Operator Platform: Requirements and Architecture

Version 8.0

28 February 2025

Security Classification: Non-Confidential

Access to and distribution of this document is restricted to the persons permitted by the security classification. This document 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

Copyright © 2025 GSM Association

Disclaimer

The GSMA 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.

Compliance Notice

The information contain herein is in full compliance with the GSMA Antitrust Compliance Policy.

This Permanent Reference Document is classified by GSMA as an Industry Specification, as such it has been developed and is maintained by GSMA in accordance with the provisions set out GSMA AA.35 - Procedures for Industry Specifications.

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 on the exposure of network and edge computing capabilities of public network deployments, including but not limited to platform developers, edge cloud providers, SDOs, Open Source Communities, industry fora, and market participants.

1.1.1 Relationship to existing standards

1.1.1.1 3GPP

Unless otherwise stated, the requirements listed in this document are based on the open and published 3GPP specifications listed in Section 1.6. 3GPP Release 17 is taken as the basis.

1.1.1.2 ETSI

Unless otherwise stated, the requirements listed in this document are based on the open and published ETSI ISG MEC specifications listed in Section 1.6. ETSI MEC Phase 3 is taken as the basis

1.2 Scope

This document covers requirements and architecture specifications that guide the entire industry ecosystem to define a common solution for exposing network capabilities and edge compute resources, referred to as Operator Platform. The ecosystem includes operators, vendors, OEMs, and service providers.

This document covers the following areas:

- Requirements for Operator Platform-based service exposure

  - **Exposure of Network and Edge Computing capabilities**: 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 networks and 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, among others.

- Architecture, functional levels and components

  - **Reference architecture for enabling edge computing**: Definition of modular architecture suitable for implementation at the network edge.
  - **Reference architecture for exposing network capabilities**: Definition of modular architecture suitable for the exposure of network capabilities.
  - **Federation:** Enable federation between Operator Platform instances allowing an Application Provider using an Operator Platform to access also the capabilities and resources exposed by the Operator Platform instances that have federated with that platform.
  - **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

  - The **Detailed specification** of architecture and interface specifications **will be defined by SDOs or Open-Source communities**, using the baseline in this document. While this document identifies at a high level the specifications and communities that might be relevant for the realisation of the Operator Platform, other PRDs such as OPG.03 cover the mapping in more detail.

    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 concept 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 and/or make use of the network capabilities exposed. Therefore, wherever possible, the OP concept reuses existing and established structures and processes.

1.3 Objective and use cases

Focusing on Edge Computing and network capability exposure, 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 application delivery should be covered:

- from Application Providers providing their applications to be deployed on Edge resources or using the system from their backend,
- to the deployment of resources in clouds and networks,
- the interaction of potentially multiple operators to deploy these applications or provide network capabilities, and
- finally, to the subscribers who will enjoy and interact with the application.

The use-cases covered by the OP concept demonstrate the benefits that access to network capabilities and 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
Aggregation Platform	A platform through which the Aggregator offers the services.
Aggregator	An actor who provides (or combines) services exposed by different Operators and exposes them for use to the Application Providers. Note: Exposed services by the Aggregator may differ from the services provided by the Operators. Synonyms: Channel Partner, Hyperscaler(one possible role)
Alternative QoS References	A prioritised list of alternative QoS profiles which refers to set of QoS parameters e.g., bit rate, packet delay budget etc. which OP should use in case specific QoS targets requested by the Application Provider cannot be met
Application Backend Part	An architectural part of an Application that is to be deployed on public or private (and central) cloud infrastructure.
Application Client	A specifically developed client component of an application.
Application Edge Part	An architectural part of an Edge Application that is to be deployed on edge compute cloudlets. An End-to-End Application may include multiple Application Edge Parts (e.g. microservices).

Term	Description
Application Identifier (Application ID)	Identification of an Application (owned by an Application Provider) towards the end user as part of the Consent capture
Application Instance	An instantiation of an Application Edge Part on a Cloudlet.
Application Provider	The provider of the application that accesses an 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, enterprise customer of an OP, or be an independent entity. Synonym: Developer
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 an Application Provider 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 an Application Provider is given a choice of Availability Zones in a Region, they are not coupled which ensures separation and resilience.
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.
Communication Service	a service that enables transmission and receipt of information between two or more points/entities
Communication Service Customer	an entity that uses communication services, e.g. tenant, enterprise customer, Application Provider
Communication Service Provider	an entity that provides communication services. Designs, builds and operates its communication services. The provided communication service can be built with or without a network slice.
Consent	The agreement of a subscriber to allow the usage of their personal data. This agreement can be revoked at any time.
Consent Capture	Process through which a subscriber grants permission to the OP to share certain personal data with an Application for processing under a defined Purpose of Data Processing.
Consent Record	The structure used for storing the captured Consent and related data.
Converged Charging System	The element within the Operator’s Network that allows to do the real time rating and charging for the services that are provided by that Operator [35]
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 an OP that extends an operator's reach beyond their footprint and subscriber base.

Term	Description
Ecosystem Party	In the context of GSMA OP, [OP] Ecosystem Party represents either an Application Provider, Aggregator, Partner OP or corresponding synonyms.
Edge Application	An Application whose architecture includes one or more Application Edge Parts (e.g. microservices) Note: an Application doesn't necessarily need to be an Edge Application to use capabilities exposed by an OP.
Edge Cloud (EC)	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 management function /domain is accessed through the Operator Platform. The phrase "located at the infrastructure edge" is not intended to define where an Operator deploys its Edge Cloud. The Edge Cloud is expected to be closer (for example, latency, geolocation, etc.) to the Application Clients than today's centralised data centres, but not on the User Equipment, and could be in the last mile network. Note: This definition is based on that in "Open glossary of edge computing", v2.0 [3].
Edge Cloud Resources	In the context of this document, resources of the Edge Cloud Service whose management function / domain is accessed through the Operator Platform SBI-CR.
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: This definition is based on that in "Open glossary of edge computing", v2.0 [3].
Edge Interconnection Network (EIN)	A direct and dynamically managed (optionally pre-existing) logical interface between two EC instances. The interface shall use existing network infrastructure for connection establishment, and may have security and rules applied based on the EC and OP requirements.
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: This definition is based on that in "Open glossary of edge computing", v2.0 [3].
Edge Node	A resource in a physical data centre. The term Edge Node used in context with the Edge Node Sharing refers to the compute resources offered by the Partner OP to the Leading OP. The Leading OP may use such resources to serve its own end users in scenarios such as not having the edge clouds

Term	Description
	footprint in locations where the end users requesting access to edge services but a Partner OP is offering edge cloud resources in those locations. [29]
Edge Resource	Sum of compute, network, and storage capabilities made available for workload deployment and processing in edge nodes.
Edge Site	A physical location where an edge node is deployed.
EIN establishment	A procedure to create an EIN connection between two ECs. The process also establishes application and traffic security and filtering rules on the EIN as part of the establishment.
EIN Termination	A closure of an existing EIN connection. It can include cleaning up the application security and traffic rules applied at the time of EIN establishment.
End-user	A human participant who uses the application. A customer of the Application Provider. Note: The End-user is not always the Subscriber.
End-to-End network slicing	Slicing concept for mobile network which include UE, RAN, Core and Transport. [33]
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 [21]. 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
Marketplace Platform	A platform where services (and APIs) are published and offered to 3rd parties (e.g. Application Providers).
Network Communication Service	The external representation of a communication service orderable by an Application Provider
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). An OP decides on the Network Resource Locations.
Network Resources	In the context of this document, the network services and capabilities provided by the Operator whose management function /domain is accessed through the Operator Platform SBI-NR.
Network Slice	A logical network that provides specific network capabilities and network characteristics [10]

Term	Description
Network Slice Instance	A set of Network Function instances and the required resources (e.g. compute, storage and networking resources) which form a deployed Network Slice [10]
Network Subscription	An entry in the network user database that regulates authentication and authorization (including policy aspects) of a user trying to access network services. It includes a mapping from its GPSIs to SUPI, only a single SUPI or IMSI/MSISDN (in case 3G/4G access) can be active at a time for the network subscription
Non-SIM User Equipment (Non-SIM UE)	Any device that is used by end-user that does not require SIM to access communication services. Examples of such devices are Tablets, Laptops, VR headsets, gaming consoles, IoT sensors etc. UCs and Application Clients are deployed on the Non-SIM User Equipment.
Northbound Interface	The interface that exposes an Operator Platform to Application Providers
Operator	In the context of GSMA OP, an Operator is an entity that exposes capabilities and/or resources of their network (IT, mobile) to Application Providers, provides connectivity to User Equipment and has an Operator Platform. Synonyms: CSP (Communication Service Provider), MNO (Mobile Network Operator)
Operator Platform	An Operator Platform (OP) facilitates access to the Edge Cloud and other capabilities of an Operator or federation of Operators and Partners. It follows the architectural and technical principles defined in this document.
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.
Privacy Management	Service within the CSP domain supporting create, read, update, and delete (CRUD) operations on Application-related Consent records. The service supports also notifying (to the interest parties) when a Consent record has changed.
Purpose of Data Processing	Reason for which processing personal information is required by the Application (owned by an Application Provider). It declares what the application intends to do with a set of personal information resources. Each Purpose of Data Processing maps to a set of Scopes (usually defined in an API specification).
Region	An OP Region is equivalent to a Region on a public cloud. The higher construct in the hierarchy exposed to an Application Provider 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

Term	Description
	scope of geography managed by a Regional Controller is up to the operator to define.
Representational Consistency	Representational Consistency means that the information elements that the Application Provider exchanges with an Edge Cloud do not change as a function of the Partner OP with which it is ultimately interacting. This implies that a function of the Exposure functional level is to provide a consistent information model.
Service Continuity	The uninterrupted user experience of a service, including in those cases where the IP address or anchoring point change
Session Continuity	The continuity of a PDU Session. For PDU Session of the IPv4 or IPv6 or IPv4v6 types, "Session Continuity" implies that the IP address is preserved for the lifetime of the PDU Session
Southbound Interface	Connects an OP with the specific operator infrastructure that delivers the network, cloud and charging services and capabilities.
Subscriber	A client/customer of the Operator, identified by a unique identifier.
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 Virtual Machines (VMs) running native or containerised applications or cover a complete server.
User Client	Functionality that manages on the user's side the interaction with an OP. The User Client (UC) 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 (UE)	Any device with a SIM used directly by an end-user to communicate. UCs and Application Clients are deployed on the User Equipment. By default, the term “UE” means UE with the explicit SIM-based Telecom wireless network connectivity throughout the document.
User Identity Token	A security token that serves as proof of authentication, confirming that a user is successfully authenticated with the parameters provided at token generation.
User Identity Token Manager	An OP (external) function that generates User Identity Tokens for UE application clients and verifies the User Identity Token when received from the AP through the NBI interface.
User-Network Interface	Enables the UC hosted in the user equipment to communicate with an 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 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

1.5 Abbreviations

Term	Description
3GPP	Third Generation Partnership Project
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
BS	Base Station
BTS	Base Transceiver Station (equivalent to BS)
CAPIF	Common API Framework
CCS	Converged Charging System
CDM	Common Data Model
CEF	Charging Enablement Function
CFSP	Customer Facing Service Portal
CHF	Charging Function
CI/CD	Continuous Integration / Continuous Development and Deployment
CISM	Container Infrastructure Service Manager
CN	Core Network
CPU	Central Processing Unit
CRUD	Create, Read, Update and Delete
CSC	Communication Service Customer
CSP	Communication Service Provider
CTF	Charging Trigger Function
D2D	Device to Device
DBaaS	DataBase as a Service
DDoS	Distributed Denial of Service
DL	DownLink
DNAI	Data Network Access Identifier
DNN	Data Network Name
DNS	Domain Name System
DoS	Denial of Service
DSCP	Differentiated Services Code Point

Term	Description
EAS	Edge Application Server
EC	Edge Cloud
ECN	Explicit Congestion Notification
ECP	Edge Computing Platform
ECS	Edge Configuration Server
ECSP	Edge Computing Service Provider
EEC	Edge Enabler Client
EES	Edge Enabler Server
EGMF	Exposure Governance Management Function
EIN	Edge Interconnection Network
eNB	E-UTRAN Node B, Evolved Node B (LTE base station)
ETSI	European Telecommunications Standards Institute
E/WBI	East/Westbound Interface
eMBB	Enhanced Mobile Broadband
FPGA	Field Programmable Gate Array
FQDN	Fully Qualified Domain Name
FSAG	(GSMA) Fraud and Security Architecture Group
GDPR	General Data Protection Regulation
GMLC	Gateway Mobile Location Centre
gNB	gNodeB
GPS	Global Positioning System
GPSI	Generic Public Subscription Identifier
GPU	Graphic Processing Unit
GST	Generic network Slice Template
GW	GateWay
HIDS	Host-based Intrusion Detection System
HPLMN	Home Public Land Mobile Network
HR	Home Routing
HTTP	HyperText Transfer Protocol
IaaS	Infrastructure as a service
ID	IDentifier
IDS	Intrusion Detection System
IEC	Immediate Event Charging
IETF	Internet Engineering Task Force
IMSI	International Mobile Subscriber Identity
I/O	Input/Output
IoT	Internet of Things
IP	Internet Protocol

Term	Description
IPsec	Internet Protocol Security
ISG	Industry Specification Group
ITU	International Telecommunication Union
KPI	Key Performance Indicator
L4	Layer 4
L4S	Low Latency, Low Loss, Scalable Throughput
LADN	Local Area Data Network
LAI	Location Area Identification
LBO	Local BreakOut
LCM	Life-Cycle Management
MCC	Mobile Country Code
ME App	Mobile Edge Application
MEC	Multiaccess Edge Computing
MEH	Mobile Edge Host
MEO	Mobile Edge Orchestrator
MEP	Mobile Edge Platform
MNC	Mobile Network Code
MR	Mixed Reality
MSISDN	Mobile Subscriber Integrated Services Digital Network Number
NAS	Non-Access Stratum
NAT	Network Address Translation
NBI	Northbound Interface
NCSaaS	Network Communication Service as a Service
NDS	Network Domain Security
NEF	Network Exposure Function
NEST	NEtwork Slice Type
NFV	Network Functions Virtualisation
NRT	Near Real Time, or Non-Real Time
NS	Network Slicing, or Network Services
NSaaS	Network Slice as a Service
NSI	Network Slice Instance
NSMF	Network Slicing Management Function
NSSI	Network Slice Subnet Identifier
NSSAI	Network Slice Selection Assistance Information
NPU	Neural Processing Units
NUMA	Non-Uniform Memory Access
NWDAF	Network Data Analytics Function
OCI	Open Container Initiative

Term	Description
OP	Operator Platform
OS	Operating System
OSC	Open Source Community
OSS	Operation Support System
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
PEC	Post Event Charging
PGW	PDN (Packet Data Network) GateWay
PII	Personally-Identifiable Information
PLS	Private LAN Service
PMIPv6	Proxy Mobile IPv6 (protocol)
PRD	(GSMA) Permanent Reference Document
PrM	Privacy Management
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RBAC	Role-Based Access Control
RI	Roaming and Interconnect (controls)
RN	Radio Network (operational controls)
RNIS	Radio Network Information Service
RRS	Resource Requirements Specification
RT	Real Time
SA3	Service and System Aspects WG3 (within 3GPP)
SAAS	Software as a service
SBI	Southbound Interface
SBI-CR	Southbound Interface – Cloud Resources
SBI-NR	Southbound Interface – Network Resources
SBO	Session BreakOut
SCEF	Service Capability Exposure Function
SD	Service Differentiator
SDK	Software Development Kit
SDN	Software Defined Network
SDO	Standards Developing Organisation
SFC	Service Function Chain

Term	Description
SH-IoT	Smart Home Internet of Things
SLA	Service Level Agreement
SLI	Service Level Indicators
SMF	Session Management Function
S-NSSAI	Single Network Slice Selection Assistance Information
SPE	Security and Privacy Enhanced (framework for UEs)
SPR	Subscriber Profile Repository
SR/IOV	Single Root I/O Virtualisation
SSC	Session and Service Continuity
SST	Slice/Service Type
SUPI	SUbscription Permanent Identifier
TAC	Tracking Area Code
TAI	Tracking Area Identification
TLD	Top-Level Domain
TLS	Transport Layer Security
TPM	Trusted Platform Module
TTL	Time To Live
TV	Threat Vector
UALCMP	User Application Life Cycle Management Proxy
UAV	Unmanned Aerial Vehicle
UC	User Client
UE	User Equipment
UE App	UE application
UL	UpLink
UNI	User to Network Interface
UPF	User Plane Function
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URSP	UE Route Selection Policy
VI	Virtualisation Infrastructure
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]	GSMA WP 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 TS 29.522	5G System; Network Exposure Function Northbound APIs, Stage 3 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]	GSMA PRD NG.126	Cloud Infrastructure Reference Model
[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
[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]	GSMA PRD FS.30	Security Manual, GSM Association Official Document FS.30, 20 April 2020.
[14]	GSMA PRD FS.31	Baseline Security Controls, GSM Association Official Document FS.31
[15]		Pasika Ranaweera, et al., Survey on Multi-Access Edge Computing Security and Privacy, to be published in IEEE Communications Surveys & Tutorials
[16]	3GPP TS 33.122	Security Aspects of Common API Framework (CAPIF) for 3GPP northbound APIs (Release 16) https://www.3gpp.org/DynaReport/33122.htm

Ref	Doc Number	Title
[17]	3GPP TS 23.558	Architecture for enabling Edge Applications https://www.3gpp.org/DynaReport/23558.htm
[18]	ETSI ISG MEC 003	Multi-access Edge Computing (MEC); Framework and Reference Architecture https://www.etsi.org/deliver/etsi_gs/MEC/001_099/003/03.02.01_60/gs_mec003v030201p.pdf
[19]	3GPP TS 29.514	5G System; Policy Authorization Service; Stage 3 https://www.3gpp.org/DynaReport/29514.htm
[20]	3GPP TR33.839	Study on Security Aspects of Enhancement of Support for Edge Computing in 5GC https://www.3gpp.org/DynaReport/33839.htm
[21]	3GPP TS 23.122	Non-Access-Stratum (NAS) functions related to Mobile Station (MS) in idle mode
[22]	ETSI ISG MEC 35	Multi-access Edge Computing (MEC): Study on Inter-MEC systems and MEC-Cloud systems coordination, V3.1.1 (2021-06). https://www.etsi.org/deliver/etsi_gr/MEC/001_099/035/03.01.01_60/gr_mec035v030101p.pdf
[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 TS 33.310	Network Domain Security (NDS); Authentication Framework (AF) https://www.3gpp.org/DynaReport/33310.htm
[25]	IETF RFC 4122	A Universally Unique IDentifier (UUID) URN Namespace https://datatracker.ietf.org/doc/html/rfc4122
[26]	3GPP TS 23.548	5G System Enhancements for Edge Computing; Stage 2 https://www.3gpp.org/DynaReport/23548.htm
[27]	3GPP TS 23.503	Policy and charging control framework for the 5G System https://www.3gpp.org/DynaReport/23503.htm
[28]	3GPP TR 23.748	Study on enhancement of support for Edge Computing in 5G Core network (5GC) https://www.3gpp.org/DynaReport/23748.htm
[29]	GSMA PRD OPG.04	East-Westbound Interface APIs
[30]	IETF RFC 5865	A Differentiated Services Code Point (DSCP) for Capacity-Admitted Traffic https://datatracker.ietf.org/doc/html/rfc5865
[31]	IETF RFC 9330	Low Latency, Low Loss, Scalable Throughput (L4S) Internet Service: Architecture https://datatracker.ietf.org/doc/html/rfc9330
[32]	GSMA PRD NG.116	Generic Network Slice Template

Ref	Doc Number	Title
[33]	GSMA PRD NG.135	E2E Network Slicing Requirements
[34]	3GPP TS 28.530	Management and orchestration; Concepts, use cases and requirements https://www.3gpp.org/DynaReport/28530.htm
[35]	3GPP TS 32.240	Charging management; Charging architecture and principles https://www.3gpp.org/DynaReport/32240.htm
[36]	3GPP TS 28.201	Charging management;Network slice performance and analytics charging in the 5G System (5GS);Stage 2 https://www.3gpp.org/DynaReport/28201.htm
[37]	3GPP TS 28.202	Charging management;Network slice management charging in the 5G System (5GS);Stage 2 https://www.3gpp.org/DynaReport/28202.htm
[38]	3GPP TS 28.533	Technical Specification Group Services and System Aspects; Management and orchestration architecture framework https://www.3gpp.org/DynaReport/28533.htm
[39]	3GPP TS 32.257	Telecommunication management; Charging management; Edge computing domain charging https://www.3gpp.org/DynaReport/32257.htm
[40]	GSMA PRD OPG.03	Southbound Interface Network Resources APIs
[41]	GSMA PRD OPG.05	User-Network Interface APIs
[42]	ETSI ISG MEC 040	Multi-access Edge Computing (MEC); Federation enablement APIs https://www.etsi.org/deliver/etsi_gs/MEC/001_099/040/03.02.01_60/gs_mec040v030201p.pdf
[43]	IETF RFC 2119	“Key words for use in RFCs to Indicate Requirement Levels”, S. Bradner, March 1997. Available at http://www.ietf.org/rfc/rfc2119.txt
[44]	3GPP TR 23.958	Edge Application Standards in 3GPP and alignment with External Organizations https://www.3gpp.org/DynaReport/23958.htm
[45]	3GPP TS 33.558	Security aspects of enhancement of support for enabling edge applications https://www.3gpp.org/DynaReport/33558.htm
[46]	3GPP TS 33.501	Security architecture and procedures for 5G System https://www.3gpp.org/DynaReport/33501.htm
[47]	IETF RFC 9113	HTTP/2 https://datatracker.ietf.org/doc/rfc9113
[48]	IETF RFC 9110	HTTP Semantics https://datatracker.ietf.org/doc/rfc9110
[49]	IG1224 v13	TM Forum NaaS Transformation v13.0.0 https://www.tmforum.org/resources/reference/naas-transformation-v13-0-0-ig1224/

Ref	Doc Number	Title
[50]	GSMA PRD WA.101	Open Gateway Channel Partner Onboarding Guide
[51]	GDPR	"General Data Protection Regulation". Available at https://gdpr-info.eu/
[52]	3GPP TR 33.867	Study on user consent for 3GPP services https://www.3gpp.org/DynaReport/33867.htm
[53]	RFC 6749	The OAuth 2.0 Authorization Framework https://datatracker.ietf.org/doc/html/rfc6749
[54]	CIBA Flow	OpenID Connect Client-Initiated Backchannel Authentication Flow - Core 1.0 https://openid.net/specs/openid-client-initiated-backchannel-authentication-core-1_0.html
[55]	GSMA PRD OPG.09	NBI APIs Realisation in the SBI
[56]	GSMA PRD OPG.10	Open Gateway Technical Realisation Guidelines

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

1.7 Conventions

The key words "must", "must not", "required", "shall", "shall not", "should", "should not", "recommended", "may", and "optional" in this document are to be interpreted as described in RFC2119 [43].

