supported.

5.1.1.3 Management Profile

The OP shall offer a uniform view of management profile(s) to Application Providers:

105. The OP shall enable application developers to request Edge Cloud in an Availability Zone (within the OP and federated OPs):

a) On a basis where the application developer reserves resources (on a relatively long-lasting basis) ahead of their usage.
b) On a basis where resources are allocated as the application instance needs them (“reservationless” or “dynamic”) and the application developer selects the degree of scaling it requires (for example, number of sessions).
c) On a basis where resources are isolated from those used by other application developers.
d) An application developer may provide the OP with information about its estimated workload, to help the OP optimise the deployment of Edge Application(s).

106. An OP shall offer a range of quality policies so that a developer can choose the performance that their application requires. These policies are defined based on objectively measured end-to-end parameters that include performance aspects of both the network and the Cloudlet, such as latency, jitter and packet loss (measured as average statistics).
107. The NBI shall enable a request-response mechanism through which the developer can state a geographical point where a typical user could be, and be informed of the mean latency performance that is expected.
108. The OP shall describe the capabilities of the Edge Cloud, for example, :

a) The geographical zones where it is provided
b) The type and “granularity” of edge cloud and network service (typically generic Compute, memory, storage and specialised compute (such as GPU and future types of resource).

Note: Optionally, an OP may present types of resource and their attributes as “flavours”. Flavours are intended to be a useful “shorthand” for Application Providers, but are optional and do not have to be used.

Note: if a federation of OPs uses flavours, then they should agree on common definitions.

Note: the NBI shall not reveal the exact geographical locations of individual Cloudlets, and shall not allow an application developer to request deployment of its application on a specific Cloudlet.

109. The OP shall offer a structured workflow for application deployment and instantiation: CRUD functions.

110. The OP shall allow a developer to specify that its Edge Applications are restricted to Cloudlets in a particular geographical zone. This restriction would ensure compliance with the applicable data privacy laws.

111. The OP shall allow an Application Developer to specify whether or not it requires service availability on visited networks (that is, when a UE roams away from its home network operator).

112. The OP shall provide an Application Developer with telemetry information concerning the performance of the Edge Cloud service, including fault reporting.

113. The OP shall allow an Application Developer to request a particular granularity for the telemetry information that they receive.

Note: Possibly using a publish-subscribe approach.

Note: Different operational profiles will require different granularity about the telemetry information (how fine-grained and how often).

114. The OP shall allow an Application Developer to require that outbound access to the internet is prohibited.

115. The OP shall offer Application Providers a registry where they store their application images and can update or delete them. The registry may be centralised or distributed, depending upon the Application Provider’s needs to reduce boot time and recovery.

116. The OP shall support Single Sign-on based on login credentials for an Application Provider.

117. The OP shall offer functionality that supports the application developer to manage its application instances. For example, to monitor operational performance, get diagnostic logs and help with debugging.

5.1.2 East-Westbound Interface

5.1.2.1 High-level requirements

118. The E/WBI is universal, meaning that all Operators and Operator Platforms provide Edge Cloud to each other through the same E/WBI.

5.1.2.2 Security Requirements

OP instances are intended to belong to different operators/players, so special requirements shall be considered for managing the relations and the resources/information sharing.

119. The E/WBI shall maintain the topology hiding policy between operators/players.

    a) Resources shall be published as “edge resources” entities, referred to a specific zone served by one or more edge servers/nodes.
    b) Specific edge node information shall not be shared.

120. An OP shall only expose the resources previously agreed with each specific federated instance.
121. It shall be possible to identify the User Clients among OP instances.
122. It shall be possible to identify the Application Providers among OP instances.
123. It shall be possible to identify the applications among OP instances.
124. An OP shall be able to act as a proxy for any interaction between operators’ networks, hiding any detail on the network architecture of both federated networks.

5.1.2.3 Application Management

The federation interface needs to replicate the behaviour and functions available on the NBI to transmit the application load, requirements, mobility decisions and policies across all the operators’ instances required to deploy the application.

