Operator Platform Telco Edge Requirements

Version 1.0

29 June 2021

This is a Non-binding Permanent Reference Document of the GSMA

Security Classification: Non-confidential

Access to and distribution of this document is restricted to the persons permitted by the security classification. This document is confidential to the Association and is subject to copyright protection. This document is to be used only for the purposes for which it has been supplied and information contained in it must not be disclosed or in any other way made available, in whole or in part, to persons other than those permitted under the security classification without the prior written approval of the Association.

Copyright Notice

Disclaimer

The GSM Association ("Association") makes no representation, warranty or undertaking (express or implied) with respect to and does not accept any responsibility for, and hereby disclaims liability for the accuracy or completeness or timeliness of the information contained in this document. The information contained in this document may be subject to change without prior notice.

Antitrust Notice

The information contain herein is in full compliance with the GSM Association's antitrust compliance policy.

Table of Contents

1 Introduction

1.1 Overview

Operators in the 5G era have a significant opportunity to monetise the capabilities of their networks. Moreover, with the existing relationships that operators have with enterprises, their vast local footprint, their ability to support digital sovereignty principles and their competence to provide high-reliability services, the missing piece is the ability to package and expose their networks in a scalable fashion across multiple operators. The Operator Platform concept, as introduced in [1], described the architecture of a generic platform to fill this gap, identifying main functional blocks and interfaces.

Subsequent whitepapers described edge services and associated commercial principles [6], and detailed technical requirements and a provisional architecture [2], inviting comments from Standards Developing Organisations (SDOs), Open Source Projects, industry fora, and market participants across the cloud services value chain.

The previous work is continued in the present Permanent Reference Document (PRD). This document defines technical requirements, functional blocks and interface characteristics. In addition, it maps the requirements and architecture to specifications from certain selected SDOs, to identify gaps between the PRD and those specifications. This mapping enables partnerships between OP and the SDOs to fill those gaps and potential partnerships with Open Source community projects that may target OP implementations.

The target audience for the PRD is all organisations working in edge computing of public network deployments, including but not limited to platform developers, edge cloud providers, SDOs, Open Source Communities, industry fora, and market participants.

1.2 Scope

This document intends to guide the entire industry ecosystem; operators, vendors, OEMs, and service providers to define a common solution for exposing network capabilities and edge compute resources. The document provides an end-to-end definition of the Operator Platform for support in edge computing environments. The scope of this document covers requirements and architecture specifications that would guide the industry ecosystem into creating a common solution for exposing network capabilities and edge compute resources. The document intends to span an end-to-end view of the Operator Platform in edge computing environments. The ecosystem includes operators, vendors, OEMs, and service providers.

This document covers the following areas:

- Operator Platform requirements
    - **Focus on Edge Computing**: The PRD should define edge computing exposure and network services integration for the Application Providers, whether within enterprises or independent third parties, to enable a simple and universal way of interacting with edge computing platforms.
    - **Open to new services**: The PRD definition should allow the platform's evolution to expose additional services in the future, such as IP Communications and networking slicing, among others.

- Architecture, functions and roles

  - **Reference architecture for enabling 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. This document focuses on Northbound, Southbound, East/West (i.e. Operator Platform Federation), and User to Network interfaces as a first approach.
  - **Mobility**: Network and terminal integration should allow service continuity against end-user mobility in the home and visited networks.

- Standardisation and Open Source communities

  - **Gap evaluation in the standards**: This document analyses gaps in current networking and edge computing standards and identifies SDOs that are appropriate to complete the OP architecture via detailed specifications, protocols and Application Programming Interfaces (API).
  - The **Detailed specification** of architecture and interface specifications **will be defined by SDOs or Open Source communities**, using the baseline in this document.

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

- Evolution from legacy

  - **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 must fit into existing structures and staging, enabling Application Providers to spin their existing capabilities into the Mobile Edge space. Therefore, wherever possible, the OP reuses existing and established structures and processes.

This version of the document focuses on the use of the Operator Platform to provide services to devices attached to their home network. However, it also includes high-level requirements beyond this scenario because they may influence future architecture choices.

Future versions of this document may cover, for example, the following areas in greater depth:

- The detailed impact of service access by devices that are attached to networks other than their home network (e.g. roaming, Wi-Fi, etc.) on the various interfaces and functions of the OP,
- Access to edge resources in the visited network when no federation exists between that network's OP and the Home OP of a subscriber,
- Seamless service continuity when users move to a different network (see sections 2.2.7.3 and 5.2.2.3.6),
- Low latency interaction between applications in different networks in a standardised manner (see section 3.3.10),

- Exposure of operator network capabilities beyond edge resources (e.g. Network as a Service features offering improved QoS on network access),
- The handling of non-SIM devices,
- Inclusion of further capabilities to allow providing a complete Platform as a Service offering,
- The management in a federation of legal constraints that restrict an application's distribution to specific regions (see section 3.3),
- Detailed requirements on the Capability Exposure Role (see section 5.2.1),
- Detailed requirements on the User Client (see sections 3.5.5.2 and 5.2.4),
- A more detailed alignment with NG.126 [9] on the information elements that can be used on the different interfaces and in the OP's data model for the Edge Application and the Resource/Node,
- Data Sharing capability, i.e. Data is 'open' for use by multiple application providers (see section B.8),
- The sharing of an Application Server between different operators (see section B.8).

1.3 Objective and use cases

Focusing on Edge Computing, this document provides a target architecture and requirements to enable an end-to-end delivery chain for different services. The interaction of the entire ecosystem involved in the Edge Computing application delivery should be covered: from developers providing their applications to the system, to the deployment of resources in clouds and networks, the interaction of potentially multiple operators to deploy the applications, and finally, to the customers who will enjoy and interact with the application.

The use-cases covered by the OP demonstrate the benefits that Edge Computing provides, such as low latency interactions between user and application, reduction of network bandwidth, and support of high bandwidth applications and location-bound services.

The use-cases appearing in this document include:

- Automotive
- Mixed/augmented reality
- High-resolution video streaming
- Cloud gaming
- Remote control

A full description of the use-cases, illustrating the benefits brought to them by OP, can be found in Annex B.

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.

Term	Description
Application Provider	The provider of the application that accesses the OP to deploy its application on the Edge Cloud, thereby using the Edge Cloud Resources and Network Resources. An Application Provider may be part of a larger organisation, like an enterprise, customer of the OP, or be an independent entity.
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-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 which ensures 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/OP Partners 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 provide a similar set of network performance.
Data collection interval	A common interval for data reporting that should be negotiated to facilitate federation.
Data Protection	Legal control over access to and use of data stored in computers.
East/Westbound Interface	The interface between instances of the OP that extends an operator's reach beyond their footprint and subscriber base.
Edge Application	The application functionality deployed on the cloudlet
Edge Cloud	Cloud-like capabilities located at the network 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's Service Resource Manager Role. 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. (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 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)

Term	Description
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 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 simplify its contractual relationships. The Federation Broker Role is optional since a federation can be performed 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 an application. Flavours can vary between operator networks.
Home OP	The Operator Platform instance belonging to the subscriber's Operator; that is, whose PLMN identity (MCC and MNC) matches with the MCC and MNC of the subscriber's IMSI, as defined in 3GPP TS 23.122. Note: non-SIM devices are for further study
Leading OP	The Operator Platform instance connected to the Application Provider and receiving the onboarding requests, sharing them to the selected federated platforms/operators.
Local Breakout	Edge Cloud Services are provided to a roamed UE by the Visited OP, rather than by the Home OP
Network Resource Location	The Network Resource Location is how near to the edge or the centre of the network an application is instantiated and Cloud resources are consumed. Whilst typically, an OP deploys an application on a Cloudlet at the edge of the network, it may choose to deploy it, for example, at a Regional level or centrally (but within the OP). The OP decides on the Network Resource Locations.
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	The 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 has an Operator Platform.
Operator Platform	An Operator Platform (OP) facilitates access to the Edge Cloud capability of an Operator or federation of Operators and Partners. It follows the architectural and technical principles defined in this document.

Term	Description
	NOTE: Future versions of this document may extend the capabilities of the Operator Platform.
Partner	An entity or other party that offers and provides a service or resource, in the context of the Operator Platform's federation, to other partners. Each partner hosts an OP and offers the resources through its E/WBI federation. For example, a partner can be an Operator that provides network, subscribers and cloud services or a hyperscaler / cloud provider that offers cloud services only.
Partner OP	An Operator Platform that federates with another Operator Platform and through the E/WBI offers its Edge Cloud capabilities to the other Operator Platforms.
Region	An OP Region is equivalent to a Region on a public cloud. The higher construct in the hierarchy exposed to a developer who wishes to deploy an Application on the Edge Cloud and broadly represents a geography. A Region typically contains 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 the scope of geography managed by a Regional Controller is up to the operator to define.
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. This role includes managing the application load over the Edge Cloud, the configuration of network capabilities, and the relationship with the User Client.
Southbound Interface	Connects the OP with the specific operator infrastructure that delivers the network, cloud and charging 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 applications or cover a complete server.
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 on 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.
Visited OP	The Operator Platform instance that belongs to the Operator providing access to a roaming subscriber; that is, whose PLMN identity (MCC and

Term	Description
	MNC) matches with the MCC and MNC of a roaming subscriber's current VPLMN. Note: non-SIM devices and non-3GPP access are for further study

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
CI/CD	Continuous Integration / Continuous Development and Deployment
CISM	Container Infrastructure Service Manager
CPU	Central Processing Unit
CRUD	Create, Read, Update and Delete
DBaaS	DataBase as a Service
DC	Data Centre
DDoS	Distributed Denial of Service
DNAI	Data Network Access Identifier
DoS	Denial of Service
EAS	Edge Application Server
ECP	Edge Computing Platform
ECS	Edge Configuration Server
EEC	Edge Enabler Client
EES	Edge Enabler Server
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
NAS	Non-Access Stratum
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
OCI	Open Container Initiative
OP	Operator Platform
OS	Operating System

Term	Description
OSC	Open Source Community
OTT	Over the Top
PaaS	Platform as a service
PCF	Policy Control Function
PCRF	Policy and Charging Rules Function
PDN	Packet Data Network
PDU	Protocol Data Unit
PGW	PDN (Packet Data Network) GateWay
PII	Personally-Identifiable Information
PRD	(GSMA) Permanent Reference Document
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
SMF	Session Management Function
SPR	Subscriber Profile Repository
SR/IOV	Single Root I/O Virtualisation
SRM	Service Resource Manager
SSC	Session and Service Continuity
SUPI	SUbscription Permanent Identifier
TAC	Tracking Area Code
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

Term	Description
URL	Uniform Resource Locator
URSP	UE Route Selection Policy
VIM	Virtualised Infrastructure Manager
VM	Virtual Machine
VPLMN	Visited Public Land Mobile Network
VPS	Visual Positioning Service
VPU	Vision Processing Unit
VR	Virtual Reality
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]	OPG.01	Whitepaper: Operator Platform Telco Edge Proposal – Version 1.0, 22 October 2020 https://www.gsma.com/futurenetworks/resources/op-telco-edge-proposal-whitepaper/
[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
[6]		Telco Edge Cloud: Edge Service Description & Commercial Principles Whitepaper, version 1.0, 27 October 2020 https://www.gsma.com/futurenetworks/resources/telco-edge-cloud-october-2020-download/
[7]		OCI Image Format Specification https://github.com/opencontainers/image-spec
[8]		Open Container Initiative Runtime Specification https://github.com/opencontainers/runtime-spec
[9]	NG.126	Cloud Infrastructure Reference Model https://www.gsma.com
[10]	3GPP TS 23.501	System architecture for the 5G System (5GS) https://www.3gpp.org/DynaReport/23501.htm
[11]	3GPP TS 23.502	Procedures for the 5G System (5GS) https://www.3gpp.org/DynaReport/23502.htm

Ref	Doc Number	Title
[12]		The rise of serverless computing, Association for Computing Machinery, Communications of the ACM, Volume 62, Issue 12 https://dl.acm.org/doi/10.1145/3368454
[13]	FS.30	Security Manual, Version 1.0, GSM Association Official Document FS.30, 20 April 2020.
[14]	FS.31	Baseline Security Controls, version 2.0, GSM Association Official Document FS.31
[15]	Ranaweera2021	Pasika Ranaweera, et al., Survey on Multi-Access Edge Computing Security and Privacy, to be published in IEEE Communications Surveys & Tutorials
[16]	3GPP TS33.122	Security Aspects of Common API Framework (CAPIF) for 3GPP northbound APIs https://www.3gpp.org/DynaReport/33122.htm
[17]	3GPP TS33.210	IP network Layer Security (Release 16), 3GPP TS 33.210, https://www.3gpp.org/DynaReport/33210.htm
[18]	3GPP TR33.805	Study on security assurance methodology for 3GPP network products https://www.3gpp.org/DynaReport/33805.htm
[19]	3GPP TS33.535	Authentication and Key Management for Applications (AKMA) based on 3GPP credentials in the 5G System (5GS), 3GPP TS 33.535 https://www.3gpp.org/DynaReport/33535.htm
[20]	3GPP TR33.839	Study on Security Aspects of Enhancement of Support for Edge Computing in 5GC (Release 17), 3GPP TR33.839 https://www.3gpp.org/DynaReport/33839.htm
[21]	ETSI ISG MEC 36	Harmonising standards for edge computing – a synergised architecture leveraging ETSI ISG MEC and 3GPP specifications, 1st edition, July 2020 https://www.etsi.org/images/files/ETSIWhitePapers/ETSI_wp36_Harmonizing-standards-for-edge-computing.pdf
[22]	ETSI ISG MEC 35	Multi-access Edge Computing (MEC): Study on Inter-MEC systems and MEC-Cloud systems coordination, Draft ETSI GR MEC 035 V3.0.0 (2021-04).
[23]	NIST P800	Ron Ross, et al., Developing Cyber Resilient Systems: A Systems Security Engineering Approach, NIST Special Publication 800-160, Volume 2, November 2019 https://doi.org/10.6028/NIST.SP.800-160v2
[24]	3GPP TS33.310	Network Domain Security (NDS); Authentication Framework (AF), 3GPP TS 33.310 https://www.3gpp.org/DynaReport/33310.htm

NOTE: Some documents in this list (e.g., [13], [14]) may not be released as public documents.

2 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 applications.
2. For each operator supporting the OP, there shall be an OP instance that has the sole responsibility for managing the resources and services that the OP exposes in that operator's network.

   a) This instance may be operated by the Operator or be outsourced.

3. The OP shall be able to effectively isolate each Tenant's applications from the applications of all the other Tenants.
4. The interfaces that an OP instance offers to other parties shall be provided using common definitions based on the requirements in this document.

2.1.2 Functionality offered to Application Providers

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

1. The OP architecture shall allow an Application Provider to use a common interface to manage edge applications deployed towards the subscribers of multiple operators subject to 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 Application Providers for the development and deployment of edge applications shall allow for easy deployment of applications developed for public clouds.
3. The OP shall allow an Application Provider to reserve resources for future application deployments, ensuring the availability of the booked capacity.
4. The OP shall hide the complexity of the OP architecture, the involved operator networks and client access to those networks from Application Providers.
5. 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.

6. The OP architecture shall allow an Application Provider deploying an application using the OP to monitor the application's usage across the networks on which it is deployed.
7. The OP architecture shall allow an application deployed within an operator network to interface securely with the application's back-end infrastructure outside of the operator network.
8. The OP architecture shall allow an application deployed within an operator network to store data in a manner that is secure and compliant with applicable local regulations.
9. The OP shall enable the utilisation of Cloud Resources that support deploying applications as VMs or Containers.
10. The OP shall support applications packaged as VMs and Containers.

Note: These are requirements on what the OP architecture shall enable. It is up to the individual parties providing an OP to decide whether they offer these capabilities in their deployment.

11. The OP shall expose network capabilities to Application Providers, including longer-term managed network services (such as for QoS) and shorter-term or transactional style services (such as SIM-derived services, such as location verification).

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:

1. The OP shall allow end-user devices to access services provided through Edge Enhanced and Edge Native Applications.
2. The OP shall allow the end-user to access Edge applications deployed on edge resources seamlessly and securely.
3. Services provided as Edge Enhanced and Edge Native 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. This latter case is subject to an agreement between the involved operators (i.e. home and visited) and the Application Provider's requirements (e.g. locality, availability when roaming).

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

2.1.4 Functionality offered to Operators

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

1. The OP architecture shall allow an operator to monitor and track the usage by an OP of its compute (including specialised compute), storage and networking resources.

2. The OP architecture shall enable operators to monitor their subscribers' usage of Edge Cloud resources (including network) in a visited network.
3. The OP architecture shall allow an operator to charge for the services and capabilities provided by OP to application providers, subscribers, and other operators.
4. The OP architecture shall allow the OP to influence the quality of service delivered by the network for the interaction between an end-user device and an application.
5. If the Operator Platform is part of the operator's security domain (see Note 2), it can access the network and cloud resources through the SBI (and any other operating interface).

   Note 1: 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. That 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 2: Security Domains administer and determine the classification of an enclave of network equipment/servers/computers. Networks using different security domains are isolated from each other. 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. For example, 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. See 3GPP TS 33.210 Network Domain Security (NDS); IP network layer security.

6. Similarly, there is a "separation of concerns" of the operators between each other and between OPs. Where the Operator Platform is not part of the operator's security domain, 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:

1. The OP architecture shall allow an OP to deploy applications provided by Application Providers on another OP (when there is a federation agreement between the OPs).

   a) Both containerised applications and applications relying on VMs shall be supported.

2. A federation of independently owned and operated Operator Platforms enables additional capabilities, such as:

a) the User Equipment (UE) can continue to use the Edge Cloud service when moving into a "visited network" and in an area where Edge Node Sharing takes effect.

3. The OP architecture shall allow a "home" OP to receive applications from "foreign" OPs to serve subscribers, whether they are home OP subscribers or visiting OP subscribers.
4. The OP architecture shall allow such an 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:

1. The OP shall allow the operator to expose compute and storage resources within the Operator or Partner network on which applications can be deployed for use by specialised and regular end-user devices.
2. The OP architecture shall allow an application deployed on cloudlets within the operator network to interact with low latency with applications deployed at nearby operator network cloudlets, 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 managing 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:

1. An OP shall provide edge compute resources on a virtualised basis to another party in the OP ecosystem (e.g. an Application Provider, a Partner OP or another operator).
2. An OP is responsible for managing the virtualised resources with which it has been provided. For example, this includes the reservation, de-reservation, allocation, de-

allocation and potentially in-life management (such as scaling) of virtualised resource to a specific Application Provider.
3. If one OP or Application Provider overloads the virtualised resource it has been allocated, this should not degrade the performance of others.
4. An OP or Application Provider does not have visibility of the resources that another has allocated or is using.
5. All parties in the OP ecosystem use the same data model for the virtualised resources.
6. It is optional for resource management to provide telemetry or other metrics from the edge node.

2.2.2.3 Resource Reservation

The OP, as manager of the edge compute resources, shall allow application providers to optionally reserve resources that they may not consume immediately. This feature allows Application Providers to ensure resource availability independently from when they may deploy/modify the different applications under their control.

1. An Application Provider shall be able to reserve a certain amount of resources that would be logically bound to them.
2. An OP shall validate the reservation based on the currently available resources and ensure that those booked resources, the amount reserved by the application provider, remain available until the Application Provider requires them.
3. An Application Provider shall assign (or modify) reserved resources to an application when deploying (or modifying) it.

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 functionality includes:

1. Support for Continuous Development through code development pipelines similar to those provided in a public cloud.
2. Support for Continuous Integration through staging in edge test sites.

2.2.4 Edge deployment enhancements

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

2.2.5 Data Protection Management

The OP shall offer Data Protection management. Specifically:

1. 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.
2. 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:

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

2.2.7 Mobility Requirements

Mobile subscribers accessing the edge resources can move to different locations within or outside their home operator's footprint, and they can do so while using the service. In all these cases, the subscribers may expect applications that depend on application functionality deployed on edge resources to provide an experience similar to what they are used to (i.e. when not mobile). The following sections detail the requirements to enable that.

2.2.7.1 Roaming Requirements

The OP shall support subscribers accessing the service from outside their home operator's footprint (i.e. roaming subscribers). For those scenarios, the following applies:

1. Roaming subscribers shall be able to access applications deployed on edge resources within the visited network.

   Note: This requires local breakout (LBO) of the subscriber's Protocol Data Unit (PDU)/Packet Data Network (PDN) connection to a User Plane Function(UPF)/PDN Gateway (PGW) in the visited network.

2. Access of roaming subscribers to edge applications in the visited network shall be subject to authorisation by the subscriber's Home OP and the Visited OP.
3. An Application Provider shall be able to indicate whether their application is available to inbound/outbound roaming subscribers and, if so, in which networks.

   Note: Availability of the applications a subscriber wishes to access is currently assumed to be covered by the federation between networks. Roaming on a non-federated operator's network is not in scope.

4. If an OP is not available in the visited network or the OP managing the resources in that network is unavailable to the subscriber (e.g. the required federation or LBO roaming agreements are missing), the subscriber shall still be routed to the most favourable location. This would be the location in the network closest to the user where the application is available and authorised. Because the visited network cannot provide the application, the subscriber shall be routed to the edge application in the subscriber's home network, i.e. the next most favourable location.
5. An Application Provider shall be able to indicate whether their application can support access by subscribers connected to visited networks, given that such access may result in significant increases in latency.

Note: As indicated in section 1.2, a seamless handover from home or visited network to another visited network is not in the scope of the current version of this document.

2.2.7.2 Requirements for defining geographical conditions on mobility

An application may wish to restrict its service to UEs in particular geographical areas or ensure that the application instance/function serving the UE is placed in the same zone. The movement of the UE out of the service area might not trigger a session anchor change of the UE.

The OP shall be able to receive an application's geographical coverage restrictions as part of the application provider's criteria. These restrictions may be driven by privacy, data retention policies, etc.

1. The OP shall be able to receive geographical UE mobility events (e.g. when leaving a pre-defined area) from the network or the UE.
2. The OP shall perform the application mobility management process to ensure that the criteria are accomplished.

Note: Section 5.2.2.2 provides more details on the instantiation process.

Note: Area restrictions should be bound to availability zones

2.2.7.3 Requirements for Application Session Continuity

The objective is that the OP offers a seamless experience to an end-user, even as they move around the network. An application's sensitivity to mobility is strongly influenced by its nature, including whether it is implemented as stateless or stateful.

The operator is responsible for the mobility management of the UE. There are four different types to be considered:

- SSC Mode 1: Preservation of IP address, PDU/PDN session and UPF/PGW
- SSC Mode 2: 'Break before make' - change of IP address, PDU/PDN session and UPF/PGW
- SSC Mode 3: 'Make before break' - change of IP address, PDU session and UPF
- Inter-operator mobility - change of IP address, PDU/PDN session, UPF/PGW, operator and OP.

Ideally, mobility is handled invisibly to the application's end-user by the mobile network operator, perhaps in conjunction with the OP and the application provider.

With Mode 1, typically, the mobility is invisible to the application and the Application Provider. It is expected for the application to continue using the same edge compute resources despite mobility events.

With Modes 2 and 3 (and occasionally Mode 1), the OP and perhaps the Application Provider must do some work to minimise the impact on the experience provided to the end-user.

In those situations where the application instance serving the user is changed, an application session may need to be maintained to ensure that the user does not notice any effect on the received experience, such as a VR video delay during application instance reattachment.

The OP shall be responsible for:

1. Deciding that a different edge compute resource can better host the Edge Application. The decision should take the Application Provider's policy into account. Such policy may depend on the application's sensitivity to a change of compute resource, required notification before a move, etc.
2. Maintaining an inventory of network and Edge Computing local resources to facilitate the mobility and enable advanced application and connection use cases, e.g. duplicating session traffic to ensure availability.
3. If the Application Provider requires, notifying them about this recommendation.
4. When required, informing the Application Provider about the mobility of the user, data session anchor change. The Application Provider is then expected to collaborate with OP in transferring the application state from one edge compute resource to another, preferably before the user's application session is routed to the new application server on the new edge cloud compute resource.
5. When required, notifying the Application Provider on a recommended change of edge compute resource, the Application Provider is responsible for determining the exact timing of the change.

Note: The end user's application experience may be compromised if the change of edge compute resource is delayed for too long.

Note: It is for further study how to solve inter-operator session continuity.