2 Architectural Requirements

This section defines the requirements that an OP's architecture should fulfil. Section 2.1 defines the high-level requirements for the platform in general that are independent of the services/capabilities exposed and section 2.3 provides the security requirements of an OP at a similar service-agnostic level. Other sub-sections provide requirements on the exposure of specific services using an OP, section 2.2 on the exposure of Edge capabilities and section 2.4 on the exposure of network capabilities.

Note: The section does not define the architecture itself nor what the platform's interfaces should support. Those aspects are introduced in sections 3 and 5 of this document respectively.

2.1 High-Level Requirements

This section defines the requirements that the OP's architecture should fulfil independent of the actual service or capabilities exposed by the platform. Next to general requirements in section 2.1.1, the remaining subsections define what an OP should offer to the different parties that might interact with it.

Note: despite the service-agnostic nature of the section for historic reasons a few requirements are included that relate to specific capabilities exposed.

2.1.1 General

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

1. An OP shall expose supported network capabilities to the Application Providers.
2. For each operator supporting the OP architecture, 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.

Note: This instance may be operated by the Operator or be outsourced.

3. An 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

An 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 the Application Providers for the development and deployment of edge applications shall allow for easy deployment of application instances developed for public clouds.

Note: An OP can only manage application instances or resources on the infrastructure under direct control. Federation with other OPs is used when applications or resources need to be managed on other infrastructure.

3. An OP shall allow an Application Provider to reserve resources for future application instance deployments, ensuring the availability of the booked capacity.
4. An OP shall allow an edge application to be deployed within an operator network where it can utilise the optimum resources.
5. An OP shall hide the complexity of the OP architecture, the involved operator networks and client access to those networks from the Application Providers.
6. There shall be a "separation of concerns" of the OP and the 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) An 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) An OP does not know how the application works (for instance, it does not know about the application's identifiers and credentials).

c) It is responsibility of the OP to provide to the Application Providers the required tools for them to operate accordingly to the local regulation.
d) The OP shall provide to the Application Providers all the required information, e.g. a clear identification of the geographical location of the Edge resources, with a level of granularity that allows the Application Providers to manage their services in a way that is compliant with local regulations.
e) It is responsibility of the Application Providers to use the tools and the information provided by the OP to operate their applications in compliance with the local regulations.

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

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

Note: It is up to the individual parties providing an OP to decide whether they offer these capabilities in their deployment.

12. An OP shall expose network capabilities to the Application Providers, including longer-term managed network services (such as for QoS, i.e. Quality of Service) and shorter-term or transactional style services (such as SIM-derived services, such as location verification).

Note: It is the responsibility of the Application Provider to ensure proper QoS support in the application when making use of end-to-end QoS mechanisms such as Low Latency, Low Loss, Scalable Throughput (L4S), traffic category and Differentiated Services Code Point (DSCP).

2.1.3 Functionality offered to End-Users/Devices

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

1. An OP shall allow end-user devices to access services provided through Edge Enhanced and Edge Native Applications.

    a) An OP shall be able to manage the service access that the device can use to reach the Edge Native Applications.

2. An 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.

4. The OP shall allow the end-user to access service in the visited network securely.
5. The OP, in coordination with the Consent Management function, located in the CSP domain, shall provide the means for the End-User to express its consent for the usage of its sensitive data with respect to the services provided by the OP, in accordance with the local regulation.

   Note: Availability in visited networks is dependent on the Operator’s roaming agreements.

2.1.4 Functionality offered to Operators

An 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 enable operators to establish EIN connections, and monitor their usage by the applications for charging purposes.
4. The OP architecture shall allow an operator to charge for the services and capabilities provided by OP to application providers, subscribers, and other operators.
5. The OP architecture shall allow an OP to influence the quality of service delivered by the network for the interaction between an end-user device and an application.
6. If an Operator Platform is part of the operator's security domain (see Note 2), it shall 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.

7. Where an Operator Platform is not part of an 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. In terms of responsibilities:

    a) It is responsibility of Operators to provide to the OP the required tools for the OP to operate accordingly to the local regulation.
    b) The Operator shall provide all the required information to the OP, e.g. a clear identification of the geographical location of the Edge resources, with a level of granularity that allows the OP to be compliant with local regulations.
    c) It is the responsibility of the OP to use the tools and the information, provided by the Operator, to offer its services, to its customers, in compliance with the local regulations.

2.1.5 Functionality offered to other OPs

An 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, operate and manage Edge Applications provided by the Application Providers with 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 shall enable additional capabilities, such as:

    a) the User Equipment (UE) continuation of use of 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 Visited OP to receive applications from Home OPs to serve subscribers, whether they are home OP subscribers or visiting OP subscribers.
4. The OP architecture shall allow a Leading OP to monitor and track resource usage of an application in the OP on which it has been deployed.
5. Similarly, there shall be a "separation of concerns" between OPs. In terms of responsibilities:

a) It is responsibility of "Partner OP" to provide to the "Leading OP" the required tools for the "Leading OP" to operate accordingly to the local regulation.
b) The “Partner OP" shall provide to the "Leading OP" all the required information, e.g. a clear identification of the geographical location of the Edge resources, with a level of granularity that allows the "Leading OP" to be compliant with local regulations
c) It is responsibility of the "Leading OP" to use the tools and the information, provided by the "Partner OP", to offer its services, to its customers, in compliance with the local regulations.

2.1.6 High-Level Roaming Requirements

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

1. UEs, while roaming, shall be able to access applications deployed on edge resources within the visited network with the specified characteristics.

   Note 1: This requires local breakout (LBO) or session breakout (SBO) of the subscriber's Protocol Data Unit (PDU)/Packet Data Network (PDN) connection to a UPF/PDN Gateway (PGW) in the visited network.

   Note 2: To allow LBO/SBO for the subscribers, the agreements between operators need to be in place. Part of the information shared between the operators is the APN/DNN/ Network Slice Selection Assistance Information (NSSAI) used for LBO/SBO.

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

   Note 3: Availability of the applications a UE 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 UE (e.g. the required federation or LBO roaming agreements are missing), the UE shall still be routed to the most favourable location. This would be the location in the network closest to the UE 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 UE'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 UEs connected to visited networks, given that such access may result in significant increases in latency.

   Note 4: Seamless handover from home or visited network to another visited network is not in the scope of the current version of this document.

2.1.7 High-Level Privacy Requirements

An OP shall:

1. Ensure that there is a legal basis under the local privacy regulation (if any) for sharing personal data on subscribers with an Application (potentially through an Aggregator) for the Purpose of Data Processing indicated by the Application Provider.
2. Ensure that no restriction on data sharing is in place when the Purpose of Data Processing is covered by a legal basis (different from Consent) that is compliant with local regulations.
3. Interface with an external Consent Manager to retrieve Application-related Consent information (if already available) or if the subscriber's Consent to share data with a particular Application Provider must be captured.
4. Ensure that authorization (see Note 1) and Consent (if applicable) are obtained prior to an API invocation that would trigger the sharing of the personal data with an Application Provider for a specific Purpose of Data Processing.
5. Support Consent (to share personal data with an Application) being revoked by the subscriber.
6. Support Legitimate Interest legal basis (when applicable) being objected by the subscriber (e.g., the subscriber may object marketing campaigns at any time).
7. Keep records of access to personal data through Logging, Tracing and Auditing functions.
8. Have the technical mechanisms in place to guarantee mutual authentication with an Application Provider so that the Consent capture can be based on reliable information elements, e.g., Application Provider ID, Application ID, Purpose of Data Processing, etc.
9. Present available APIs, Scopes, Purpose of Data Processing to the Application Provider at any time (and to other OPs in federated environments)
10. If the integration of an external party (e.g., Application Provider) includes more parties, e.g., an Aggregator in-between, the identity of the party triggering the capture of the Consent (if applicable legal basis) shall be the identity of the Application Provider.

Note 1: Whether that authorization is obtained in batch or as part of individually triggered requests depends on the application logic and similarly what identifiers are used to identify the subscriber on whom they would obtain the data.

Editor's Note: Further architecture considerations and requirements to verify the Application ID are FFS.

2.1.8 Functionality offered to Aggregators

An OP needs to expose information to enable an Aggregator to find the Home OP where to route an API call to on behalf of an Application Provider's application. The Application Provider's application typically has the MSISDN, IP address or a User Identity Token that will be provided to the Aggregator in the API call. Such a User Identity Token should allow to identify both the Network Subscription and the Application using the subscribed network services. The OP architecture shall fulfil the following requirements related to the functionality offered to Aggregators.

1. An Aggregator shall be able to identify the Home OP based on the identifiers in an API request.
2. The OP architecture shall allow an Aggregator to use a common interface to find the target OP with a single API call to identify the target Operator ID.
3. The OP architecture shall enable the Aggregator to discover the End-user's Home OP when roaming.
4. The OP architecture shall enable the Aggregator to identify the target OP for non-subscription-related APIs

2.2 Edge Enabling Requirements

This section defines the requirements that the OP architecture should fulfil for the exposure of edge computing capabilities. Section 2.2.1 defines the high-level requirements for exposing these capabilities. Section 2.2.2 goes into the management and reservation of compute resources. Section 2.2.3 defines the requirements for integration with development environments for Edge Applications. Section 2.2.4 provides requirements on edge specific enhancements while section 2.2.5 defines data protection requirements. Requirements on what an OP should enable regarding the life-cycle management of the Edge parts of Edge Applications are defined in section 2.2.6 and section 2.2.7 covers the requirements on what an OP should provide to support Edge Compute capabilities serving subscribers that may be mobile. Section 2.2.8 defines what an OP should support to manage the interconnection network between Edge Cloudlets and to enable Edge Application access to that interconnection network.

2.2.1 High-Level Requirements

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

1. An OP shall allow an 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 an 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 OP-enabled Edge Resource management requirements

2.2.2.1 General principles

"Edge Resource" refers to edge compute resources (processing and storage), associated networking, associated container resources and edge application resources.

The general principles for OP-enabled Resource Management are as follows:

- An 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 with.

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 Edge Resource management accessed through the OP

An OP and its architecture shall fulfil the following requirements related to enabling access to the management function/domain for edge compute resources (processing and storage) and associated networking:

1. An OP shall provide access to edge compute resources to another party in the OP ecosystem (e.g. an Application Provider, an Aggregator, a Partner OP or another operator).
2. An OP shall facilitate access to the management function of the virtualised resources. For example, this includes the reservation, de-reservation, allocation, de-allocation and potentially lifecycle management (such as scaling) of virtualised resources to a specific Application Provider.
3. If one OP Ecosystem Party (e.g., Application Provider)overloads the virtualised resources assigned to it, this should not degrade the performance of other resources assigned to the other OPs or Ecosystem Parties.
4. An OP shall allow an Edge Application to be relocated to another appropriate edge cloud for optimum resource utilisation.
5. An OP or an Application Provider does not have visibility of the resources that another OP or Application Provider has allocated or is using.
6. All parties in the OP ecosystem shall use the same data model for the virtualised resources.
7. It shall be optional for an OP to expose telemetry or other resource-related metrics from the edge node to Application Providers or other OPs.

2.2.2.3 Edge Resource Reservation

An OP shall allow OP Ecosystem Parties 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 OP Ecosystem Party (e.g., Application Provider) shall be able to reserve a certain amount of resources that would be logically bound to them.
2. An OP shall allow to validate the reservation based on the currently available resources and to 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

An 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

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

2.2.5 Data Protection Management

An 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. An 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 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.

The following connectivity models have been specified in 3GPP TS 23.548 [26] to enable Edge Computing in 5G networks:

1. Distributed Anchor Point
2. Session Breakout
3. Multiple Protocol Data Unit (PDU) Sessions

Each of these connectivity models may be used to optimise the user plane data routing towards the Edge cloud. During UE mobility between networks of different operators, the requested level of QoS shall be provided with connectivity to the appropriate Edge Applications instance(s) irrespectively of the connectivity models and session continuity modes. Mapping a QoS model to network implementation and the methods for Edge Application server discovery shall be the responsibility of each operator OP and operator deployments may differ but should be compatible with 3GPP specifications.

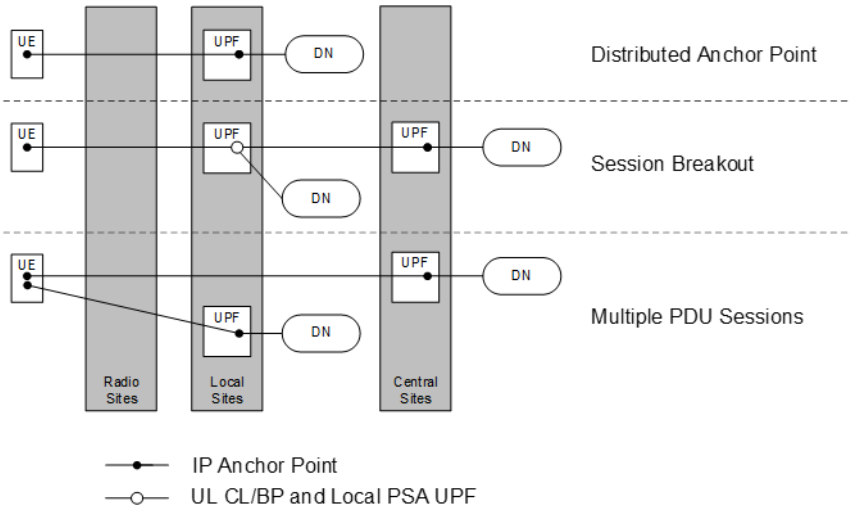


Figure 1: 5GC Connectivity Models for Edge Computing, REF 3GPP TS 23.548 [26]

When Distributed Anchor Point is used, all applications connect to the closest anchor point; this method is device independent, including 4G devices. However, it also means that all sessions with the same Data Network Name (DNN) and Single Network Slice Selection Assistance Information (S-NSSAI) are directed to the distributed User Plane Function(UPF), not only edge traffic.

Session Breakout works for all 5G terminals with 5G coverage; however, it is not supported in 4G.

Multiple PDU sessions are supported in 4G and 5G networks from 3GPP Rel-15 onwards, and the UE support is gaining momentum. UE Route Selection Policy (URSP) rules are sent to the UE when connecting to 5G so that traffic flows for edge applications can be separated from other traffic and routed to a local UPF/Edge Data Network.

The network shall support at least the Distributed Anchor Point connectivity model. It is recommended to consider Multiple PDU Sessions support as soon as commercially viable.

The following requirements apply for all three connectivity models listed above

- An OP should be able to influence the URSP rules sent to the UE related to the applications managed by that OP.

Note: Some limitations exist when interworking with EPS, such as described in Annex E of 3GPP TS 23.548 [26].

- An OP, networks and terminals shall support all Session and Service Continuity (SSC) modes.
- An OP shall be able to control UPF reselection in its own network via existing methods (Influence on traffic routing) to influence options for application distribution
- The network shall be able to notify the OP in case of UPF reselection. The NEF API is available and notifications need to be requested for every session of interest.

The following requirements apply for the Multiple PDU Sessions connectivity model

- The Home OP shall be able to influence the URSP rules sent to the UE related to the applications managed by the OP
    - The Multiple PDU sessions connectivity model allows for flexible (and dynamic) mapping of Application traffic to PDU sessions. Without URSP support to control and update the application to PDU session mapping, this flexibility will be unmanageable.

2.2.7.1 Void

2.2.7.2 Requirements for defining geographical conditions on mobility

An Edge 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.

An OP shall be able to receive an Edge 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. An 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. An OP shall perform the Application Instance mobility management process to ensure that the criteria are accomplished.

Note: Section 5.2.4.2 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 an 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.

An 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.

2.2.8 Edge Interconnection Network (EIN) management

An OP shall provide a way to establish and manage EIN connections between two ECs. This may include the following:

- Communicating connection information of the peer ECs
- Enforcing security guidelines and providing security parameters to ECs
- Setting up rules to allow/restrict application access – access control, traffic filtering, application filtering etc.
- EIN Termination

Note: Actual EIN setup and termination may be delegated to underlying network infrastructure controller.

Note: Handling of EIN across operators involving federation establishment is case for further future study.

2.3 High-Level Security Requirements

The OP architecture shall comply with the following security requirements:

1. An OP shall expose network capabilities and resources data (e.g., compute and storage) to Application Providers and federated Partner OPs following the 'need-to-know' principle and only for the legitimate scenarios expected in the PRD.
2. An OP shall not expose its configuration data and internal topology (referred to as topology hiding) to Application Providers and federated Partner OPs.
3. An OP shall apply data protection mechanisms to assure data availability, confidentiality, authenticity, and integrity. Data shall be protected both during storage and processing and be transported 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.

4. The permitted data (i.e., data that may be shared on the need-to-know principle as in requirement 1) shall be exposed only to authorised and authenticated entities.
5. An OP shall implement role-based access control for configuring users, with policies defined and enforced, ensuring a secure binding between services and authorised entities.
6. An OP shall adopt an integrity protection mechanism for the various identifiers in use (such as resource IDs, user/subscriber IDs, session IDs, application IDs).
7. An OP shall adopt certificate-based authentication and security protocols as described in 3GPP TS 33.310 [24].
8. An OP should support TCP proxies to avoid server IP address guessing and TCP connection hijacking.
9. An OP should support flow-control on invoking application services control plane APIs to protect federated services from abuse of these APIs.
10. An OP should support different role-based privileges for such roles as OP tenants and network/infrastructure operators to control unauthorised access to network slice management of shared/virtualised resources.
11. Security mechanisms (e.g., certificate authorities) used to protect tracking and logging of information on an OP's different interfaces should be independent of each other.