125. The E/WBI shall be able to forward the instantiation requests to any federated OP whose footprint has been selected to be covered.
126. An OP receiving an instantiation request through its E/WBI shall get in charge of the management of the application:

    a) An OP receiving an instantiation request through its E/WBI shall be able to apply its own policies and criteria for processing the request and managing the application.
    b) An OP receiving an instantiation request through its E/WBI shall get in charge of the operator deployment criteria management.
    c) An OP receiving an instantiation request through its E/WBI shall get in charge of the edge node selection based on the application criteria and its own operator criteria.
    d) An OP receiving an instantiation request through its E/WBI shall get in charge of the application mobility management.

127. The E/WBI shall forward the application mobility notifications and procedures towards the Leading OP, to be managed with the Application Provider.

128. The E/WBI shall forward the management procedures, information and statistics to be shared with the Application Provider on the Leading OP.
129. The E/WBI shall be employed for managing the service continuity on visited networks.

5.1.3 Southbound Interface to Network Resources

5.1.3.1 General

The SBI-NR connects the OP with the specific operator infrastructure that will deliver the network services and capabilities to the user.

When an end user accesses an edge service from a network, the OP shall be able to reach some basic network capabilities through the SBI-NR interfaces of the operator. However, an operator need not implement the NEF/SCEF interfaces, in which case these capabilities have to be achieved in some other way, or else may not be available.

OP integration to network resources shall allow:

- The OP to authenticate and authorise the end users to access the services in the home and visited network scenarios.
- The OP to access location information of the end users in the network.
- The OP to access policy control capability exposed by the network, e.g. for charging or quality of service handling.
- The OP shall be made aware of the data connection status (e.g. if a user has a data session or not).
- The home network OP shall be the only entity able to access home network resources.

5.1.3.2 OP integration to 5G Core/4G Core via Exposure Functions

5.1.3.2.1 Introduction

The NEF/SCEF APIs [4] [5] are a set of APIs defining the related procedures and resources for the interaction between NEF/SCEF and AF/Services Capability Server (SCS). The APIs allow the AF/SCS to access the services and capabilities provided by 3GPP network entities and securely exposed by the NEF/SCEF. Some APIs are applicable for both 5G Core and 4G Core.

Figure 7 shows a functional mapping that describes how an OP accesses features and services that are exposed by the NEF/SCEF.

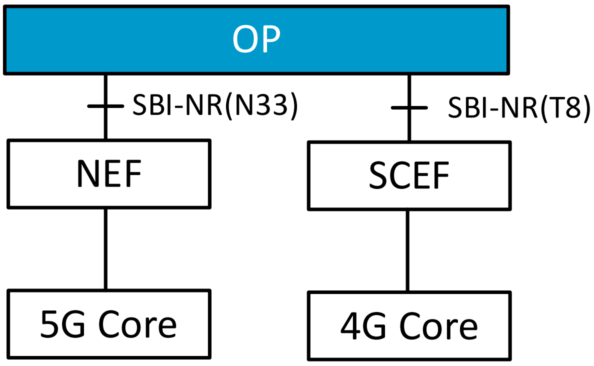

Figure 7: Functional mapping between OP and NEF/SCEF

General Requirements

130. An OP's SBI-NR shall be able to interact with 5G Core/4G Core via the NEF or SCEF to access network capabilities.

5.1.3.2.2 131. An OP's SBI-NR shall support the exposure interface [4] [5] for interacting with the 5G Core/4G Core.

132. If the NEF/SCEF returns an error response to an OP's SBI-NR, the OP's SBI-NR shall perform error-handling actions.

133. An OP's SBI-NR may be able to configure the user traffic to be routed to the applications in the local Data network with NEF API [4].

134. An OP's SBI-NR may be able to collect the information of network congestion or access concentration on a specific area with NEF API [4] or SCEF API [5].