2.3 High-Level Security Requirements

The OP architecture shall comply with the following security requirements:

1. The OP shall expose operator network functions and resources data (e.g., compute and storage) following the 'need-to-know' principle and only for the legitimate scenarios expected in the PRD.
2. The OP shall expose operator network functions and resources data (e.g., compute and storage) following the 'need-to-know' principle and only for the legitimate scenarios expected in the PRD.
3. The OP shall not expose its configuration data and internal topology (referred to as topology hiding).
4. The OP shall apply data protection mechanisms to assure data availability, confidentiality, authenticity, and integrity. Data shall be protected both during storage and processing.
5. The permitted data (i.e., data that may be shared on the need-to-know principle) shall be exposed only to authorised and authenticated entities in a secure way. This means:

    a) protecting the data in transit, via encrypted and integrity protected channels, to prevent data interception and manipulation, as well as to prevent intervening attacks, while also assuring user privacy protection;

b) in storage and execution, via technological means, e.g., log file or database access controls, trusted enclaves.

6. The OP shall implement role-based access control for configuring users, with policies defined and enforced, ensuring a secure binding between services and authorised entities.
7. The OP shall adopt an integrity protection mechanism for the various identifiers in use (such as resource IDs, user/subscriber IDs, session IDs, app IDs) to prevent user and resource usage tracking, ensuring privacy protection.
8. The OP shall require operational procedures to carry out security hardening. This hardening includes, e.g., auditing to ensure that software patches are up to date, publishing regular security audits.
9. The OP shall adopt certificate-based authentication with a federation certificate authority, as described in [24].
10. The OP shall apply protection mechanisms to ensure service availability to prevent attacks targeting the availability of exposed applications/services, e.g., denial of service attacks and brute force attacks.
11. Telemetry for intrusion detection should be supported.
12. The OP shall adopt best practices of 3GPP SA3 on the selection of security protocols, certificate authorities, as described in [24] and elsewhere, as provided in the References list in Annex E.
13. Services, processes, and tenants running in containers and virtual machines, and their data, shall be protected. Approaches to protecting them include process isolation via name-spacing or hypervisor controls and trusted enclaves.
14. Best practices for DevSecOps (i.e., the practice of introducing security practices into DevOps), as described in [14], should be followed.
15. Security Controls related to Edge computing, as described in [14], should be followed.
16. The OP shall employ telemetry and analytics to detect and report application security policy violations at runtime to localise and isolate malicious application behaviour.
17. The OP shall employ telemetry and analytics to detect Distributed Denial of Service (DDoS) attacks against the network and enable rate-limiting and traffic isolation in network segments and endpoints.
18. The OP should support hardware-root-of-trust (e.g. TPM) based security keys for platform integrity checks, mutual authentication, and the establishment of secure tunnels with tenants/application service providers.

Note: A future phase of this work will investigate defining security levels between operators.

19. The OP should support a secure DNS service to avoid attacks that exploit DNS, such as impersonation attacks.

Note: A future phase of this work will investigate secure DNS options and options for including a DNS service in an Edge architecture.

20. The OP should support TCP proxies to avoid server IP address guessing and TCP connection hijacking.
21. The OP should support flow-control on invoking application services control plane APIs to protect federated services from abuse of these APIs.

22. The OP should support different role-based privileges for such roles as OP tenants and network/infrastructure operators to control unauthorised access to slice management of shared/virtualised resources.
23. The OP should enable network/slice resource management through allocation, isolation, telemetry, analytics, and quota enforcement to meet slice/shared network resource SLA requirements.
24. The OP should enable resource isolation, sharing authorisation, and residual data clean-up to protect shared network resources/slices from tampering and data theft.
25. The OP should employ message filtering of HTTP control plane signalling and firewall configurations to protect network resources from spoofing attacks from roaming interconnections.
26. The OP should enable security audits on the access privilege management to avoid identity theft or fraud.
27. The OP should employ secure storage of account credentials to avoid identity theft or fraud.
28. The OP shall employ secure initialisation and secure configuration data storage to avoid the exploitation of network configuration data weaknesses.
29. The OP should provide hardware root-of-trust based tools to guard network configuration status.
30. The OP should support centralised and unified log management to protect from any tampering, whether malicious or inadvertent,
31. The OP should support the automation of security operations.

3 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 by employing intermediate integration platforms.

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

To satisfy its goals, the OP shall enforce a multi-layer architecture with multi-role separation of the complete functionalities and requirements presented in Chapter 2. For a system as complex as the OP, a target architecture is needed to localise and inter-relate the requirements. Such a target architecture presented in this section.

The target architecture is described at a relatively high level. Where OP-specific concepts are specified, they are defined as roles, functionalities, and interfaces. This is done to capture the essential behaviour needed by OPs without constraining the ability of the architecture to conform to prevailing standards or the ability of vendors to innovate.

There are certain exceptions to this rule where more concrete architectural descriptions are provided:

- Containers and Virtual Machines: In the application development ecosystem with which OP must interact, deploying applications in containers and virtual machines is a well-established practice. The OP does not intend to create a new framework for application development and lifecycle management. Therefore, separate sections relating OP requirements to containers and VMs are provided. In recognition of prevailing trends in application development, these sections are somewhat specific about container management, operating systems, and other system components.
- Serverless computing: In previous work (e.g., the whitepapers published in earlier phases of the OP project), the serverless computing architectural pattern was identified as a high priority for monetising edge computing in the OP environment. Analogously to the cases of containers and VMs, serverless computing presupposes interactions between users and applications that are somewhat specific, and so OP requirements become more specific here.

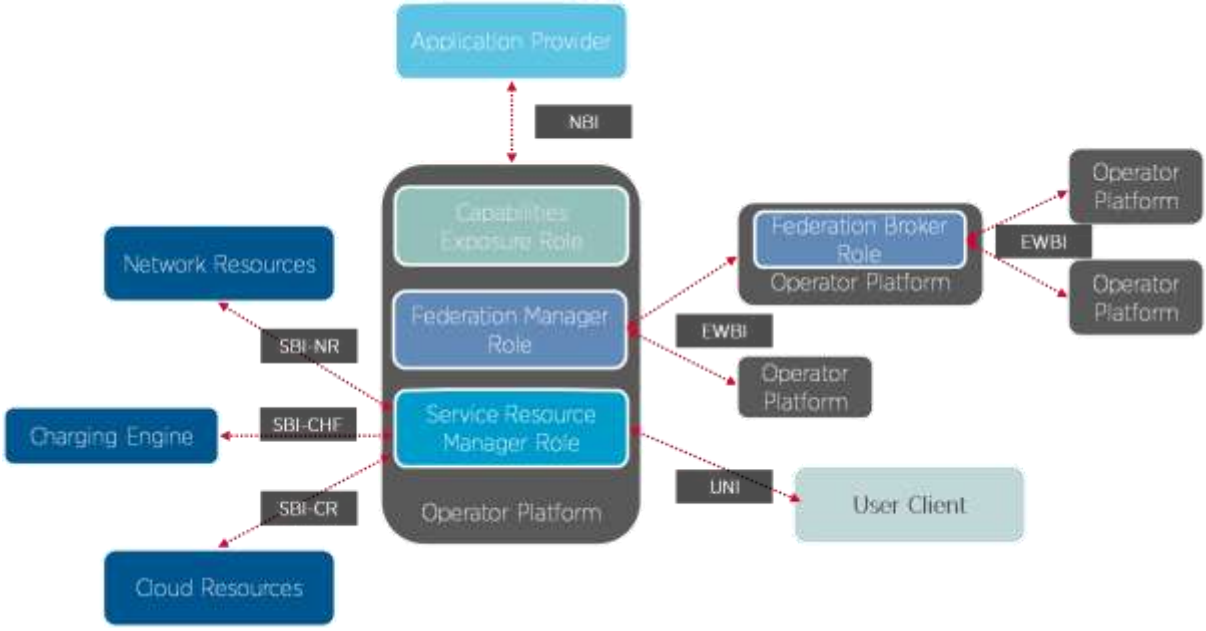

Figure 1: OP Roles and Interfaces Reference Architecture

The following sections cover the functionalities and role separation and 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 enable an OP instance to interact with and execute scenarios from/towards other actors in the OP ecosystem, namely Application Providers, other OP instances, the Cloud Resources and the Network Resources.

A single OP can implement the roles and their functionality, or separate instances provide different roles (for example, an OP instance provides the Service Resource Manager Role and another OP instance provides the Capability Exposure Role).

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:

- Edge Cloud Infrastructure Endpoint Exposure;
- Application Onboarding;
- Application Metadata/Manifest Submission;
- Application CI/CD Management DevOps;
- 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 managing 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 instances on the Edge Cloud Infrastructure via the SBI-CR interface;
    - Cloud resource reservation managed by the OP,
    - Configuring UE traffic management policies to accomplish the application's requirements, e.g. as described in 3GPP TS 23.502 [11], or the UE's IP address shall be maintained;

  Note: UE Route Selection Policy (URSP) rules influenced by the OP may also be considered a solution.

    - Exposure of usage and monitoring information to operator's charging engine via the Southbound Interface – Charging functions (SBI-CHF) to enable operators to charge for the OP's services.
    - Interacting with the Mobile Network via the Southbound Interface – Network Resources (SBI-NR), for example to:

  - Fetch Cloudlet locations based on the mobile network data-plane breakout location;
  - Subscribe and receive notifications on UE Mobility events from the network to assist applications.
  - Configure traffic steering in the Mobile Network towards Applications orchestrated in Edge Clouds;
  - Receive statistics/analytics, e.g. to influence Application placement or mobility decisions.
  - Receive information related to the network capabilities, such as QoS, policy, network information, etc.

- **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 and Federation Manager Roles

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 the federation management 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 for the Application Providers to a larger footprint of Edge Clouds, a more extensive set of subscribers and multiple Operator capabilities

The following are prerequisites to enable the federation model:

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

Federation Management provides the Management plane. The Management Plane covers the set of functionalities offered to Application Providers and OPs to control and monitor the resources and applications within the federation under their responsibility.

The Management Plane functionality is realised via the multiple functional blocks within an OP instance listed in the subsections below. The management actions are relayed between these different functional blocks using the NBI, SBI and E/WBI interfaces that have been defined for communication between them in section 3.1.

The Management plane works at two domain levels: application and infrastructure (resources). Each of these domains supports management at two distinct stages in the managed entities life-cycle: the configuration and the run time management. Table 1 lists the functionality provided by the Management Plane in each domain and stage.

Domain	Stage	Management Functionality
Resources	Configuration	Federation Interconnect Management
		Resource Catalogue Synchronisation and Discovery
		Edge Node Sharing
		Partner OP Provisioning
		Authentication and Authorisation
		Resource sharing policies
		Automation of Orchestration
	Run Time	Edge Cloud resource monitoring
		Lifecycle Automation
Application	Configuration	Application Management
		Service Availability on Visited Networks
		Automation of Orchestration
	Run time	Operational visibility
		Lifecycle Automation

Table 1: Management Functionalities

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 the federation is 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 establishing and sustaining 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 federation agreement between the partnering OPs/Operators.

3.3.2 Resource Catalogue Synchronisation and Discovery

Operators can include the edge resources in the OP's set of available resources using the SBI.

The OPs shall exchange and maintain the types of resources offered to each other (E/WBI).

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, Bandwidth in/out);
- Specialised Compute Offered: catalogue of add-on resources, e.g. Graphic Processing Units (GPU), Vision Processing Units (VPU), Neural Processing Units (NPU), and Field Programmable Gate Arrays (FPGA).
- Network QoS supported by the zone: maximum values of latency, jitter, packet loss ratio.
- Supported virtualisation technology: only VMs, only containers, both.
- Costs associated with the use of resources. This information can influence the Availability Zone selection (e.g. the use of several small zones, that combined, cover the needed Region and are offered by different partners, instead of a more extensive and expensive zone offered by another partner)

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 is supported over the E/WBI to achieve the above.

3.3.3 Application and Resources Management

This procedure corresponds to the forwarding of a northbound request from one operator to accommodate an Edge Application or a resource booking in another operator's Cloudlets. Operators authorise the deployment or reservation 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. Therefore, the E/WBI interface must provide capabilities to support resource reservation and 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, indicating the intended geographical Regions that they want to target. The OP instance translates these NBI interactions to the E/WBI.

The Application Provider request contains mandatory information (required CPU, memory, storage, bandwidth…) defined in an application manifest. It may also include other optional characteristics indicating the application's needs (latency, prioritisation, reservation, etc.).

There may be multiple models possible for performing application orchestration via the E/WBI.

On a federation relationship, the Partner OP decides which Edge Cloud(s) to deploy the applications on or which Cloudlet provides the resources available for a reservation based on the Availability 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.

The application provider's criteria about Availability Zone / Region are considered, but, in the end, it is the Operator Platform that decides which edge cloud resources provide the better fit with the application requirements (QoS) and the costs of using those resources.

3.3.4 Service Availability on Visited Networks Management

When a User Client (UC) requires accessing the Edge Cloud service of a visited network, the federation model facilitates service availability for this UC. The service should be provided via local Edge Cloud resources of the Visited OP if local breakout is available for roaming UEs.

Note: It is highly recommended that when entering into a federation agreement, MNOs also agree to enable Local Breakout for the data connections towards the edge cloud resources in visited networks.

Note: When enabling local breakout, MNOs need to consider regulatory requirements on the home and visited network (e.g. lawful interception).

If local breakout is not possible, the UC may be served via the Home OP. For that reason, and considering the credentials and authoritative ownership of the users to the home operator, the authentication and authorisation of the first register request shall always be made 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 network, the E/WBI shall allow the Home OP to provide the Visited OP with the necessary information to perform authorisation and grant the service access (e.g. a token). When the UC tries to access a service when on visited networks, the Visited OP authorises the UC using the authorisation 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 presence. Using the figure below as an example, the mobile network of both operators covers the whole country. However, Partner A deploys edge sites in the country's North Region and operator B in the South Region. In this case, Operator B might deploy an application on Partner A's edge node while providing connectivity to the end-user over their own radio network.

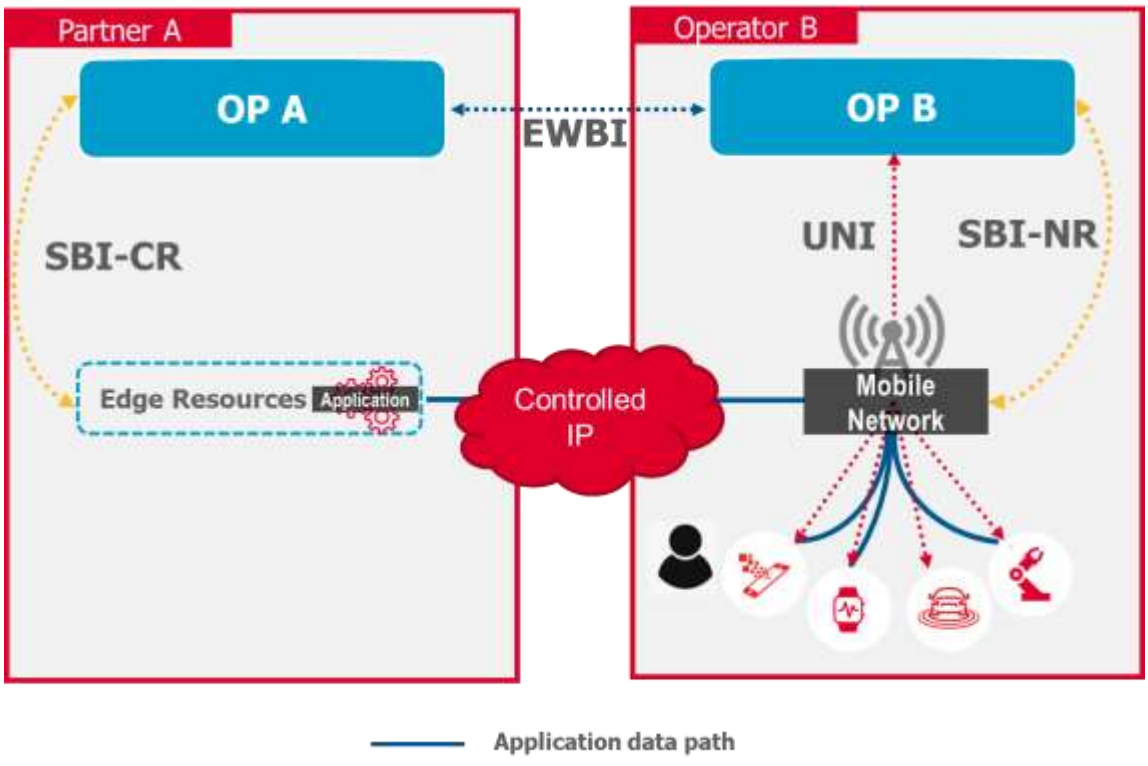

Figure 2: Edge Node Sharing

Figure 2 above shows an end-user who is a subscriber of Operator B's OP services and is currently connected to Operator B's network in the country's north. Edge node sharing enables this end-user to access the Edge Cloud service, even though Operator B does not have their own edge resources in this Region; the Operator B Edge Cloud service is hosted

on Partner A's edge node. The connectivity between the two OPs is over the E/WBI interface.

The East/Westbound interface enables Operator B's OP to retrieve the application instance access information and provide it to the user. This approach allows performing service discovery and delivery in the same way as when the application was delivered from a Cloudlet in Operator B's own network.

A subscriber of Operator B accesses its home network/operator platform and asks for the required Edge-Enhanced or Edge-Native Application. When Operator B's OP identifies that the most suitable edge node is in Partner A, Operator B's OP requests the Edge Cloud service through the E/WBI to Partner A's OP. In this example, since the OPs have a long-running partnership, they have pre-established commercial agreements, security relationships and policy decisions (for instance, QoS-related). Thus (assuming enough edge resource is available), Partner A can reply with the application endpoint (e.g. FQDN) on the Cloudlet at which the subscriber can connect to the application.

Note that network resources remain managed by Operator B, the operator providing the actual mobile network connection to the user, and IP connectivity between Partner A's edge node and Operator B is managed to ensure end-to-end QoS delivery for the subscriber. Responsibility for the management of the edge cloud resources depends on the agreement between the partners. Most likely, Operator B has a long-term allocation of resources in Partner A's cloudlets and manages them amongst its subscribers wanting access to the edge service.

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 used for Federation Interconnect establishment and management. This information would include:

- The Partner Name;
- The Partner's geographical area (e.g. Country of operation);
- The Partner identifiers;
- The 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 that partner OP. This authentication requires that authentication information (e.g. digital certificate or passphrase) is provisioned in the OP. This mechanism can be mutually agreed between the involved 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 a federation agreement) or specific Availability Zone(s) where they have 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 Edge Cloud resource monitoring

The OP shall offer to application providers and operators the capability to monitor the resources by:

- Usage: compute, memory, storage, bandwidth ingress and egress
- Events, alarms/faults, logs
- Metrics performance

Usage data about resources consumed, per partner or by application, are the parameters that are monitored by default. However, specific events, alarms, logs and metrics should be defined by the application provider (those related to the applications) or the federation contract between the operators (those related to the shared resources).

An OP monitors Edge Cloud resource consumption by the Edge Applications, including applications from the partner OPs. In addition, an OP informs the partner OP of the resource consumption statistics of its applications via the E/WBI.

The resource usage 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.

3.3.8 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.

This fault and performance management is based on the information obtained through the monitoring described in section 3.3.7.

Due to the amount of exchanged information, a subscription/notification mechanism should be available to allow the above filtering for the information relevant for Fault and Performance management.

3.3.9 Automation Capabilities

The OP shall offer application providers the automation of the everyday actions related to the resources' lifecycle management across a federation. The information assets used in a federation should be harmonised to enable this (see Common Data Model, section 3.4).

There are a few essential scenarios considered for automation:

- starting new application instances
- the reconfiguration of resources and network to maintain SLAs
- the execution of application policies
- the reservation and release of resources

3.3.10 Low latency interaction between UCs and applications in different networks

The end to end latency between a UC and corresponding edge application on an OP's edge cloud may play a vital role in the user experience, e.g. for AR/VR based applications or V2X applications for automotive and many others.

Through Edge Node sharing or in a roaming scenario (without LBO), an Application Client may get serviced from Operator A, for example, in the context of edge services. At the same time, the UE is attached to a different mobile network of, say, Operator B, as shown in Figure 2. In such cases, the MNOs in a federation relationship need to manage the inter-operator IP connectivity carrying application traffic. They need to do this to meet the required SLAs demanded by edge applications sensitive to latency and other QoS attributes, e.g. throughput, jitter, packet loss, latency etc., averaged over time.

Note1: The inter-operator IP interconnect carrying application traffic between two operators corresponds to the data plane and is different from the E/WBI interface carrying the OP control plane communication for applications and federation management.

MNOs wishing to participate in edge node sharing or offering a home routed scenario involving inter-operator IP connectivity in different networks may agree to set up specific IP transport. This transport may include but is not limited to dedicated connections, IPX or colocation services, to name a few possible options. These IP interconnects and the technologies to be used can be mutually agreed and preconfigured to provide the agreed IP services with the required QoS.

The Service Resource Manager (SRM) could be statically configured to be aware of such inter-IP connectivity aspects with the partner OPs and the associated QoS supported over the IP interconnect.

The IP interconnect between MNOs could be monitored by the operators to assess its performance. However, the OP is not expected to be directly involved in any management, control or monitoring functions. The division of control over the set of relevant QoS attributes of IP interconnect can be a mutual agreement between the OP and the operator to provide such network services to Application Providers.

Note2: Inter-operator IP connectivity in this phase is assumed to be a pre-established dedicated connection between the MNOs that the OP could utilize as a network resource to enable edge node sharing or home-routed scenarios.

Note3: Aspects like standardized interfaces or dynamic interaction between the OP and the network controller (or management plane) of such inter-operator IP network are for further study in a subsequent phase.

3.4 Common Data Model

The Common Data Model (CDM) introduces standardised data schemas for describing characteristics of the elements of an OP system. The 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 defines the information elements required to deploy and manage an OP system.

The data model defines a minimum set of mandatory information elements and allows reasonable default values for these elements where they make sense.

The data model accommodates 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).

GSMA PRD NG.126 [9] provides, in its sections 2 and 4, a more detailed overview of information elements that can be covered for the Edge Application and the Resource/Node.

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

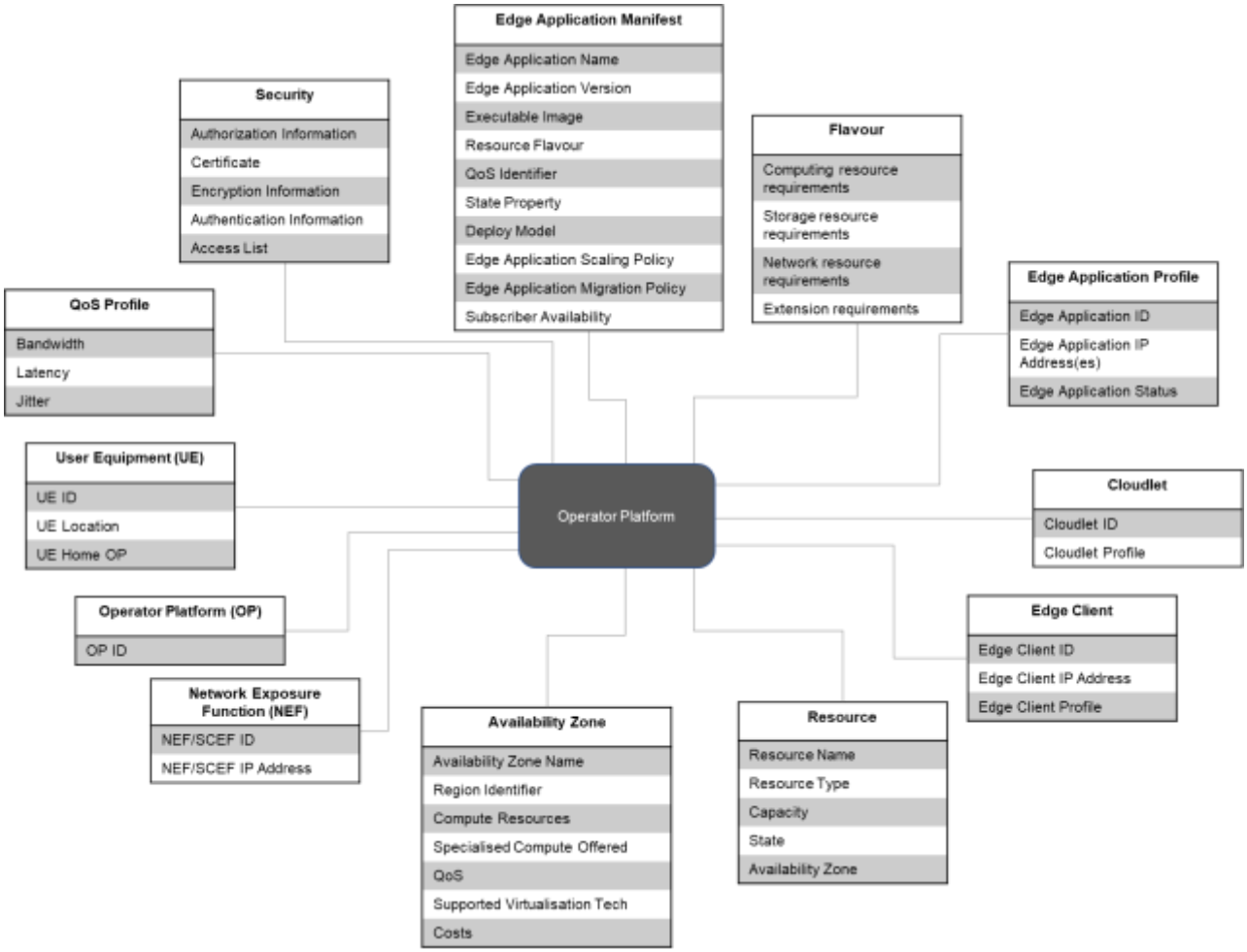

Figure 3: Common Data Model

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. Therefore, the following table shows the information elements maintained by a role (e.g., OP, Application Provider) about other trusted domains.