2.4 Network Capability Exposure requirements

This section provides the requirements that an OP's architecture should fulfil for the exposure of Network Capabilities to Application Providers. Section 2.4.1 provides high-level requirements for that exposure and section 2.4.2 defines requirements the different types of network capabilities that could be exposed and their management as well as their usage by an OP. Finally, the aspects around mobility of subscribers for which an OP is managing capabilities are covered in the requirements in section 2.4.3.

2.4.1 High-Level requirements

The following requirements apply for an OP related to the exposure of network capabilities:

1. An OP shall allow an Operator to expose network capabilities within the Operator or Partner network on which applications consuming such capabilities can be deployed.
2. An OP shall allow the exposure of operator network capabilities beyond edge resources based on the integration of 3GPP functions like:

    a) Network Exposure Function (NEF)
    b) Network Data Analytics Function (NWDAF)

3. The OP architecture shall allow an Application Provider to consume network capabilities for their end-to-end application.

Note: The capabilities are those of the network domains used by the IP flows between the Application Clients and the corresponding Application Instances associated with that Application Provider. Potentially this includes different network domains, such as

- RAN
- Core Network
- Transport Network (including Intra-Cloudlet network and Inter-Cloudlet network)
- Non-3GPP Access networks

4. An OP needs to handle the situation that specific capabilities are not or differently available for all networks or network domains between the respective Application Clients and the corresponding Application Instances at any time and in any federated network.
5. An OP shall allow an Operator to expose network capabilities based on network functions exposure and resources sharing.
6. The OP architecture shall allow an Operator to expose, when available, network related Statistics and Analytics to application providers regarding different information types (descriptive, predictive and prescriptive ones).
7. The OP architecture shall allow an Operator to expose, when available, network related Statistics and Analytics to Partner OPs, regarding information types (descriptive, predictive and prescriptive ones).
8. An OP shall allow an Operator to manage network capabilities exposure depending on different factors (type, interface, application provider and operator).
9. The OP architecture shall allow an Operator to expose the end user's profile data to the Application Provider.

Note: This applies to the end users managed by the given Application Provider, and some examples are applicable 5QI, S-NSSAI, etc.

10. The OP shall support the exposure of network capabilities to the federated Partner OPs.

Note: For clarity, availability of a network capability in a visited network is only possible by federation between networks. Roaming on a non-federated operator's network does not allow exposing the visited network's capabilities.

11. Access to the network capabilities in the visited network shall be subject to authorisation by the Home OP and the Visited OP.

2.4.2 Capability Management requirements

2.4.2.1 General principles

An OP shall be able to access network capabilities exposed by an Operator network and use those capabilities either to optimise its OP specific services or expose those to the Application Provider.

An OP shall be able to access network capabilities exposed by Partner OPs that offer such capabilities (indicated through the catalogue), and use those capabilities either for optimising its own OP-specific services or expose those to the Application Provider.

An OP shall be able to expose information related with analytics when available (5G core) filtering what information is offered depending on the requester (Application Provider or Partner OP).

An OP shall allow the consumption of the network capabilities by the Application Client and the Application Instance as applicable.

An OP shall allow an application to request network capabilities

- for any data session established between any Application Client and Application Instance,

Note: This may be done as part of the application onboarding process, e.g. using the Application Manifest.

- for specific data sessions between a subset of Application Clients or a subset of Application Instances;

Note: This could be, for example, requesting a QoS profile for a specific set of premium users

- for a particular time period of data sessions between Application Client and Application Instance.

Note: This could be, for example, to increase the QoS profile for a video analytic application only while specific events occur.

An OP shall support the exposure of four categories of network capabilities:

- Control capabilities: these are network capabilities that allow controlling certain behaviours or characteristics of the network, e.g. applying a QoS profile to a session or Traffic Influence

- Event-based capabilities: these are typically notification-based network services allowing to inform the application upon specific network-related events, e.g. UE reachability
- Transactional information capabilities: these are capabilities allowing to obtain information from the network in a request-response pattern
- Analytics Capabilities: these are capabilities allowing to obtain analytical information from the network through notifications or in response to a request.

2.4.2.2 Control Capabilities

An OP shall be able to expose control capabilities to the Application Provider. Those capabilities may be exposed by matching a declared intent expressed by the application provider for using specific capabilities.

An OP shall expose the available service access with their QoS options to the Application Provider for the Application Provider to choose the wanted connection options and their priority. An Application Provider must be able to discover which QoS mechanisms are supported in the network where the UE is consuming its service.

Note: Any control capability includes a specific set of Service Level Indicators (SLIs) used to exchange information on particular levels of a network capability between the operator network, partner networks, the respective OPs and the application. An example of SLIs is the QoS profile as defined in Table 21.

2.4.2.3 Event-based capabilities

An OP shall be able to request to receive notifications about network events. The information elements obtained may be used for other use within the OP, e.g. for the orchestration of Application Instances, or be exposed to the Application Provider and to a Partner OP.

An OP shall provide a generic event notification mechanism for those events exposed to the Application Provider as well as those exposed to a Partner OP.

An OP shall be able to provide as part of the Network Event-Based Capabilities, “status type” information to the Application Provider regarding, e.g. UE mobility, UE communication, UE Statistics, Status Reports or QoS sustainability.

2.4.2.4 Transactional capabilities

An OP shall allow using transactional capabilities for purposes within the OP itself or exposure through appropriate APIs to the Application Provider.

2.4.2.5 Analytics capabilities

An OP shall be able to request network analytics information and to be notified about events.

Note: The information elements obtained may be used within the OP, e.g., for the orchestration of Application Instances, or be exposed to the Application Provider or a Partner OP.

An OP shall provide a generic event notification mechanism for those events exposed to the Application Provider as well as those exposed to a Partner OP.

Note: An OP can have access to the 3GPP Nnwadaf_AnalyticsInfo NWDAF service to obtain through the NEF a specific analytic Request.

2.4.3 Mobility Requirements

2.4.3.1 Bearer Change Requirements

The OP shall track the network bearers within the different network domains and ensure that requested capabilities are maintained. E.g. if a session between an application client and a UPF over a 5G is using network capabilities exposed to the OP via the Network Exposure Function (NEF) and this session is transferred to 4G, the OP shall manage the same capability over the Service Capability Exposure Function (SCEF), if available.

2.5 Network Communication Service Enabling Requirements

An OP needs to expose control mechanisms for network connectivity if an Application Provider wants to influence this for their service. This section addresses the high-level requirements on what the OP shall be able to expose to the Application Provider and other OPs for such services. These services address 4G and 5G network connectivity capabilities.

1. An OP architecture shall allow an Operator to expose Network Communication Services.
2. The OP architecture should be able to expose the existence of the Network Communication Service(s) to the Application Provider and other OPs.

Note: In 5G the Network Communication Services can be realised with a network slice, and in earlier network generations this could be realised using APNs and/or applying QoS.

3. The OP architecture should be able to access network slice lifecycle management and/or other needed capabilities of an Operator and include these in exposed Network Communication Services to the Application Provider.

Note: This capability is dependent on the agreement between the Operator and the Application Provider.

4. An OP should be able to expose KPIs to the Application Provider. Network performance attributes such as throughput, latency and reliability could be used to assure that the Application Provider’s requirements are met.
5. An OP shall allow an Application Provider to request authorisation for the end user to access the Network Communication Service.
6. The OP architecture shall allow the Application Provider to manage the end user’s profile data related to the Network Communication Service.

Note: This applies to the end users managed by the given Application Provider.

7. An OP should be able to request Network Communication Services in federated networks that support the requested SLA by the Application Provider.

3 Target Architecture

This section defines the architectural model for the OP. This model is introduced in section 3.1 with the OP functional levels and components being covered in section 3.2 and the interfaces in section 3.5. Section 3.3 defines how the model is enabling the OP's Federation capabilities. The data model that underpins the OP's functionality is covered in section 3.4 and sections 3.6, 3.7 and 3.8 define the architecture and support for offering different edge computing models.

3.1 Introduction

The primary goal of the OP's architecture is providing a global and common way of exposing an Operator's services to external Application Providers or Aggregators, whether through a direct connection from the resource owner towards the final consumer or by employing intermediate integration platforms (e.g. Marketplace).

The OP environment hosts multiple business 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. The business actors may be:

- Operator: The owner of an OP. The Operator may act as an Aggregator or even an Application Provider.
- Application Provider: The owner of an Application.
- Aggregator: The owner of a Marketplace. The Aggregator may act as an Application Provider.
- End-User: The user of an Application.

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

The target architecture is described at a relatively high level. Where OP-specific concepts are specified, they are defined as OP functional components 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 an OP must interact, deploying applications in containers and virtual machines is a well-established practice. The OP architecture 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.

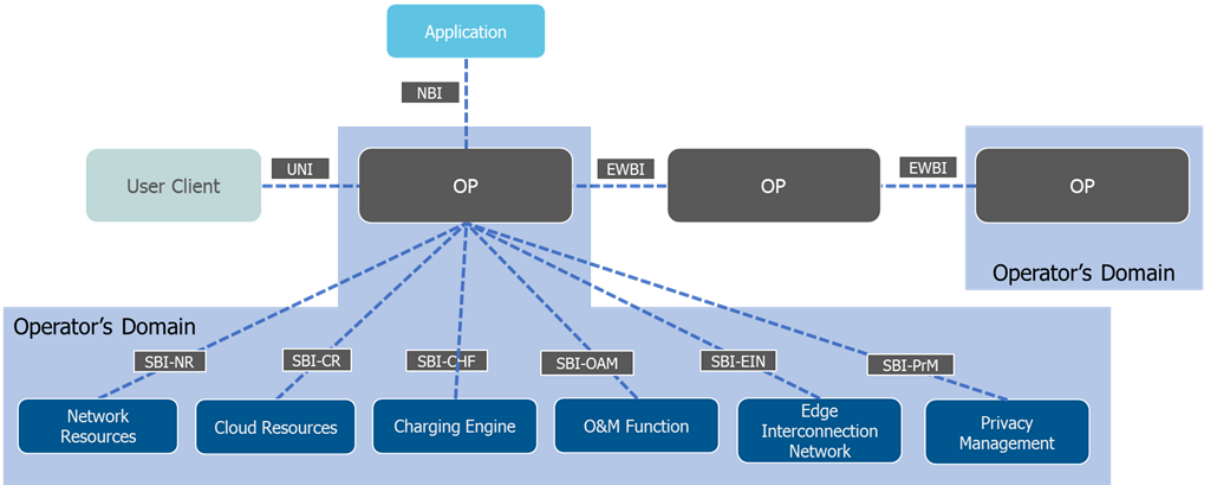


Figure 2: OP Interface Reference Architecture

The Operator's Domain contains all the endpoints for the South Bound Interfaces (SBI) needed for a given service.

GSMA PRD WA.101 [50] describes different models how the service can be exposed and offered to the Application Provider, either directly or via a Marketplace. Figure 3 depicts how an Application can consume capabilities exposed by the OP. The services and APIs exposed by the Aggregator, who owns the Marketplace, may differ from the services provided by the Operators.

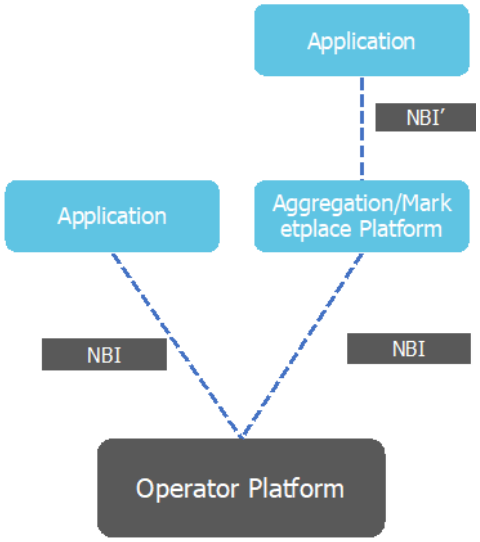


Figure 3: Different Deployment Options for Application

Note: NBI' is not in the scope of this document.

The following sections cover the OP functional levels, functional components and interfaces.

3.2 Functional Levels and Components

3.2.1 General

The OP is realised via multiple functional components (or functions). These functions enable an OP instance to interact with other end-points in the OP ecosystem, namely other OP instances, the Cloud Resources and the Network Resources and execute scenarios from/towards Application Providers or Aggregators. Figure 4 shows the high-level architecture and functions in an OP.

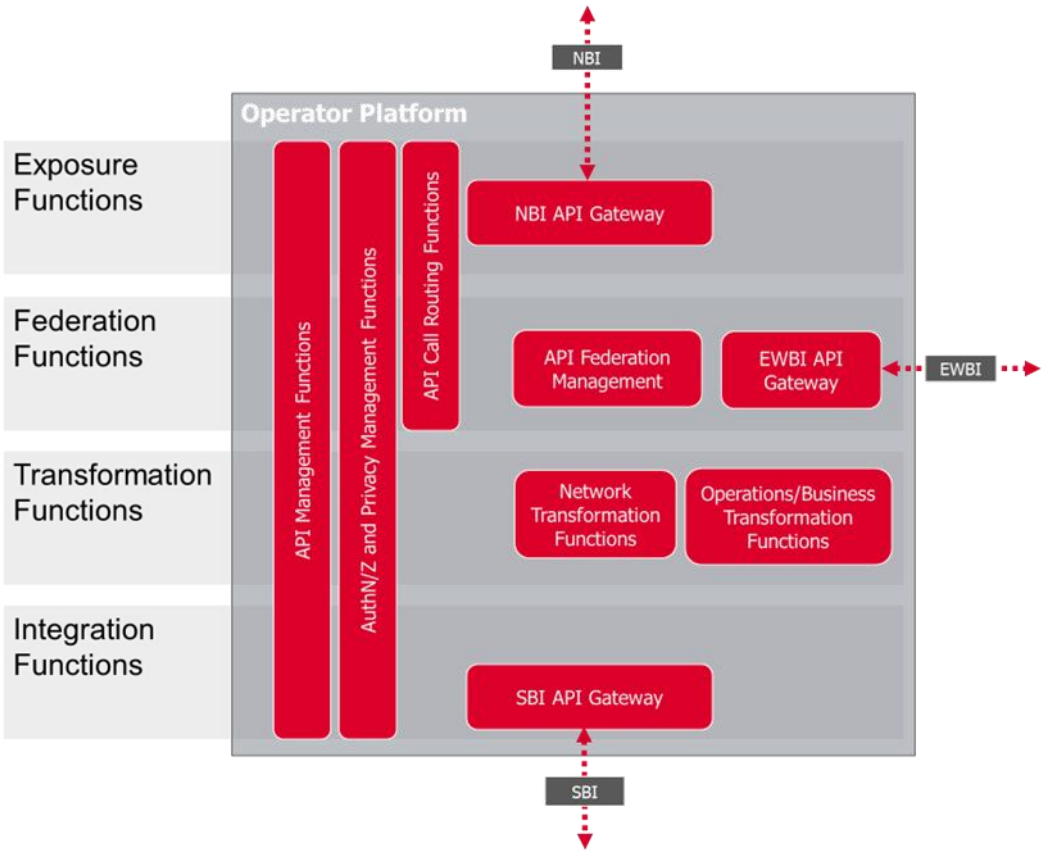


Figure 4:OP Functional levels and components

As shown in Figure 4, the functions can be grouped into four functional levels:

1. Exposure Functions,
2. Federation Functions,
3. Transformation Functions, and
4. Integration Functions.

It is worth mentioning that some “common” functions may span multiple functional levels (see e.g., API Management or API call Routing Functions in Figure 4).

The functional components in Figure 4 may be deployed in a distributed manner (as an architectural pattern that goes beyond monolithic realisations) enabling also flexible functional composition (for instance, if federation is not a scenario to be considered, the Federation-related functions do not need to be deployed).

This section elaborates on the functional levels and components depicted in Figure 4.

3.2.2 Common Functions

This category groups functions that may apply to more than one functional level.

3.2.2.1 API Management Functions

Providing (among others) the following functions:

- API Catalogue
- Application Provider management
- Application Onboarding
- API Subscription management
- API Usage management
- API Monitoring
- API SLA management
- API Provider management
- API Lifecycle management
- API Access Policy management

3.2.2.2 API Gateway Functions

API Gateway Functions are available in all of the interfaces in Figure 4, namely NBI, EWBI and SBI. They include (among others) the following functions:

- API Registry
- API Access Control / Security enforcement
    - Authentication (see below clause 3.2.2.3)
    - Authorisation (see below clause 3.2.2.3)
    - Plan control
- API Usage Data Generation
- API Logging and Tracing
- API Metrics Generation
- API Audit Logging
- API Traffic Management
    - Spike arrest
    - Usage throttling / Rate limiting
    - Traffic prioritization
- Interface translation
    - Format translation (e.g., from XML to JSON)
    - Protocol translation (e.g., from SOAP to REST)
- Caching
- Handling NB/SB/EWB interface connectivity aspects
    - For instance, providing Heartbeat/Keep-Alive mechanisms over the interfaces

3.2.2.3 Authentication, Authorization and Privacy Management

Providing (among others) the following:

- Authentication and Authorization (server side).
- Identity Management (if applicable)
- User Identity Token Management (if applicable)
- Privacy Management (if applicable)
    - key and certificate management
    - whenever Consent is the applicable legal basis:
        - Consent enforcement point (for NBI or EWBI)
        - Caching relevant Consent configuration retrieved from the Privacy Management function in the CSP domain (if allowed by local regulations)
        - Triggering Consent capture by the Privacy Management in the CSP domain
        - In federated scenarios, triggering Consent capture by the Privacy Management function in the CSP domain of the federated partner

Note: OP may relay procedures regarding Authentication / Authorization / Identity / Privacy management to servers already in place in the CSP domain via SBI-PrM interface.

3.2.2.4 API Call Routing Functions

The API call routing functions provides (among others) the following:

- Load balancing
- Telco Finder service which is responsible for resolving the operator associated with a target user identifier (e.g. based on a specific phone number) and returning information about the associated operator. More details regarding Telco Finder can be found in GSMA PRD OPG.10 [56].

3.2.3 Exposure Functions

The Exposure Functions enable exposing Service APIs (to Applications or Aggregation/Marketplace/Enterprise Platforms) via the NBI-Service interface, and Operate APIs (to Aggregation/Marketplace/Enterprise Platforms) via the NBI-Operate interface.

Note: The names NBI-Service and NBI-Operate may change in the future based on further discussions taking place across several groups.

The Exposure Functions in an OP enable an Application Provider or an Aggregator to operate applications. Operating an application includes discovering the capabilities of the OP, both in functionality (e.g., how an application may be onboarded or instantiated) and in range of functionality (e.g., where may an application be run, and what QoS attributes are possible). The Exposure Functions also enable the exposure of other Operator Capabilities that can be consumed, via APIs, by the Application Providers or the Aggregators.

The Application accesses the Exposure Functions via the North Bound Interface (NBI). The Application Provider or the Aggregator expects to use APIs implemented at the NBI to carry out required functions.

Capabilities are provided in part by actions that the Exposure Functions carry out on behalf of the Application Provider or the Aggregator and by data models for application manifests and resource catalogues. Data models may be used by multiple functional levels in an OP and extend across multiple federated OPs.

The data models available via the NBI are a subset of the data models used elsewhere in the OP. Still, they must be kept representationally consistent with the other data models, both in structure and in interpretation of individual data elements in a data model.