135. An OP's SBI-NR may be able to retrieve UC location information with NEF API [4] or SCEF API [5].

136. An OP's SBI-NR may be able to control the transfer of data in the background for UCs with NEF API [4] or SCEF API [5].

137. An OP's SBI-NR may be able to configure QoS session parameters to communicate with a UC with a guaranteed level of QoS (e.g. low latency, priority, maximum bandwidth) with NEF API [4] or SCEF API [5].

138. An OP's SBI-NR may be able to configure service-specific parameters for UCs (e.g. network slice) with NEF API [4].

139. An OP's SBI-NR may be able to initiate a device trigger to a UC for performing application-specific actions (e.g. starting communication with the OP's SBI-NR) with NEF API [4] or SCEF API [5].

5.1.4 Southbound Interface to Cloud Resources

5.1.4.1 Cloud Resources Management

The integration with cloud resources APIs on SBI allows OP to support the needed functionalities for application and resources management.

The Operator Platform shall be able to access cloud resources of the operator/cloud provider. The OP shall be able to do this not only for fulfilling request/response transactions regarding an application's lifecycle but also to be able to catalogue the resources/capabilities
5.1.4.1.1 and get feedback about the status of the different Cloudlets or edge nodes.

Integration with Cloud Orchestrator

A cloud provider/operator may want to expose the cloud resources through an orchestrator. This integration will not expose the whole set of functionalities that an Operator Platform may need to provide. In this case, only a serverless approach would be available where the provider’s orchestrator performs the instantiation of the application based on the request from the OP, instead of the OP taking up the responsibility of the application Life-Cycle Management (LCM).

With this orchestrator integration, an OP shall be able to integrate with the orchestrator for:

- Application onboarding/instantiation on specific edge/cloud site (Cloudlet);

- Image management;
- Application lifecycle management;
- Limited resources management;
- Retrieval of limited resource usage statistics for settlement.

The capabilities exposed by the Orchestrator will not allow the OP to enlarge or reduce the resources reserved for edge purposes. Furthermore, the limited information provided will not enable the OP to ensure an application's instantiation until the orchestrator performs the internal infrastructure procedures. These limitations endorse the serverless approach of this integration.

The resource management and the statistics offered by an orchestrator to an OP is limited to the amount of resources and the scope of the assigned orchestrator's tenant.

OP SBI-CR integration shall be able to adapt industry standards for orchestrator integration, including but not limited to OSM/MANO, ONAP, VMware VCloud Director.

5.1.4.1.2 Integration with Infrastructure Manager

If the integration with the cloud resources is done directly using the Virtualised Infrastructure Manager (VIM) or Container Infrastructure Service Manager (CSIM), an OP will have additional functions. These functions will include, for example, resource management, reservation and detailed statistics, resource catalogue and load reporting.

An OP having direct access to cloud resources can support more functions than an OP accessing the resources through a Cloud Orchestrator. These additional functions include infrastructure exposing to Application Providers, analytics retrieval from the Cloudlets for the instantiation selection procedures, resources scaling based on traffic.

With direct VIM/CISM integration, an OP shall be able to integrate with an infrastructure manager for:

- Application onboarding/instantiation on specific edge/cloud site (Cloudlet);
- Image management;
- Application lifecycle management;
- Resources management;
- Retrieval of resource usage statistics for settlement;
- Resources/Services catalogue retrieval;
- The catalogue shall include the availability of, at least:
  - Edge site identification;
  - Location;
  - CPU;
  - Memory;
  - Storage;
  - GPU;
  - NPU/FPGA;
  - I/O;
  - Cloudlet load reporting.

The OP SBI-CR integration shall be able to adapt industry standards for VIM/CISM integration, including but not limited to ETSI-MEC (Mm4/Mm6 interfaces), Openstack, Kubernetes and VMware vCenter.

5.1.4.1.3 Integration with Hyperscalers

When using a hyperscaler as a cloud infrastructure provider, the OP shall be able to support the APIs that those providers currently expose.