Data type	Description	Interface Applicability
Authorisation information	Authorisation information of the Application Provider	UNI/East/West/North
Certificate	The certificate of the Application Provider	UNI/East/West/North
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/North
Authentication information	Certified identities of other trusted domains	UNI/East/West/North

Access List	For information elements that an API may request between trust domains, the list of identities authorised to make a request	UNI/East/West/North

Table 2: Common Data Model – Security

3.4.2 Edge Application

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

An OP instantiates an application. 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. Therefore, an OP that instantiates an application from the application manifest should expect the manifest from the Application Provider. This requirement implies that Partner OPs should be provided, if needed, with the application manifest by 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 managed Edge Cloud. 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 additional attributes to be defined at the Application Provider's or OP's discretion. A possible realisation of optional elements is key-value pairs, as is used in various data models.

Data type	Description	Interface 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/North	Mandatory
Edge Application version	The version of the Edge Application.	East/West/North	Mandatory
Executable Image	A URI (or similar name) of the VM or Container image to be installed and executed by the OP.	East/West/North	Mandatory

Data type	Description	Interface Applicability	Optionality
Resource Flavour	The "name" or identifier of the Flavour that should be used to instantiate the application, as selected by the Application Provider. "Flavour" is defined below.	East/West/North	Mandatory
QoS Identifier	A "name" or identifier of the QoS description for network traffic, as selected by the Application Provider. The default value is "best-effort".	East/West/North	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 session is relocated). The default value is "stateless".	East/West/North	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/North	Optional
Edge Application scaling policy	Indicates whether a backend application can be scaled up or down based on observed traffic. The default value is "not scalable".	East/West/North	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 Protection Regulation (GDPR)).	East/West/North	Optional
Subscriber Availability	Indicates which subscribers the application is available to (e.g. only to subscribers on Home OP, to inbound/outbound roaming subscribers in a specific operator or country, all subscribers, etc.). If not provided, no restrictions on availability should be assumed.	East/West/North	Optional

Table 3: Information elements in the Application Manifest data model

The following table is the model of the Edge Application profile

Data type	Description	Interface Applicability
Edge Application ID	The ID of the Edge Application running on the edge node	East/West/North
Edge Application IP address(es)	The IP address(es) of the Edge Application running on the edge node	East/West/North
Edge Application status	The status of the Edge Application running on the edge node	East/West/North

Table 4: Edge Application profile

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

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 can successfully run if instantiated into a cluster containing at least the resources specified.

Flavours are not standardised (at this time) in this document. Federated Operators and OP Partners should undertake to produce and maintain a consistent Flavour catalogue.

Data type	Description	Interface Applicability	Optionality
Computing resource requirements	The computing resource requirements of the Edge Application, including whether the resource should support Containers or VMs	East/West/North	Optional
Storage resource requirements	The storage resource requirements of the Edge Application	East/West/North	Optional
Network resource requirements	The network resource requirements of the Edge Application	East/West/North	Optional
Extension requirement	The extension requirements of the Edge Application	East/West/North	Optional

Table 5: Flavour profile parameters

In the data model, a QoS description characterises the traffic between an Application Client and an Edge Application 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 common (or are expected to be common) in 5G networks. The reader is directed to 3GPP 23.501 [10], 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, the length of which 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 essential.

Based on this discussion:

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

Note: Considerations of QoS from UE to UPF, and the definition of QoS classes from UPF to backend application, require further investigation.

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

Table 6: QoS profile

3.4.3 Cloudlet

The Cloudlet is where the application is deployed. The Cloudlet data model (Table 7) provides the required parameters to deploy applications in the Cloudlet. Therefore, the Common Data Model of Cloudlet involves Cloudlet ID -for the OP to manage the Cloudlet- and Cloudlet Profile.

Data type	Description	Interface Applicability
Cloudlet ID	The FQDN defining the Cloudlet of where the Edge Client shall connect.	East/West
Cloudlet Profile	Gathers the Cloudlet information (e.g. ID) and characteristics (e.g. storage, GPU support, etc.)	East/West

Table 7: 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), etc. The data model of the edge application includes Edge Client ID, Edge Client IP address and Edge Client Profile. There may be multiple Edge Clients on a single UE, and a separate data module may exist for each.

Data type	Description	Interface 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 8: Common Data Model of Edge Client

3.4.5 Resource

A resource can be provided by cloud and edge. The Common Data Model of resource properties includes the resource's type, capacity, location, and state.

Data type	Description	Interface Applicability
Resource name	The name of the resource	East/West/North
Resource type	The type of resource	East/West/North
Capacity	The capacity of the resource	East/West/North
State	The state of the resource (e.g. running, hibernated)	East/West/North
Availability Zone	The associated availability zone	East/West/North

Table 9: Common Data Model of resource properties

3.4.6 Availability Zone

The Common Data Model of Availability Zone includes the compute resources, the supported virtualization technology, the QoS parameters supported and the associated costs.

Data type	Description	Interface Applicability
Availability Zone Name	The name of the availability zone	East/West/North
Region identifier	Geographical identifier	East/West/North
Compute resources	Flavours (e.g. CPU, memory, storage, in/out bandwidth)	East/West/North
Specialized compute offered	Particular compute resources (e.g. GPU, VPU, FPGA, NPU)	East/West/North
QoS	Maximum values of latency, jitter, packet loss ratio	East/West/North
Supported virtualization technology	VMs, containers, both	East/West/North
Costs	Costs associated with the use of the resources	East/West/North

Table 10: Common Data Model of availability zone properties

3.4.7 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	Interface 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 a different format (e.g. a token) bound by the OP to ensure user privacy.	UNI/North/East/West/South
UE location	UE location indicates where the UE connects to the network. For a UE in a mobile network, this is expected to be tied to a relatively static element, such as a data session anchor or mapped Availability Zone, rather than a granular location identifier. 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/North/East/West/

Data type	Description	Interface Applicability
UE Home OP	The ID of the Home OP of the UE	UNI/East/West/

Table 11: Common Data Model of UE

3.4.8 OP

The Common Data Model of Operator Platform includes the OP ID.

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

Table 12: Common data model of Operator Platform

3.4.9 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	Interface 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-NR
NEF/SCEF IP address	The IP address of the SCEF or NEF against which the operator platform shall connect	South-NR

Table 13: 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 with the advantage of better QoS. Notably, the Edge Cloud provides lower latency in a geographical Region or Regions that 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. In addition, the NBI allows an Application Provider to reserve a set of resources or request an Edge Cloud service with the resources and features that they require 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 is 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.

1. 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.
2. 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).
3. 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.
4. An Application Provider shall not have visibility of the exact geographical locations of the individual Cloudlets and shall not be able to request deployment of its application on a specific Cloudlet. Instead, the OP shall offer to Application Providers the edge cloud service in Availability Zones. The OP chooses each Availability Zone's size and which and how many Cloudlets it would use to provide its edge cloud service in each Availability Zone.
5. The NBI shall provide a request-response mechanism through which the Application Provider can state a geographical point where a typical user would be and then be informed of the expected mean latency performance. 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.
6. 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).
7. 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 resource types as they are defined.
8. 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 the current availability or usage of the resources.
9. The NBI allows the OP to accept or reject the request but not to negotiate.
10. The NBI allows an Application Provider to upload its application image to the OP. In addition, the NBI enables an Application Provider to delete its application image.
11. 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 are Created, Read, Updated and Deleted (CRUD).
12. 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.

13. The NBI allows an Application Provider to specify whether their edge application requires service availability on visited networks (that is, when a UE roams away from its home network operator) and on which visited networks the service should be available.
14. The NBI allows an Application Provider to specify whether service availability should be provided to non-roaming subscribers (that is, to UEs in their home network).
15. 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 require (and different OPs offer) different degrees of performance information (how fine-grained and how often), the NBI shall provide a request-response mechanism to allow an Application Provider to request a particular granularity of the telemetry. Similarly, the NBI shall provide an Application Provider with information about faults that (may) affect its edge cloud service.
16. 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 Availability Zones 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.

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

    a) Resource reservation (or pre-reserved resources association to the new Application Instance) 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 auto-scaling if required.

18. The other processes of lifecycle management of Edge Applications should follow a similar pattern.
19. 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 application 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 specified by the Application Provider and used by the Orchestrator.

20. 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.

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

3.5.1.2 Resource Requirement Specification

1. The OP shall enable Application Providers to express the resource (e.g., compute, networking, storage, acceleration) requirements of an application running on a Cloudlet.
2. 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 the 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. Thus, 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 upon among the operators in the federation.
ii. The flavour definitions shall be defined in the 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 an OP should provide consistent "fallback" behaviour when Flavour catalogues between OPs are not consistent.
v. The protocols and APIs provided by an OP should provide consistent "fallback" behaviour when 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. For example, these attributes could support sequestering 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

1. The NBI shall allow applications providers to access the resource catalogue.
2. The Resource catalogue shall consider local resources.
3. Resources footprint shall be abstracted to Availability Zones, preserving the network topology hiding as stated in sections 2.1.2 and 2.1.4.
4. 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 a 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. Therefore, 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).

GSMA PRD NG.126 [9] provides, in its sections 2 and 4, a more detailed overview of data elements that can be covered for the Edge Application Profile.

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

1. **Executable Image**

   A URI (or another similar name) identifying the executable image that should be deployed on a VM or as containers and be installed and executed by the OP.

2. **Resource Flavour**

   A Flavour is a description of a set of resource requirements used by an application instantiation. It should have a name that identifies the description uniquely and globally across OPs in an OP system.
   A resource description should be consistent with those appearing in Flavours available in public clouds. This requirement means that a Flavour should specify CPU, memory, storage, I/O bandwidth, CPU architecture, special hardware (e.g., accelerators), and, for VMs, the Hypervisor supported.
   A Flavour definition ensures that if an Application Provider selects a Flavour for a manifest, the application should successfully run if provided with at least the resource described in the Flavour.
   Flavours are not standardised (at this time) in this document. Therefore, 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.**

3. **QoS Requirements (optional)**

   A QoS description characterises the traffic between an Application Client and an Edge Application 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**, the 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.

4. **Application Session Migration Policy (optional)**

   The NBI allows an Application Provider to specify their support for a stateful or stateless Edge Application, i.e. whether the Edge Application can be moved from one

edge compute resource to another and this with or without prior notification. In addition, the NBI allows an Application Provider to specify additional mobility-related policy requirements:

- Application mobility allowed/restricted
- Application mobility prior notification required

5. **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).

6. **Application Scaling Policy**

   A scaling policy indicates whether an application can be scaled up or down based on observed traffic.
   The NBI shall support setting the scaling policy, based on the Application Provider's criteria, when creating an application instance and the ability to switch to another scaling policy when it is necessary.

7. **Edge Application Mobility 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).

8. **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:

1. Create application instances;

   The input parameters of an application instance include:

   a) URL of the image for the Application that is to be deployed <required>;
   b) Deployment related constraints, e.g. Availability Zone, multiple instances (for resilience), etc. <optional>.

2. Update application instances;
3. Query application instances;
4. Delete application instances.

3.5.1.6 Image Management

An Application Provider deploys the application by providing an image for containers (per section 3.6) or VMs (per section 3.7). They upload the image to an image repository and use its URL to deploy as containers or VMs.

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

1. **Upload images;**
2. **Update images;**
3. **Download images;**
4. **Query images;**
5. **Delete images.**

3.5.1.7 Network Event Support

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

The NBI shall expose network information towards Application Providers and application instances so 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.

3.5.1.8 CI/CD functionalities

An OP shall allow Application Providers to integrate the edge environment in their existing development pipelines.

The services exposed by the OP shall include in the API:

- Support cloud-native deployment systems, e.g. Helm.
- Expose internal repository API to:
    - Update application version
    - Update application image
    - Update application deployment artefact
- Support for multiple deployment strategies, for instance:
    - Basic deployment (all services and instances updated)
    - Rolling deployment (phased update of instances and services)
    - Blue-green deployment (staging-production update)
    - Canary deployment (only one small segment of final users updated)
    - Any other requested by the Application Provider.
- Support for following and controlling the deployment process, allowing KPIs monitoring and rollback.

- Support of additional services like GitOps, for facilitating application provider CI/CD integration.

3.5.1.9 Cloud Infrastructure as a Service (optional)

The Northbound interface may support additional exposure of the cloud infrastructure managed by the OP so that Application Providers can access similar infrastructure services to those provided in a traditional public cloud. Then, the OP enables a distributed cloud service with the same features as a traditional cloud but with more granular deployments.

The OP may get in charge of securing the access and controlling the amount and type of resources that can be retrieved, based on their availability. Therefore, the specific features, infrastructure type, and APIs that should be used depend on the OP's SBI-CR and the available resources in each situation.

Note: It is clear that all the enhanced features that the OP is providing to the edge service, such as mobility, federation or smart allocation, cannot be available on this kind of IaaS.

3.5.1.10 Resource Reservation

Independently of the applications that they are deploying, an Application Provider may require reserving a specific set of resources so that the OP guarantees its availability in any situation, even in resource congestion due to punctual application overuse. The OP shall ensure that the Application Provider can deploy any application within the limits of their reserved resources in a particular availability zone.

1. The OP shall enable Application Providers to express the resource (e.g., compute, networking, storage, acceleration) requirements that the Application Provider wants to be guaranteed.
2. The NBI shall allow an Application provider to request a set of resources to be booked, specified as Resource Requirements Specification (RRS), including the availability zones where the resources shall be located.
3. The NBI allows an Application Provider to reserve resources ahead of the application onboarding and unrelated to any specific application, only related to the Application Provider themselves. The NBI allows an Application Provider to consume the reserved resources when onboarding a new application, creating the association between the resources and the application (resources allocation). The NBI allows an Application Provider to delete their reservation.

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. Therefore, the SBI includes interfaces for:

- Infrastructure manager functions of a cloud or edge cloud infrastructure (e.g. resource management for compute and network resources);

- Orchestrator functions facilitating the application lifecycle management and scheduling;
- Service management functions (e.g. platform services, network services, mobility support, etc.);
- Other external functions that are providing services to the OP.

In many cases, close interworking between resource management, application 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 uses 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 use 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 systems to be consumed.

In addition to the management of the virtualised resources, 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.

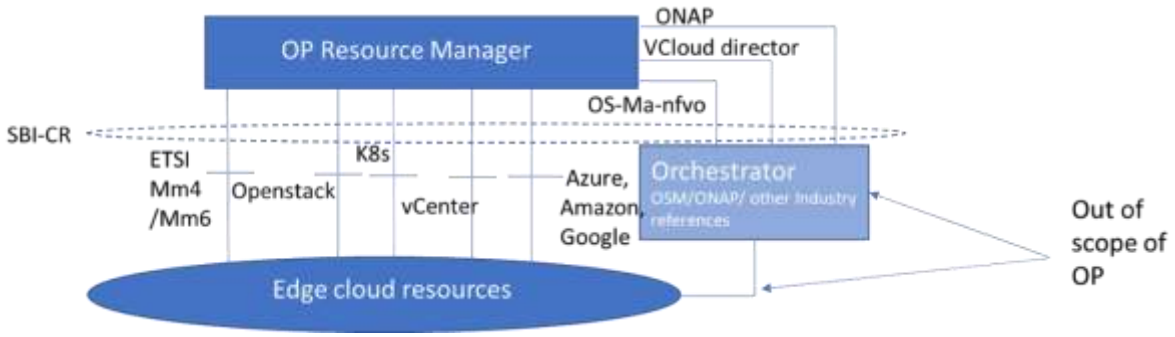

Figure 4: Possible SBI CR integrations

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 can be implemented internally by the OP or externally, e.g. consuming ETSI ISG MEC or ETSI ISG NFV via the SBI or by a combination.

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 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 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 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 (indirect and abstracted) access to some of the APIs to authorised applications. For example, some services, namely the Location Service, Radio Network Information Service (RNIS) defined by ETSI ISG MEC and available over the ETSI APIs, can be exposed in simplified abstractions to applications that provide location-aware features to end-users.

3.5.2.3 SBI-CHF

The operator that runs the OP decides on its commercial model and how it charges for OP services. There are many potential choices. Two simple examples are subscription-based and pay per use, whilst a more complex example is demand-based pricing. The OP architecture, therefore, defines various information to support a variety of commercial models. However, a particular commercial model may only require a subset of the information, while another may require additional details. When a service uses federated

resources, the two operators need to agree in advance on what charging information to report. Note that this is independent of the commercial model between the application provider and its OP.

Finally, OP shall expose all of that information to an external charging engine through an SBI for charging (SBI-CHF) under Operator or resource owner control so that each stakeholder can define its commercial strategy, models and offers. This interface shall be exposed from the Service Resources Manager role, as it is the cloud and network resources manager.

3.5.3 User to Network Interface

3.5.3.1 General Requirements

1. 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.
2. The UNI shall allow the communication between the User Client on the user equipment and the Operator Platform.
3. The User Client should be implemented on User Equipment software, e.g. through an SDK or OS add-on.
4. The UNI shall allow the User Client to discover the existence of an Edge Cloud service.
5. The OP's UNI shall allow the user client registration process with the Operator Platform's 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 the operator.
    c) It enables the User Client’s usage tracking. For example, to support integration with the network operator’s billing infrastructure.

6. The OP's UNI shall allow the user client to trigger the selection of a Cloudlet by the OP.
7. The OP's UNI shall allow the user client to trigger the instantiation of an application instance on the selected Cloudlet.
8. The OP shall measure network performance metrics for tracking the average latency characteristics of the edge network.
9. Based on metrics and location information, the User Client may request through 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:

1. the UE and the OP;
2. the User Client and the OP;
3. the Application Client and the Edge Application;
4. the operator Network and the Edge Application;
5. 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 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 learning the end user’s UE location, which must be done via the operator. The SRM is a service trusted by the operator network, allowing it to learn authenticated identity and location.

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

The mechanism shall ensure security, privacy and commercial confidentiality. An obfuscation technique, such as opaque tokens, shall be used to support the end-user's privacy.

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 establishes a chain of trust.

An example of such a service is “verify location”. The "verify location" input 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 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.

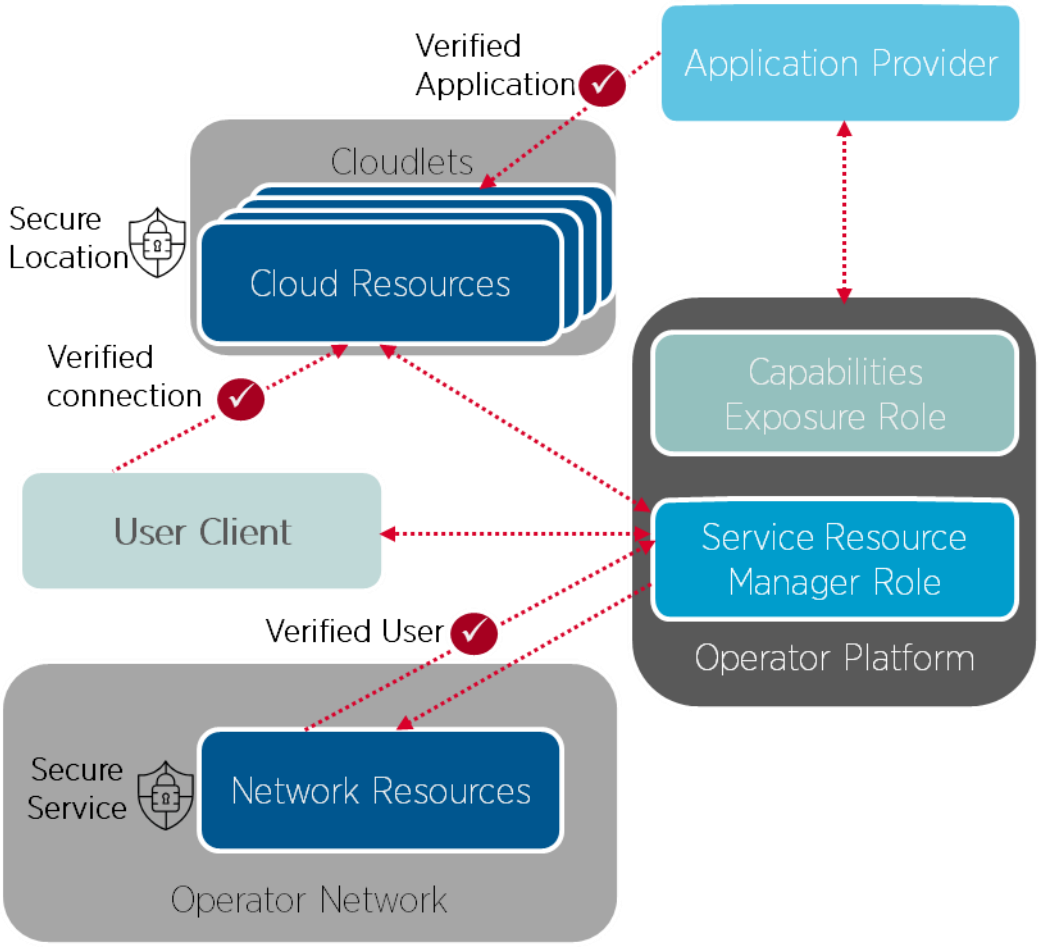

Figure 5: SRM as a trusted service: High-level Diagram

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 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 specific information on the Availability Zone corresponding to one OP's Edge Cloud resources provided to another.

The Availability Zone information shared over E/WBI shall provide a partner OP information about which zones are shared with that OP, where 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 and Resources 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 application images (container per section 3.6 or VMs per section 3.7) and Application Provider criteria towards a partner OP. The procedure may also request the launch of application instance(s) in partner OP edge clouds as a follow-up action after onboarding.
- Transfer of other application-specific files, e.g. application manifest, specifying the workload information like mobility strategy, QoE and privacy policies, also other optional characteristics indicating the application's needs (flavours, latency, prioritization, reservation)
- Publishing of application information to support the Edge Node Sharing scenario (as described in Section 3.5.4.3.3).

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

- Submitting 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 application information towards a Partner OP (e.g. application versions, Application Provider criteria, target availability zones).

3.5.4.3.2 Resources Reservation Management Service

An OP E/WBI shall use the Resources Reservation Management Service over E/WBI to reserve resources towards another OP.

The reservation service shall include transferring the Resource Requirements Specification of the Application Provider towards the Partner OP.

Note: Using this service by operators to reserve resources for their own purposes is for further study. E.g. ensuring SLA to certain Application Providers or roaming assurance.

3.5.4.3.3 Edge Node Sharing 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). Like the scenario discussed in section 3.3.5, this decision may be due to the Operator's policy controls, specific Application Provider restrictions, due to constraints originating from the federation agreement between the Operators and others.

An E/WBI service is required to support the publishing of application and Availability Zone information to enable specific applications to be served from a Leading OP’s Edge Cloud 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 need to be served from the Home OP for consumption by roaming UCs;
- In a non-roaming scenario where an OP needs to allow its own UCs, the consumption of applications published by a Partner OP served from that partner’s Edge Cloud.

The E/WBI service shall support the following information:

- Publish Application, including application metadata information (including information about the policies controlling application distribution restrictions)
- Availability Zones;
- Unpublish application; to cancel the availability of published application(s)
- Get a list of Applications; for an OP to retrieve the list of published application instances with specific criteria (e.g. edge location, availability zone, etc.)
- Get Application instance information; for an OP to retrieve the application instance information in the “Edge Application profile” as part of the Common Data Model in section 3.4.2. Then, the OP serving the subscriber can use that information for sharing connection parameters with the User Client (e.g. application IP address or access token).

Note: this document assumes that the application deployment information (i.e. manifest, criteria, and flavour profile) is available on the partner OP.

3.5.4.3.4 Application Deployment Management Service

An OP shall use the Application Deployment Management Service to control the 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 OPs;
- Termination of running application instances from select Partner OPs.