3.2.3.1 NBI API Gateway

In addition to the common API Gateway functions provided in above clause 3.2.2.2, the NBI API Gateway supports the following functions:

- Providing termination points for Service API calls from Applications (owned by Application Providers) or Aggregation/Marketplace/Enterprise Platforms (owned by an Aggregator or a 3rd party)
- Mapping to Transformation functions / SBI Gateway
- Routing to API Federation Management function / EWBI Gateway in case of API call Federation

Additionally, the NBI API Gateway supports:

- Providing termination points for Operate API calls from Aggregation/Marketplace/Enterprise Platforms
- Mapping to Operations and Business Transformation Functions / SBI Gateway

3.2.3.2 Exemplary scenarios

Based on the aforementioned capabilities, the NBI is expected to enable the following non-exhaustive list of scenarios:

- Edge Cloud Infrastructure Endpoint Exposure: the Application Provider uses an authenticated and authorised endpoint to carry out scenarios involving application instances on edge clouds;
- Application Onboarding: the Application Provider directly or via the Aggregator uses the NBI to provide application images and metadata to the OP Federation Functions (see clause 3.2.2.1);
- Application Metadata/Manifest Submission: the Application Provider directly or via the Aggregator uses the NBI and the metadata model to submit application metadata to the OP and follows defined procedures to extend the metadata model specification;
- Application CI/CD Management DevOps: the Application Provider directly or via the Aggregator integrates the CI/CD framework used to create an application with the OP via NBI APIs (which implies an integration between a CI/CD framework and Application Onboarding and Lifecycle Management) (see clause 3.2.2.1);
- Application Lifecycle Management: the Application Provider directly or via the Aggregator observes and changes the operational state of application instances, including the geographical/network extent of the OP on which application instances may run;

- Application Resource Consumption Monitoring: the Application Provider directly or via the Aggregator observes resource consumption of application instances, using the resource data model;
- Edge Cloud Resource Catalogue exposure: the Application Provider directly or via the Aggregator inventories edge cloud resources nominally available to application instances.
- Operator Capabilities exposure: the Application Provider or the Aggregator inventories the operator's capabilities, like Network Analytics, nominally available to applications as telco services via Service APIs (aka OpenGateway Services) (see clause 3.2.2.1).
- Privacy management for accessing Operator Capabilities: certain operator services consider personal information of the subscribers, the management of which shall remain under the Operator's control, based on functionalities exposed by the OP on the NBI (see clause 3.2.2.3).
- Operator Capabilities OAM: the Aggregator or the Application Provider, when integrated through an Aggregator, will require certain functionalities to configure and control the Operator Capabilities, nominally available to an Aggregator via Operate APIs.
- Telco routing information exposure: the Aggregator or the Application Provider use the NBI of different operators to identify which operator is responsible for handling an Operator Capability, as related to a specific operator's subscriber (see clause 3.2.2.4).
- Network Communication Service capabilities: the Application Provider directly or via the Aggregator observes and can manage the lifecycle of a Network Communication Service or the resources allocated to the Network Communication Service.
- Roaming capabilities: the Application Provider directly or via the Aggregator uses the NBI to consume information about service availability in the visited networks.

Note: The Aggregator consumes the OP services on behalf of the Application Provider. This interaction only refers to a technical service consumption, but the final service consumer remains in the Application Provider, e.g. for the privacy management.

Note: An Application Provider accessing OAM capabilities is considered as implementing an Aggregator's functionalities, therefore the Application Provider is indeed taking an Aggregator role in that case.

3.2.4 Federation Functions

3.2.4.1 API Federation Management

Providing (among others) the following services supporting the E/WBI:

whenever a Leading OP is originating an E/WBI request:

- For incoming NBI-service requests, routing to the Partner OP selected to fulfil the request
- For incoming API responses from other OPs, routing to the NBI API Gateway.

whenever Partner OP is originating the E/WBI request

- For incoming API requests from Partner OPs, routing to the relevant Transformation function to map to SBI API calls through SBI API Gateway. This includes:
    - Forwarded Partner OP NBI-service requests
    - Partner OP Service and Resource Management requests, for instance, for reservation / monitoring of Edge Cloud Resources

3.2.4.2 EWBI API Gateway

In addition to the common API Gateway functions provided in above clause 3.2.2.2, the EWBI API Gateway supports (among others) the following functions:

- Providing initiation / termination of E/WBI API calls to / from other OPs
- Routing to API Federation Management Function (when applicable)

3.2.4.3 Exemplary scenarios

Typical scenarios enabled by the Federation Functions across the E/WBI role are:

- Edge Cloud Resource Exposure and Monitoring towards partner OPs;
- Network and Analytics Capabilities Exposure 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.

Most of the above scenarios are supported by operator's OSS capabilities (Service and Resource domains) which would need exposure of corresponding APIs (e.g., TMF Open APIs for Service / Resource management) through the OP. Exposure would mean publishing these APIs in the OP Catalogue or creating an operational transformation function if one wants to expose simpler APIs and publish those APIs instead.

Depending on the deployment choice, an OP may perform a brokering function to simplify the interaction across multiple OPs, for which a subset of functions introduced in this section may be considered.

3.2.5 Transformation Functions

The OP Transformation Functions enable NBI and E/WBI API calls to be mapped to the required SBI API calls. One incoming API call may be mapped to one or more API calls on the SBI to provide a suitable response.

Transformation may concern mapping of the following request:

- NBI-service API requests towards the underlying network capabilities
- Service and Resource Management requests towards the corresponding Service and Resource Management domain(s). A Transformation Function may need information about available Services and Resources from those domains before performing any transformation logic.
    - Service and Resource Management requests may originate from the NBI-service, NBI-operate or EWBI.

Depending on the destination of the SBI calls, the Transformation Functions can be split into:

3.2.5.1 Network Transformation Functions

Providing the following services:

- Transformation Functions for the realisation of the Service APIs in the lower levels of the architecture (e.g., as in GSMA PRD OPG.09 [55]).

3.2.5.2 Operations and Business Transformation Functions

Providing the following services:

- Transformation Functions for the realisation of the TM Forum Operate APIs in the lower levels of the architecture (e.g., on the SBI-OAM interface).

3.2.5.3 Exemplary scenarios

Typical scenarios enabled by the Transformation Functions 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 through the OP,
    - Configuring UE traffic management policies through the OP 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: 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.
    - Interaction 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;
        - Request 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.
        - Receive the end user’s profile data (e.g. S-NSSAI, DNN, etc.)
    - Management of network slice lifecycle via SouthBound Interface – Operation and Maintenance (SBI-OAM)

- UNI:
    - Application Instantiation/Termination, e.g. based on triggers from the UNI;
    - Application Endpoint exposure towards User Clients (UC) via the UNI;
    - Application Placement decisions, e.g. based on measurements/triggers from the UNI.

3.2.6 Integration Functions

3.2.6.1 SBI API Gateway

In addition to the common API Gateway functions provided in above clause 3.2.2.2, the SBI API Gateway provides (among others) the following functions:

- Termination of the SBI towards:
    - Network Resources (SBI-NR)
    - Operations and Management systems (SBI-OAM)
    - Authentication, Authorization and Privacy Management functions in the CSP domain (SBI-PrM)
    - Cloud Resources (SBI-CR)
    - Edge Interconnection Network (SBI-EIN)
    - Charging (SBI-CHF)

3.3 Federation Management

The Federation Management functionality within an 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 and Network resources;
- Operators need to agree on an Edge Cloud and Network 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 Functions (see Figure 4) in an OP deal 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 Functions of the participating OPs need to verify each other's identities through mutual authentication (being the security enforcement point the EWBI API Gateways).

Federation Functions (see Figure 4) also ensure 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 and network resources in the OP's set of available resources using the SBI.

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.
- Network Analytics supported by the zone: catalogue of capabilities offered (abnormal behaviour, user data congestion, UE communication, UE mobility, service experience, network performance, QoS sustainability, load level information).
- 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 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 (using the API Federation Management Functions in Figure 4) and application onboarding, deployment and monitoring in partner OP Edge Clouds. Capabilities that overlap with NBI capabilities, such as for application onboarding, shall be maintained consistently with E/WBI capabilities. This is achieved via common OP functions spanning over the Exposure and Federation functional layers as shown in see Figure 4.

In scenarios in this category, an Application Provider interacts with an individual OP instance (the “Leading OP”) via the NBI. The Exposure Functions provide the Application Provider with a means of identifying and expressing geographic regions in which application instances should be run. The OP instance forwards the request and related information through interactions over the E/WBI as required.

The Application Provider request contains mandatory criteria (e.g. required CPU, memory, storage, bandwidth) defined in an application manifest. The Application Provider may optionally provide criteria such as QoS requirements (e.g. latency, prioritisation, reservation).

There may be multiple models possible for performing application orchestration via the E/WBI. The Exposure Functions should convey intent (from the Application Provider) and

result (from the OP) but should not require knowledge on the part of the Application Provider of the model or algorithms used.

For federated OPs (here, “Leading” and “Partner”), the Partner OP decides on which Edge Cloud(s) to deploy the applications and 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 provided to the Leading OP are transferred via the E/WBI to the Partner OP and used to deploy the application through the Partner OP.

The application provider's Availability Zone / Region criteria are considered, but, in the end, it is the Leading and Partner OPs that decide which edge cloud resources provide the better fit with the application requirements (QoS).

3.3.4 Service Availability on Visited Networks Management

When a 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 LBO/SBO for the data connections towards the edge cloud resources in visited networks.

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

If LBO/SBO 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 OP URL. e.g. http://register.op.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.

The Exposure Functions enable an Application Provider whose Leading OP is OP B to perform lifecycle management for their application instances without regard to whether the resources are controlled through OP B or OP A.

The Exposure Functions enable an Application Provider whose Leading OP is OP B to inventory resources available to their application instances, without regard to whether the resources are controlled by OP B or OP A, for resources controlled through OP A that are shared with OP B and to the Application Provider.

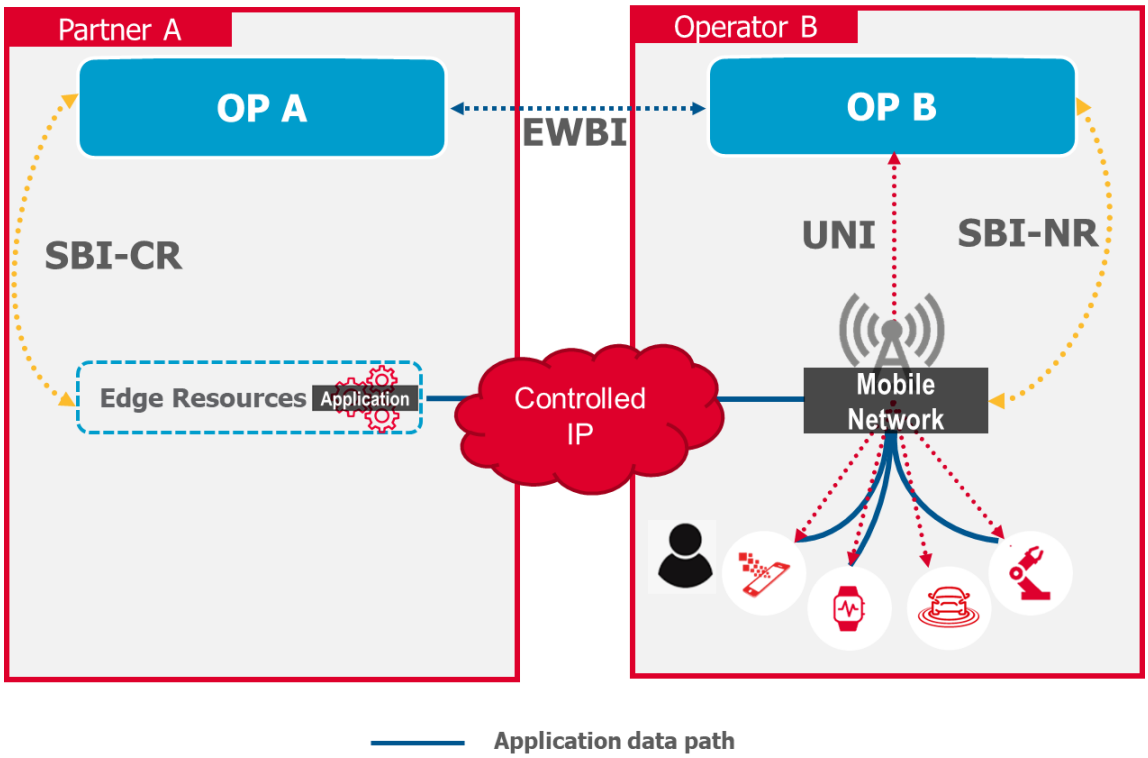


Figure 5: Edge Node Sharing

Figure 5 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.

Note: It is possible that in South Region, where Operator B has edge deployments, in addition the Operator B may also have edge node sharing agreements with multiple Partners (e.g. Partner X) who also have edge deployments in South Region. Hence, in South Region, Operator B can provide the edge services to the UCs in their own radio network either by their own edge deployments or via the partner edge deployments closest to the UCs.

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 through 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.

The information shared between OPs and the information visible to the Application Provider via its Leading OP NBI is subject to federation agreements between the Operators.

3.3.6 Configurations

An OP shall provide various configuration capabilities to establish and manage the Federation Interconnect (e.g., via the API Federation Management function in Figure 4).

API Federation Management Functions interact with the Exposure Functions and the NBI. Cases of this interaction are detailed as appropriate in the following subsections.

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, at a minimum, the following:

- The Partner Name;
- The Partner's geographical area (e.g. Country of operation, Regions, and Availability Zones);
- The Partner's description of shared resources ;
- The Partner identifiers;
- The Partner's federation interconnect E/WBI endpoint;
- The federation agreement validity duration.

Between any two Partner OPs, the provisioning information shall be mutually consistent.

The Exposure Functions of a Leading OP (of an Application Provider) are responsible for providing a representationally consistent view of Regions, Availability Zones, and Resources as might be required by the Application Provider to perform application lifecycle actions.

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.

Authentication and authorisation between partner OPs do not reach an Application Provider via the NBI. An Application Provider is expected to authenticate and authorise with its Leading OP, but not with any Partner OPs on which their applications run. The “chain of trust” required for an Application Provider to deploy an application on a Partner OP is composed of the authentication of the Application Provider on the Leading OP and the authentication of the Leading OP to the Partner OP.

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.

The information elements and data model used to represent Availability Zones in the Partner OP shall be representationally consistent with the NBI data model.

3.3.6.4 UE Provisioning of URSP rules

To correctly provision the UE with URSP rules, the serving network deployment and configuration need to be considered. The home network defines the URSP rules, while the visited OP needs to direct the visited network to connect the application to the correct DNN/NSSAI.

Operators need to exchange information to populate the URSP rules that map an application to the right DNN/NSSAI; please refer to 3GPP TS 23.503 [27] for more details. The OP may facilitate the exchange of relevant deployment and configuration information between the serving and home networks to influence the construction of the URSP rules to be provided to the UE. This can apply to all connectivity models.

3.3.7 Edge Cloud resource monitoring

An OP shall offer to Application Providers and Operators the capability to monitor resources (i.e., collect telemetry data) by the following categories:

- Usage: compute, memory, storage, bandwidth ingress and egress
- Events: Alerts raised by alarms/faults, log file entry search
- Performance Metrics: hardware and software counters
- Aggregate statistics: data sources from Usage and Metrics, aggregated and summarised via statistical methods (to reduce the network overhead of transmitting data). Data may be aggregated over time ranges or geographic ranges such as Availability Zones.

It is expected that Usage data is enabled by default.

It is expected that Event and Performance Metric data must be enabled by a consumer of those data sources. This enabling may be done via a Partner OP (of its corresponding Partner OP) or by an Application Provider for application instances that it owns or resources that the instances use.

The APIs by which resource monitoring data is managed, and the data models followed by the collected data, shall be representationally consistent between the E/WBI and the NBI.

The monitoring of any resources required for the functioning of the Operator’s charging engine shall be enabled.

Data collection shall be subject to the security requirements of section 3.4.1 and Annex E.

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 notification mechanism should be available to allow the above filtering for the information relevant for Fault and Performance management.

3.3.9 Automation Capabilities

An 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 service 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 5. 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.

Note: 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 willing 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 API Federation Management Function could be 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, an 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.

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

Note: Aspects like standardised 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.

3.3.11 Network Capability Exposure in a visited network

The exposure of network capabilities in a federated/visited network, such as applying QoS or obtaining certain network information, is crucial for the edge service to provide the desired quality of experience to the Application Client in the roaming scenario. Therefore, the goal is to provide the same network capabilities and Service Level Indicators (SLIs) in the visited network as in the home network. To achieve that, the Visited OP has to inform the Home OP about the network capabilities available, including the SLIs. This may be subject also to the specific federation agreement.

If the visited network cannot fulfil a requested network capability, the Home OP shall provide this information to the Application Provider.

The NEF's Network Capabilities access will not be exposed directly on the E/WBI. The catalogue of available network capabilities that can be exposed to the federated applications of the Leading OP shall be shared over the E/WBI.

3.3.12 Routing of Requests

An OP may have E/WBIs with multiple other OPs, as shown in Figure 2. Therefore, when an OP needs the support of another OP, it will need to make a routing decision determining what Partner OP's support would be required and what E/WBI would be best suited to reach that OP. Depending on the nature of the request, this decision could be based on combining information related to the request with information that the OP has on its Partner OPs. For example:

- For UE or Subscriber-related requests,
    - On the UE's IP address combined with information on the IP Address ranges that a Partner OP is covering,
    - On the MSISDN associated to a subscription combined with either information on the MSISDN ranges covered by the Partner OP or MSISDN-specific information from a Number Portability Database,
    - Domain information from a token or General Public Subscription Identifier (GPSI) related to the Subscriber combined with the domain(s) covered by the Partner OP,
    - As the Home OP of an UE or Subscriber roaming on another network, information on what network they are roaming on combined with information on what Visited OP would serve that network.
- For resource-related requests,

- On the Availability Zone that the request refers to combined with information on the Availability Zones supported by the Partner OP.

Note: An OP may have multiple routes towards a Partner OP, direct and indirect where the request might pass through one or more other Partner OPs. The criteria to select the most suitable route in that case are out of scope of this specification.

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 conceptual data model presented in this document offers a big-picture view of all entities, the OP system and their properties.

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

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.

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 functional components 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	Optionality
Authorisation type	Authorisation type supported by an OP	UNI, E/WBI, NBI	Mandatory
Certificate	The certificate of the Application Provider	UNI, E/WBI, NBI	Mandatory
Application Provider	Identifies the Application Provider to whom Certificate belongs. As defined in 3.4.18.	NBI, E/WBI	Mandatory

Data type	Description	Interface Applicability	Optionality
Access List	For information elements that an API may request between trust domains, the list of identities authorised to make a request	UNI, E/WBI, NBI	Mandatory
Operator Platform	Identifies the OP that exposes capabilities to the Application Provider and the Partner OP. As defined in Table 10.	E/WBI, NBI	Mandatory
Application Client	Identifies the Application Client as defined in Table 5.	UNI	Mandatory

Table 2: Common Data Model – Security

3.4.2 Edge Application Manifest

The data model of the Edge Application Manifest contains the information about the application to be instantiated, the Application Provider responsible for managing it, and the capabilities that the application may require .

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 ID	The ID of the Edge Application running on the edge node	E/WBI, NBI, SBI-CHF	Mandatory
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.	E/WBI, NBI	Mandatory
Edge Application version	The version of the Edge Application.	E/WBI, NBI	Mandatory
Edge Application Image	A URI (or similar name) of the VM or Container image to be installed and executed by the OP.	E/WBI, NBI	Mandatory
QoS Profile	The identifier of the QoS description for network traffic, as selected by the Application Provider. As defined in Table 21.	E/WBI, NBI	Optional
Alternate QoS References	A prioritised list of identifiers to the alternate QoS References for network traffic	E/WBI, NBI	Optional
Network Capability	A list of network capabilities requested by the application. As defined in Table 12.	E/WBI, NBI	Optional
Cloudlet Capabilities	A list of cloudlet capabilities requested by the application.	E/WBI, NBI	Optional
Deploy model	Indicates whether an application may be located freely by the OP or whether the Application Provider specifies the edge cloud on which it is to be deployed. The default value is "free".	E/WBI, NBI	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".	E/WBI, NBI	Optional
Edge Application migration policy	Indicates whether a backend application may be moved from its current operator network or current geographic Region (i.e., without violating the General Data Protection Regulation (GDPR)).	E/WBI, NBI	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.	E/WBI, NBI	Optional

Data type	Description	Interface Applicability	Optionality
Edge Application FQDN	Indicates if cloudlet-specific FQDNs needs to be assigned to backend application instances	NBI	Optional

Table 3: Common Data Model – Edge Application Manifest

3.4.3 Cloudlet

The Cloudlet is where the application is deployed. The Cloudlet data model (Table 4) provides the required parameters to deploy applications in the Cloudlet. Therefore, the Common Data Model of Cloudlet involves Cloudlet and the resources needed.