The OP shall be able to access to the same capabilities that are enabled to Application Providers through those interfaces, in an IaaS/PaaS manner that provides the full set of needed functionalities, limited to the offered amount of resources that the hyperscaler provides.

5.1.5 User to Network Interface

5.1.5.1 High-Level Requirements

140. The UNI shall be universal, meaning that the Application Provider does not have to modify its applications for different Operators or OPs.

141. The UNI between the User Client (typically located in the UE) and the Operator Platform should be kept to a minimum and not overlap with, or have an impact on, the existing UNI interfaces:

a) between the application client and the Application Provider;
b) between the mobile UE and the operator.

142. In this proposal, we assume that the UE attaches to a trusted network (such as the 4/5G network) so that the OP can utilise AAA services provided by the operator. Where the UE accesses via an untrusted network (such as public Wi-Fi), then the OP needs to undertake its own AAA services.

5.1.5.2 User First Attachment

5.1.5.2.1 General

When a UC requests access to an Edge Application the OP receiving the request shall authorise/authenticate the user as well as the requesting application. Once the OP has authorised the request, it gather all the necessary data to redirect the request to the most suitable edge node. UC connectivity should be available to allow initiating this request. UC connectivity is out of the scope of this document.

5.1.5.2.2 Edge Cloud service discovery

The UC shall be able to reach the OP, so that it can request Edge Cloud services, using the UNI:

143. An OP shall expose a connection reachable by any customer on the operator network.

144. An OP shall offer a general URL that can be constructed based on operator information available to the UE, e.g. MCC/MCN, to which a User Client can request an Edge Cloud service.

145. A UNI UC request shall include identity information and parameters:

a) UE ID, e.g. MSISDN, GPSI;
b) Application ID;
c) Location, e.g. cell-ID, TAI. This information does not need to be included, in the case where the OP knows the UE’s location.

5.1.5.2.3 Authentication and Authorisation

The OP shall authenticate the UC and authorise the application request received through the UNI:

146. Where the UE is attached to the 4/5G network, then the OP may optionally rely on user authentication by the operator.
147. Otherwise, the OP shall interact with the network authentication elements, for instance, Authentication, Authorisation and Accounting (AAA) or Application Authorisation Framework (AAF), to authenticate the UC.
148. An OP shall authorise the usage of the application by the UC, for example checking that the particular application is part of the user’s ‘package’. The OP shall provide a mechanism, such as a token, to allow efficient authorisation of subsequent interactions.

5.1.5.2.4 Cloudlet selection

The OP process all the information from the UC, network and application requirements to select the most appropriate Cloudlet where the Edge Application is deployed:

149. An OP shall be able to obtain the UE's location by SBI interaction to operator core network elements, e.g. Gateway Mobile Location Centre (GMLC)/Access and Mobility Management Function (AMF)-NEF, and also obtain the UPF /PGW associated with the UE.
150. An OP shall select an appropriate Cloudlet that:

a) depending on the actual UE’s location (See 149. above), and/or the geographical zone that the Application Developer has previously determined where its Application Clients will be,
b) satisfies the Application Developer’s statement about the requirements for data privacy,
c) meets the Application Developer’s input on requirements for QoS, and/or the User Client’s selection of QoS (including bandwidth and latency),
d) Takes account of the capacity and usage of the Cloud Resources (e.g. CPU and memory) at the various Cloudlets and the Network Resources (e.g. congestion),
e) The choice of Cloudlet may result in the UE needing to be redirected to a different UPF /PGW.

151. An OP shall request, through the SBI, the application to be available on the selected Cloudlet.

5.1.5.2.5 Service Provisioning

The OP shall enable the requested Application and provide over the UNI the parameters and configuration needed so that the Application Client can connect to the selected Cloudlet:

152. If necessary, the OP shall deploy the application image and create an instance on the selected Cloudlet,
153. The OP shall inform the application client of how to reach the Edge Application on the Cloudlet chosen (for example, a URL or IP address),
154. The UE shall be able to test the connectivity characteristic towards the selected Cloudlet.