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 the information about the resource levels 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 consumption reaches 80% of the agreed threshold value);
- Imminent Federation Agreement expiry.

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.
- Authorisation information for User Clients.

Note: In this version of the document, it is assumed that the applications available to roaming subscribers have been provided to the Visited OP through a federation including both OPs. Future versions of this document may extend to roaming outside of a federation.

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

- Setup Service Availability in Visited Network related parameters towards partner OPs;
- Update Service Availability in Visited Network related parameters towards partner OPs;
- Enable UC authentication information and provide authorisation for a visiting UC from the Home OP.

3.5.5 Local interface on an end-user device

Using edge computing through the Operator Platform should be as easy as possible from a developer's perspective. As envisioned in the OP architecture, the UNI interface between UCs and the OP exposes specific APIs needed for, for example, discovering and connecting Application Clients to the edge nodes and enabling the requested services. However, most of these procedures require multiple interactions that are not specific to the application (e.g. registration). Thus, these procedures would benefit from being provided through a common implementation; the Application Client accesses that through a device-local interface (see Figure 6).

Note: By nature, such a common implementation would be device platform-specific; see section 3.5.5.2 for some considerations.

The requests to these UNI APIs may also contain specific privacy-sensitive parameters, e.g. location of the UE (Latitude/Longitude), network attachment location information CellID/Tracking Area Code (TAC), etc. (see also section 3.5.5.1). These parameters are typically maintained within the device platform (e.g. Android, iOS etc.). Based on the platform design, application permissions and philosophy, the applications on the device get access to some of these parameters.

Thus, implementing the OP UNI would require access to some of these parameters available from the underlying device platform. However, if the OP UNI is exposed to the Application Clients through common libraries or a runtime, access to those parameters can be handled within that common implementation which may avoid exposing sensitive information to the Application Client. The interface between the Application Client and this common, device platform-specific implementation is referred to as "local interface on an end-user device".

There can be different ways an Application Client developer can be provided with access to the UC to consume OP services using UNI APIs. Examples could be:

- having an OP Edge Client SDK for building UNI APIs and functions that a developer can integrate with their application business logic or
- a thin client application on the device aggregating the UNI access (UNI aggregation) of different Application Clients.

Note: Use of a common runtime aggregating the UNI may not be possible on all platforms without the support of the platform provider, but may be required to fulfil (potential future) requirements such as a single registration to the OP per UE rather than registering every UC separately. Therefore, cooperation with the platform providers is recommended for the long term, even if common implementations would have to handle existing platform limitations for the short term.

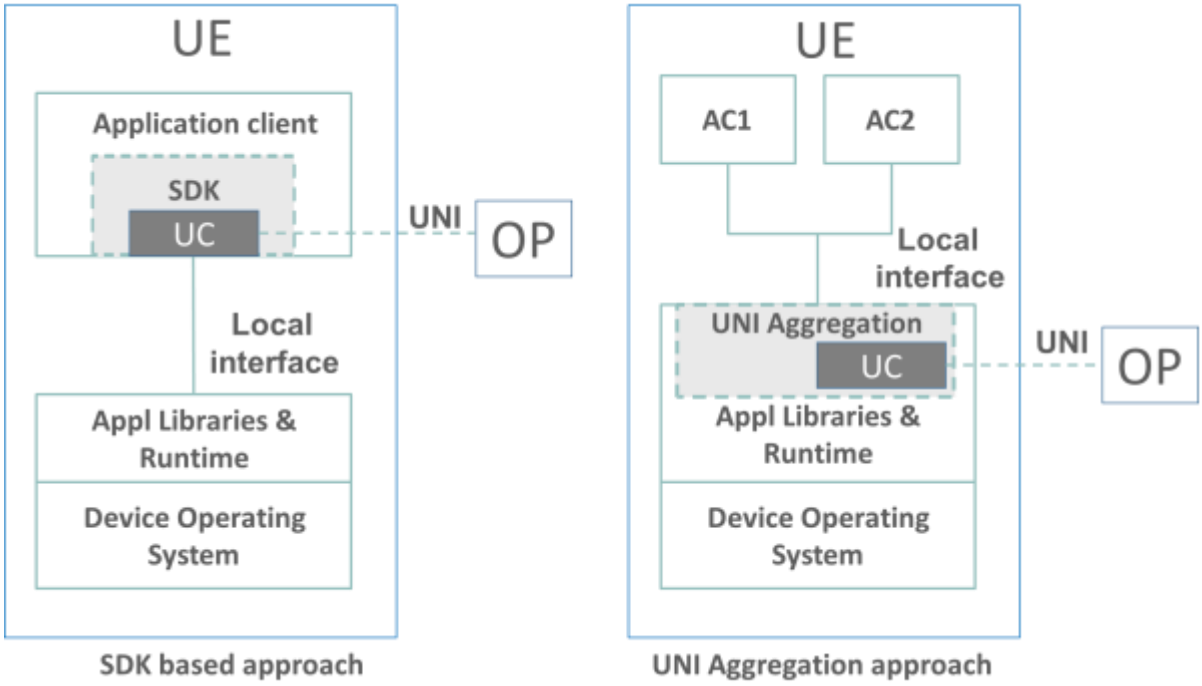

Figure 6: OP Device side architecture (local interface)

Note: As per two of the possible approaches for building UNI support for Application Clients, i.e. SDK and the UNI aggregation, Figure 6 represents the conceptual placement of the two enabler components in relation to the device platform without elaborating on the merit of one over the other. There could be other approaches, but not all have been explored yet.

3.5.5.1 Privacy sensitive parameters for UNI

The UNI requests from Application Clients on end-user devices, as described above, require access to specific privacy-sensitive parameters available from the device platform or the OP. The following list provides an indicative, non-exhaustive overview of such parameters:

- Subscriber identity and credentials for authentication, e.g.
  - MSISDN,
  - GPSI,
  - Token for authentication,
  - SIM credentials
- Geo-Location information
  - Latitude/Longitude
- Network Information
  - Home MCC/MNC,
  - Visited MCC/MNC,
  - Cell-ID, TAC etc.,
  - Wi-Fi SSID and Access Point MAC address

Note: Some of these parameters would be available to the OP through the SBI-NR. So it is up to the detailed UNI definition whether they are required in the UNI requests.

These parameters would be used in the UE's UNI API requests to perform functions, e.g. edge discovery, application endpoint exposure, application location verification, measuring and reporting network performance metrics, etc.

3.5.5.2 Key considerations for architectural requirements on the local interface

The client applications or UCs on the end-user device would need access to the OP UNI interface for consuming OP provided edge services. There are various possibilities for providing this access using a common implementation where each possibility would come with associated advantages and shortcomings. When designing and developing a feasible solution for this common implementation and the local interface that it offers to the Application Clients, there would be main guiding principles to be taken into account:

- Functional parity across multiple device platforms
- Short evolution cycles
- Must meet OP security and data privacy principles on the UNI interface
- Keeping Application Client developers agnostic to mobile network-related aspects

Note: Support for features like mobility, roaming, network slicing, session continuity etc. in the context of device clients is for further study

3.6 Containers

3.6.1 Description

The OP intends to provide developers with a consistent application deployment environment independent of the network and OP platform in which they deploy their applications. The goal is to establish requirements for interoperability and federation between OPs.

The following areas and their requirements have been identified as the baseline to ensure a consistent environment across OP platforms:

- Container Image
- Container runtime compliance
- Cloudlet Host OS
- Cloudlet CPU architecture

3.6.2 Container Image and Repository format

The OP shall support the Open Container Initiative (OCI) Image-spec [7], specifying how container images are bundled.

3.6.3 Container runtimes

The OP shall support the Open Container Initiative (OCI) Runtime-spec [8] for container applications on Cloudlets. The Runtime Specification outlines how to run an “OCI Image bundle” unpacked on a disk.

3.6.4 Cloudlet Host OS

A Cloudlet shall support a Linux Kernel as Host OS to run containers.

3.6.5 Supported Architectures

A Cloudlet shall support x86_64 CPU architectures to run containers.

3.7 Virtual Machines

3.7.1 Description

As indicated in section 2.1.2, the OP shall support applications relying on VMs. The OP intends to provide developers with a consistent application deployment environment for VMs independent of the network and OP platform in which they deploy their applications. The goal is to establish requirements for interoperability and federation between OPs.

Next to some more generic requirements covered in the following subsections, a minimum alignment is needed between the OPs in a federation on the following areas to ensure a consistent environment across OP platforms regarding Virtual Machine support:

1. VM based application Image & metadata format
2. VM runtime environment
3. Accelerator support: SRIOV, DPDK
4. Specific HW features support: GPU, FPGA, etc.
5. Performance Optimisation Capabilities: NUMA, CPU Pinning, use of dedicated core, Affinity/non-affinity, etc.

3.7.2 Guest OS support

The Guest OS shall be assumed to be part of the VM Image.

3.7.3 CPU Architecture support

A cloudlet shall support x86_64 CPU architectures to run the VMs.

3.8 Serverless

3.8.1 Description

Serverless computing is a platform that hides server usage from developers and runs code on-demand automatically scaled and billed only for the time the code is running [12].

The OP intends to provide developers with a consistent serverless application deployment environment independent of the network and OP platform in which they deploy their applications. The goal is to establish requirements for interoperability and federation between OPs for serverless containerised applications. In this context, ‘workload’ refers to the application component deployed on the serverless compute.

3.8.2 Serverless Architecture

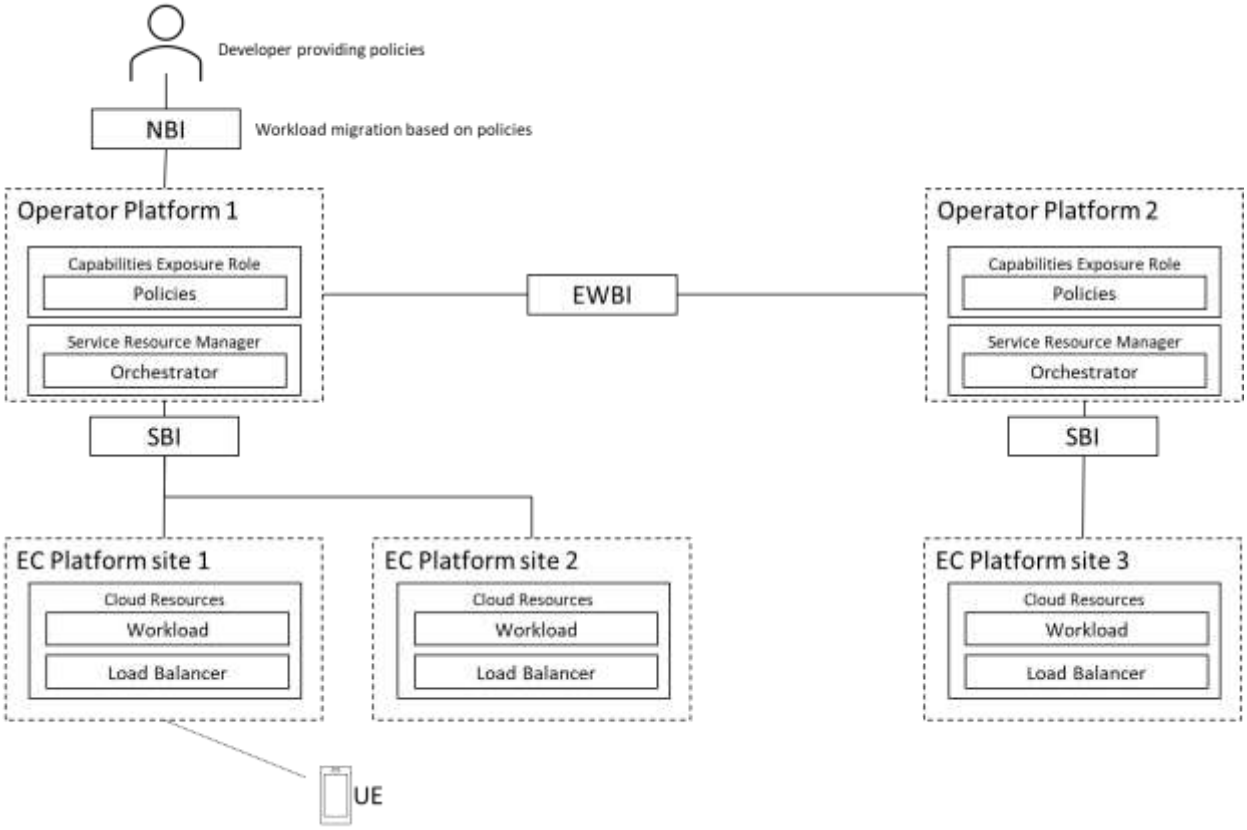

Figure 7: Serverless architecture

The following are the main components of a Serverless solution:

- Policies

  Ingesting and controlling policies set by the developer to establish scaling/migration thresholds. See section 3.8.4.4 Policies.

- Orchestrator

  Scaling in/out of container applications from zero based on developer and OP policies. Migrating workloads to the appropriate point of presence on an Edge Computing Platform, again based on policies.

- Load Balancer

  Load Balancer of connections. The Load Balancer is part of the OP's SRM. It is physically located in the Edge Computing Platform to act as a proxy and gateway, forwarding a workload request to the Point of Presence and the Orchestrator. That can be potentially extended to listen to a broader set of events and traffic.

- Edge Computing Platform

  Edge Computing Platform (ECP) has the point of presence sites that are discoverable by the UC. It hosts the Load Balancer. The ECP point of presence has one or more Cloudlets.

One ECP point of presence is used as a serverless application's “homebase”. The Orchestrator and policies are provided in the "homebase". The location of the “homebase” is solution dependent and may be defined by the developer or by the OP.

Note: It is assumed that the traffic from the UE is directed to the closest ECP point of presence.

Note: It is assumed that there is network connectivity between ECP sites.

3.8.3 Lifecycle

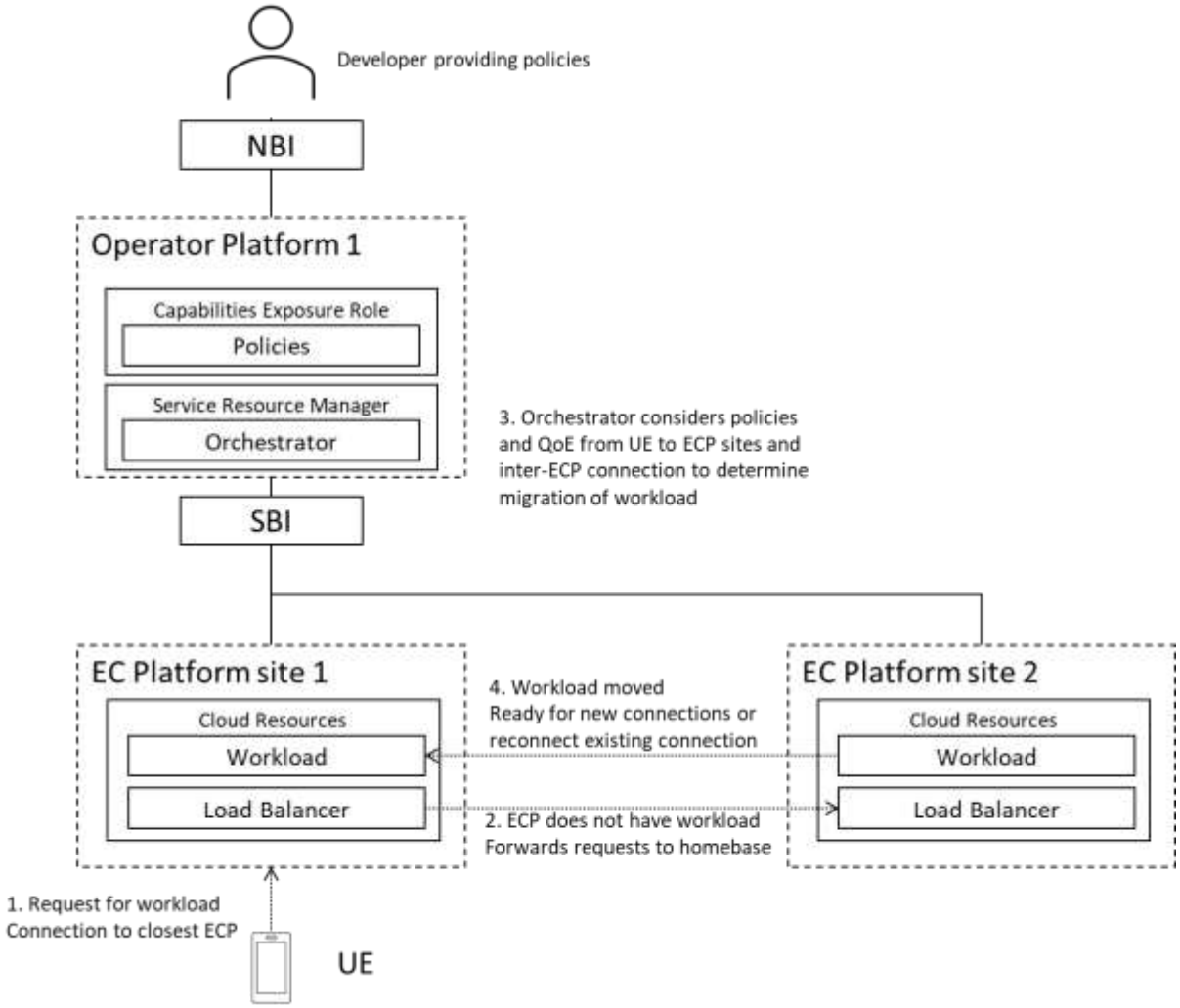

Figure 8: example sequence of a serverless lifecycle

An example sequence of a serverless lifecycle:

Note: The sequence below may be changed.

1. Application Provider providing policies for the application.
2. Connections reaching the closest ECP point of presence (ECP site 1).

3. The requested workload is not present on the closest ECP point of presence, so the request is forwarded to the “homebase” ECP point of presence (with the ECP Load Balancer acting as a proxy forwarder).
4. At first, the application on the “homebase” ECP point of presence (ECP site 2) starts to serve the UE through the target proxy. Secondly, based on developer policies, the Orchestrator determines the need for migration of the application to the target ECP point of presence (ECP site 1).
5. Based on the policies, the Orchestrator migrates the application to the closest ECP point of presence (ECP site 1). From then onwards, the target proxy Load Balancer serves the UE from the application instance at the local (closest) target ECP point of presence.

3.8.4 Architectural Components & Considerations

3.8.4.1 Application Packaging

Serverless applications shall be packaged as containers according to the container definition in section 3.6.

3.8.4.2 Serverless event

The OP shall support connection events to determine the number of concurrent sessions and devices.

3.8.4.3 Orchestrator

The Orchestrator shall be capable of instantiating and scaling applications/containers based on the Application Providers' and OP policies.

3.8.4.4 Policies

Developers shall create policies for the orchestrator to define the scale-in/out and migration of serverless applications.

The following developer policies shall be supported:

- The number of concurrent connections per application instance. Informed through connections request on the Load Balancer.
- The number of concurrent sessions on an ECP point of presence (as seen by the Load Balancer proxy).

4 Service flows

4.1 User Client (UC Registration) - Home Operator platform

This procedure describes the registration between the user client and the Operator Platform, allowing the user client to be authenticated and authorized to access the service.

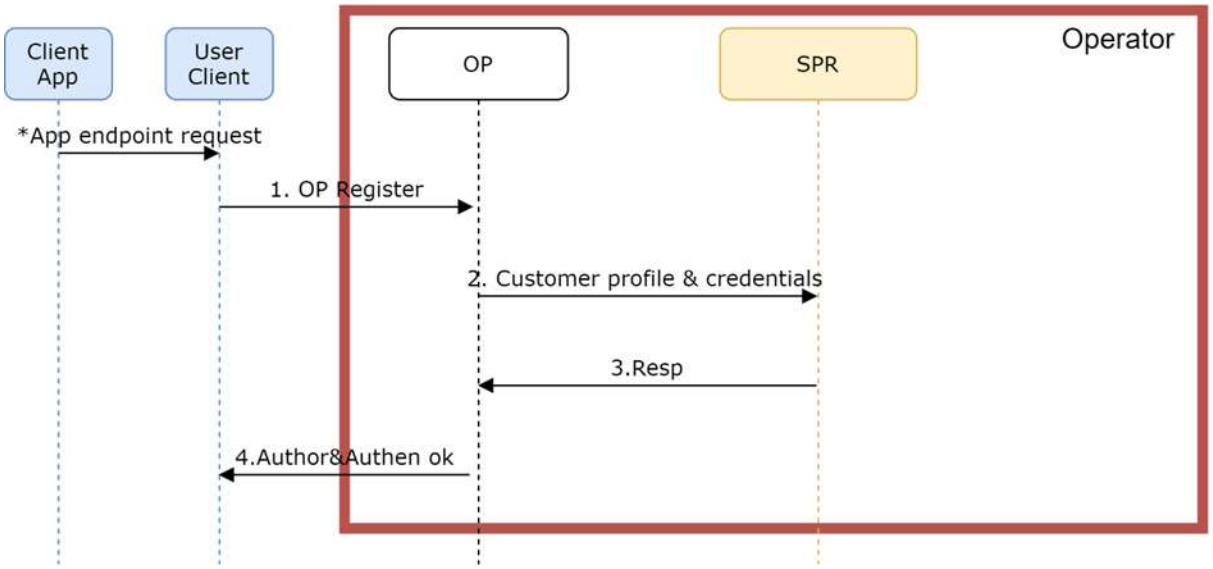

Figure 9: User Client (UC Registration)- Home Operator platform

1. A user client (UC) on a UE tries to register on its home OP. This request can be triggered by a cloudlet discovery from the application on the device. The register request is driven to the OP UNI of the operator hosting the user, whose URL is composed using the unique network operator identifiers, MCC & MNC. E.g. config.edge.mnc<MNC>.mcc<MCC>.3gppnetwork.org
2. From this registration request, the OP derives a request for profile and credentials to the operator's Subscriber Profile Repository (SPR) endpoint, accessed through the SBI.
3. The OP validates the user access based on the information and credentials retrieved from the operator's SPR endpoint and the information and identities received from the UC in the registration request.
4. The User Client receives the authentication validation and is authorized to request OP services from that moment onwards (e.g. cloudlet discoveries).

Note: Other authentication/authorization methods like UC redirection to an external entity can also be considered.

4.2 User Client (UC Registration) - Visited Operator platform

This procedure describes the User Client registration with an Operator Platform while accessing the service from a visited network. For such cellular roaming, two models exist as defined in section 3.3.4:

1. Home routing, for scenarios where edge services provided by the visited network cannot be supported.

   The Home OP is the only OP involved in this case, with registration handled as defined in section 4.1. Figure 10 shows the relations between the networks in this case. This scenario comes with limitations on application availability due to increased latency (see section 4.5).

2. Local breakout, to access edge nodes available in the visited network. This model is preferred because the edge cloud service is provided closer to the User Client then.

In this case, the Home OP is involved managing the subscriber's authentication and authorisation, with the Edge Discovery provided by the Visited OP. While not a service flow because detailed interface impact hasn't been studied yet (see section 1.2), Figure 11 shows the relations between the networks in this case with the following clarifications:

- The black path (long dashes). Device registers on OP-A. OP-A steers the user to OP-B since the user is attached to Operator B, and the operators have agreed that LBO can be used.
- The yellow path (short dashes). The device is redirected to OP-B, gets authorised there and can request access to edge services (see section 4.5) provided based on the user's location.
- The red path (dotted). Federation connection for enabling the application availability on Operator B, sharing user's authorisation information
- The blue line (continuous): User access to the edge on Operator-B, accessing through the UPF-PGW in Operator B.