Data type	Description	Interface Applicability	Optionality
Cloudlet ID	The FQDN defining the Cloudlet of where the Edge Client shall connect.	SBI-CR	Mandatory
Availability Zone	Identifies the Availability Zone, as defined in Table 7.	E/WBI, NBI	Mandatory
Resources	Identifies the resource the application will need. As defined in Table 6.	E/WBI, NBI, SBI-CR	Mandatory

Table 4: Common Data Model – Cloudlet

3.4.4 Application Client

The Application 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 Application Client ID, Application Client IP address; Application Client Profile, Security and the UE/Non-SIM UE. There may be multiple Application Clients on a single UE, and a separate data module may exist for each.

Data type	Description	Interface Applicability	Optionality
Application Client ID	A unique identifier of the Application Client.	UNI, E/WBI, SBI-CHF	Mandatory
Application Client IP	The IP address of the Application Client	UNI, E/WBI	Mandatory
Operator Platform	The OP for the Application Client. As defined in Table 10.	UNI	Mandatory
Application Provider	The Application Provider(s) who manages the Application. As defined in Table 18.	UNI	Mandatory
Edge Application Profile	The application associated with the Application Client. As defined in Table 19.	UNI	Mandatory
Security	Security details that are supported by the Application Client. As defined in Table 2.	UNI	Mandatory

Data type	Description	Interface Applicability	Optionality
UE	The UE where the Application Client is installed. As defined in Table 8.	UNI	Optional
Non-SIM UE	The Non-SIM device where the Application Client is installed. As defined in Table 9.	UNI	Optional

Table 5: Common Data Model – Application 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	Optionality
Resource name	The name of the resource	E/WBI, NBI, SBI-CR	Mandatory
Resource type	The type of resource	E/WBI, NBI, SBI-CR, SBI-CHF	Mandatory
Capacity	The capacity of the resource	E/WBI, NBI, SBI-CR	Mandatory
State	The state of the resource (e.g., running, hibernated)	E/WBI, NBI, SBI-CR	Mandatory
Availability Zone	The associated availability zone, as defined in Table 7.	E/WBI, NBI, SBI-CHF	Mandatory

Table 6: Common Data Model – Resource

3.4.6 Availability Zone

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

Data type	Description	Interface Applicability	Optionality
Availability Zone Name	The name of the availability zone	E/WBI, NBI	Mandatory
Region identifier	Geographical identifier	E/WBI, NBI	Mandatory
Flavour	Flavours (e.g., CPU, memory, storage, in/out bandwidth) as defined in Table 20.	E/WBI, NBI, SBI-CR	Mandatory
Specialised compute offered	Particular compute resources (e.g. GPU, VPU, FPGA, NPU)	E/WBI, NBI, SBI-CR	Mandatory

Data type	Description	Interface Applicability	Optionality
QoS Profile	The identifier of the QoS description for network traffic, as selected by the Application Provider. As defined in Table 21.	E/WBI, NBI, SBI-NR	Mandatory

Table 7: Common Data Model – Availability Zone

3.4.7 UE

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	Optionality
UE ID	A unique identifier that can be used to identify a UE. 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 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, E/WBI, NBI, SBI-NR, SBI-CHF	Mandatory
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, E/WBI, NBI	Mandatory
OP	The ID of the UE's Home OP as defined in Table 10	UNI, E/WBI	Mandatory
Network Slice Profile ID	Network Slice Profile ID identifies the network slice and service that the end user can access. See Table 17 for more details.	E/WBI, NBI, SBI-NR	Optional

Table 8: Common Data Model – UE

3.4.7.1 Non-SIM UE

The Common Data Model of Non-SIM UE includes the Non-SIM UE ID, Non-SIM UE location. Non-SIM UEs are mostly non-mobile, or so considered as per the current scope of this document.

Data type	Description	Interface Applicability	Optionality
Non-SIM UE ID	A unique identifier that can be used to identify a Non-SIM UE by the OP. For a Non-SIM UE, this is a unique ID to identify the Non-SIM UE permanently. It is generated at the Non-SIM UE's first registration with the OP.	UNI, NBI, E/WBI, SBI-CHF	Mandatory
Non-SIM UE location	Non-SIM UE location indicates where the non-SIM UE connects to the network. The OP will perform the Non-SIM UE's location identification with the help of the Non-SIM UE's network information. The Non-SIM UE Location is expected to be tied to a relatively static element, such as a mapped Availability Zone, rather than a granular location identifier.	UNI, NBI, E/WBI	Mandatory
OP	The ID of the Home OP of the non-SIM UE as defined in Table 10. This will be preconfigured in non-SIM UE through SDK or UC.	UNI, E/WBI	Mandatory

Table 9: Common Data Model – Non-SIM UE

3.4.8 OP

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

Data type	Description	Interface Applicability	Optionality
OP ID	The ID of the Operator Platform. This ID shall be unique per OP domain	UNI, NBI, E/WBI, SBI-CHF	Mandatory

Table 10: Common Data Model – Operator Platform

3.4.9 NEF/SCEF

NEF/SCEF, as a 5G/4G network capability opening function, provides secure disclosure services and capabilities provided by 3GPP network interfaces.

Data type	Description	Interface Applicability	Optionality
NEF/SCEF ID	The FQDN of the NEF/SCEF against which the OP shall connect. The ID shall be unique per OP domain	SBI-NR	Mandatory
NEF/SCEF IP address	The IP address of the SCEF or NEF against which the operator platform shall connect	SBI-NR	Mandatory

Table 11: Common Data Model – NEF/SCEF

3.4.10 Network Capability

Network capabilities are accessed by an OP through the SBI-NR and consumed by the OP or exposed through the NBI (to the Application Provider) or E/WBI (to the Leading OP) as described in section 3.3.11. Network Capabilities are enumerated and described via their SLIs and SLOs to support the federation of network capabilities.

Note: The realisation of a specific network capability in a network is up to the individual operator; that is, the same capability may be achieved by different means (i.e., using other SBI-NR interfaces/parameters).

Data type	Description	Interface Applicability	Optionality
Capability ID	The ID of the enumerated network capability	SBI-NR, E/WBI, NBI	Mandatory
Network Capability Profile	The profile describes the Service Level Indicators (SLI) and Objectives (SLO).	SBI-NR, E/WBI, NBI	Optional

Table 12: Common Data Model – Network Capability

3.4.11 Void

3.4.12 Cloudlet Network and QoS Topology

Cloudlets, hosting compute resources for edge applications are interconnected with the mobile network and could provide different levels of QoS based on location and infrastructure capabilities. OP would need to manage the information described in Table 13 via the SBI-NR interface to provide the requested QoS level for the application in conjunction with a Cloudlet.

Data type	Description	Interface Applicability	Optionality
Edge Network Location Information	It may include location information referred within NEF/SCEF APIs, e.g. Cell IDs, Tracking Area Code (TAC), Registration Area (RA) etc.	SBI-NR	Mandatory
Edge Local Data Network IDs	Data Network Access Identifiers (DNAIs) representing networking and routing information associated with cloudlets	SBI-NR, SBI-EIN	Mandatory
QoS Profile	The identifier(s) of the QoS for network traffic, as defined in Table 21, that a mobile network can provide to a cloudlet	SBI-NR	Mandatory

Table 13: Common Data Model – Cloudlet Network and QoS Topology

3.4.13 Network Analytics

Network Analytics capabilities are accessed by an OP through the SBI-NR and consumed through the NBI (to the Application Provider) or E/WBI (to the Leading OP) as described in section 3.3.11 regarding Network Analytics. Those capabilities shall be enumerated and described via their SLIs to support federation and be classified by type. Each set of SLIs, SLOs, type and granularity shall have a unique ID.

Data type	Description	Interface Applicability	Optionality
Network Analytics ID	ID of enumerated analytics capability	SBI-NR, E/WBI, NBI	Mandatory
Network Analytics Profile	The profile describes the Service Level Indicators (SLI) and Objectives (SLO).	SBI-NR, E/WBI, NBI	Optional
Type	Type of Analytics Capability (e.g. event based, transactional)	SBI-NR, E/WBI, NBI	Mandatory
Granularity Scope	Defines the requested granularity.	SBI-NR, E/WBI, NBI	Optional

Table 14: Common Data Model – Network Analytics Capabilities

3.4.14 Void

3.4.15 NSaaS Lifecycle Status

The Common Data Model of NSaaS Lifecycle Status includes the network slice identifier and a lifecycle state of the network slice. An OP accesses the data through SBI-OAM interface or E/WBI from its partners and exposes the data through NBI to the Application Provide.

Data type	Description	Interface Applicability	Optionality
Network Slice ID	Identifier of a network slice	SBI-NR, SBI-OAM	Mandatory
Network Slice State	Indicates the network slice state	SBI-OAM	Mandatory
Requested Action	Indicates the requested action for network slice lifecycle change	SBI-OAM	Optional
Application Provider	Identifies the Application Provider(s) who is the network slice customer and can manage the network slice. As defined in Table 18.	SBI-OAM	Mandatory

Table 15: Common Data Model – NSaaS Lifecycle Status

3.4.16 Network Communication Service lifecycle management

The common data model of Network Communication Services includes the service identifier and a lifecycle state of the Network Communication Service. An OP exposes the data through NBI to partners and EWBI to federated OPs.

Data type	Description	Interface Applicability	Optionality
Network Communication Service ID	Identifier of a Network Communication Service	NBI, EWBI	Mandatory
State	Indicates the Network Communication Service state	NBI, EWBI	Mandatory
Requested Action	Indicates the requested action for Network Communication Service change	NBI, EWBI	Optional
Geographical region	Indicates the geographical region where the request is applicable	NBI, EWBI	Optional
Timeframe	Indicates the time when the service should be available. Note: this can be in the future	NBI, EWBI	Optional
Edge Application Profile	Identifies the edge application that is associated with the Network Communication Service. As defined in Table 19.	NBI, EWBI	Optional
Application Provider	Identifies the Application Provider(s) who is the Network Communication Service customer and can manage the service. As defined in Table 18.	NBI, EWBI	Mandatory
Target operators	Specifies the operators where the Network Communication Service needs to be provided.	NBI, EWBI	Optional

Table 16: Common Data Model – Network Communication Service Lifecycle Status

3.4.17 Network Slice Profile

A Network Slice Profile provides information related to the network slice.

Data type	Description	Interface Applicability	Optionality
S-NSSAI	S-NSSAI is used to uniquely identify a network slice that is subscribed for the end user.	E/WBI, NBI, SBI-NR	Mandatory
DNN	DNN that the end user uses to access the service	E/WBI, NBI, SBI-NR	Mandatory
QoS Profile ID	As defined in Table 21	E/WBI, NBI, SBI-NR	Optional

Table 17: Common Data Model – Network Slice Profile

3.4.18 Application Provider

The Common Data Model of the Application Provider.

Data type	Description	Interface Applicability	Optionality
Application Provider ID	The identifier of the Application Provider.	NBI, SBI-CHF, E/WBI	Mandatory
OP	The leading OP for the Application Provider. As defined in Table 10.	NBI	Mandatory
Edge Application Manifest	The application to be instantiated and managed by the Application Provider. As defined in Table 3.	NBI	Optional
Security	A set of security rules are supported by the Application Provider. As defined in Table 2.	NBI	Mandatory

Table 18: Common Data Model – Application Provider

3.4.19 Edge Application Profile

The following Table 19 is the model of the Edge Application Profile.

Data type	Description	Interface Applicability	Optionality
Edge Application ID	The ID of the Edge Application running on the edge node	E/WBI, NBI, SBI-CR	Mandatory
Edge Application IP address(es)	The IP address(es) of the Edge Application running on the edge node	E/WBI, NBI, SBI-CR	Mandatory
Edge Application FQDN(s)	The FQDN(s) that an OP needs to create for edge application instances on OP-managed availability zones for usage by application clients	NBI, E/WBI	Optional

Data type	Description	Interface Applicability	Optionality
Edge Application status	The status of the Edge Application running on the edge node	E/WBI, NBI, SBI-CR	Mandatory
Edge Application Traffic Flow Rules	The traffic flow rules describing application traffic characteristics (e.g., IP, Port, Protocol etc.) for filtering and routing of traffic to cloudlets	NBI, SBI-EIN, SBI-CR, SBI-NR	Mandatory

Table 19: Common Data Model – Edge Application Profile

3.4.20 Flavour

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	E/WBI, NBI, SBI-CR	Optional
Storage resource requirements	The storage resource requirements of the Edge Application	E/WBI, NBI, SBI-CR	Optional
Network resource requirements	The network resource requirements of the Edge Application	E/WBI, NBI, SBI-CR	Optional
Memory resource requirements	The memory requirements of the Edge Application.	E/WBI, NBI, SBI-CR	Optional
GPU resource requirements	The GPU requirements of the Edge Application.	E/WBI, NBI, SBI-CR	Optional
Virtualisation options	The deployment options.	E/WBI, NBI, SBI-CR	Optional

Table 20: Common Data Model – Flavour

3.4.21 QoS Profile

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	SBI-NR, E/WBI, NBI	Optional
Latency	The round trip delay between UE and Edge Application measured end-to-end with a "loopback" application	SBI-NR, E/WBI, NBI, SBI-CR, SBI-EIN	Optional
Jitter	The variance of round-trip delay between UE and Edge Application measured end-to-end with a ""loopback"" application	SBI-NR, E/WBI, NBI, SBI-CR, SBI-EIN	Optional

Table 21: Common Data Model – QoS profile

3.4.22 Operator

The key properties of the Operator are listed in the Common Data Model in the Table 22.

Data type	Description	Interface Applicability	Optionality
Operator ID	The identifier of the Operator. Usually it is a (sub)brand identity + ISO country code.	SBI-NR, NBI	Mandatory
OP	The OP that belongs to the Operator. As defined in Table 10.	E/WBI, UNI, NBI	Mandatory
NEF/SCEF	Identifies the NEF or SCEF, as defined in Table 11.	SBI-NR	Optional
Resources	Identifies the resources, as defined in Table 6.	SBI-NR, E/WBI, NBI, SBI-CR, SBI-CHF	Optional
Network Capabilities	All the Network Capabilities supported by the Operator, as defined in Table 12.	SBI-NR, E/WBI, NBI	Optional
Cloudlet Capabilities	All the Cloudlet Capabilities supported by the Operator.	SBI-CR, E/WBI, NBI	Optional
OSS/BSS	Identifies the OSS/BSS, as defined in Table 23.	SBI-OAM	Optional
Cloudlet	Identifies the Cloudlet, as defined in Table 4.	SBI-CR	Optional
Availability Zone	Zone(s) covered by the Operator, as defined in Table 7	NBI, E/WBI	Optional
CCS	Identifies the Operator's Charging Engine, as defined in Table 24	SBI-CHF	Optional
IP Range	Identifies the IP Address Range(s) used by the Operator	E/WBI	Optional
MSISDN Range	Identifies the MSISDN Range(s) served by the Operator Note: in regions where number portability applies this property might not be used.	E/WBI	Optional
Token Domain	Identifies the domain(s) used by the Operator in tokens or external GPSIs referring to end-users.	E/WBI	Optional

Table 22: Common Data Model – Operator

3.4.23 OSS/BSS

The key properties of OSS/BSS Common Data Model are listed in the table below.

Data type	Description	Interface Applicability	Optionality
OSS/BSS ID	The ID OSS/BSS that is unique per OP domain	SBI-OAM	Mandatory

Data type	Description	Interface Applicability	Optionality
OSS/BSS IP address	The IP address of the OSS or BSS.	SBI-OAM	Mandatory

Table 23: Common Data Model – OSS/BSS

3.4.24 CCS

The key properties of CCS Common Data Model are listed in the table below.

Data type	Description	Interface Applicability	Optionality
CCS ID	The CCS ID	SBI-CHF	Mandatory
CCS IP address	The IP address of the CCS.	SBI-CHF	Mandatory

Table 24: Common Data Model – CCS

3.4.25 Consent Record

The key properties of Consent Record Data Model are listed in the table below.

Data type	Description	Interface Applicability	Optionality
CM ID	Consent Management Function ID	SBI-PrM, NBI, EWBI	Mandatory
Consent ID	Identifier (on the Privacy Management Function) of the Consent entry	SBI-PrM, NBI, EWBI	Mandatory
Authorizing Party ID	Identity of the party granting the Consent for processing personal data	SBI-PrM, NBI, EWBI	Conditional (see Note 1)
Matching Criteria	Individual or list of Device ID(s), or PDU filter(s), or Subscription ID(s) for which the personal information processing is allowed	SBI-PrM, NBI, EWBI	Mandatory
Application Provider ID	Unique identifier of the Application Provider	SBI-PrM, NBI, EWBI	Mandatory
Application ID	Unique identifier for the Application requesting access to personal information	SBI-PrM, NBI, EWBI	Mandatory
Purpose of Data Processing	Predefined/standardized Purpose of Data Processing	SBI-PrM, NBI, EWBI	Mandatory
Legal basis	Predefined/standardized applicable legal basis	SBI-PrM, NBI, EWBI	Mandatory
Scope(s)	Reference to a set of resources being protected defined in an API specification	SBI-PrM, NBI, EWBI	Mandatory
Capture Method	Mechanism by which consent was obtained (Batch, Frontend based, SMS, API calls, e-mail, etc).	SBI-PrM, NBI, EWBI	Mandatory

Data type	Description	Interface Applicability	Optionality
Status	Granted, Denied, Revoked, Pending	SBI-PrM, NBI, EWBI	Mandatory
Consent Grant Timestamp	Timestamp at which the Consent was granted	SBI-PrM, NBI, EWBI	Mandatory
Revocation Method	Mechanism by which revocation was requested (Batch, Frontend based, SMS, API calls, e-mail, etc.)	SBI-PrM, NBI, EWBI	Mandatory
Revocation Timestamp	Timestamp at which the Consent was revoked	SBI-PrM, NBI, EWBI	Mandatory
Retention Period	Duration of time for which the personal data needs to be retained following receipt of revocation request	SBI-PrM, NBI, EWBI	Optional

Table 25: Common Data Model – Consent Record

Note 1: On federation scenarios it might not be needed or allowed to share information about the party who granted the Consent.

3.4.26 Privacy Management function (within the CSP domain)

The key properties of Consent Manager Data Model are listed in the table below.

Data type	Description	Interface Applicability	Optionality
PrM ID	The Application-related Consent Manager ID	SBI-PrM	Mandatory
PrM IP address	The IP address of the Privacy Management Function within the CSP domain	SBI-PrM	Mandatory

Table 26: Common Data Model – Privacy Management function (within the CSP domain)

3.5 Interfaces

3.5.1 Northbound Interface (NBI)

An Edge Cloud is similar to a traditional Cloud, but, in an Edge Cloud, the geographical location of Cloudlet resources provides additional capabilities and imposes additional constraints, compared to a traditional Cloud.

New capabilities provided by an Edge Cloud include satisfying more demanding QoS requirements, notably for latency. Use cases such as autonomous driving, which may not be feasible in a traditional Cloud, can be supported in an Edge Cloud. Application instances may be located in multiple Edge Clouds to accomplish this, whose locations are selected to satisfy QoS requirements. The application context, the information relating a UE with an application instance, may migrate from one application instance to another.

New constraints imposed by an Edge Cloud include more complex application migration. These constraints can arise because an Edge Cloud is deployed with a smaller physical footprint than a data centre-based cloud. A traditional cloud may respond to a change of

workload or traffic by scaling an application instance within a cluster. An Edge Cloud may need to locate application instances in different cloudlets.

The capabilities and constraints apply not only to the edge application but to the Application Provider. In the context of a CI/CD DevOps environment, the Application Provider and developer may be the same person/team, and the distinction between an application scaling itself, and an Application Provider scaling it, may be blurred.