5.2 Functional Elements

5.2.1 Capabilities Exposure Role

Detailed requirements on the Capabilities exposure role will be provided in a future version of this document.

5.2.2 Resource Manager Role

5.2.2.1 Network/Operator Criteria

When several edge nodes meet the Application Provider criteria, and to support operator policies, the platform shall be able to support the following operator requirements to select the edge where to deploy the application:

155. Edge weight matrix, for determining the importance of each requirement on the final selection decision.
156. Edge node load.
157. Network load.
158. Network usage forecast.
159. Edge usage forecast.
160. Application availability (already deployed/onboarded on edge node).
161. UE mobility supported.
162. Network mobility supported (integration with data packet core).
163. Specific constraints/barring for users, application or edge nodes selection.
164. Specific considerations to abide by commercial agreements between involved parties.

5.2.2.2 Instantiation Strategy

Considering the Application Provider requirements and policies, and the operator restrictions and preferences over the application instantiation, the OP shall be able to request instantiation over the edge resources:

165. An OP shall be able to request the static instantiation of the application on a specific edge node.
166. An OP shall be able to request the static instantiation of the application on all the available edge nodes.
167. An OP shall be able to determine the minimum amount of edge nodes to select for covering the footprint and onboarding requirements.
168. An OP shall be able to request dynamically the instantiation of an Edge Application based on a user's request.

5.2.2.3 Mobility Management

5.2.2.3.1 General principles for mobility management

In the context of this document, mobility management deals with the movement of the Edge Application from one edge compute resource to another, and a change of the application client's IP address, port or both. These may happen together or independently.

As general principles:

- The operator is responsible for mobility management of the UE (end user's device) (through standard 3GPP mobility management mechanisms);
- These standard mobility management mechanisms may involve a change in the IP address used by the application client – the operator informs the application about such a change.

  Note: the application cannot reject or delay the change.

- Because of this UE mobility, or because of the OP's measurements or knowledge, or hints from the application about performance degradations, the OP may decide that a different edge compute resource can better host the Edge Application.

  Note: In this section, the term "OP" is intended to leave open which party(s) within the OP does something.

  Note: the term "application" in the bullet point above is intended to leave open which part of the application is involved (Edge Application, application in the central cloud etc).

- The OP should be cognisant of the policy indication from the Application Provider about its sensitivity to a change of the edge compute resource hosting the Edge Application.
- In the case where the policy is that a change of edge compute resource can be done with a prior notification: the OP decides that a change of edge compute resource is needed and selects the new edge compute resource, whilst the application chooses the exact timing of the move and is responsible for the transfer of application state from one edge compute resource to another.
- During a period when a non-optimal edge compute resource is used, the service provided by the OP may be at a lower quality, or even have to be ended.
- From a requirements perspective, mobility management includes support for a change of operator and OP.

5.2.2.3.2 Mobility triggers

Many different elements shall monitor and control the end-to-end service delivery for detecting any modification and trigger a change on the path:

169. Mobility triggers from the OP:

   a) Related to the movement of the UE which causes a change in the application client's IP address;

b) Related to the movement of the UE (for instance, for each Edge Cloud location, the operator identifies the set of base stations that it most naturally supports);
c) Related to lifecycle management of its edge compute resources(for example, the overload of an edge compute resource, a failure or planned maintenance, a new or expanded edge compute resource, an issue with the network for its edge compute resource);
d) Related to usage forecasts about its edge compute resource and network;
e) Related to its measurements of application performance.

Note: this seems less likely, as it is hard for the OP to measure application-level performance accurately, but some simple measures such as packet drops may be possible.

170. Mobility triggers from the application:

a) Related to its measurement of QoS parameters (such as latency, jitter and bandwidth);
b) Related to its measurement of application-level QoE parameters;
c) The application should note that QoS and QoE might temporarily degrade in a mobile network, due to the UE having inadequate radio coverage (i.e. unrelated to the Edge Cloud service).
d) The application should not over-report mobility triggers.