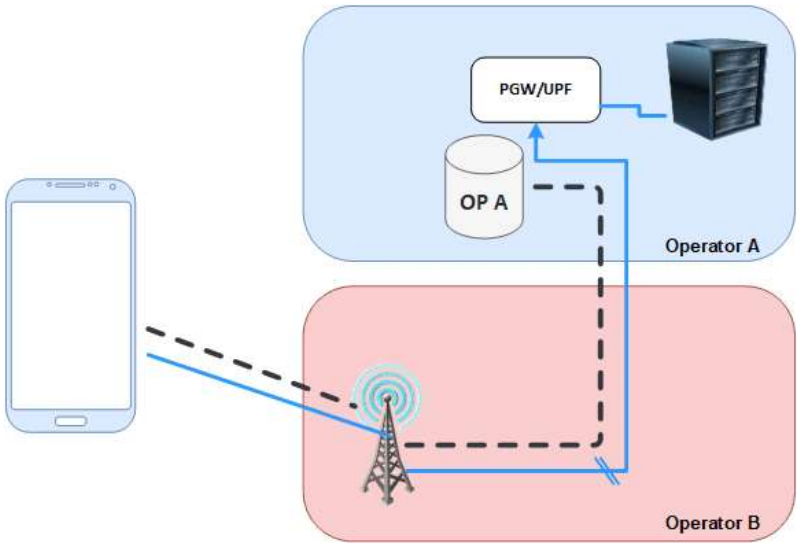

Figure 10: Roaming access to OP and edge resource - home routing

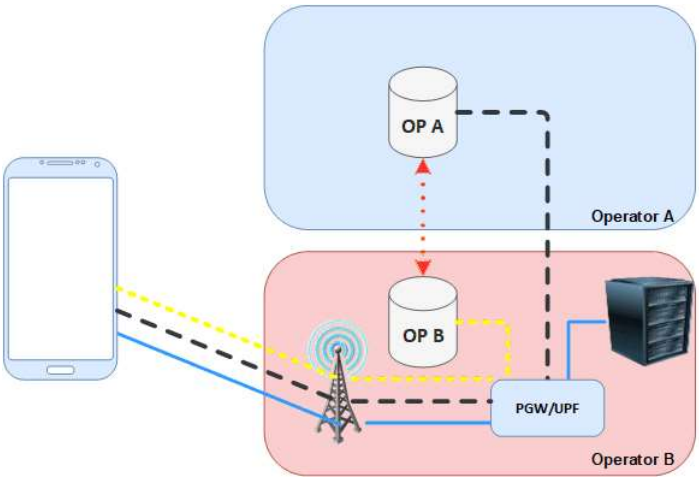

Figure 11: Roaming access to OP and edge resource – local breakout

4.3 Edge discovery in the home network

This procedure describes the edge discovery by a UC when the most suitable cloudlet is in the home network and may be provided in a future version of this document.

4.4 Edge discovery in an edge-sharing partner network

This procedure describes the edge discovery when the UE is physically attached to the home operator, but the most suitable cloudlet is in an "edge-sharing" Partner OP.

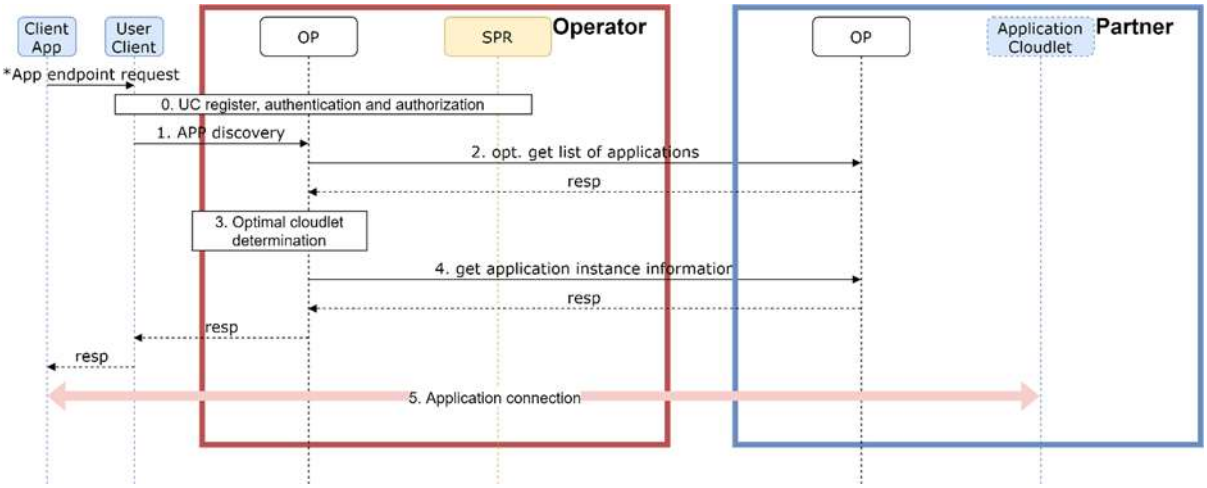

Figure 12: Edge discovery in an edge-sharing partner network

1. A user client (UC) on a UE requests a discovery query for a particular application. The UC previously registered with the OP as in the procedure described in section 4.1.
2. Optional. Operator's OP (Home OP) may trigger a discovery request for the applications available on the Partner's resources.

Note: The Partner OP may also publish those available applications independently of the User Client's interactions.

3. The Home OP determines the most optimal application locations, based on local and federated resources from the Partner, and determines that the user is best served by an application instance provided by the Partner OP.
4. The Home OP requests the Partner OP for the application instance information to allow the Home OP to provide the connection data to the UC.
5. The UC is provided with the connection data of the application instance and connects to it.

4.5 Edge discovery in a visited partner network

This procedure describes the edge discovery when the UE is physically attached to a visited operator and the most suitable cloudlet is in the Visited Partner OP. The two cases for the Registration in the visited network (see section 4.2) also apply to Edge Discovery. When using home-routing, the discovery is similar to the case described in section 4.3. The only difference is that some applications may not be available because their latency constraints cannot be satisfied in this home-routing case. For local breakout, the Visited OP handles the discovery using the authorisation information provided by the subscriber's Home OP.

4.6 Application deployment In the Home Operator Domain

This procedure describes the application deployment in a cloudlet of the operator domain and may be provided in a future version of this document.

4.7 Application deployment In the Federated Operator Domain

This procedure describes the application deployment in a cloudlet of a federated operator domain and may be provided in a future version of this document.

5 Requirements on interfaces and functional elements

This section defines the requirements of the interfaces and functional elements that make up the OP architecture. They should be fulfilled by solutions developed in SDOs (see section 6) and implementations provided by the open-source community.

5.1 Interfaces

5.1.1 Northbound Interface

5.1.1.1 High-level requirements

1. All Operators and Operator Platforms shall offer the Edge Cloud service through the same NBI.
2. 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.

3. 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. For example, profile-based access control such as 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 have access to all the functions listed below. For example, monitoring information would not necessarily be accessible during onboarding. In addition, 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 more granular profiles are likely in practice.

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 be 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 to meet the Application Provider’s request.

Thus, the deployment management shall allow onboarding and instantiating the application while meeting different criteria provided by the Application Providers and the operators that own the OP instance and the underlying resources.

The OP's NBI shall support applications depending on Containers and VMs that comply with the format criteria specified in sections 3.6 and 3.7, respectively.

5.1.1.2.2 Application Provider Criteria

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

1. Footprint/coverage area selection;
2. Customer reach/ operator selection;
3. Infrastructure resources:

    a) CPU;
    b) Memory;
    c) Storage;
    d) Hypervisor (for VM based applications);
    e) Networking definition used by the application.

4. Specific and optional requirements definition, for example:

    a) Use of GPUs;
    b) Use of FPGAs;
    c) Accelerator support: SRIOV, DPDK;
    d) Any other set of accelerators;
    e) Performance Optimisation Capabilities: NUMA, CPU Pinning, use of dedicated core, Affinity/non-affinity, etc.

    GSMA PRD NG.126 [9] provides, in its sections 2 and 4, a more detailed overview of data elements that can be covered for the Edge Application Profile.

5. Edge-Cloud requirements:

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

6. Type of application instantiation:

    a) Static: the application shall be deployed in several edge sites 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 requests 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 needs to be an instance per user or one instance per specific number of users.

7. Policies that allow the Application Provider to manage circumstances where user conditions do not comply with the deployment criteria.
8. Support for telemetry information from the operator.
9. 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.

10. 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:

1. 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).

2. An OP shall offer a range of quality policies so that an Application Provider 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).
3. The NBI shall enable a request-response mechanism through which the Application Provider can state a geographical point where a typical user could be and get informed of the mean latency performance expected.
4. 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 resource types).

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.

Note: The definition of geographical Regions should be aligned among the partners in a federation, ensuring a shared understanding of a Region.

5. The OP shall offer a structured workflow for application deployment and instantiation: CRUD functions.
6. The OP shall allow a developer to specify that its Edge Applications should be restricted to a particular geographical zone. This restriction would ensure compliance with the applicable data privacy laws.
7. 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).
8. The OP shall provide an Application Developer with telemetry information concerning the performance of the Edge Cloud service, including fault reporting.
9. The OP shall allow an Application Developer to request a particular granularity for the telemetry information they receive.

   Note: Possibly using a publish-subscribe approach.

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

10. The OP shall allow an Application Developer to require that outbound access to the internet is prohibited.
11. The OP shall offer Application Providers a registry to store their application images and update or delete them. The registry may be centralised or distributed, depending upon the Application Provider’s needs to reduce boot time and recovery.
12. The OP shall support Single Sign-on based on login credentials for an Application Provider.
13. 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.
14. The OP shall offer functionality that supports the Application Provider in managing the application development, integration and deployment.

5.1.1.4 Resource Reservation Profile

5.1.1.4.1 General

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

- receiving the request,
- authorising/authenticating the Application Provider, and
- gathering all the necessary data to reserve the resources based on the Application Provider criteria.

Thus, the reservation management shall allow reserving resources meeting different criteria defined by Application Providers. The operator owns the OP instance and underlying resources.

5.1.1.4.2 Application Provider Criteria

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

1. Footprint/coverage area selection;
2. Infrastructure resources:

    a) CPU;
    b) Memory;
    c) Storage;
    d) Networking resources.

3. Specific requirements definition:

    a) Use of GPUs.
    b) Any other set of hardware accelerators

4. Expiration time.

5.1.1.5 Security Requirements

Based on the attack surface analysis provided in Annex E, the following security requirements shall be considered:

1. The NBI shall provide an authentication mechanism to enable access only to authenticated and authorized entities. All interactions over the NBI interface shall use an application layer security protocol that runs over a reliable transport and guarantees mutual authentication between the OP (the Capabilities Exposure Role) and the Application Provider. This authentication shall rely on public-key based digital signatures backed by certificates issued by a commonly trusted certification authority.
2. The NBI shall provide an authorization mechanism to grant access to only the necessary authorised services and data. The NBI shall provide a fine-grained authorization mechanism to grant authenticated entities selective access to the NBI exposed services and functionalities. NBI shall use profile-based access control to provide appropriate restrictions on the amount of functionality that the NBI offers to a particular Application Provider, according to their operational profile and the type of access requested. When defining and assigning the authorisation profiles, the principle

of **least privilege** shall be applied, ensuring that any entity should have only the minimum **profile** roles necessary to perform its function.

3. The NBI shall provide security mechanisms to guarantee the confidentiality, integrity and authenticity of the exchanged data. The security protocol used over the NBI shall also guarantee security properties such as perfect forward secrecy and mechanisms to prevent intervening attacks, such as replay, relay, and man-in-the-middle attacks.
4. Given the external exposure of this interface, the NBI shall provide security mechanisms to counteract/prevent attacks aimed to undermine the availability of the interface, such as DoS and DDoS attacks, reconnaissance attacks (attempts to identify service or API vulnerabilities) and brute force attacks.
5. The NBI should provide isolation between resources of different Application Providers (e.g. when providing telemetry data or when accessing and managing Edge Applications configuration data).
6. The NBI should provide security mechanisms to protect accounting and guarantee safe logging (e.g. integrity, non-repudiation, etc.).

5.1.2 East-Westbound Interface

5.1.2.1 High-level requirements

1. 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

Based on the attack surface analysis provided in Annex E, the following security requirements shall be considered.

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

1. 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 Availability Zone served by one or more edge servers/nodes.
    b) Specific edge node information shall not be shared.
2. An OP shall only expose the resources previously agreed with each specific federated instance.
3. An OP shall be able to identify the User Clients among OP instances.
4. An OP shall be able to identify the Application Providers among OP instances.
5. An OP shall be able to identify the applications among OP instances.
6. An OP shall be able to act as a proxy for any interaction between operators’ networks, hiding any detail on the network architecture of the federated networks.
7. The E/WBI shall provide an authentication mechanism to enable access only to authenticated and authorized entities. Therefore, mutual authentication shall be provided between the instances of the OP.
8. The E/WBI shall provide an authorization mechanism to grant access only to the necessary authorised services and data.
9. The E/WBI shall provide a security mechanism to safeguard the confidentiality, integrity and authenticity of the exchanged data

10. The E/WBI shall provide security mechanism to counteract attacks aimed to prevent the availability of the interface, such as Denial of Service (DoS) attacks
11. The E/WBI shall support the adoption of strong security algorithms that guarantee forward secrecy and prevent intervening attacks such as replay, relay, and man-in-the-middle attacks.
12. The extended trust model of 3GPP TS 33.310 [24], calling for direct cross-certification of entities across different trust domains, should be followed across E/WBI.
13. The E/WBI should provide security mechanisms to protect accounting and guarantee safe logging (e.g., integrity, non-repudiation, etc.)

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.

1. The E/WBI shall allow forwarding the instantiation requests to any federated OP whose footprint has to be covered.

    a) The E/WBI shall support instantiation requests for applications depending on Containers and VMs that comply with the format criteria specified in sections 3.6 and 3.7, respectively.

2. 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 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 be responsible for the operator deployment criteria management.
    c) An OP receiving an instantiation request through its E/WBI shall be responsible for the edge node selection based on the Application Provider criteria and its operator's criteria.
    d) An OP receiving an instantiation request through its E/WBI shall be in charge of the application mobility management.

3. The E/WBI shall forward the application mobility notifications and procedures towards the Leading OP for management with the Application Provider.
4. The E/WBI shall forward the management procedures, information and statistics to be shared with the Leading OP of the Application Provider.
5. The E/WBI shall be employed for managing the service continuity on visited networks.

5.1.3 Southbound Interface to Cloud Resources

5.1.3.1 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 the cloud resources of the operator/cloud provider. This access shall allow the OP to fulfil request/response transactions regarding an

application's lifecycle, catalogue the resources/capabilities and get feedback about the status of the different Cloudlets or edge nodes.

5.1.3.1.1 Integration with Cloud Orchestrator

A cloud provider/operator may want to expose the cloud resources through an orchestrator. However, this integration does 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 do not allow the OP to enlarge or reduce the resources reserved for edge purposes. Furthermore, the limited information provided does 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 that an orchestrator offers to an OP are limited to the amount of resources used and the assigned orchestrator's tenant's scope.

OP SBI-CR integration shall allow adopting industry references for orchestrator integration, including but not limited to OSM/MANO, ONAP, VMware VCloud Director.

5.1.3.1.2 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 has additional functions. These functions 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 exposure 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 a 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 allow adopting industry standards for VIM/CISM integration, including but not limited to ETSI-ISG MEC / ETSI ISG NFV, Openstack, Kubernetes and VMware vCenter.

5.1.3.1.3 Integration with Hyperscalers

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

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

5.1.3.2 Security Requirements

Based on the attack surface analysis provided in Annex E, the following security requirements shall be considered:

1. The SBI-CR shall provide an authentication mechanism to enable access only to authenticated and authorized entities. Therefore, mutual authentication shall be provided between the OP (Service Resource Manager Role) and the CR.
2. The SBI-CR shall provide an authorization mechanism to grant access to only the necessary authorised services and data.
3. The SBI-CR shall provide a security mechanism to safeguard the confidentiality, integrity and authenticity of the exchanged data.
4. The SBI-CR shall provide security mechanisms to counteract attacks aiming to prevent data availability, such as DoS attacks.
5. The SBI-CR shall support the adoption of strong security algorithms that guarantee forward secrecy and prevent intervening attacks such as replay, relay, and man-in-the-middle attacks.
6. The SBI-CR shall provide security mechanisms to protect the live migration of services.
7. The SBI-CR shall provide security mechanisms to prevent attack from containers or VMs, of which Docker or VM Escape attacks are examples.
8. The SBI-CR shall provide security mechanisms for the SDN control plane.
9. The SBI-CR shall safeguard the protection and integrity of traffic steering parameters and controls.

5.1.4 Southbound Interface to Network Resources

5.1.4.1 General

The SBI-NR connects the OP with the specific operator infrastructure that delivers 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 access 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 provided 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 the 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 control home network resources.

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

5.1.4.2.1 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 13 shows a functional mapping that describes how an OP accesses features and services exposed by the NEF/SCEF.

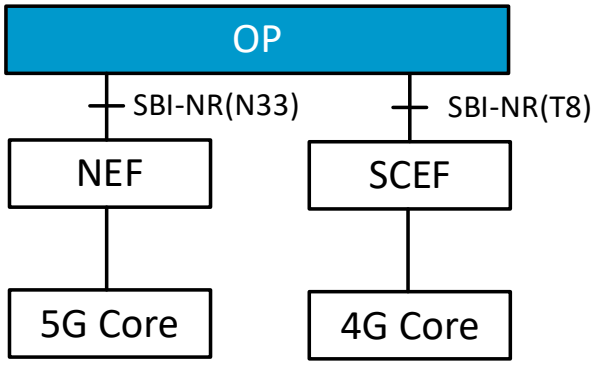

Figure 13: Functional mapping between OP and NEF/SCEF

5.1.4.2.2 General Requirements

1. An OP's SBI-NR shall be able to interact with 5G Core/4G Core via the NEF or SCEF to access network capabilities.
2. An OP's SBI-NR shall support the exposure interface [4] [5] for interacting with the 5G Core/4G Core.
3. If the NEF/SCEF returns an error response to an OP's SBI-NR, the OP shall perform error-handling actions.
4. An OP's SBI-NR shall be able to report the functionality available from the network.
5. An OP shall be able to deal with the situation where the network is not providing the expected functionality.
6. An OP's SBI-NR may be able to configure the user traffic to be routed to the applications in the local data network.
7. An OP's SBI-NR may be able to interact with the NEF for configuring and influencing the traffic routing policies.

    a) An OP may be able to specify the request for routing, influencing network mobility and routing, including but not limited to:

        i. UE and application identities
        ii. Traffic filtering and routing criteria,
        iii. Possible locations of the application instances
        iv. Whether the UE network data plane can be relocated.
        v. Whether validation on UE network data plane relocation is required.
        vi. Whether the UE IP address shall be preserved in data plane relocation
        vii. The type of SSC mode
        viii. Whether inter-operator handover is required.

    b) An OP may be able to subscribe to UE data plane mobility events.
    c) An OP may be able to receive UE data plane mobility events, receiving the target node identifier where the UE should re-attach because of the network mobility process.
    d) An OP may be able to receive UE data plane mobility events, receiving and processing the target IP of the UE that will be assigned.
    e) An OP may be able to negotiate the UE data plane mobility process based on the application instance relocation process.

8. An OP's SBI-NR may be able to collect information on network congestion or access concentration in a specific area.
9. An OP's SBI-NR may be able to retrieve UE status reports (e.g. location information, reachability, roaming status).
10. An OP's SBI-NR may be able to control the transfer of data in the background for UCs.
11. An OP's SBI-NR may be able to configure QoS session parameters to communicate with a UC with an improved QoS level (e.g. low latency, priority, maximum bandwidth).
12. An OP's SBI-NR may be able to receive QoS relevant notifications based on UE connection statistics.
13. An OP's SBI-NR may be able to configure the charging party of the UE data sessions.
14. An OP's SBI-NR may be able to configure service-specific parameters for UCs (e.g. network slice).

15. 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).

5.1.4.3 Security Requirements

Based on the attack surface analysis provided in Annex E, the following security requirements shall be considered:

1. The SBI-NR shall provide an authentication mechanism to enable access only to authenticated and authorized entities. Therefore, mutual authentication shall be provided between the OP (Service Resource Manager Role) and the NR.
2. The SBI-NR shall provide an authorization mechanism to grant access to only the necessary authorised services and data.
3. The SBI-NR shall provide security mechanisms to safeguard the confidentiality, integrity and authenticity of the exchanged data.
4. The SBI-NR shall provide security mechanisms to counteract attacks aimed to prevent the availability of the interface, such as DoS attacks.
5. The SBI-NR shall support the adoption of strong security algorithms that guarantee forward secrecy and prevent intervening attacks such as replay, relay and man-in-the-middle attacks.
6. The SBI-NR shall support security mechanisms to protect network functions discovery procedure.
7. The SBI-NR shall safeguard the protection and integrity of traffic steering parameters.

5.1.5 Southbound Interface to Charging Function

5.1.5.1 General

The OP resource manager and OP federation broker, as roles in charge of resources management, shall expose the charging information, parameters and events related to resource consumption:

1. The OP shall log all the service resource consumption events and data involved in any transaction required for the operator to charge and bill for the service.
2. The OP shall expose consumptions and event data required for charging purposes through an interface (SBI-CHF) to an external charging engine.
3. The OP shall maintain security and data/topology privacy requirements when reporting federated consumption.

5.1.5.2 Charging information

The consumption reports shall include any information usable by a charging engine to address the final billing of the services. This information shall also include the identities of the chargeable parties, from the Application Provider to the user client.

Consumption reports shall be exposed to the operator based on the agreed data collection interval.

Note: in the context of this section, the following terms are used to capture consumption:
Effective Usage: the effective usage of workloads. For example, Network I/O over a time period

Subscribed Capacity: The requested capacity of workload. For example, 2vCPU, 2 GB of memory. That capacity is subscribed independently from the Effective Usage.

The following applies concerning the consumption data that shall be collected:

1. The OP shall report the subscribed compute capacity

    a) vCPU
    b) Memory
    c) Network Resource Location
    d) Availability zone

    Note: This includes used and reserved compute capacity.

2. The OP shall report the effective compute usage

    a) vCPU
    b) Memory
    c) Network Resource Location
    d) Availability zone

3. The OP shall report the subscribed storage capacity

    a) Storage
    b) Type
    c) Network Resource Location
    d) Availability zone

4. The OP shall report the effective storage usage

    a) Storage
    b) Type
    c) Network Resource Location
    d) Availability zone

5. The OP shall report the subscribed Network capacity

    a) Input
    b) Output
    c) Label (Internet traffic, Intra-cluster traffic, Inter-Edge Cloud traffic, etc.)

6. The OP shall report the effective Network usage

    a) Input
    b) Output
    c) Label (Internet traffic, Intra-cluster traffic, Inter-Edge Cloud traffic, etc.)

7. The OP shall report the subscribed accelerators capacity

    a) Accelerator name (Example: GPU)
    b) Type
    c) Network Resource Location

d) Availability zone

8. The OP shall report the effective accelerators usage

   a) Accelerator name (Example: GPU)
   b) Type
   c) Network Resource Location
   d) Availability zone

9. The OP shall report the API Usage

   a) API Name (Example API: Verify Location)
   b) Number of requests
   c) Request type (Example: GET, POST, PUT, DELETE)

10. The OP shall identify the parties involved in each charging transaction: (Metadata)

    a) (mandatory) OP ID
    b) (mandatory) Application provider ID
    c) (when available) Edge application name (including Application provider namespace).
    d) (when available) Edge application ID
    e) (when available) Operator ID
    f) (when available) Availability Zone

Note: It is for further study to include the application customer's perspective next to the application providers'.

5.1.5.3 Security Requirements

Based on the attack surface analysis provided in Annex E, the following security requirements shall be considered:

1. The SBI-CHF shall provide an authentication mechanism to enable access only by authenticated and authorized entities. Therefore, mutual authentication shall be provided between the OP (Service Resource Manager Role) and the Charging Engine element.
2. The SBI-CHF shall provide an authorization mechanism to grant access to only the necessary services to which previous authorisation has been granted.
3. The SBI-CHF shall provide security mechanisms to safeguard the exchanged data's confidentiality, integrity, and authenticity.
4. The SBI-CHF shall provide security mechanisms to counteract attacks aiming to prevent data availability, such as DoS attacks.
5. The SBI-CHF shall support the adoption of strong security algorithms that guarantee forward secrecy and prevent intervening attacks such as replay, relay, and man-in-the-middle attacks.
6. All OPs creating or sharing charging data shall guarantee the security, integrity, availability, and non-repudiation of charging data.
7. The OP shall provide security mechanisms to guarantee a robust subscriber ID assignment and tracing (e.g., to prevent guessable IDs).

8. Personally identifiable information (PII) of subscribers shall be protected while in transit or storage.
9. Role-based access control (RBAC) policies shall be in effect to regulate access to charging information.
10. The OP shall maintain security and data/topology privacy requirements when reporting federated consumption.
11. The OP shall provide secure tracing and logging of charging and billing data requests.

5.1.6 User to Network Interface

5.1.6.1 High-Level Requirements

1. The UNI shall be universal, meaning that the Application Provider does not have to modify its applications for different Operators or OPs.
2. 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.

3. In this document, 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. On the other hand, where the UE accesses via an untrusted network (such as public Wi-Fi), the OP needs to undertake its own AAA services.

5.1.6.2 User First Attachment

5.1.6.2.1 General

When a UC requests access to an Edge Application, the OP receiving the request shall authorise/authenticate the user and the requesting application. Once the OP has authorised the request, it gathers 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.