For example, an application that takes advantage of low latency may do so to enhance an end-user experience (such as a game) or to support a mission-critical use-case (such as autonomous driving). An Application Provider may want to improve the application with latency statistics collection to tune its performance in a given environment or know when mission-critical QoS constraints are broken to take remedial action.

An application whose operation involves scaling to handle variation in traffic or migration to follow an end-user geographically may need to know the constraints of the Cloudlet in which the application is running.

Exploiting the capabilities and coping with the constraints both add burden to the Application Provider. One of the OP's goals is to reduce this burden as much as possible.

In an OP, this is accomplished by ensuring that the OP's Exposure and Federation Functions provide an appropriate subset of the capabilities exposed by the E/WBI. For federated OPs to work together, they must share significant amounts of information about cloud and network resources, availability zones, and configuration information. And they must create secure links to protect themselves and each other from attacks. To satisfy the scenarios listed in section 3.2.3, the Application Provider needs some of this information. However, the Application Provider may not to be an active participant in orchestration and migration decisions (apart from delivering the intent).

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 a workload profile. Application Providers may estimate workload profile parameters a priori or construct workload profiles from collected telemetry data. Application Providers provide workload profiles to Orchestration Services to automate the deployment of Application Instances. Application Providers may retrieve constructed workload profiles from the OP for offline use, such as in operational analytics.

The deployment of Edge Applications can be independent of network mobility or specific device attachment.

The NBI is the interface between the Application Providers and an OP.

1. To allow an Application Provider to “write once, deploy anywhere”, the NBI is a standard, universal interface. In other words, a developer does not need to rewrite their 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, an OP shall offer to Application Providers the edge cloud service in Availability Zones. An 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 an OP to advertise the (relatively) static information about the types of resource that it offers (“flavours”) but does not allow an OP to indicate the dynamic information about the current availability or usage of the resources.
9. The NBI allows an 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 the edge application should be provided in the 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 the service/application should be provided to home and visiting end-users in the network served by the OP.
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 OP-enabled resource management: orchestration and resource control:

    a) Orchestration View: Operators and Application Providers interact through the OP to create a running Edge Application. The Application Provider specifies application requirements, and the Operator uses them (with other information) to enable orchestration of an Edge Application.
    b) Resource Control View: The resource provider manages its Cloudlets in response to Orchestration requests which may include creating collections of resources as Flavours specified by the Application Provider.

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.

22. The NBI shall allow Application Providers to access network capabilities available in a specific Operator network.
23. The NBI shall allow requesting those network capabilities either as part of the Application Manifest or dynamically.
24. The NBI of the Leading OP shall expose the network capabilities of the federated Partner networks.
25. It is essential to provide the UE with the correct network details. The NBI shall allow the Application Provider indicate which UEs shall have access to the specific Network Communication Service.
26. The NBI shall allow an Application Provider to request that their applications requires cloudlet-specific FQDNs that can be resolved to the respective IP that application clients can access.

3.5.1.2 Resource Requirement Specification

1. An 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 application 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 Application 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 Leading 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 an 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 representationally 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.
   - **Alternative QoS References,** refers to the QoS descriptions that an Application Provider can optionally provide along with the specific QoS (refers to the QoS Identifier as defined in Application Manifest data model in Table 3) .

Note: when the specific QoS is not available for a PDU session, an OP may request the mobile network over the SBI-NR to provide the QoS for that PDU session from this alternative QoS set.

4. **Network Capability Requests (optional)**

   The Application Provider can specify a list of network capabilities consumed by the application; that is, capabilities exposed by the Operator for the data sessions between the Applications Client and the Application Instances. Each network capability request includes the following properties:

   - **ID**, a unique identifier of that specific capability to ensure using the same capability over different networks;
   - **Service Level Objectives**, the application requirements for the SLIs of that network capability;
   - **Request scope**, the definition for which of the data sessions this capability shall be requested; this may be a subset of all data sessions or provide a time/event-bound scope for the network capability request.

5. **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

6. **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).

7. **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.

8. **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).

9. **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

10. Network Analytics Requests (optional)

The Application Provider can specify a list of network analytics consumed by the application; that is, capabilities exposed by the Operator for the data sessions between the Applications Client and the Application Instances. Each network capability request includes the following properties:

- **ID**, a unique identifier of that specific capability to ensure using the same capability over different networks;
- **Service Level Objectives**, the application requirements for the SLIs of that network capability;
- **Type**, to request for a type of analytic capability, depending if it is based on a transactional or an event-based (notification) network analytics capability.
- **Granularity scope**, the definition of granularity of capability requested, depending on the type (e.g. event/notification based).

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, including the following abilities:

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

3.5.1.7 Network Capability Exposure Support

The NBI shall expose network capabilities towards Application Providers and Application Instances, allowing to use them alongside the provided edge service.

An Application Provider may use exposed network capabilities, that is

- Setting network behaviour for the data sessions between the UC and the Application Instance, e.g. QoS;
- Receiving notifications about network events or requesting specific information about UE, network status or information.

The request to use exposed network capabilities may be done statically through the application manifest or dynamically through a dedicated NBI.

3.5.1.7.1 Requesting network capabilities through the application manifest

The Application Provider can declare the requested capabilities as part of the Application Manifest. Depending on the selected value for the request scope, this declaration applies to all data sessions associated with that application or only to a subset.

3.5.1.7.2 Requesting network capabilities dynamically

The NBI shall allow the Application Provider to request network capabilities dynamically; that means any or any combination of the following:

- Enabling a specific capability at a specific time (e.g. ‘now’, ‘in 10minutes’) or a time-bound (e.g. ‘for the next 5 minutes’); this also includes the option to stop consuming a capability
- Enabling a specific capability for a particular connection between the Application Client and Application Instance, defined by the source or destination port and protocol, e.g. applying QoS only for the traffic between port range 10000-11000 on the Application Client and port range 12000-13000 on the application instance for UDP
- Enabling a specific capability for a specified subset of data sessions defined by a set of Application Clients
- Enabling a specific capability for a specified subset of Application Instances
- Optionally, requesting new Service Level Objectives

3.5.1.7.3 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 capabilities, 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;
- UE S-NSSAI.

An OP shall provide through the NBI a publish and notification framework for the Application Provider to request notifications about any network-related events. Events may occur due to

an explicit request to receive notifications from an event provider service (e.g. connectivity change events). They may also occur due to a request to use network capabilities affecting the network behaviour. For example, an Application Provider could request QoS and then receive an event notification that the desired QoS level cannot be maintained (e.g. due to a change of the connectivity bearer).

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 an 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 an 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.

An 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 an 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. An OP shall ensure that the Application Provider can deploy any application within the limits of their reserved resources in a particular availability zone.

1. An 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 an OP includes all interfaces the OP is consuming from other parts of the service provider's infrastructure to create the capabilities of the different functional components described in section 3.2. Therefore, the SBI may support access to:

- Cloud infrastructure, container and application resource management functions;
- Cloud resource orchestration functions facilitating the application lifecycle management and scheduling;
- Service management functions pertaining to the edge resource management (e.g. platform services, network services, mobility support, etc.);
- Other 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 defined here via the interfaces produced by the consumed functions.

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

The picture below illustrates two possible SBI-CR integrations between an OP and the cloud resources. Whether the integration is to a) cloud resource orchestration functions or to b) cloud infrastructure, container and application resource management functions is considered a deployment choice.

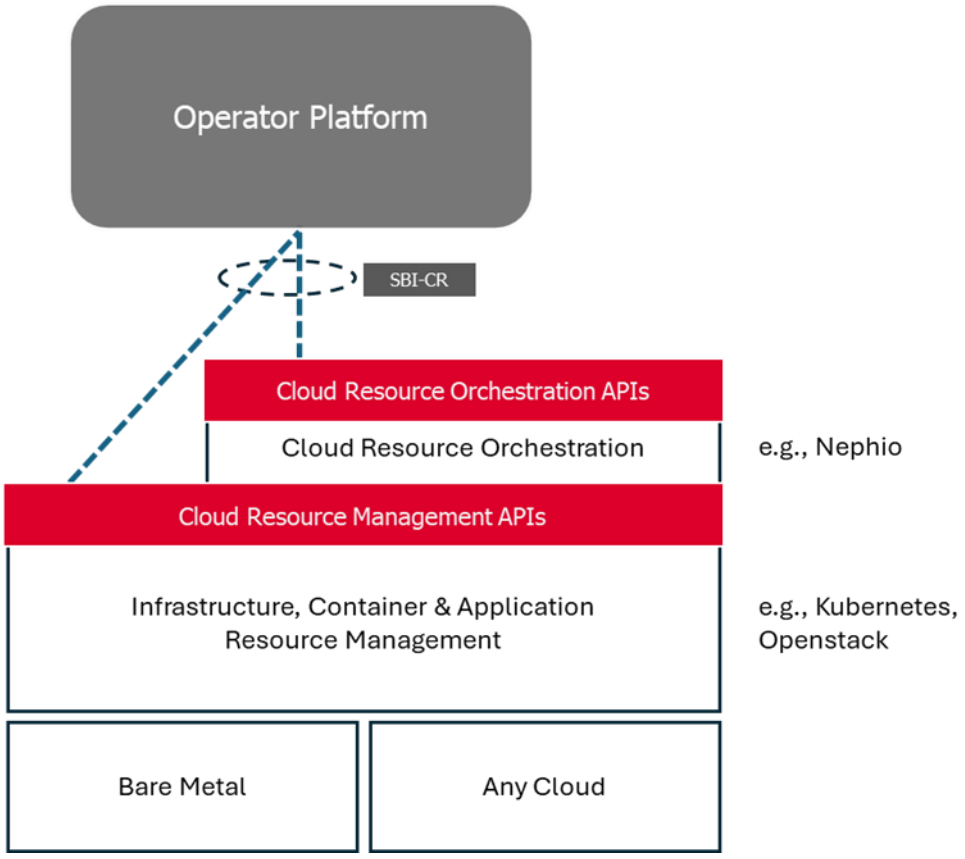


Figure 6: Possible SBI CR integrations

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

3.5.2.1.2 SBI Cloud infrastructure, container and application resource management functions

An OP may use cloud infrastructure management. An OP is expected to work over key industry reference cloud infrastructures. There are several options in the industry, e.g., OpenStack® or Kubernetes.

3.5.2.1.3 SBI Orchestrator functions

OP may use resource management via an orchestrator function, e.g. Nephio. In these cases, both resource management and workload management are consumed via the orchestrator function.

3.5.2.2 SBI-NR

3.5.2.2.1 Network

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

- UC location information retrieval;
- Request specific Quality of Service (QoS);

- Apply local routing and traffic steering rules for LBO/SBO for the Edge Application traffic;
- Application relocation on most adequate edge nodes;
- Influence Data plane attachment point (re)selection for service continuity;
- Collect radio network information, e.g. cell change notification, measurement reports etc. for mobility decisions;
- Support the profile data for the end user.

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 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, an 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, an 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. For enabling that, the Transformation functional level (and the respective Transformation Functions) depicted in Figure 4 are involved.

3.5.2.3 SBI-CHF

The operator that runs an 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. As shown in Figure 4, the OP

termination of the SBI-CHF interface is within the Integration functional level (e.g., in a SBI API gateway instance).

3.5.2.4 SBI-EIN

To execute operations where ECs or edge applications hosted on the ECs can communicate directly with each other, an OP shall enable EIN establishment between ECs. Example of such operations are:

1. Application relocation to a new EC.
2. Application context relocation to a new EC.
3. Application load sharing or failover handling.

The above example operations can be executed over the EIN by ECs and Edge Applications running on them. The OP will enable the ECs and Edge Applications to communicate over the EIN by providing the right information and applying appropriate rules over the SBI-EIN interface.

3.5.2.5 SBI-OAM

The APIs exposed on the SBI-OAM interface can help an OP to determine the status of a network slice in its life cycle. The details about the network slice lifecycle are in Annex H. In some cases, the OP needs to inform the Application Provider if a network slice status has changed or can request such change. An Operator can make a choice on which APIs to use for network slice lifecycle management, this is depending on where the Network Slice Management Function (NSMF) sits in the Operator's architecture, i.e. whether this function is located outside OP.

3.5.2.6 SBI-PrM

The SBI-PrM is the interface between the OP and the Privacy Management Function in the CSP domain. The Privacy Management Function plays a role when e.g., Consent is the applicable legal basis for an API call. The interactions and information flows on the SBI-PrM interface shall allow an OP:

1. To securely and confidentially retrieve, and update the Consent-related configuration (in the shape of Consent Records),
2. Trigger the capture of the Consent from the end user via Privacy Management Function within an authentication and authorization context,
3. Request to get notifications from the Privacy Management Function related to any changes in the Consent-related configuration,
4. Get notified about any changes in the Consent-related configuration applicable for a specific API call.

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 UC to interact with an 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 UC on the user equipment and the Operator Platform.

3. The UC should be implemented on User Equipment software, e.g. through an SDK or OS add-on.
4. The UNI shall allow the UC to discover the existence of an Edge Cloud service.
5. An OP's UNI shall allow the UC registration process with the Operator Platform, which entails the following:

    a) It enables the end-user device to establish an encrypted communication channel with the Operator Platform.
    b) Authentication and authorisation of UEs.

Note: 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) Authentication and authorisation of Non-SIM UEs.
    d) For the case of non-SIM UEs, the OP may not be aware of the Non-SIM UE's details and its authentication information when Non-SIM UE connects for the first time. The Non-SIM UE shall register with OP on the first connection and exchange identity and security information. Subsequent connections shall use recorded information from this first registration for authentication and authorisation.
    e) It enables the UC's usage tracking. For example, to support integration with the network operator's billing infrastructure.

6. An OP's UNI shall allow the UC to trigger the selection of a Cloudlet through the OP.
7. An OP's UNI shall allow the UC to trigger the instantiation of an application instance on the selected Cloudlet.
8. An OP shall measure network performance metrics for tracking the average latency characteristics of the edge network.
9. Based on metrics and location information, the UC may request through the UNI that the OP considers a change of Cloudlet.

Note: Some of these capabilities might be offered also through the NBI (see section 3.5.1) allowing to provide OP services to Application Clients on UEs that do not support the UNI.

3.5.3.2 Establishing Chain-of-Trust between architectural elements

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

1. the UE and the OP;
2. the UC 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 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 OP is located in the operator’s trust domain, which allows it to learn authenticated identity and location.

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

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 UC. These services may have a chain of trust established with the OP. If they have a chain of trust established with the OP, 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.

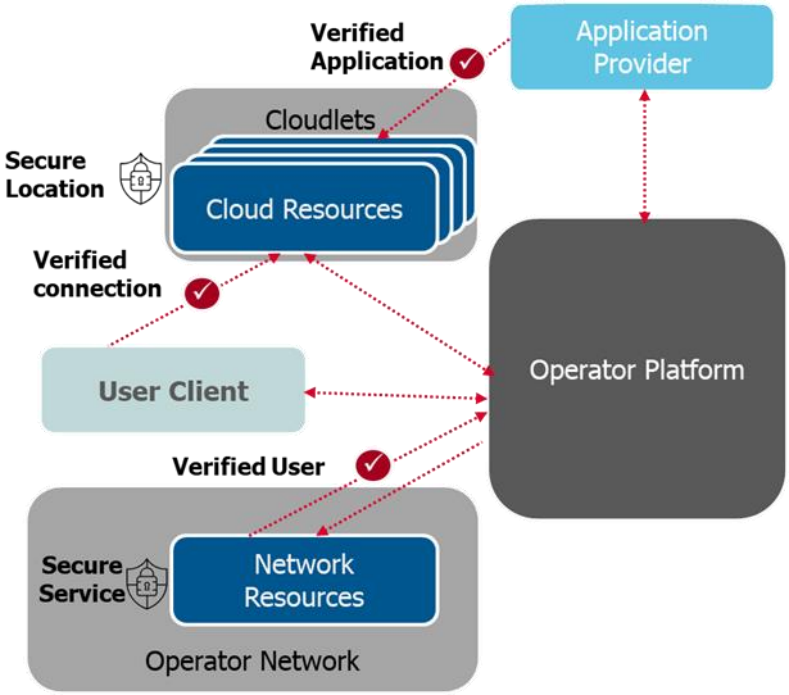


Figure 7: Chain of Trust: High-level Diagram

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 managed by the Federation Functions within the OP (see Figure 4).

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 support over E/WBI to request Availability Zone information update notifications;
- Request over E/WBI to stop receiving Availability Zone information update notifications;
- Update the request 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, prioritisation, reservation, cloudlet-specific FQDNs)
- 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 UC (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 OP 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 UEs 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 UCs.

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 an Operator Platform should be as easy as possible from an Application Provider’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 8).

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

Note: An option when the UNI is not available/supported is FFS.

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 an 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.

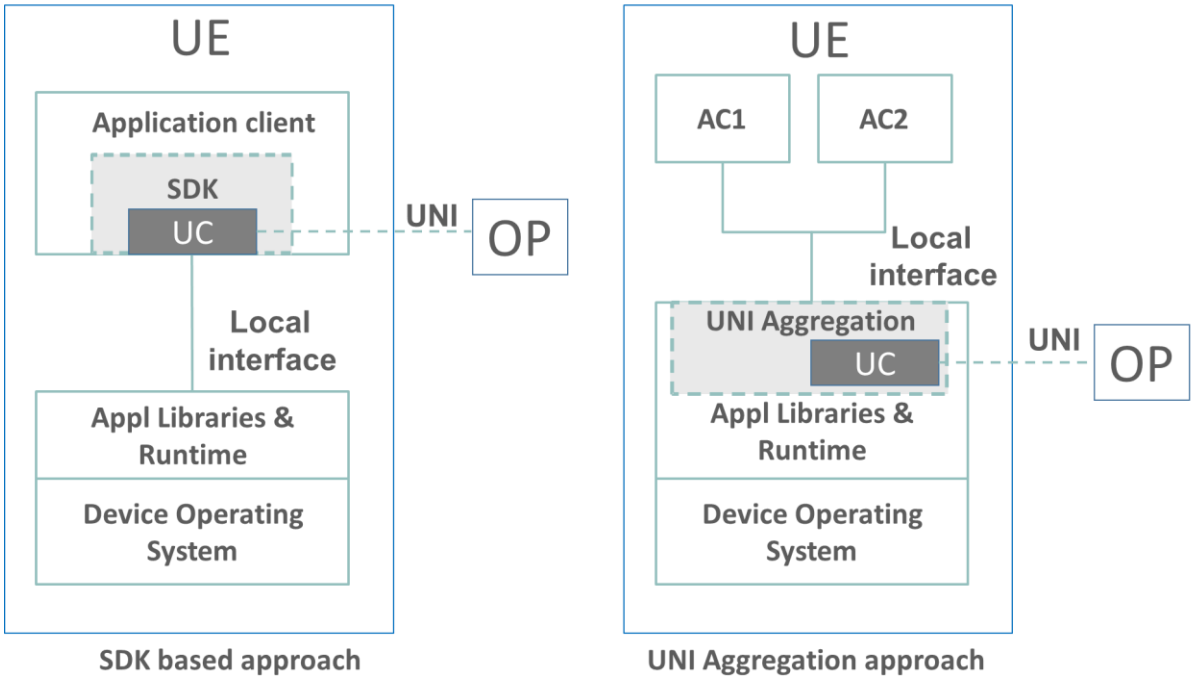


Figure 8: OP Device side architecture (local interface)

Note: As per two of the possible approaches for building UNI support for Application Clients, i.e. SDK and the UNI aggregation, Figure 8 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 (UE or non-SIM UEs), as described above, require access to specific privacy-sensitive parameters available from the device platform or the OP. 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.

3.5.5.1.1 UNI parameters for UEs

The following list provides an indicative, non-exhaustive overview of such parameters for a UE:

- 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.

3.5.5.1.2 UNI Parameters for non-SIM UEs

The following list provides an indicative, non-exhaustive overview of such parameters for non-SIM UEs:

- Non-SIM UE identity and credentials for authentication, e.g.
    - UUID (RFC 4122 [25] based) or equivalent.
    - Token for authentication
- Geo-Location information
    - City/State (If available)
    - Public IP address of the non-SIM UE's network
- Network Information
    - Wi-Fi SSID, Public IP and MAC address
    - Internet service provider information (If available through network information).