Note: it is left open which part or parts of the application are involved in this (application client, Edge Application, application in the central cloud)

5.2.2.3.3 Application Conditions/Restrictions

The Operator Platform shall be able to consider the application-specific requirements for managing mobility over different edge nodes.

171. An OP shall get in charge of managing the application mobility for all the edge services associated with each UC.
172. An OP shall consider the mobility sensitiveness of the applications.
173. An OP shall take into account the active Edge Application on the UC for considering the mobility.

a) An OP shall ensure that all the active Edge Applications are moved correctly when network mobility is required.
b) An OP shall not perform a network relocation in case an active application does not support mobility.
c) An OP shall perform a network relocation if an application requires mandatory mobility.

5.2.2.3.4 Application Mobility (Server-Side)

The OP needs to get in charge of managing the reconfiguration of the Edge Application environment, selecting a new edge node to have the application available.

174. An OP shall be able to ensure that the selected edge node has enough capacity.

175. An OP shall be able to request the instantiation of the Edge Application on the target edge node, if not previously available or if the capacity is not enough.
176. An OP shall ensure that the resources are released on the original edge node.

5.2.2.3.5 Session Mobility (User Side)

Application session mobility is mandatory for maintaining the session continuity on stateful applications, where the Edge Application moves from one edge compute resource to another. This section concerns cases where the Application Provider has indicated as part of the initial policy phase that it requires notification in advance of a change of which edge compute resource hosts the Edge Application.

177. An OP shall be able to notify the application about the forthcoming mobility procedure if required.
178. An OP shall inform the application on what it needs to know so that the application can move the application-related state from the old edge compute resource to the new one.
179. The application indicates to the OP when it is ready for the move to the new edge compute resource. This approach means that the application is generally in charge of the timing of the movement (since it knows best, for example, when the end user's experience of the application will be least affected). Note that KPIs may be suspended during this period.
180. The application may indicate that it cannot currently handle mobility. The OP shall be able to cancel the mobility procedure. Note that the service may be degraded or even lost. Note also that as part of the initial policy phase the application may give a permanent indication that it cannot handle mobility.
181. The application shall confirm the completion of the mobility of the Edge Application onto the new Cloudlet to the OP.
182. Movement of the UE may require that the operator changes the IP address used by the application client.
183. The operator shall notify the application about a change of IP address

5.2.2.3.6 Mobility Enforcement

184. An OP shall be able to request a network GW relocation (if possible), based on location and network statistics.
185. An OP shall be able to request an Edge Application relocation, based on application requirements and different information, e.g. network and physical location or edge resources usage.
186. An OP shall be able to request an application session relocation, based on the application requirements.
187. An OP shall be able to handle the previous relocation requests, ensuring the service and session continuity.

a) The OP shall coordinate the different procedures with the Edge Application.
b) The OP shall coordinate the different procedures with the Edge Application, from the original node to the target.
c) The OP shall coordinate the different procedures with the application client on the UC.

188. An OP shall ensure that the UC is enforced to apply the mobility procedures.

189. Network GW location may not be needed in case of service degradation due to an edge node saturation, for instance.

5.2.2.4 Service Availability on Visited Networks

5.2.2.4.1 General

For allowing the user clients to enjoy edge service outside of their operator network, service availability on visited networks shall be considered. This condition includes not only international situations but also inter-operator handovers, for instance when connecting to the end-user's home Wi-Fi network which may be provided by a different operator.

With no service availability interaction, the edge service would be delivered from home network resources, with the inherent latency and service degradation.

5.2.2.4.2 Requirements

190. When a user client first attaches to a visited OP, there shall be messaging between the user client, home OP and visited OP. The purpose of the messaging is for the Home OP to authenticate the User Client and to authorise it to use the Edge Cloud on the visited OP.

   a) The messaging shall not be repeated for each application session or each application.
   b) The authorisation shall be valid for a finite period.
   c) The home OP and visited OP shall have a separate process to agree about charging /settlement for the use of Cloudlets by user clients of the Home OP. It is not intended to define a mechanism for granular charging /settlement ("granular" meaning, for example, per user client or per application instance).