5.1.6.2.2 Edge Cloud service discovery

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

1. An OP shall expose a connection reachable by any customer on the operator network.
2. 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.
3. 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. The UNI request does not need to include this information if the OP knows the UE's location.

5.1.6.2.3 Authentication and Authorisation

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

1. If the UE is attached to the 4/5G network, the OP may rely on user authentication by the operator.
2. 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.
3. An OP shall authorise the usage of the application by the UC, for example, by checking that the particular application is part of the user's 'package'. In addition, the OP shall provide a mechanism, such as a token, to allow efficient authorisation of subsequent interactions.

5.1.6.2.4 Cloudlet selection

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

1. 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 as well as the UPF/PGW associated with the UE.
2. An OP shall select an appropriate Cloudlet that:

    a) depending on the actual UE's location (See 1. above) and the geographical zone that the Application Developer has previously determined where its Application Clients would 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 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.

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

5.1.6.2.5 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:

1. If necessary, the OP shall deploy the application image and create an instance on the selected Cloudlet,
2. 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),
3. The UE shall be able to test the connectivity characteristic towards the selected Cloudlet.

5.1.6.3 Security Requirements

Based on the attack surface analysis provided in Annex E, the following security requirements shall be considered:

1. The UNI shall provide an authentication mechanism to enable access only by authenticated and authorized entities. Therefore, mutual authentication shall be provided between the UC and the OP.
2. The UNI shall provide an authorization mechanism to grant access to only the previously authorised services. The authorization mechanism shall ensure that the EC is authorized to access the provisioned services and that the UE can access the edge data network.
3. The UNI shall provide secure communication between the UC and the OP, assuring integrity protection, replay protection and confidentiality protection.
4. Given the external exposure of this interface, the UNI shall provide security mechanisms to counteract attacks aimed to prevent the availability of the interface, such as DoS or DDoS attacks.
5. The UNI shall support the adoption of strong security algorithms that guarantee forward secrecy and prevent intervening attacks such as relay, replay and man-in-the-middle attacks.
6. Privacy and tracking protection: Information originating in the UE should be protected for integrity, privacy, confidentiality, nonrepudiation.
7. Security mechanisms (e.g., certificate authorities) used to protect tracking, logging, and charging information should be independent of those used to protect UE access to the OP via UNI.

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:

1. Edge node resources and load.
2. Network resources and load.
3. Network usage forecast.
4. Edge usage forecast.
5. Application availability (already deployed/onboarded on edge node).
6. Reserved resources availability.
7. UE mobility supported.
8. Network mobility supported (integration with data packet core).
9. Specific constraints/barring for users, application or edge nodes selection.
10. Specific considerations to abide by commercial agreements between involved parties.

5.2.2.2 Instantiation Strategy

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

1. An OP shall be able to request the static instantiation of the application on a specific edge node.
2. An OP shall be able to request the static instantiation of the application on all the available edge nodes.
3. An OP shall be able to determine the minimum amount of edge nodes to select for covering the footprint and onboarding requirements.
4. 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, 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 use of the term "OP" intentionally leaves open which party(s) within the OP does something.

  Note: The term "application" in the bullet point above intentionally leaves 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.
- When the policy is that a change of edge compute resource can be done with prior notification, the OP decides that a change of edge compute resource is needed and selects the new edge compute resource. In this case, the application chooses the exact timing of the move and is responsible for transferring the 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 of 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:

1. Mobility triggers from the OP:

    a) Related to the movement of the UE that causes a change in session anchor (PGW/UPF) network point;
    b) Related to the movement of the UE that causes a change in the serving network (i.e. PLMN change);
    c) Related to the movement of the UE that causes a change in the application client's IP address;
    d) 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);
    e) 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);
    f) Related to usage forecasts about its edge compute resource and network;
    g) 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.

        Note: additional triggers can be considered, e.g. 3GPP 23.501 section 6.3.3.3

2. 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.

1. An OP shall be able to interact with the SBI-NR to configure the network to meet the application's requirements or restrictions on mobility, e.g. mobility not supported, session continuity (SSC Mode 3) required, UE IP address preservation.
2. An OP shall manage the application mobility for all the edge services associated with each UC.
3. An OP shall consider the mobility sensitiveness of the applications.
4. 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 not perform application mobility to another Operator's network domain if an active application does not support roaming.
    d) 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 manage the reconfiguration of the Edge Application environment, selecting a new edge node to have the application available.

1. An OP shall be able to ensure that the selected edge node has enough capacity.
2. An OP shall be able to request the instantiation of the Edge Application on the target edge node if not previously available or if capacity is insufficient.
3. 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.

1. An OP shall be able to notify the application about the forthcoming mobility procedure if required.
2. An OP shall inform the application about what it needs to know to move the application-related state from the old edge compute resource to the new one.
3. The application indicates to the OP when it is ready to 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 is least affected). Note that KPIs may be suspended during this period.
4. The application may indicate that it cannot currently handle mobility. Then, 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.
5. The application shall confirm the completion of the mobility of the Edge Application onto the new Cloudlet to the OP.

6. Movement of the UE may require that the operator changes the IP address used by the application client.
7. The operator shall notify the application about a change of IP address

5.2.2.3.6 Mobility Enforcement

1. An OP shall be able to request a network gateway relocation (if possible) based on location and network statistics.
2. 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.
3. An OP shall be able to request an application session relocation based on the application requirements.
4. 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.
    d) The OP shall coordinate the different procedures with the Network through the SBI-NR.

        Note: It is for further study how to provide session continuity between different OPs or network domains.

5. An OP shall ensure that the UC is forced to apply the mobility procedures.
6. Network GW location may not be needed in case of service degradation due to an edge node saturation.

5.2.2.4 Service Availability on Visited Networks

5.2.2.4.1 General

Service availability on visited networks shall be considered to allow the users to enjoy edge service outside of their operator network. This condition includes international situations and the inter-operator handovers that occur, for example, when connecting to the end-user's home Wi-Fi network, which a different operator may provide.

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

1. When a user client first attaches to a visited OP, there shall be messaging between the user client, home OP and visited OP. The messaging's purpose is for the Home OP to authenticate the User Client and 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 the intention to define a granular charging /settlement mechanism ("granular" meaning, for example, per user client or per application instance).

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

   a) If no local breakout is available or there is no service availability agreement among operators, the User Client receives service from home resources and Home OP without Visited OP interaction.

3. The visited OP may be capable of obtaining the application image (and any associated policies) directly from the Application Provider (typically if it has an NBI with it); otherwise, it shall request it from the Home OP via the E/WBI.
4. Based on the information received from Home OP and the internal policies, the visited OP shall instantiate the Edge Application on a Cloudlet for use by the user client.
5. The Visited OP shall be in charge of selecting the Cloudlet within the Visited OP best placed to host the Edge Application (including when the user device moves within the visited OP).

   Note: User client mobility management is handled with existing mobility management mechanisms.

5.2.2.5 Operation and Management

The OP shall offer a centralised management plane for the operator to manage the infrastructure. This management plane shall offer an operator

1. The capability to

   a) Create Cloudlets within an Availability Zone
   b) a) Create Cloudlets in a Public Cloud
   c) Manage Edge sites in a federated operator

2. The capability to manage security groups and privacy policies at each Cloudlet

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

3. The capability to manage the compute footprint

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

4. The capability to manage Availability Zones across the geographical sites within the operator’s domain
5. Capabilities for the operator to monitor Cloudlet usage in terms of compute, memory, storage and bandwidth ingress and egress
6. The capability to monitor the above metrics per tenant.
7. Capabilities for automation, with some associated requirements like

   a) Transactions related to automation shall be atomic transactions (i.e. if not all steps of a transaction are completed, then no steps are completed, and no side effects

of those steps remain). Possible methods of achieving atomic transactions include:

i. Two-phase commit (prepare and commit): in a Prepare phase, services carrying out an atomic transaction notify a Coordinator that they are ready to complete the transaction. In a Commit phase, the Coordinator issues a Commit command to all services that must complete their transaction or a Rollback command if the transaction must not be completed.
ii. Eventual consistency and compensation: A service that updates its state (e.g., updating data that it owns) publishes an event, and other services that subscribe to that event, receive it. Subscribing services updates their corresponding data. If a failed transaction event, the subscribing service can perform a compensating transaction (e.g. emitting a delete event, rolling back processing steps).

b) Event notifications related to milestones, status changes, changes in the infrastructure or resource availability changes should be used.
c) The Service Resource Manager shall provide resilience support such as timeouts, support for atomic transactions, and other features that allow a system to be maintained in a consistent state.
d) The Service Resource Manager shall release reserved resources after the reservation expires (in case of reservation).

8. The capability to monitor Cloudlet event, alarms logs
9. The capability to monitor Cloudlet performance metrics
10. The capability to offer operator interfaces to federated partner to monitor usage across Cloudlets

5.2.3 Federation Manager Role

5.2.3.1 Federation and Platform Interconnection

5.2.3.1.1 General

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

The communication between federated entities shall support a distributed tracking mechanism that allows end-to-end tracking across these federated entities. For example, requests may contain identifiers that are propagated and used in every communication.

5.2.3.1.2 Authentication/authorisation

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

1. There shall be a mechanism to register and authenticate different OP instances.
2. An OP shall be able to identify unequivocally any federated OP instance.

3. An OP shall be able to authorise a registration request from another OP instance.
4. An OP shall exchange a token or “federation key” on the association handshake, identifying each federation integration.
5. 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 allows the charging systems of each operator to account for the services consumed.

1. An OP shall share usage statistics through the E/WBI for the services requested by the federated connection.
2. An OP shall provide any needed information that is 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 sharing the Resource Catalogue between instances.

1. An OP shall be able to share (publish) the Availability Zones available on its footprint/resources:

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

2. An OP shall allow the operators/resource owners to select the resources to be shared via federation.
3. An OP shall be able to push an Availability Zones catalogue update based on:

    a) Resources specification change, e.g. adding GPU support on a zone;
    b) Resources are no longer available;
    c) New resources/zone availability.

4. An OP shall allow operators to request 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.

6 External fora conclusions and collaboration model

A consistent set of standards is required to realize Operator Platform (OP) services supporting federation among operators. These standards must be well-supported by Standards Development Organizations (SDOs) and cover the requirements identified and documented in this current

The OPG also recognizes that Open Source communities (OSCs) exist with API specifications and software blueprints that may approach the OP requirements. The OPG believes that, for operators to develop a federated edge computing platform such as the OP, requirements must be enforceable in contracts by a published set of standards.

To this end, the OPG proposes selecting ETSI ISG MEC and 3GPP to provide a standard reference for an edge service end to end definition.

We note that EDGEAPP architecture and ETSI ISG MEC architecture could complement each other in a way that is acceptable to OPG:

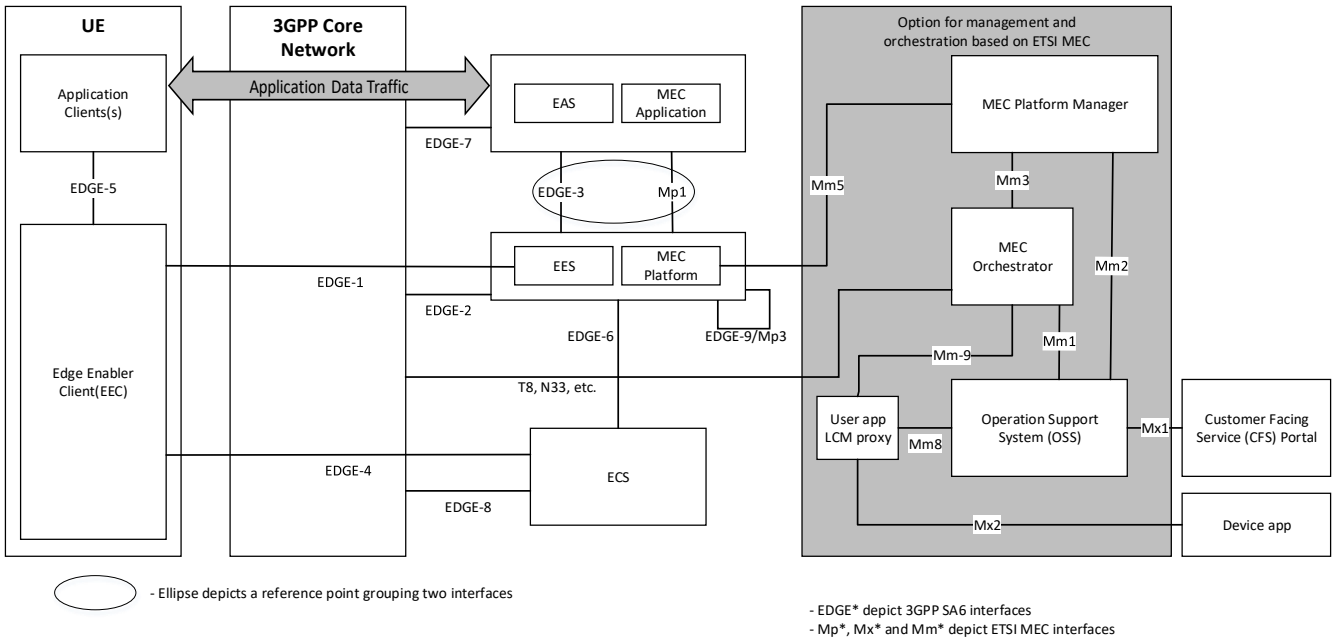

Figure 14: Relationship between ETSI ISG-MEC and 3GPP EDGEAPP architectures (Source: Informative Annex C in 3GPP TS 23.558)

Both EAS and MEC Application are application servers and can provide similar application-specific functionalities. EAS utilizes the services of the EES, whereas MEC Application utilizes the services provided by the MEC platform as specified in ETSI GS MEC 003. The EAS and MEC Application can be collocated in an implementation.

ETSI ISG MEC specifies handling application-specific management for MEC Apps, while 3GPP SA5 provides application-specific management aspects concerning the 3GPP EDGEAPP architecture. OPG encourages collaboration between 3GPP and ETSI for harmonising the application management.

Applied to the OP interfaces, the following mappings and gaps exist by selecting ETSI ISG MEC and 3GPP. Additional OSC implementation could be adopted as and where needed and contributed based on gaps and needs of OPG requirements.

- **UNI**: Device communication interface needs to be part of the 3GPP architecture and relates to how devices connect to the network.
- **SBI-NR**: Relationship from the edge platform to the core network, as defined by 3GPP.
- **SBI-CHF**: Currently only standardized for charging engine/function/system on 3GPP and related interfaces to connect to other elements.
- **SBI-CR**: Although it shall remain open to multiple architectures (as indicated in section 3.5.2.1.2) and considering the current industry solutions, the ETSI ISG NFV architecture is used as a reference for cloudlet functionality, because of its current alignment with 3GPP, but also because of the similarity of Edge Computing and Network Function Virtualisation.
- **E/WBI**: Already in the scope of ETSI ISG MEC with a close relationship with applications deployment and developers. 3GPP SA6 provides some details on this aspect but will need to provide the details for the network-related use cases, such as roaming.
- **NBI**: No single SDO covers a complete interface NBI as required by OPG to handle the application provider relationship. Since this is an area where application developers and OSCs are very active, we propose a parallel task to align them with the selected SDOs and provide convergence. ETSI ISG MEC and 3GPP both handle the application side interactions required to host NBI and shall align the capabilities exposure. OPG proposes for ETSI ISG MEC to host the NBI standard.

Note: Management plane functionalities will be covered and aligned with proper standard and industry groups in a later phase.

Annex A Mapping of Requirements to External Fora

A.1 ETSI

A.1.1 ETSI ISG MEC

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

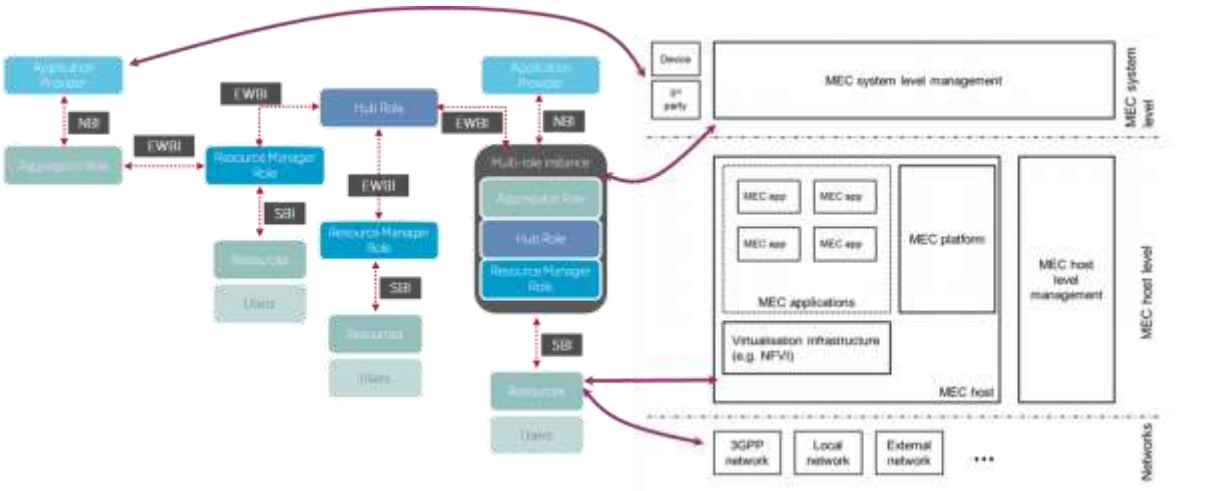

Figure 15: OP to ETSI ISG MEC mapping

A.1.2 ETSI ISG MEC specifications relevant to the NBI and the SBI

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

A.1.3 ETSI ISG MEC specification relevant to the UNI

- ETSI ISG MEC 016: UE Application Interface

A.1.4 ETSI ISG MEC specifications relevant to OP optional capabilities

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

A.1.5 ETSI ISG MEC activities relevant to the E/WBI interface

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

A.2 3GPP

A.2.1 3GPP SA6 EDGEAPP

3GPP defines a core network-compatible architecture for the edge, including the relationship with UEs and the edge network configuration.

Edge Enabler Server (EES) and Edge Configuration Server (ECS) are introduced as key elements for communicating with the device Edge Enabler Clients (EEC) and the core network elements, including provisioning the edge service and enabling application management (instantiation, session mobility). The Edge Application Server (EAS) discovery may be performed through an interaction between EEC and EES, extended with the UE location. The interaction with the network includes policy requests to Policy Control Function (PCF)/Policy and Charging Rules Function (PCRF), application traffic configuration APIs, and service APIs exposed by SCEF/NEF.

NOTE: The EEC(s) may be provisioned with the ECS address(es) information also by the Session Management Function (SMF) at Protocol Data Unit (PDU) Session establishment or modification via Non-Access Stratum (NAS) signalling. The SMF may derive the ECS address(es) information based on local configuration, the UE's location, or UE subscription information.

EES (and ECS) map to the Capability Exposure, Service Resource Manager and Federation Manager as defined on OP, except for cloud resource management.

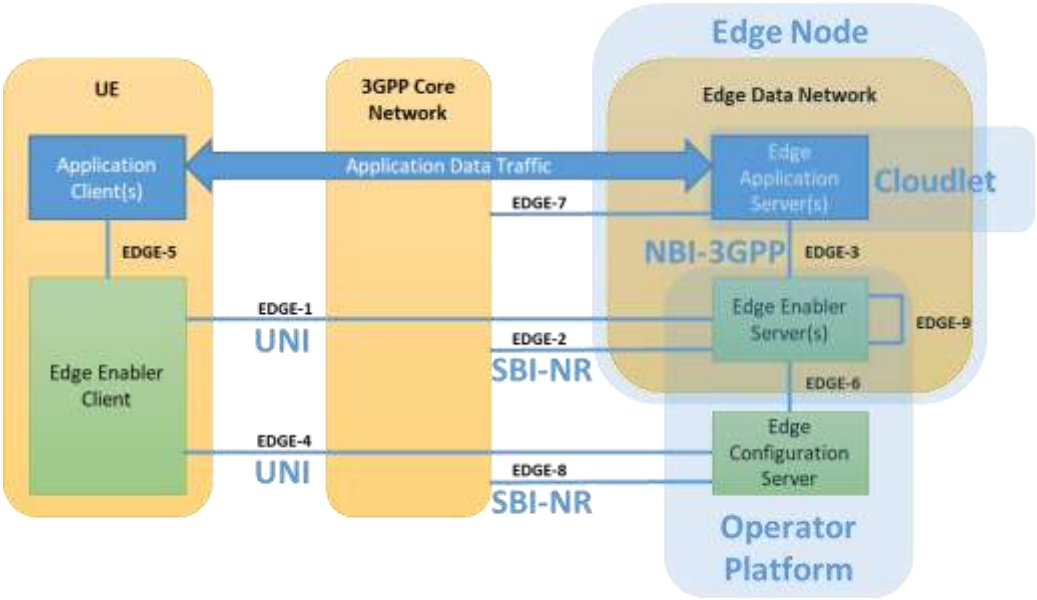

Figure 16: OP to 3GPP EDGEAPP mapping

Note: EDGE-3 is shown as NBI-3GPP because it covers the OP NBI Network Event Support, but it does not cover the complete OP NBI scope.

A.2.2 3GPP EDGEAPP Interfaces

- 3GPP SA6 defines the EDGE-1 and EDGE-4 interfaces for the device clients to communicate with the edge platform, as defined for the UNI.
- 3GPP SA6 defines the EDGE-2 and EDGE-8 interfaces for the interactions from the edge platform to the network, with functionalities related to SBI-NR. 3GPP SA5 also defines more details on the cloudlet management aspects corresponding to SBI-CR.
- 3GPP SA6 defines the EDGE-3 interface for the cloudlets to communicate with the edge platform, as defined for the NBI Network Event Support.
- 3GPP SA6 defines the EDGE-9 interface for the Operator Platforms to communicate with each other, as defined for E/WBI.
- 3GPP SA5 defines the Nchf interface for charging related to SBI-CHF.
- 3GPP SA3 defines the security details of all the EDGEAPP interfaces.

A.2.3 3GPP Exposure Interfaces

3GPP SA2 defines the interfaces N33 and T8 for 5G and 4G, respectively, enabling the following APIs that may map with the SBI-NR:

- 3GPP TrafficInfluence NEF API [4].
- 3GPP ReportingNetworkStatus NEF API [4] and SCEF API [5].
- 3GPP Monitoring NEF API [4] or SCEF API [5].
- 3GPP AsSessionWithQoS NEF API [4] or SCEF API [5].
- 3GPP ChargeableParty NEF API [4] or SCEF API [5].
- 3GPP DeviceTriggering NEF API [4] or SCEF API [5].
- 3GPP ServiceParameter NEF API [4].

Annex B Use Cases

This section introduces a set of use cases that the Operator Platform Group developed to verify whether gaps exist in the requirements proposed in OPG.01 [2]. The OPG has selected these use cases for their breadth of functional coverage rather than embark on the impossible journey of defining an exhaustive set of use cases that benefit from federated edge computing. Collectively, the use cases illustrate some of the critical capabilities that an OP has to provide.

B.1 UC1 - Automotive - Advanced Horizon

B.1.1 Description

A driver gets “look ahead” information about the local vicinity – for example, a patch of ice, a slow-moving tractor or red traffic lights. A driver’s ability to see “around the corner” could help safer and more economical driving.

The driver could be a human – as seen in today’s Advanced Horizon products from Bosch™ and Continental™ – or, in the future, it could be an automated driver.

B.1.2 OP Dependency

The service could be delivered through an application server on a cloudlet that gathers information from roadside sensors and nearby vehicles. The application server would

aggregate this data and analyse it to send updates to vehicles in the vicinity. These updates can be more accurate and timely if the application server gets information from all nearby vehicles, potentially on several mobile operators. A federation of OPs would enable such information exchange either by direct access from the devices or between application servers on different operators.

Next to that, this service has essential security and trustworthiness requirements – both for the information reported by roadside sensors and other cars and the analysis performed by the application server. An operator platform that authenticates the parties supplying the data, verifies applications and is involved in their discovery would provide the guarantees required for such a service.

B.2 UC2 - Automotive – Remote Driving

B.2.1 Description

The second use case is remote driving or flying one or more vehicles or drones. This use case involves someone at a distance controlling the vehicle based on detailed information of its surroundings. Other vehicles might then follow the path set by the one driven or flown remotely without requiring control on an individual basis.

B.2.2 OP Dependency

This use case has similar requirements on trustworthiness and communication to other operators than the use case discussed in section B.1.

The scenario requires strong guarantees on service assurance – about the network and compute's responsiveness, reliability, and security. Deploying the supporting application at the edge using the Operator Platform for discovery, potentially combined with Network Slicing that the Operator Platform intends to support in a future iteration, may provide those guarantees.