Note: Non-SIM UE may not support all the parameters; some of the parameters will be generated at first registration and shared with non-SIM UE by OP. The parameters supported are up to detailed UNI definition, the OP and the non-SIM UE.

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 and other 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

Note: Applications may not provide QoS support on Non-SIM UEs due to the device type or network limitations. Application Providers shall take note of this and accommodate it in their design and expectations accordingly.

3.6 Containers

3.6.1 Description

The OP architecture intends to provide Application Providers 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

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

3.6.3 Container runtimes

An 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, an OP shall support applications relying on VMs. The OP architecture intends to provide Application Providers 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 Application Providers and runs code on-demand automatically scaled and billed only for the time the code is running [12].

The OP architecture intends to provide Application Providers 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 Computing

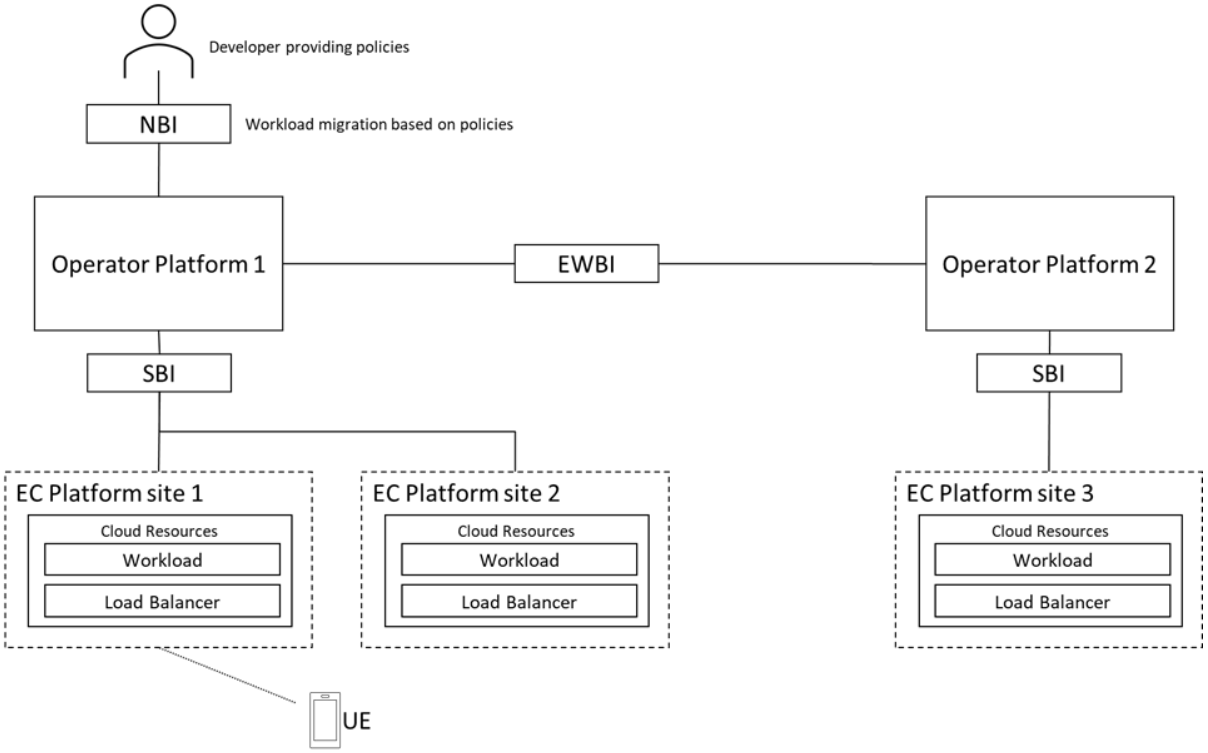


Figure 9: Serverless Computing

The following are the main enablers of a Serverless computing solution:

- Policies

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

- Orchestrator

  Scaling in/out of container applications from zero based on Application Provider 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. 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 (ECP)

  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 Application Provider 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 Serverless Computing Lifecycle

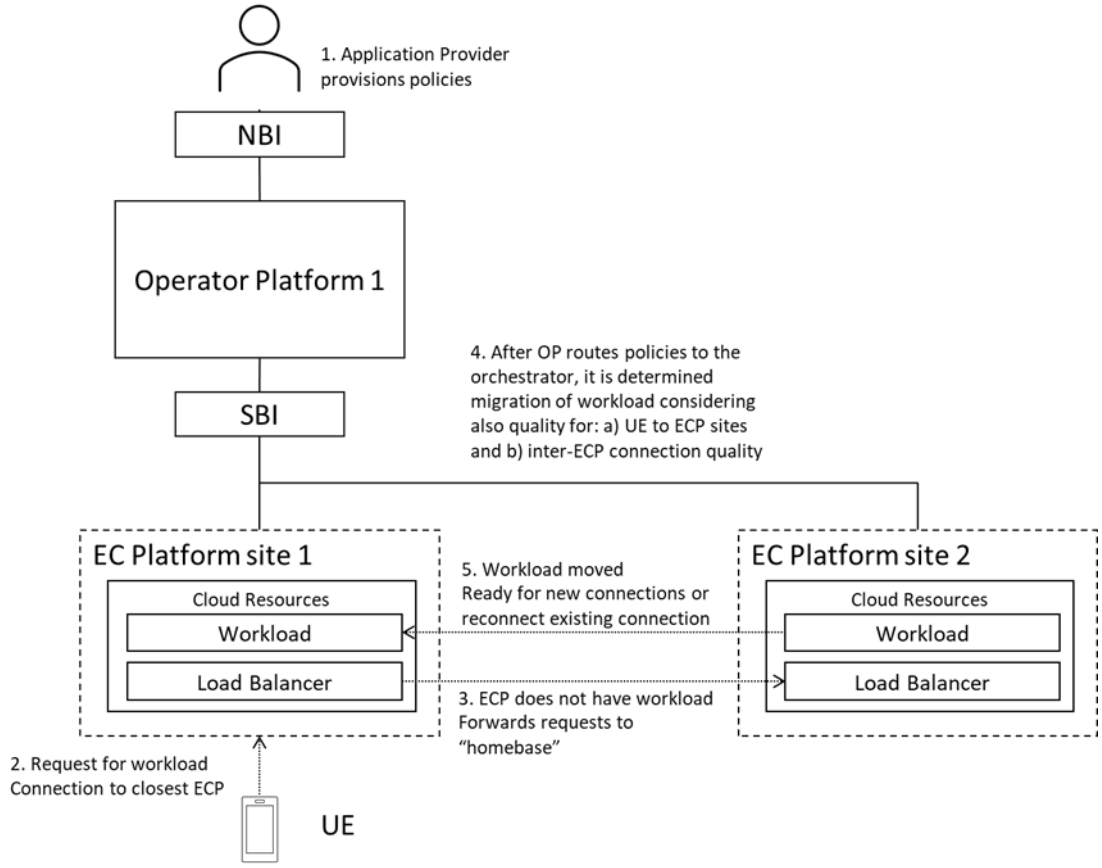


Figure 10: example sequence of a serverless lifecycle

An example sequence of a serverless lifecycle:

Note: The sequence below may change depending on implementation choices.

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 Application Provider 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

An 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

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

The following Application Provider 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

This section describes how an Operator Platform could interact with network elements, UEs and other parties to realise various service use cases that it enables and supports.

4.1 UC/UE Registration to the OP using UNI

4.1.1 UC Registration to the OP - Home Operator Platform

This procedure describes the registration between the UC and an Operator Platform, allowing the UC to be authenticated and authorised to access the service.

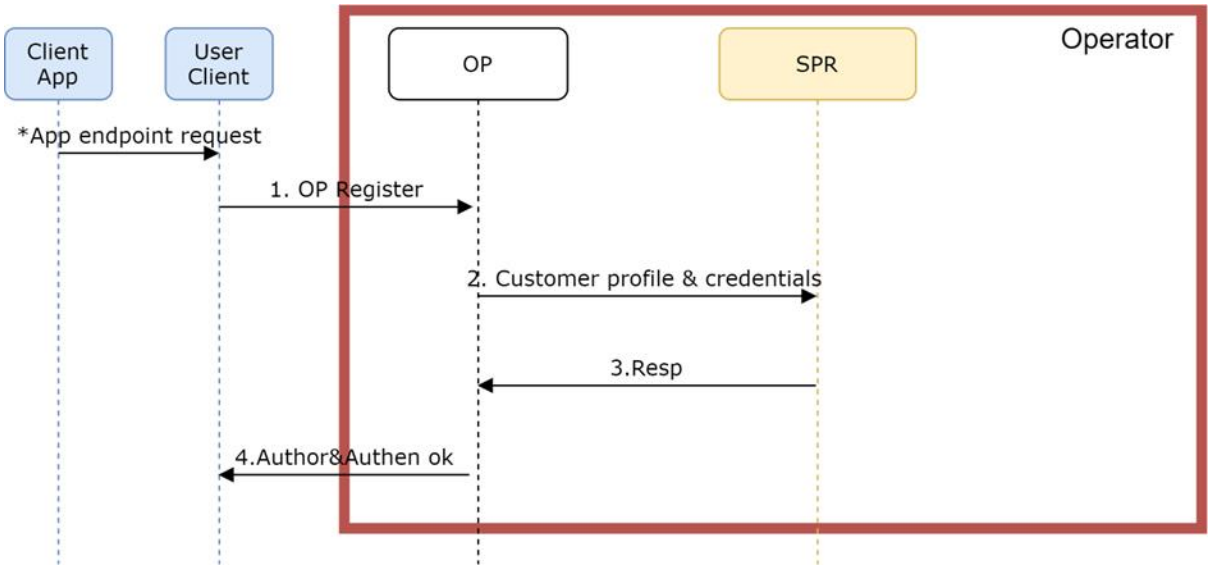


Figure 11: UC Registration to the OP - Home Operator platform

1. A 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 UC receives the authentication validation and is authorised to request OP services from that moment onwards (e.g. cloudlet discoveries).

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

4.1.2 UC Registration to the OP - Visited Operator Platform

This procedure describes the UC 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 4.1.1. Figure 12 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 UC then.

In this case, the Home OP manages 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 13 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.

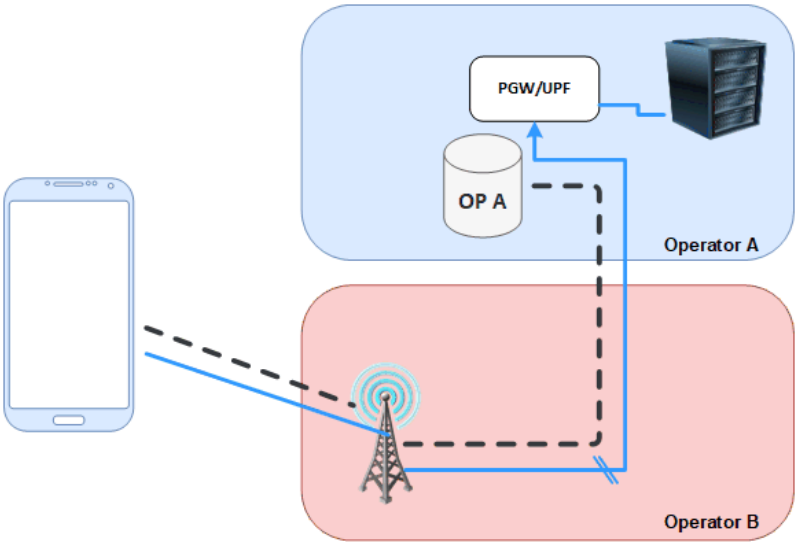


Figure 12: Roaming access to OP and edge resource - home routing

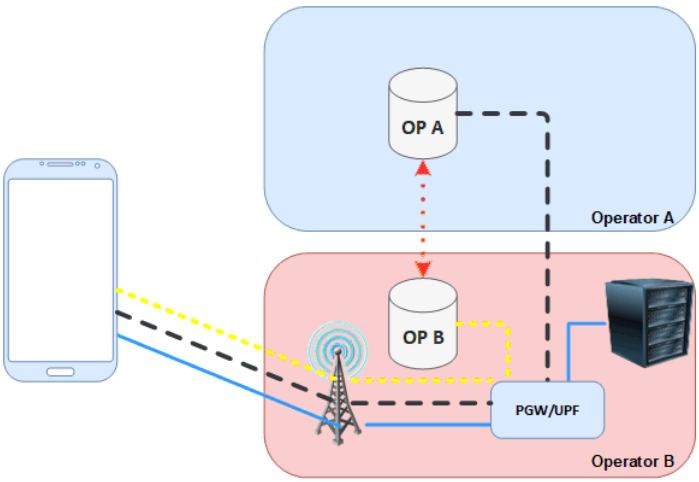


Figure 13: Roaming access to OP and edge resource – local breakout

4.2 Service delivery by the OP without UNI

4.2.1 Service delivery to UE attached to the Home Network

In case the UNI is not used, in most cases there will be no UE registration directly with the OP. The UE will register with its Application Backend and that Application Backend will be authenticated and authorised by the OP to use the services.

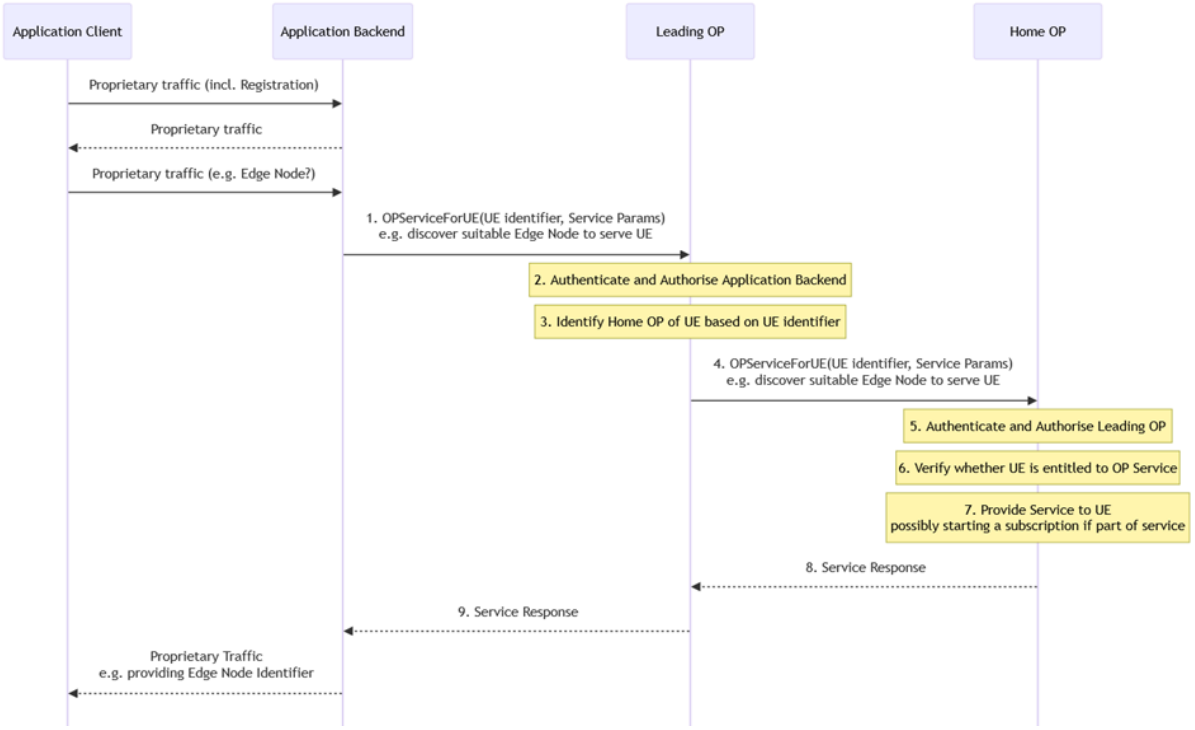


Figure 14: UE Registration to the OP without UNI

In this case, after some proprietary application-specific exchanges, the Application Backend may contact the Leading OP for the Application to obtain an OP-provided service related to the UE that would support the service that the Application is providing. The steps would then be as follows:

1. The Application Backend contacts the Leading OP, identifying itself and the UE to which the service should be delivered (e.g. providing its IP address or MSISDN).
2. The Leading OP authenticates and authorises the Application Backend
3. The Leading OP determines the target network to serve the request. In this non-Roaming case that would be the Home OP of the UE.

Note: The Leading OP is not necessarily the same as the Home or Visited OP for the UE. Therefore, the flow separates these.

4. The Leading OP provides the Application Backend's request to the UE's Home OP over the E/WBI between those OPs identifying itself and the target UE.
5. The Home OP authenticates the Leading OP.
6. The Home OP verifies whether the subscription of the UE that is referred to is entitled to use the requested service and whether the Application is authorised to request it for that Subscriber.

Editor's Note: Privacy Management is for further study.

7. The Home OP delivers the requested service. Depending on the request, this may involve starting an event subscription to inform the Application of any events related to the service delivery to the UE.
8. The Home OP provides a response on the service delivery to the Leading OP (e.g. indicating a suitable Edge Node.that was discovered)
9. The Leading OP provides a response on the service delivery to the Application (e.g. indicating a suitable Edge Node.that was discovered).

The Application Backend can then use the information in the response received from the Leading OP to enhance its service based on the agreements that the Application Provider has with the Leading OP for using the OP services (e.g. indicating to the UE in an application-specific exchange how to reach the most suitable Edge Node).

4.2.2 Service delivery to UE attached to a Visited Network

This procedure describes the UC (UC-A) registration with an Operator Platform (Operator A) without UNI while accessing the service from a visited network (Operator B). In this case the flow for requesting the service will be similar to the scenario described in section 4.2.1.

Depending on the service requested and capabilities available, the following roaming approaches are considered:

- **Home routing (HR)** – for services that depend entirely on the home network (e.g. subscription-related requests) and scenarios where service delivery provided by the visited network cannot be supported (e.g. edge resources in the visited network cannot be accessed).

  The Home OP is the only OP involved in this case, with the service request handled as defined in section 4.2.1. Figure 15 below shows the relations between the networks for the access to edge resources in this case. This scenario comes with limitations on application availability due to increased latency (see section 4.5).

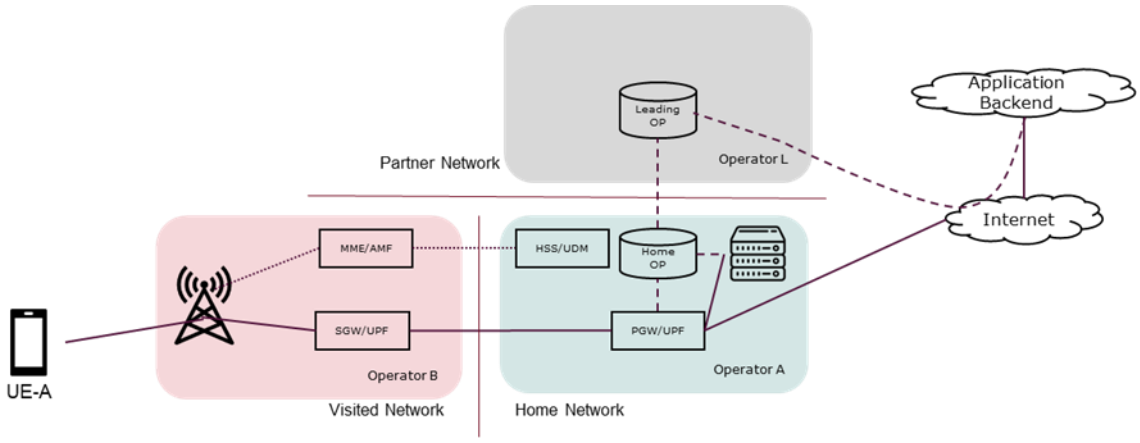


Figure 15: Roaming access to OP and edge resource - HR

- **Local breakout (LBO)** – for services that depend (partly) on service delivery by the visited network. After receiving the service request as defined in section 4.2.1, the Home OP will identify the visited network on which their subscriber is roaming and

determine what Partner OP to contact to support the service delivery. The Home OP will then request the service(s) to be delivered by the Partner OP through its EWBI with that Partner OP identifying its Subscriber as a roaming user.