191. User plane local breakout shall be available for the user client in the visited network.

   a) In case no local breakout is available, or there is no service availability agreement among operators, the user client will receive service from home resources and home OP with no visited OP interaction.

192. The visited OP may be able to obtain the application image (and any associated policies) directly from the Application Provider (typically if it has a NBI with it); otherwise, it shall request it from the Home OP via the E/WBI.

193. The visited OP, based on the information received from Home OP and the internal policies, shall instantiate the Edge Application on a Cloudlet for use by the user client.

194. The visited OP shall be in charge of which Cloudlet within the visited OP is best placed to host the Edge Application (including as the user device moves within the visited OP).

Note: User client mobility management will continue to be handled with existing mobility management mechanisms.

5.2.2.5 Operation and Management

195. The OP shall offer a centralised management plane for the operator to manage the infrastructure.
196. Create a Cloudlet at following levels

    a) Edge sites within a region
    b) Edge sites across federated operator
    c) Public Cloud Peering site

197. Capability to manage security groups and privacy policies at each Cloudlet

    a) Ability to provide isolation between workloads at run time:

198. Capability to manage the compute footprint

    a) Create, report, update, delete functions for compute, Memory, storage using the underlying IaaS stack

199. Capability to manage Availability Zones across the geographical sites within the operator's domain
200. Capabilities for the operator to monitor Cloudlet usage in terms of compute, memory, storage and bandwidth ingress and egress
201. Capability for the operator to monitor the above metrics per tenant.
202. Capabilities for automation.
203. Capability to monitor Cloudlet event, alarms logs
204. Capability to monitor Cloudlet performance metrics
205. Capability to offer operator interfaces to federated partner to monitor usage across Cloudlets

5.2.3 Federation Broker Role

5.2.3.1 Federation and Platform Interconnection

5.2.3.1.1 General

One of the Operator Platform's primary purposes is offer to customers an extended operator footprint and capabilities through interconnecting with other operators' resources and customers. This is achieved by the federation E/WBI interface; to interconnect entities of OP belonging to different operators, enterprises or others.

5.2.3.1.2 Authentication/authorisation

Federating OPs are likely to belong to different entities in different security domains. The capability to exchange authentication and authorisation between federated OPs is required:

206. There shall be a mechanism to register and authenticate different OP instances.
207. An OP shall be able to identify unequivocally any federated OP instance.
208. An OP shall be able to authorise a registration request from another OP instance.
209. An OP shall exchange a token or "federation key" on the association handshake, identifying each federation integration.

210. User authentication/authorisation shall remain independent from the OP to OP authentication/authorisation.

5.2.3.2 Settlement

Federation interfaces shall expose management and settlement data. This data will allow the charging systems of each operator to account for the services consumed.

211. An OP shall share usage statistics through the E/WBI for the services that are requested by the federated connection.

212. An Op shall provide any needed information useful for billing/settlement among operators, e.g.:

a) Type of resources used;
b) Quantity of resources employed on the service.
c) The number of application instances used.
d) The number of user sessions served.
e) Usage time of the resources.
f) Additional services employed, e.g. network location query.

5.2.3.3 Resources management via interconnection

One of the essential points to be solved through the federation interfaces is the way of sharing the Resource Catalogue between instances.

213. An OP shall be able to share (publish) the edge zones available on its footprint/resources:

a) Zone covered;
b) Specific resources, e.g. GPU, any FaaS, etc.

214. An OP shall allow the operators/resource owners to select the resources to be shared via federation.

215. An OP shall be able to push an edge zones catalogue update based on:

a) Resources specification change, e.g. adding GPU support on a zone;
b) Temporal unavailability;
c) New resources/zone availability.

216. An OP shall allow operators to request for the provision of virtualised resources on a federated OP.

5.2.4 User Client

Detailed requirements on the User Client will be provided in a future version of this document.