Furthermore, a vehicle may have to pass borders and operate in a geographical region that requires other operators for coverage. The Operator Platform would help to ensure that the supporting edge application is available on those networks.

B.3 UC3 - Multiplayer Augmented Reality Game

B.3.1 Description

The following use case is a multiplayer augmented reality game. Players participate in the real world, supplemented by online features, for example, a role-playing game. The players are thus all nearby but can be on different operators.

B.3.2 OP Dependency

For such a game, preference is that the players share the same application server, which is on a local cloudlet. A "shooter" game, for example, is moderately latency-sensitive, and fairness between players is crucial, requiring that the players all get similar server processing performance and similar network performance. An Operator Platform enabling the sharing of edge nodes between operators would be able to support this.

Some games need specialist compute (e.g. GPU). As indicated in the TEC whitepaper [6], a federated model to deliver an Operator Platform may require alignment between the federated operators to ensure that they offer similar resources. Thus, the party developing the game can expect the same specialist compute capabilities in all networks and consider them in their application design and dimensioning.

B.4 UC4 - Privacy-preserving Health Assistant

B.4.1 Description

The following use case is a privacy-preserving health assistant. Already there are health-related personal monitors, such as smartwatches, in use today. There are many more personal IoT services, perhaps including actively controlled devices to adapt an insulin dose based on its measurements automatically.

These devices all provide data to their dedicated backends without much user control over the access to the provided data from that point onwards. An edge-based health assistant's appeal could be that it can act as a trusted third-party intermediate capable of aggregating the data from different devices and providing control over the access to that data. By design, the local cloudlet could store data only temporarily. For instance, an application in the cloud would be allowed only specific request types on the cloudlet (e.g. restrict exporting the complete data set).

B.4.2 OP Dependency

When the user roams onto another network, one solution approach is that the (trusted) home operator installs its application server on the local cloudlet.

B.5 UC5 - Infrastructure sharing

B.5.1 Description

Infrastructure sharing is a technical use case where one operator uses infrastructure provided by the other. Possible examples could include:

- Two operators, each with a mobile network covering the whole country, agree to share edge compute infrastructure (say: one covering the North of the country and the other the South) – this similar to today's sharing of radio masts.
- An OP provider that provides OP services to subscribers but doesn't have their own compute infrastructure and networking capacity, sourcing those services from another OP instead.
- An OP has its own 'basic' edge infrastructure, but not the specialist compute or specialist hardware security that some application providers require.
- An OP whose edge compute is currently short of resource temporarily offloads new requests to another OP.

B.5.2 OP Dependency

The main requirement to enable this is for a commercial agreement between the involved OPs covering topics including security and trust, service level agreements and billing.

Note that the whitepaper defines home network control in the roaming case.

B.6 UC6 - High-resolution media streaming service

B.6.1 Description

The use case is to provide a high-resolution media streaming service. Next-generation broadcasting services (e.g. ATSC 3.0) plan to deliver media streams over the 5G/4G network. With added edge-based environments, very low-latency, high-resolution media transfer can be guaranteed. Next to that, personalized services can be added based on the user's location or subscription options.

B.6.2 OP Dependency

This service can be supported through a media delivery system on a cloudlet, including encoding and decoding functionalities. Traditionally, media transmission is via a single centralized system. Still, edge-based media services, located close to the user’s location, can provide enhanced streaming through content caching, fast media processing, and delivery optimization. OP can mainly provide related resources (such as network and storage resources) and computing capabilities on an edge environment for a high-resolution media streaming service.

B.7 UC7 – Visual Positioning Service (VPS)

B.7.1 Description

The use case is to provide Visual Positioning Service (VPS). VPS uses the camera on the user's device, e.g. smartphones, wearables, vehicles, to instantly determine the user's accurate position and orientation anywhere in the covered city before AR usage. The VPS can provide the user's exact outdoor location and indoor location, which the current GPS cannot support well. As it provides the precise user location and orientation, VPS may be used in combination with other AR services, e.g. AR advertisement, AR entertainment, AR navigation, AR tourism, and may become necessary for AR devices and services in the future.

B.7.2 OP Dependency

In general, VPS uses real-time computer vision matching for 3D recognition as a key process. Edge Cloud and 5G connectivity are necessary to make Low Latency and High CPU power available. Furthermore, VPS may become an essential functionality for future AR services. Therefore, VPS will rely on the OP for its federation capabilities, e.g. common NBI, Roaming and UE/App Mobility, Edge Node Sharing, etc., in addition to the application distribution function.

B.8 Use Case Overview

Capability	**Interface**	**Document section**	**UC 1** "Advance horizon" info for assisted driving	**UC 2** Remote control of a vehicle (or drone etc.)	**UC 3** Multiplayer AR location-dependent game	**UC 4** Privacy-preserving Health assistant	**UC 5** Infrastructure sharing	**UC 6** High-resolution media streaming service	**UC 7** Visual Positioning Service (VPS)
Application Provider **request** for Edge Cloud service	NBI	5.1.1.3 #1	Y	Y	Y	Y	N	Y	Y
Provide info on UE's **location**	SBI-NR	5.1.3	Y	Y	Y (& verify location)			Y	Y
Handover (UE moves in a mobile network) *(Implementation likely to require a move of the application server to a new cloudlet)*	SBI-NR	5.1.1.2.2 #9 5.1.4.2.2 #15	Y	Y	N				Y
Inter-network **Roaming** (UE roams to another operator) *(Preferably with local breakout, so application server on cloudlet in the visited operator)*	E/WBI	5.2.2.4 5.1.2.3 #5	Y preferably	Y	Y	Y			Y

Capability	**Interface**	**Document section**	**UC 1** "Advance horizon" info for assisted driving	**UC 2** Remote control of a vehicle (or drone etc.)	**UC 3** Multiplayer AR location-dependent game	**UC 4** Privacy-preserving Health assistant	**UC 5** Infrastructure sharing	**UC 6** High-resolution media streaming service	**UC 7** Visual Positioning Service (VPS)
Application Provider requests **QoS** (typically latency)	NBI	5.1.1.3 #2	Y	Y - critical	Y & 'fair'	Y - weak		Y	Y
Establish a chain of **trust** between the elements	UNI & OP	3.5.3.2	Y	Y		Y - critical	Extend over E/WBI		
Security Comms Compute Storage	 UNI OP OP	2.1.4, 3.4.1 & missing	 Y Y .	 Y Y		 Y Y Y			
Inter-OP Security		5.2.3.1.2					E/WBI		
Data sharing (Data is 'open' for use by multiple application providers)		missing	Y			Y but highly filtered			Y
Specialist compute	SBI-CR	5.2.2.3			Y				Y
Shared Application Server	SBI-CR	missing			Y				

NOTE: Y – indicates that the requirement is of particular importance in the use case

N – indicates that the requirement is not essential or not needed in the use case

Blank cell - indicates that the requirement is somewhat helpful for the use case but not central to it

Annex C Deployment Scenario

This section provides an overview of deployment options of the Operator Platform.

C.1 Relationship with OP and Operator

The OP's deployment scenario can have two options depending on whether each Operator has its OP.

In Figure 17, the OP manages at least the resources of a single Operator. OP A run by Operator 1 can federate with OP B run by Operator 2.

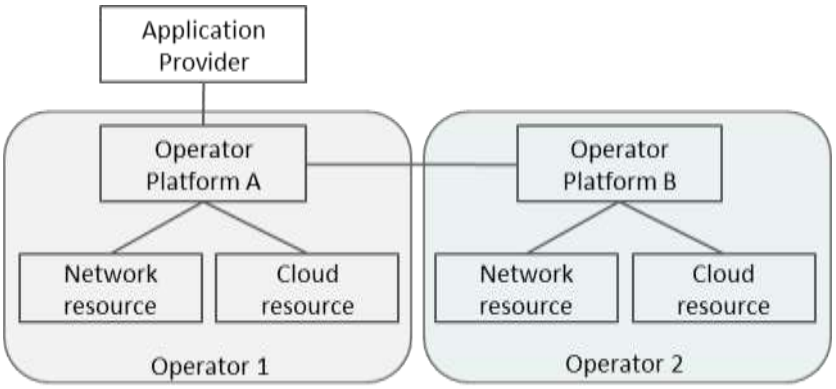

Figure 17: Each operator has an own Operator Platform

In Figure 18, an OP manages multiple Operators' resources. Because one OP manages the resources of multiple operators, when receiving a federation request from OP B or a deployment request from an Application Provider, Operator 1 or Operator 2 is selected based on OP A's policy.

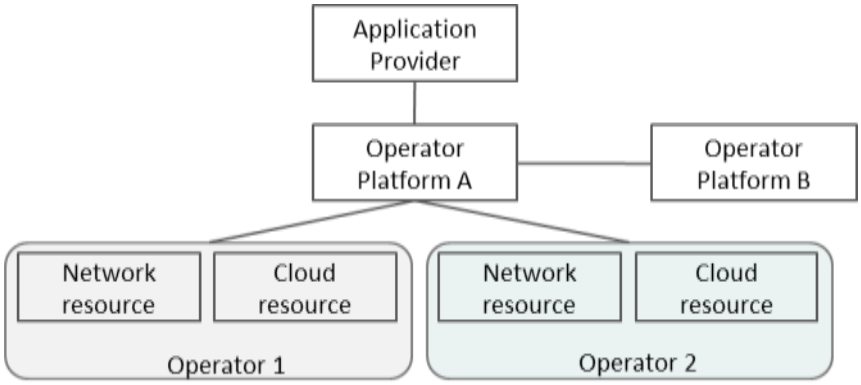

Figure 18: Multiple operators share the same OP

C.2 Relationship with hyperscalers from a single Operator perspective

An operator can have their own cloud resource and collaborate with hyperscaler simultaneously. The OP can integrate hyperscalers with the same features as it does with its own cloud resources and support APIs of hyperscalers, as described in section 5.1.3.1.3.

There are two ways for Hyperscaler integration via OP. First, hyperscalers can be considered enterprise customers to OP and can interact via the NBI. The second is that hyperscalers can implement an OP and become a Partner connecting via the E/WBI.

The SBI-CR is likely to match the interface that hyperscaler is exposing to its direct customers (i.e. Application Provider 2). In addition, Hyperscaler resources can be available for OP A to offer its customers (i.e. Application Provider 1).

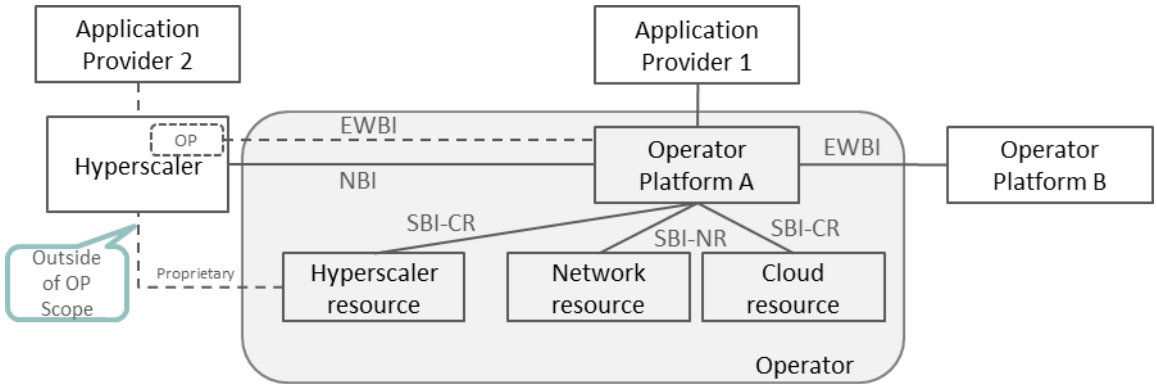

Figure 19: Relationship with hyperscalers

Annex D OP Marketplace

OP Marketplace is a store for storing and distributing Application Providers' applications and providing APIs of Operators Platform. In addition, there may be OPs that offer additional edge cloud services beyond those specified in this document, for example, specialist AI or media encoding, targeted at Edge-Enhanced Applications or Edge-Native Applications. The purpose of the OP Marketplace would be to enable Application Providers to discover such additional services and possibly buy them.

The following are potential functionalities supported by OP Marketplace:

- OP Marketplace authenticates and authorizes Application providers
- OP Marketplace aggregates the additional APIs offered by OPs and exposes them to Application Providers
- OP Marketplace receives requests from Application Providers for the additional services and requests the appropriate OP to fulfil them
- OP Marketplace provides a public repository for storage and validation of the application package that Application Providers upload for deployment.

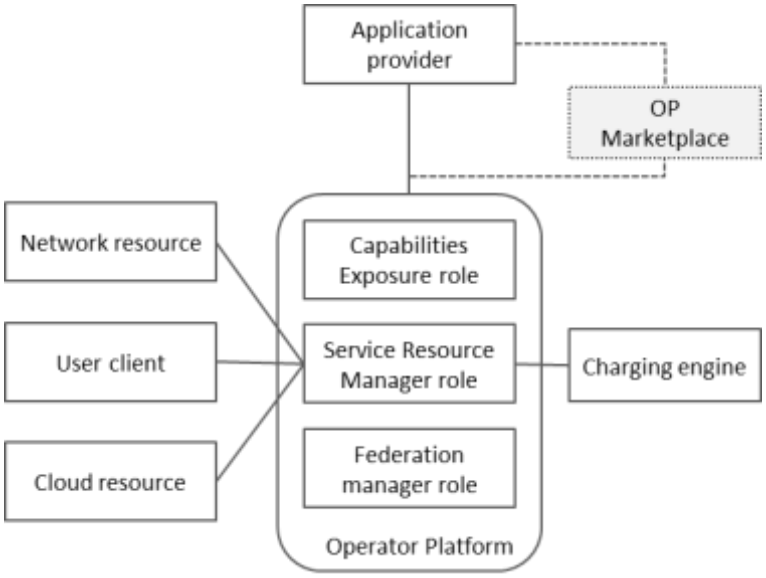

Figure 20: Operator platform with Marketplace

Note: The OP Marketplace is an optional role and is for further study beyond the scope of this PRD. It does not impact the interfaces defined by the PRD.

Note: The OP Marketplace is for B2B, not for B2C, and it is different from mobile app stores, such as the App Store and Google Play.

Annex E Analysis of Operator Platform Security

E.1 Introduction

This Annexe aims to use prior art in security technology to derive applicable security requirements for OP.

This Annex contains informative text that supplements the security requirements appearing in several sections of the PRD. Its purpose is to ensure that those requirements provide adequate coverage for security issues that may arise in the Operator Platform architecture by surveying a suitable corpus of prior art and mapping security concerns and solutions onto the OP architecture.

The security analysis reported in the present Annex is to be considered work in progress. In particular, Section E.3 is an initial mapping of the threat vectors affecting the Operator Platform architecture and the countermeasures available to address the threat vectors. The threat vectors and countermeasures are derived from the available prior art, as described in the Annex. In turn, they were used to derive the current version of the security requirements provided to the PRD. This work will be refined in future releases of the PRD.

Prior art relevant to the OP architecture is based on attack surface characterization. The attack surface of a software system consists of

"…the points on the boundary of a system, a system element, or an environment, where an attacker can try to enter, cause an effect on, or extract data from, that system, system element, or environment." [23].

Methods for compromising the attack surface are called threat vectors, and attack surface characterization consists of forming a comprehensive list of threat vectors and points on the attack surface where they apply. For the OP architecture, threat vectors may be identified in functional elements and at interfaces between functional elements.

The next step after characterization is to identify countermeasures corresponding to the threat vectors. Countermeasures vary in nature, including hardware, software, protocol design, and best practices carried out by engineering and operations personnel. For the OP architecture, countermeasures are expressed as security requirements applying to functional elements and interfaces.

In Section E.2, the primary sources (listed in E.1.1) are surveyed to produce lists of threat vectors. Subsections of E.2 deal with each of the primary sources. The threat vectors in the list are paraphrases of the threat vectors from the sources.

In Section E.3, the threat vectors are mapped to the OP architecture. The mapping is shown in Figure 23, labelled by the identifiers provided for the threat vectors of section E.2.

The threat vectors are in various categories, and each category is covered in a separate subsection of E.3. In these subsections, countermeasures for each category are provided in tables. These countermeasures are used as a guide to create the Security Requirements in the main body of the PRD.

The countermeasures of E.3 do not directly appear as security requirements, as they must be “translated” from the original text in the sources to meaningful requirements in the context of the PRD. However, the reader should see a relationship between the countermeasures mapped to a particular interface or functional element of OP and the requirements that appear in the corresponding section of the PRD.

The threat vectors and countermeasures identified in this analysis, even though they arise from the related fields of edge computing, cloud computing, mobile networks, and network functions virtualization, require a bit of interpretation before applying directly to the OP architecture.

E.1.1 Sources

The previous section explained that several sources from prior art in security are used to characterize the OP architecture attack surface. These sources are:

- In Annex A of this PRD, a provisional mapping of ETSI ISG MEC and 3GPP architectures onto the OP architecture is provided. The mapping is high-level and requires interpretation in the context of OP, but it allows threat vectors for the OP architecture to be identified provisionally.
- Reference [15] provides a detailed attack surface characterization of the ETSI ISG MEC architecture, including some 3GPP 5G architecture elements associated with ETSI ISG MEC. Therefore, this Annex uses [15] as a starting point for OP attack surface characterization.
- The GSMA Fraud and Security Architecture Group (FSAG) has published a set of recommendations for security controls [14] to apply to mobile telecommunications networks. This document covers a wide area of security issues and contains

numerous recommendations applicable as countermeasures to this PRD. This Annex notes the relevant recommendations.

- 3GPP SA3 is studying the security aspects of edge computing support in the 5G Core (e.g., [20], [16]). The approach this study follows is similar to that of [15]. It identifies security issues, maps them to reference points or elements of the 3GPP architecture, and identifies potential solutions or countermeasures.
- The ETSI ISG MEC working group are actively working on security requirements for the ETSI ISG MEC architecture. A technical report on this subject is currently in progress but is not yet publicly available, but it is possible to identify threat vectors from [22].

E.1.2 Procedure

The rest of this Annex follows the procedure:

- Survey the sources listed above, and derive lists of threat vectors. Then, use the threat vector model of [15] to provide identifiers for these threat vectors. Next, these identifiers are used to map them to the OP architecture in the following steps.
- Use the ETSI ISG MEC – OP mapping of Annex A (and Figure 21) to associate the threat vectors to the OP architecture directly.
- Create tables of countermeasures for each of the threat vector identifiers appearing in Figure 23. These tables are provided in Section E.3.
- Use the tables of Section E.3 as inspiration for Security Requirements in the main body of the PRD. This step appears in the main body of the PRD, not in this Annex.

The output of this procedure will evolve in future releases of this Annex. The recommended countermeasures re-appear in the main body of the PRD as requirements.

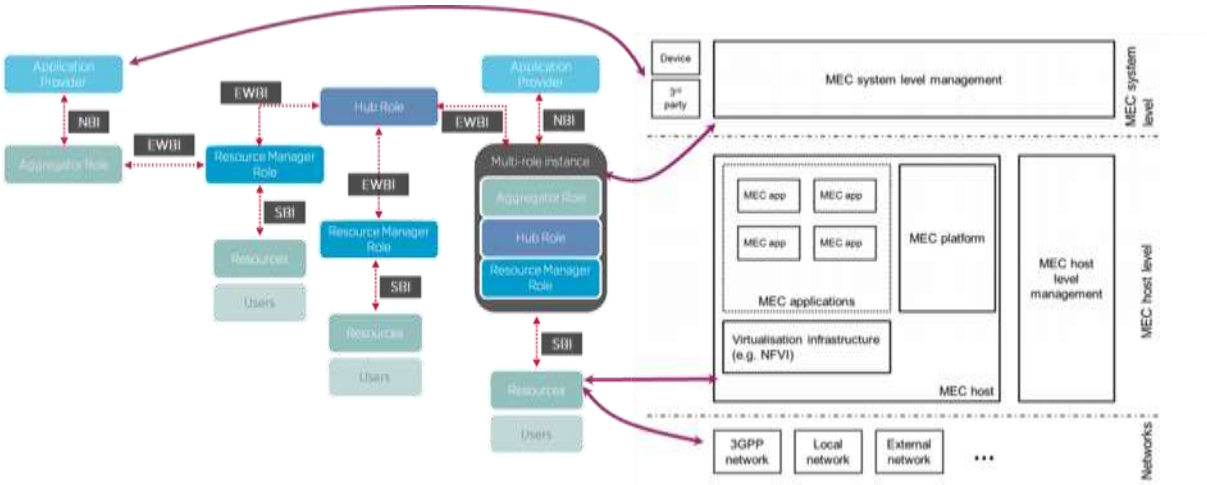

Figure 21: OP to ETSI ISG MEC mapping

E.2 Threat Vector Identification

In this section, the sources described in the previous section are surveyed to identify threat vectors and countermeasures.

The first of these sections covers [15], as this reference is a survey that characterizes the attack surface of the ETSI ISG MEC architecture. The ETSI ISG MEC architecture is

mapped to the OP architecture of Annex A, and therefore this attack surface characterization provides an initial attack surface characterization for the OP architecture.

Following that, parallel sections surveying threat vectors from 3GPP SA3, ETSI ISG MEC, and GSMA FSAG supplement the threat vectors from [15] and create a comprehensive list.

E.2.1 Threat Vectors Identified from [15]

The following figure, taken from [15], identifies and categorizes threat vectors. Because the analysis takes the ETSI ISG MEC architecture as a default, they are depicted in an ETSI ISG MEC deployment. The figure categorises the threat vectors as Access, Architectural, Core, Edge, "Other", and Privacy.

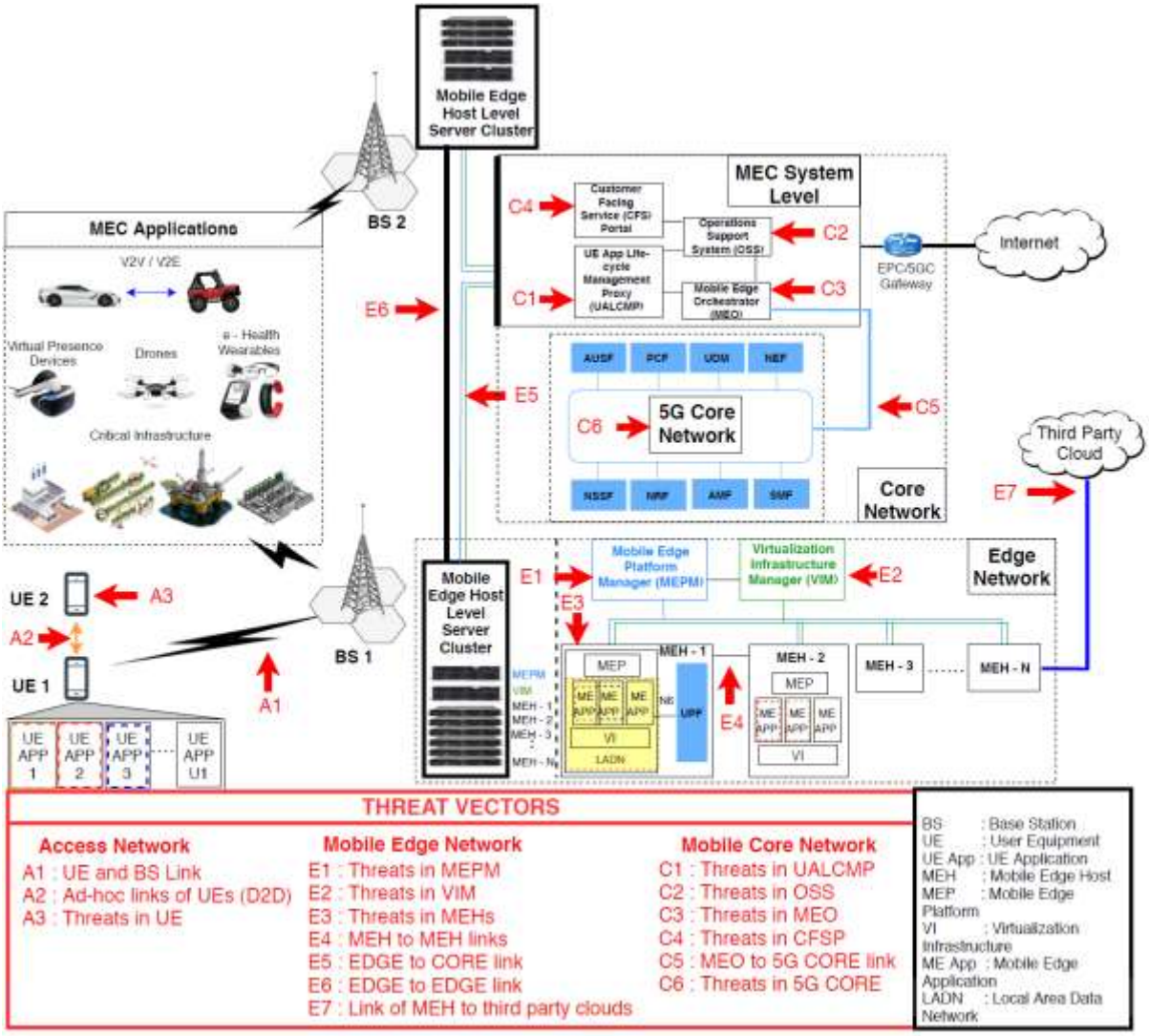

Figure 22: ETSI ISG MEC Access, Edge and Core Threat Vectors (from [15])

Table 14 summarizes the threat vector addressed in this Annex. The table contains the threat vectors noted in [15], as well as threat vectors identified by 3GPP SA 3, from [20] (with tags "SA") and the threat vectors identified by ETSI ISG MEC. The SA threat vectors, and the threat vectors related to ETSI ISG MEC, are discussed in a later section but are summarized in the table for convenience.

Privacy threats are also examined in the [15] paper and may be considered in the next version of the present Annex.

Threat Vector (TV) ID	Description
A1	Link between UE and a BTS
A2	Ad-hoc connectivity between UE
A3	UE vulnerabilities
AR1	Network Slicing (NS)
AR2	Traffic Steering
AR3	Service Migration
AR4	Mobility Management
C1	User App lifecycle management (LCM) Proxy (UALCMP)
C2	Operation Support System (OSS)
C3	Mobile Edge Orchestrator (MEO)
C4	Customer Facing Service Portal (CFSP)
C5	Connectivity of MEO and 5G Core Network
C6	5G Core Network
E1	Mobile Edge Platform Manager (MEPM)
E2	Virtualization Infrastructure Manager (VIM)
E3	Mobile Edge Host (MEH)
E4	Connectivity between MEHs
E5	MEC platform connectivity between Edge and Core
E6	Connectivity between MEC apps operated under hosts at different BTSs
E7	Link of MEH to third party clouds
MEC1	Required signalling for secure inter-MEC systems
MEC2	MEC system discovery supporting authentication, authorization, identity management, etc.
MEC3	MEC platform discovery supporting authentication, authorization, identity management, etc.
OTV1	Charging and billing for MEC subscriptions
OTV2	Service impeding/delaying threats
OTV3	Mobile offloading
OTV4	Virtualization and orchestration of the edge
Privacy	Privacy-related threats
SA1	Authentication and authorization between EEC and EES – EDGE-1
SA2	Authentication and Authorization between EEC and ECS – EDGE-4
SA3	Authentication and Authorization between EES and ECS – EDGE-6
SA4	Edge Data Network authentication and authorization
SA5	Edge Data Network user identifier and credential protection

SA6	Transport security for the Edge-1-9 interfaces
SA7	Security of network information provisioning to local applications with low-latency exposure
SA8	Authentication and authorization in EES capability exposure – SCEF/NEF northbound APIs
SA9	Security of EAS discovery procedure
SA10	Authorization during edge data network change

Table 14: Threat Vector Descriptions (adapted from [15], [20], [22])

E.2.2 Threat Vectors Identified by 3GPP SA3

3GPP Study Area 3 (SA3) is responsible for specifying security requirements for the 5G architecture. They have published numerous specifications, a few of which are provided in the references section 1.6. The requirements contained in these specifications largely apply to security, privacy, confidentiality, and other security attributes of the 5G architecture. This area is out of scope to the Operator Platform architecture, but we note that it is a Best Practice for OP owners to secure their access and core networks. We have captured this Best Practice by listing it as a countermeasure for threat vector AR4 in Table 19.

SA3 has recently engaged in studying edge computing security aspects in the 5G core network in [20]. They identified security gaps, locations, and solutions, in an approach similar to that of [15]. Table 15 summarizes the gaps from that study, extracted as threat vectors, and indicates the location of the threat vectors in the 3GPP core architecture. The threat vectors from this work are annotated in Figure 22 and summarized in Table 14 (a composite table of all threat vectors identified from all sources).

Threat Vector (TV) ID	Description	Location
SA1	Authentication and Authorization between EEC and EES	EDGE-1
SA2	Authentication and Authorization between EEC and ECS	EDGE-4
SA3	Authentication and Authorization between EES and ECS	EDGE-6
SA4	Edge Data Network Authentication and Authorization	edge data network
SA5	Edge Data Network User Identifier and Credential Protection	edge data network
SA6	Transport security for the Edge 1-9 Interfaces	EDGE-1 through Edge 9
SA7	Security of Network Information Provisioning to Local Applications with low latency exposure	UPF, AF, NEF
SA8	Authentication and authorization in EES capability exposure	SCEF/NEF northbound APIs, CAPIF
SA9	Security of EAS discovery procedure	EAS
SA10	Authorization during Edge Data Network Change	edge data network

Table 15: Threat Vectors derived from [20] with a location indication

E.2.3 Threat Vectors Identified by ETSI ISG MEC

While other information sources use the ETSI ISG MEC architecture as a starting point, the ETSI ISG MEC working group has also undertaken to study aspects of federated edge platforms [22]. This study is primarily about coordination between MEC systems (of which OP-like federated systems are a subset), not primarily about security. The use-cases studied, the gaps identified, and the solutions proposed include security topics, but most are not about security.

Table 16 is extracted informally from [22] to align the security gaps and solutions with the threat vector/name/countermeasure approach of other sources. The threat vector tags are applied to figures depicting threat vectors, and the countermeasures are adapted from the proposed solutions.

In this table, “MEC system” refers to the architectural building blocks “below the business level”, i.e., below the application level of a typical network hierarchy. On the other hand, “MEC Platform” refers to a network's application level, including services, identities, application and service access policies, and other similar behaviour.

Threat Vector (TV) ID	Description	Solution
MEC1	Required signalling for secure inter-MEC systems	Creation of Federation Manager network element to provide secure signalling
MEC2	MEC system discovery supporting authentication, authorization, identity management, etc.	Definition of a new reference point (Mff-fed) to support secure interaction between Federation Managers
MEC3	MEC platform discovery supporting authentication, authorization, identity management, etc.	Support of authentication, authorization, identity, etc., to be supported at application level. Possibly different keys, certificates, CAs, from those for MEC system discovery.

Table 16: Derived Threat Vectors and Solutions from [22]

E.2.4 Threat Vectors Identified by FSAG Recommendations [13], [14]

The GSMA Fraud and Security Architecture Group (FSAG) has studied security requirements for mobile communications, NFV, edge computing, and other related areas.

They identified numerous vulnerabilities and countermeasures in [14]. Table 17 lists vulnerabilities in the “threat vector” summary form. This table nor Table 19 includes countermeasures because they are thorough and extensive. Instead, references to the corresponding identifiers in [14] are provided for reference.

Threat Vector (TV) ID	Description	[14] reference
FS1	Interception and alteration of network traffic	RN-001
FS2	User tracking via device identities	RN-002

Threat Vector (TV) ID	Description	[14] reference
FS3	unspecified intrusion into or disruption of network	RN-003
FS4	unauthorized access to data in RAN	RN-005
FS5	unspecified vulnerabilities in base stations	RN-006
FS6	attacks on roaming and interconnect messaging	RI-001
FS7	unauthorized access to interconnect network elements	RI-002
FS8	need for roaming log information	RI-003
FS9	vulnerabilities in provisioning and decommissioning of users	CN-001
FS10	attacks on network traffic in core network	CN-002
FS11	eavesdropping and modification of voicemail content	CN-003
FS12	Attacks on customer identity on network	CN-004
FS13	unsolicited messaging traffic to customers	CN-005
FS14	inconsistent system state	CN-006
FS15	counterfeit, stolen, or substandard devices	CN-007
FS16	incomplete control of access policies	CN-008
FS17	inadvertent leaking of network data from network capability exposure	EC-001
FS18	access policy vulnerabilities from third parties	EC-002
FS19	compromised virtualization infrastructure and/or hardware	EC-003
FS20	Attacks on MEC platform/system from applications	EC-004
FS21	Attacks on applications by other apps	EC-005
FS22	lack of isolation of MEC network services	EC-006
FS23	Physical attacks on MEC platform	EC-007
FS24	Lack of traceability information for anomaly detection	EC-008, EC-014
FS25	Attacks on NEF availability	EC-009, EC-016
FS26	NEF confidentiality and integrity vulnerabilities	EC-010
FS27	data leakage from NEF	EC-011, EC-015
FS28	attacks on repudiation and fraud prevention of NEF	EC-012
FS29	NEF API vulnerabilities	EC-014
FS30	Container image vulnerabilities	CC-001, CC-003
FS31	Container registry/marketplace vulnerabilities	CC-002
FS32	orchestration vulnerabilities	CC-004
FS33	container runtime vulnerabilities	CC-005

Table 17: Threat vectors identified in [14]

E.3 OP Threat Vectors and Countermeasures

Annex A of this PRD provides mappings ETSI ISG MEC and 3GPP to OP (in Figure 15 and Figure 16, respectively) and is repeated here as Figure 21.

Figure 23, below, is derived from the mappings of Figure 15 and Figure 16 and from Figure 21 in this annex. It depicts the threat vectors identified by this analysis in the OP architecture

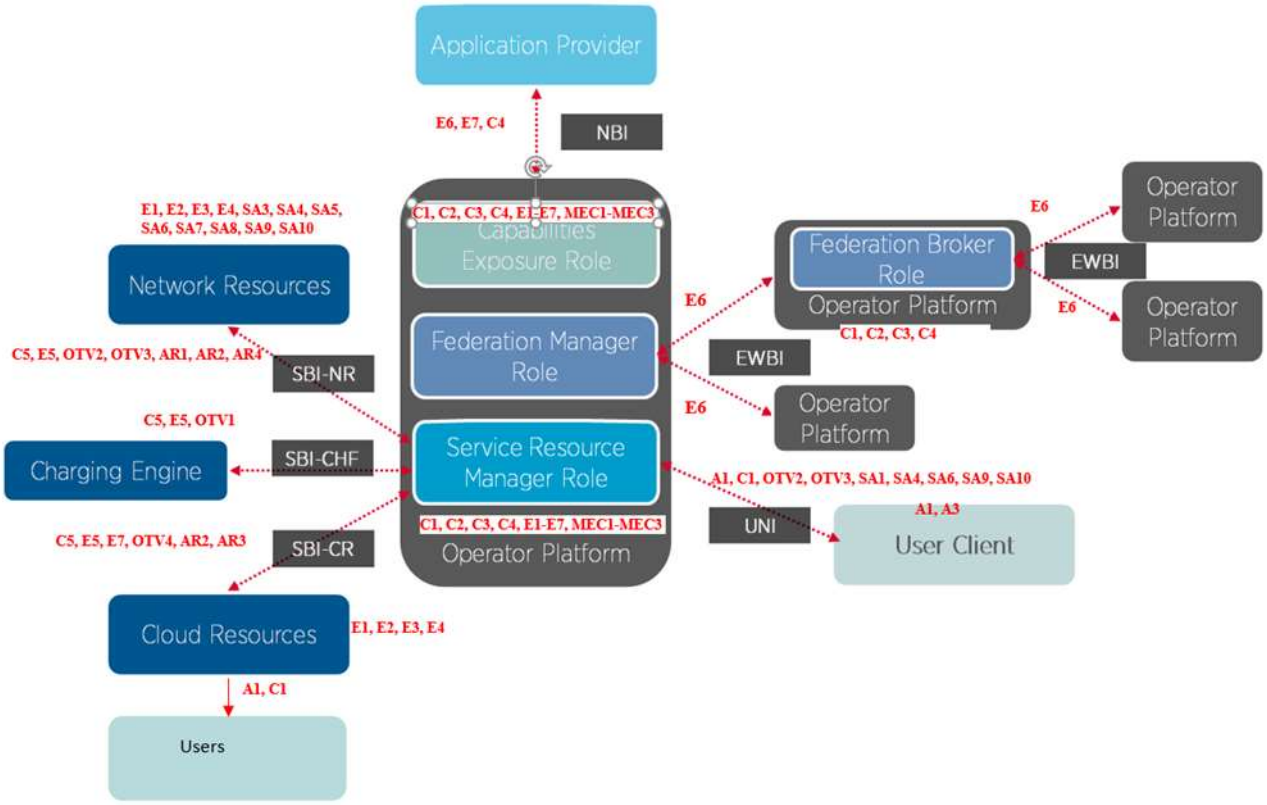

Figure 23: OP Threat Vectors

E.3.1 Access Threat Vectors

According to Figure 22, access threat vectors are at locations that connect a UE to the OP system. In ETSI ISG MEC, the vulnerabilities are on the RAN link from the UE to the BTS/eNB/gNB, between the UE app and the UE client and in the UE itself.

For OP, the RAN access link is present but is out-of-scope of the OP architecture. However, the UNI, over which control plane interactions between the UE and the OP system take place, is relevant. Internal UE vulnerabilities, particularly for app and user client, are also relevant.

The countermeasures identified for these threat vectors are listed in the following table.

Threat Vector (TV) ID	Countermeasure Recommendation
A1	Encrypting payload with AES 256-bit and securing signalling with OWS
A1	5G wireless security architecture
A1	Private LAN Service (PLS) model for multi-tier HCN
A1	RT-based channel model for 5G mmWave small cell
A3	Anomaly detection using machine learning

Threat Vector (TV) ID	Countermeasure Recommendation
A3	Security and Privacy Enhanced (SPE) framework for UEs and intent-based validation policy

Table 18: Access Threat Vectors and Countermeasure Recommendations (from [15])

E.3.2 Architecture Threat Vectors

Architecture threat vectors are vulnerabilities that occur in the overall architecture of a system or its components. Therefore, those vulnerabilities may manifest themselves in OP functions as well as in reference points.

These threat vectors were not explicitly labelled in Figure 22 (from [15]). Instead, they were added in Figure 23.

The significant categories of threat vectors have to do with validating containers and VMs, both in a particular platform and upon migration to other platforms and with performing traffic steering to applications in a secure manner.

We have proposed additional countermeasures to those presented in [15]. Some are implied in discussion within that paper but are not called out as a countermeasure. Another set of countermeasures is included by referring to work that 3GPP SA3 has done to refer. This work is not to research or forward-looking but would be items that are in a standards roadmap.

Vulnerabilities enumerated in [14] are currently categorized as architectural and so appear in this table. Because of the large number of items identified in [14], they are summarized by their identifiers in Table 17.

The countermeasures identified for these threat vectors are listed in the following table.

Threat Vector (TV) ID	Countermeasure Recommendation
AR1	Adapting mutual authentication among network slice and host network entities
AR1	Authenticating NSMs
AR1	Auditing and validating VM based slice instances
AR1	Isolation and application of diversified security for different slices
AR1	Secure service-oriented authentication framework
AR2	SFC based MEC architecture for SFs
AR2	Reactive Security framework
AR2	Standardizing on traffic steering components, e.g., AF, PCF (additional countermeasure)
AR2	Integrity of security and traffic steering parameters in packet headers (elaborated from paper)
AR3	Layered framework for VM and container migration (paper only mentions a gap, not an actual countermeasure)
AR3	Employing blockchain for establishing trust in migration

Threat Vector (TV) ID	Countermeasure Recommendation
AR4	Dynamic tunnelling method for PMIPv6
AR4	PMIPv6 based security protocol for SH-IoT
AR4	Study on PLS random models for mobility secrecy (elaborated from paper)
AR4	Monitor security levels on access networks (elaborated from paper)
AR4	Adopt best practices from 3GPP SA3
AR1	RN: Radio Network Operational Controls, FS-1 – FS-5
AR4	RI: Roaming and Interconnect Controls, FS-6 – FS-8
AR4	EC: Edge Computing & Network Exposure Functions, FS-17 – FS-29
AR4	Core Network Management Controls, FS-9 – FS-16
AR2	Virtualization Controls, FS-30 – FS-33
AR1	NS: Network Services Controls, [14] 2.2.8

Table 19: Access Threat Vectors and Countermeasure Recommendations (from [15], [14])

E.3.3 Core Threat Vectors

Core threat vectors affect the core 5G network, orchestrators, resource managers, controllers, and applications. In OP's case, where implementations of these components map onto Capabilities Exposure and Service Resource manager roles, all of the Core threat vector types appear to be relevant.

The countermeasures identified for these threat vectors are listed in the following table:

Threat Vector (TV) ID	Threat Vector (TV) ID
C1, C2, C3, C4, C5, C6	SELinux kernel and tools
C1, C2, C3	Linking remote attestation with host and system levels
C1, C2, C3	Security framework for SDN/NFV deployments (in IoT)
C1, C2, C3	Framework for adaptive trust evaluation and trusted computing technologies
C1, C3, C5, C6	Security orchestrator, security management in ETSI NFV
C1, C2, C3, C5, C6	Carry out threat analysis and security requirements in the context of NFV
C5, C6	Security Issues in SDNs when virtualized as VNFs
C5, C6	Evaluate the feasibility of extending NFV orchestrator to manage security mechanisms
C5, C6	Present integration approaches of network and security policy management into NFV
C5, C6	Provide a method of identifying the first HW unit attacked by a security attack, and security mechanism for NFV-based networks

Table 20: Core Threat Vectors and Countermeasures (from [15])

E.3.4 Edge Threat Vectors

Edge threat vectors cover platform managers, VIMs, MEC platform connectivity and connectivity of MEC apps operated at non-local base stations. These threat vectors appear to map to the EWBI.

The countermeasures identified for these threat vectors are listed in the following table:

Threat Vector (TV) ID	Countermeasure Recommendation
E1, E2	Trusted Platform Module (TPM) for validating resource exhaustion
E1, E2, E3, E4, E5, E6, E7	Form DMZs to apply access control and firewall policies at Virtual Infrastructure (VM)
E1, E2, E3	Hypervisor introspection tools serving as a HIDS
E1, E2, E3	Policy based VM IDS framework
E1, E2, E3	Encrypting VNF hard disks
E1, E2, E3	Signing VNF images
E1, E2, E3	Using a remote attestation server
E1, E2, E3, E4, E5, E6	Security framework for SDN/NFV deployments in IoT
E1, E2, E3, E4, E5, E6, E7	On-demand dynamic SFC based security service model

Table 21: Edge Threat Vectors and Countermeasures (from [15])

E.3.5 Other Threat Vectors

“Other” threat vectors (OTVs) cover areas that do not fit at a specific reference point and which manifest because of functionality, not architecture. For example, charging/billing is an OTV threat because generating events, logging and archiving them, and processing them for billing while maintaining secure subscriber IDs among the records could be associated with a charging function; but is not explicitly fixed architecturally.

These threat vectors are not explicitly labelled in Figure 22. Instead, they are provided in Figure 23.

Some countermeasures in this category were extracted from [15] rather than listed explicitly in the paper. However, it is also noted that several of them appear to be forward-looking work, and adopting best practices from 3GPP SA3 is recommended.

The countermeasures identified for these threat vectors are listed in the following table:

Threat Vector (TV) ID	Countermeasure Recommendation
OTV1	ETSI charging and billing specifications
OTV1	Security and integrity for logging and archiving of charging data (elaborated from paper)
OTV1	Security in subscriber ID assignment and tracing (elaborated from paper)

Threat Vector (TV) ID	Countermeasure Recommendation
OTV2	Blockchain
OTV2	Fuzzy logic
OTV2	Leveraging edge algorithms to mitigate IoT-DDoS attacks
OTV2, OTV3	Genetic Algorithms
OTV2	Leveraging edge computing to mitigate IoT-DDoS attacks
OTV2	Hardening resource management (elaborated from paper)
OTV2	Anomaly detection on QoE requests (elaborated from paper)
OTV3	Private LAN Service (PLS) model for multi-user multi-carrier MEC channels
OTV3	Secure UE (modified from “UAV” in paper) edge computing offloading
OTV3	MEC offloading with secure data and resource allocation
OTV4	Security service orchestration centre for SDN control plane
OTV4	SPLM for secure live migration of services
OTV4	Access control policies and deployment guidelines for Docker
OTV4	Docker escape attack defence
OTV4	Hardening network links and components (elaborated from paper)
OTV3, OTV4	Adoption of best practices from 3GPP SA3

Table 22: Core Threat Vectors and Countermeasures (from [15])

E.3.6 Privacy Threat Vectors

[15] described privacy-related threat vectors but did not map them to the ETSI ISG MEC architecture. However, because they are relevant to the OP architecture, the corresponding countermeasures have been extracted from the source to provide them in this section. For the sake of completeness, we also report here the privacy-related threat vectors from [15]:

Privacy TV	Description
P1	Data Privacy
P2	Location Privacy
P3	Identity Privacy
P4	Authorized and Curious Adversaries
P5	Computational Offloading privacy threats
P6	Service Migration privacy threats.

Table 23: Privacy Threat Vectors (derived from [15])

Privacy issues could be investigated in a future version of the present document and should probably be contextualized to the specific Use Cases defined by OP.

The authors of [15] propose the following privacy objectives for MEC:

Privacy Objectives	Recommendations
O1	Global compliance for privacy policies
O2	Responsibility of MEC service providers and consumers
O3	Privacy compliance on integrating technologies
O4	Data portability
O5	Accountability and transparency of Data Handling
O6	Declaring minimum specification requisites of UE for subscribing Mobile Edge Services
O7	Optimal utilization of UE resources with embedded privacy-enhancing mechanisms
O8	Comply with GDPR legislation.

Table 24: Privacy Objectives and Recommendations (derived from [15])

Some privacy-preserving solutions for the MEC are also proposed:

- Task Offloading based solutions: employ Constrained Markov Decision Process (CMDP) based scheduling algorithm, proposed as an approach to the task offloading process.
- Privacy partitioning, where data or devices that include information are partitioned into various layers where different privacy-preserving techniques can be applied effectively.
- Mitigation of privacy leakages in big data
- Chaff service-based privacy-preserving
- The use of privacy-preserving security protocols to guarantee anonymity, unlinkability, untraceability, non-repudiation, and confidentiality and new privacy protection schemes (such as based on blockchain approaches) for novel MEC applications.

E.4 Abbreviations and Acronyms Used in Annex E

Abbreviation/Acronym	Definition
3GPP	Third Generation Partnership Project
AF	Application Function
BS	Base Station
BTS	Base Transceiver Station (equivalent to BS)
CAPIF	Common API Framework
CFSP	Customer Facing Service Portal
D2D	Device two Device
DDOS	Distributed Denial of Service
DOS	Denial of Service
EAS	Edge Application Server
ECC	Edge Configuration Client
ECS	Edge Configuration Server

Abbreviation/Acronym	Definition
EEC	Edge Enabler Client
EES	Edge Enabler Server
eNB	E-UTRAN Node B, Evolved Node B (LTE base station)
FSAG	(GSMA) Fraud and Security Architecture Group
gNB	Next Generation Node B
HIDS	Host-based Intrusion Detection System
IDS	Intrusion Detection System
IoT	Internet of Things
LADN	Local Area Data Network
LCM	Life Cycle Management
ME App	Mobile Edge Application
MEH	Mobile Edge Host
MEO	Mobile Edge Orchestrator
MEP	Mobile Edge
NEF	Network Exposure Function
NFV	Network Functions Virtualisation
NRT	Near Real Time, or Non-Real Time
NS	Network Slicing, or Network Services
OSS	Operation Support System
PCF	Policy Control Function
PLS	Private LAN Service
PMIPv6	Proxy Mobile IPv6 (protocol)
RAN	Radio Access Network
RBAC	Role-Based Access Control
RI	Roaming and Interconnect (controls)
RN	Radio Network (operational controls)
RT	Real Time
SA3	Study Area 3 (within 3GPP)
SCEF	Session Control Exposure Function
SDN	Software Defined Network
SFC	Service Function Chain
SH-IoT	Smart Home Internet of Things
SPE	Security and Privacy Enhanced (framework for UEs)
TPM	Trusted Platform Module
TV	Threat Vector
UALCMP	User App Life Cycle Management Proxy
UAV	Unmanned Aerial Vehicle
UE	User Equipment

Abbreviation/Acronym	Definition
UE App	UE application
UPF	User Plane Function
VI	Virtualization Infrastructure
VIM	Virtualization Infrastructure Manager
VM	Virtual Machine

7 Document Management

E.5 Document History

Version	Date	Brief Description of Change	Approval Authority	Editor / Company
1.0	29 Jun 2021	New PRD OPG.02, based on requirements proposed in OPG.01.	ISAG	Tom Van Pelt / GSMA

E.6 Other Information

Type	Description
Document Owner	Operator Platform Group
Editor / Company	Tom Van Pelt / GSMA

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. You may notify us at prd@gsma.com

Your comments or suggestions & questions are always welcome.