Annex A Mapping of Requirements to External Fora

A.1 ETSI ISG MEC

ETSI ISG MEC supports aspects of the OP architecture and some interacting blocks. In this section, the intention is to highlight where ETSI MEC plays a role in OP and areas of interest. All the documents are available for the public at ETSI site https://www.etsi.org/committee/1425-mec.

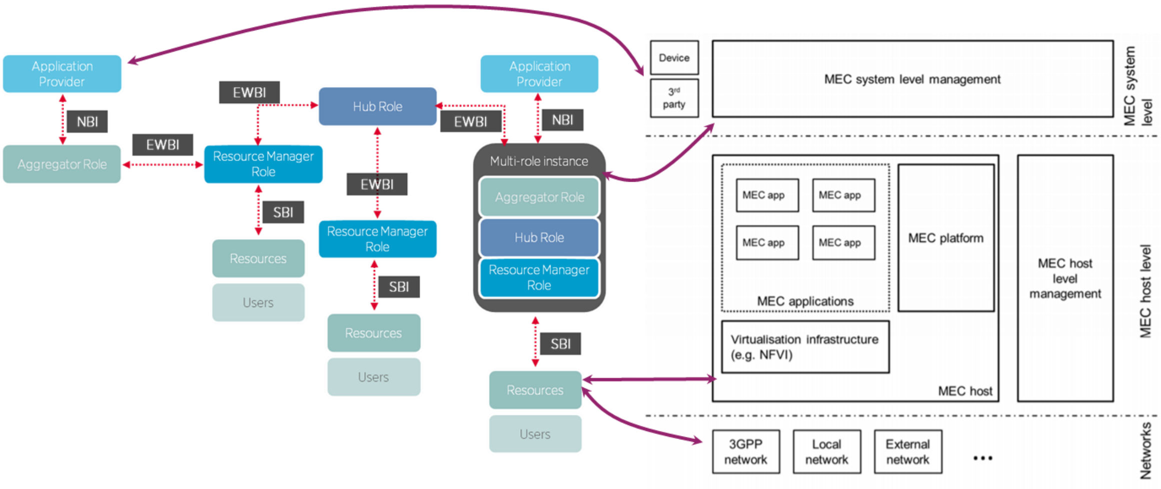

Figure 8: OP to ETSI MEC mapping

A.2 ETSI MEC specifications of interest for the NBI and the SBI

- ETSI MEC 003: The framework and reference architecture describing application placement on an edge compute resource cover certain aspects included in the NBI requirements.
- ETSI MEC 011: Edge Platform Application Enablement provides details of services that applications deployed in the MEC Platform could derive out of the network side. It has technical specifications for requirements in the SBI-NR
- ETSI MEC 012: Radio network information API provides specifications related to radio network events and fetching them.
- ETSI MEC 013: Specification describes the location API
- ETSI MEC 021: Specification provides application mobility service APIs
- ETSI MEC 029: Specification provides fixed access information API

A.3 ETSI MEC specification of interest for the UNI

- ETSI MEC 016: UE Application Interface

A.4 ETSI MEC specifications relevant to OP optional capabilities

- ETSI MEC 014: UE Identity API
- ETSI MEC 009: General principles for MEC service APIs
- ETSI MEC 015: Bandwidth management API

A.5 ETSI MEC activities relevant for the E/WBI interface

Inter MEC communication work is planned in ETSI ISG MEC under the Inter-MEC communication work item. It is believed that this work will be relevant to the area of the E/WBI.

Annex B Document Management

B.1 Document History

Version	Date	Brief Description of Change	Approval Authority	Editor / Company
1.0	October 2020	Initial draft.	Technology Group	Tom Van Pelt / GSMA

B.2 Other Information

Document owner: GSMA Operator Platform Group.

It is our intention to provide a quality product for your use. If you find any errors or omissions, please contact us with your comments at futurenetworks@gsma.com.