Note: What service(s) should be delivered by the Partner OP depends on the service that the Home OP was requested to provide and would thus be service-specific.

Note: How to identify the Subscriber to the Partner OP is for further study.

In some cases, the service to be delivered might involve the UE accessing resources (e.g. edge nodes) available in the visited network directly. This model is preferred because the service depending on these resources is provided closer to the UE then. It is expected that the UE will have one data session routed to the home network (Operator A). In this case, the Home OP will receive the identifiers for UE access to those resources over the E/WBI as part of a discovery (e.g. Edge Discovery) or other service invoked on the Visited OP through the E/WBI and pass them on to the Leading OP and the Application Backend over the E/WBI and/or the NBI. Figure 16 below shows the relations between the networks in this case with the following clarifications:

- The solid lines: UE data session (PDU)/PDN) that carries the OP and application identifiers. It is also used for any data traffic and UC access to the resources on Operator B (through the UPF/PGW in Operator B network).
- The dotted lines: Signalling traffic between the visited network (Operator B) and home network (Operator A) as defined by 3GPP TS 23.501 [10]. It is used for the UE registration into the network
- The dashed line: NBI, SBI and EWBI connection for enabling the access to Operator B's resources.

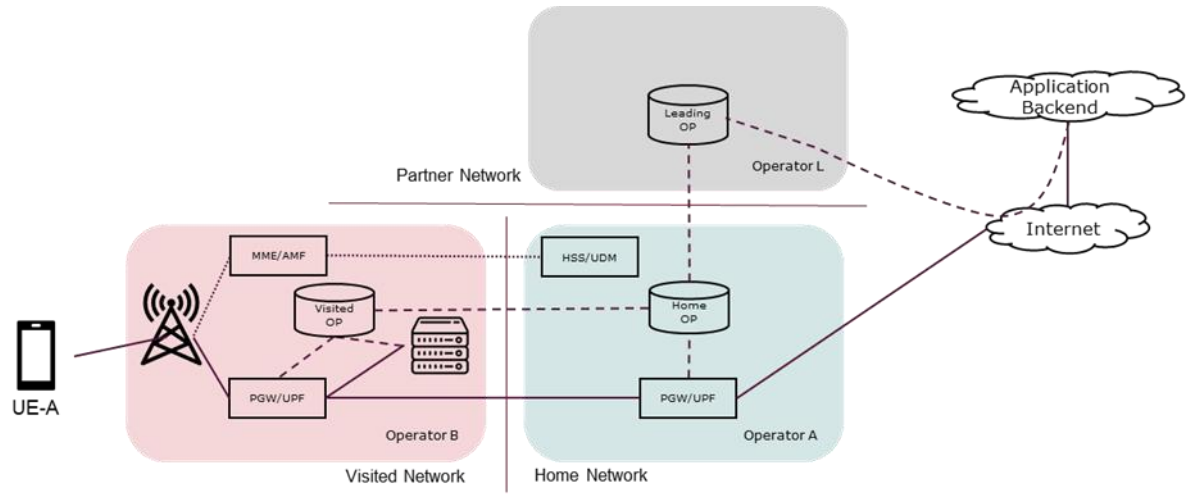


Figure 16: Roaming access to OP and edge resource – HR+LBO

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.

Note: Edge discovery for the case without UNI would use the flow provided in section 4.2.1.

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.

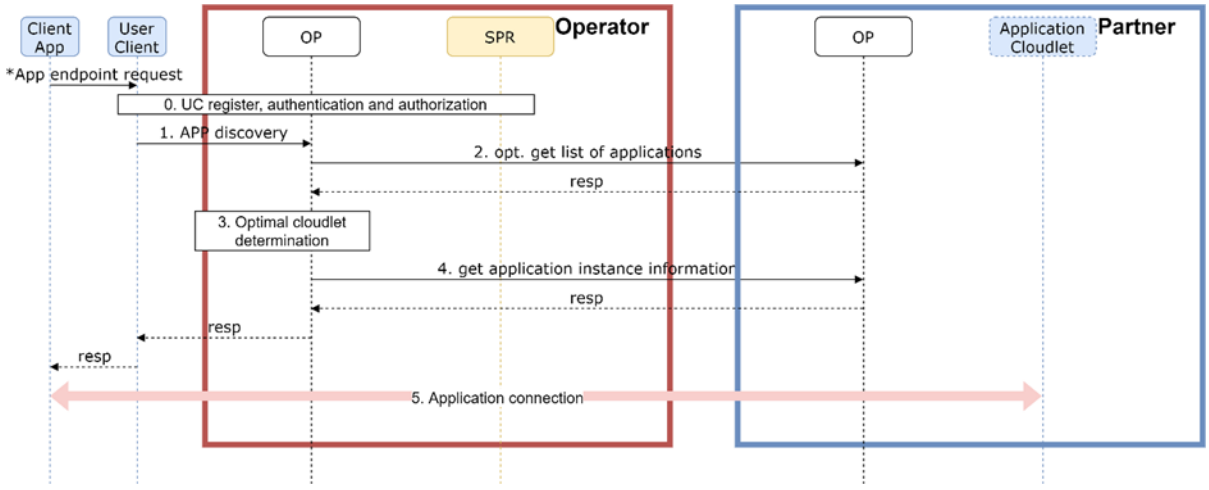


Figure 17: Edge discovery in an edge-sharing partner network

1. A 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 UC'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 (Operator B), and the most suitable cloudlet is in the Visited Partner OP. The two cases for the registration in the visited network (see sections 4.1.2 and 4.2.2) 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.

Note: Edge discovery for the case without UNI would use the flow provided in section 4.2.2.

4.6 Application deployment In the Home Operator Domain

This procedure describes the application deployment in a cloudlet of the operator domain, the edge discovery by UCs and an optional interaction of an OP with the 5G core network over the SBI-NR interface.

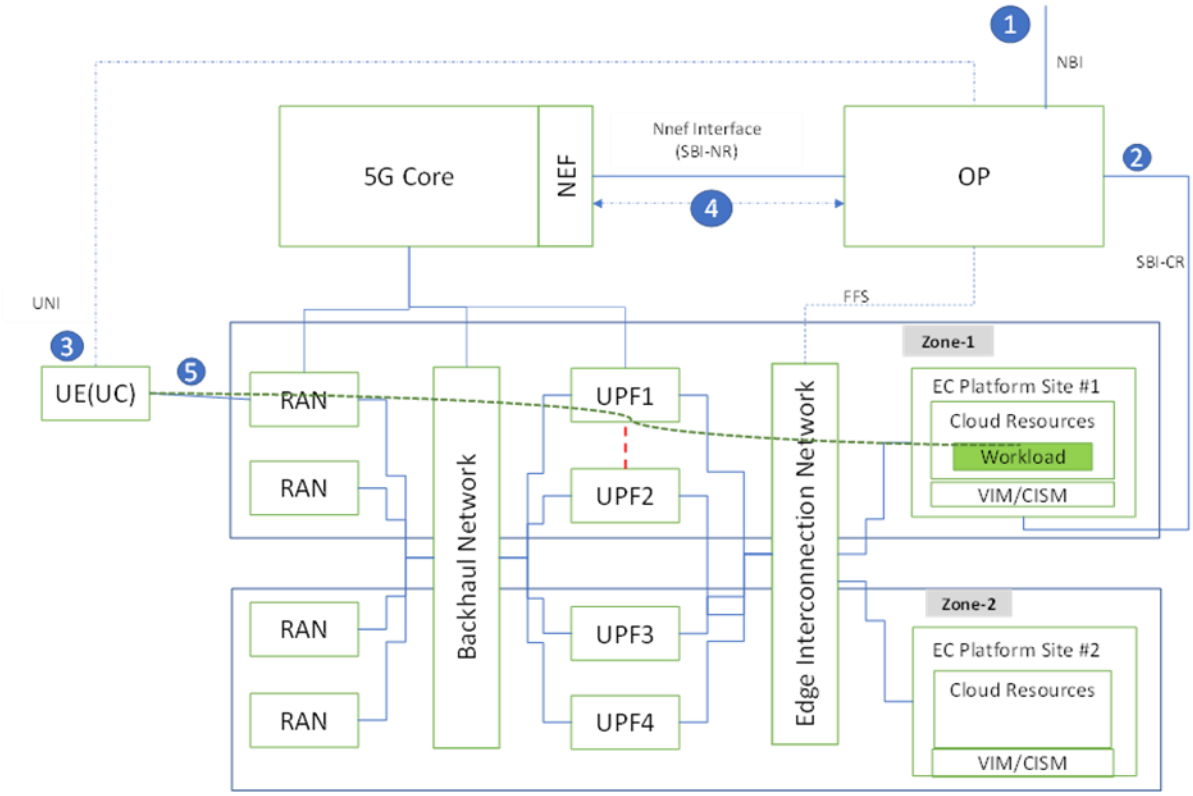


Figure 18: OP Application Deployment flow with 5G core network interaction

Note: Network layout is only for illustration purposes showing the role of various entities to support application Session Continuity in relation to an OP.

Following are the details of the various steps marked in numbers in Figure 18, highlighting the role and objectives of the different interfaces:

1. The Application Provider provides the Application Manifest with criteria indicating the application's sensitivity for Session Continuity and QoS profile, etc., and specifies the zones where an application can be deployed, e.g., Zone-1 and Zone-2.
2. The OP uses the information from Application Provider and orchestrates an Application Instance of the application on “EC Platform Site #1” in Zone-1, providing sufficient resources as required by the Application Provider.

3. When a mobile subscriber attached to the network launches the Application Client, and the UE invokes the edge discovery over the UNI, the OP returns the application communication end points in Zone-1 for the indicated Application Instance.
4. The OP may request the 5G core network via the SBI-NR interface to receive notifications related to mobility events for this UE and may request the QoS level required for the application session as per the information mentioned in the Application Provider's criteria. The OP also provides the application traffic steering rules using the SBI-NR for the mobile network to route the edge traffic to the “EC Platform Site #1”.
5. The User client (UC) connects through UPF1 with the instance on EC Platform Site #1 using the communication endpoints returned in step 3.

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.

4.8 Application Service and Session Continuity in the home network

Figure 19 provides a logical view of the various network interfaces, entities, and sequence of events on the different interfaces required for a coordinated effort to enable Session Continuity for edge applications in conjunction with the mobile network. The figure assumes the following pre-requisites,

- The OP provides services in Zone-1 and Zone-2.
- Cloudlet “EC Platform Site #1” belongs to Zone-1, and “EC Platform Site #2” belongs to Zone-2
- The OP has been configured with the network topology information and network routing infrastructure information for dynamically generating the application traffic filtering and routing rules for the SBI-NR interface

The sequence of events shown in Figure 19 assumes that the application onboarding has been completed by the OP and that notifications related to mobility events were requested to be provided by the 5G core network (see section 4.6).

Figure 19 covers an OP's possible tasks to support application Service and Session Continuity. The flow assumes that the 5G Core, based on the UE's subscription data and Operator's configured policies, selects the SSC mode 2 “Break-Before-Make” for the UE PDU Session when its location changes due to mobility.

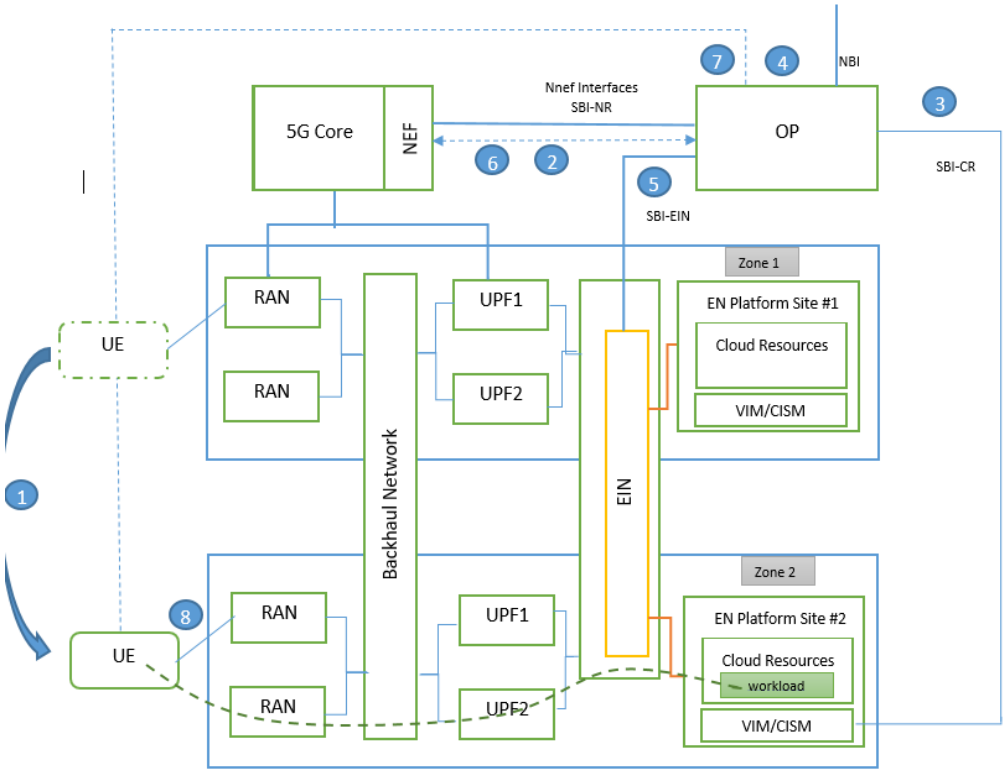


Figure 19: OP Application Relocation flow

Following are the details of the various steps marked in numbers in Figure 19, highlighting the role and objectives of the different interfaces:

1. As the UE moves, it changes to a location covered by Zone-2 from its earlier site in Zone-1. This causes the mobile network to assign a new User Plane in the subscriber's current location to maintain the agreed level of session QoS.
2. Because the OP has requested to be notified about mobility events (via the Traffic Influence Service APIs), the NEF informs the OP about the User Plane (UP Path) change event via the NEF APIs. The 5G Core based on the SSC mode can preserve the UE IP address or assign another network attachment point without preserving the IP address as described in section 2.2.7.3 Requirements for Application Session Continuity. Figure 19 assumes that the network has selected SSC Mode 2 for this session.
3. Acting on the early notification from the network on the possible change in User Plane for the UE for ongoing PDU session, the OP performs the new edge selection for the UE, i.e. “EC Platform Site #2” in Zone-2, which is deployed to provide edge services in the current location of the UE. The OP can create a new Application Instance or select an existing Application Instance in the selected zone.
4. When an Application Instance has been chosen, the OP can initiate synchronisation of the application states using the EIN.
5. The OP can check if an EIN connection is already established between the “EC Platform Site #1” and “EC Platform Site #2 ”. If not, then the OP will request and establish the EIN connection between the two ECs. The OP will then push application traffic steering rules on the EIN. The Application session state synchronisation may involve the application itself requesting and receiving notifications related to NBI APIs

and, when needed, synchronising session states between source and target edge Application Instances. The tasks mentioned are indicative and depend on how an OP implementation is designed.

Note: For an OP's interactions with the 5G Core over the SBI-NR, any of the activities described above are not necessarily dependent on when the OP needs to confirm to the mobile network to complete the User Plane change procedures. OP implementations may use, for example, predictive analytics to estimate the possible future locations for the UE based on the location events received and may prepare early for the application relocation management tasks outside of the scope of this document.

6. The OP informs the UC of the new end points for the application while the UC may continue the ongoing application session with the initial Application Instance.
7. The OP acknowledges that the 5G core network can complete the User Plane change process (UPF-4) and provides the traffic filtering and routing rules for the new Application Instance on Zone-2. The 5G Core instructs the UE to complete the handover to the new User Plane for data communications. If the edge application has requested to be informed about User Plane change events, the OP shall also notify the edge application.
8. Based on the User Plane change confirmation event on the SBI-NR interface from the mobile core network, the OP notifies the UC to switch communication to a new application instance using the application end points provided in step 6.
9. The UE continues the ongoing session with the new application instance on “EC Platform Site #2” in Zone-2, i.e. the new location of the UE, using the new endpoints received in step 6 via UPF2.

Note: The flow does not provide the coverage of other mobility events, e.g., an early notification from the SBI-NR for User Plane change events which could also be notified to UC over the UNI to enable UC apps to prepare for any application-level relocation tasks

Note: The edge interconnection network and interface with an OP is for future studies but can be used by the OP or the edge applications for application state management across the Cloudlets.

4.9 Charging Concepts

The following flows describe how charging factors can be used for different services in non-federated scenarios. Please see section 5.1.5 for the requirements related to the SBI-CHF and Annex I.5 for the charging factors and the service categories underpinning the scenarios.

4.9.1 Charging for Service API Invocation

This flow describes concepts of charging for different type of Service API invocations based on API type and payload. Depending on the API type, a subset of the parameters included in the payload will be of interest to enable rating and charging. This model also supports rating and charging based on the API plus the operation used without payload analysis.

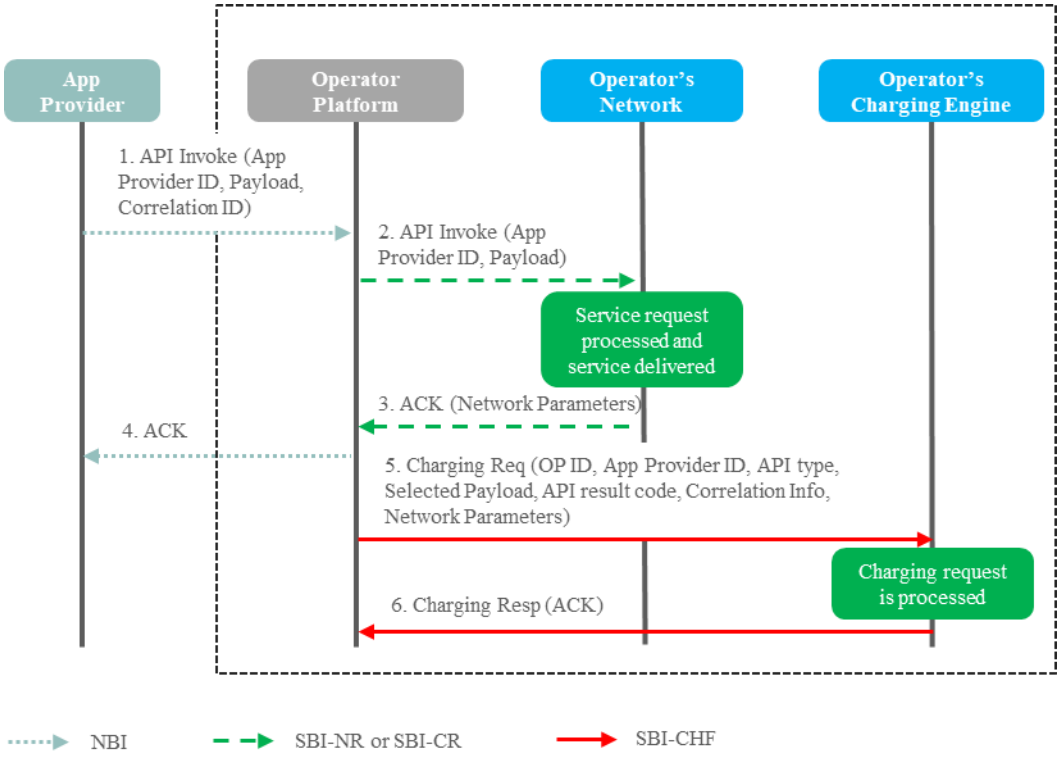


Figure 20: Charging for Service API Invocation

1. The Application Provider invokes an API. Information may include:
    - Party identifiers: App ID, Application Provider ID, Customer device ID
    - API Payload
    - Correlation ID
2. The Operator Platform invokes the corresponding API(s) using SBI-NR.
3. The service request is processed in the Operator's Network and the service is delivered. An ACK is sent back to the Operator Platform. The network response can include relevant parameters for rating and charging.
4. The Operator Platform sends back the response (ACK) to the Application Provider using the NBI
5. The Operator Platform sends a charging request to the Operator's Charging Engine using the SBI-CHF. A charging request includes at least:
    - Party Identifiers: OP ID, App ID, App Provider ID, Customer device ID
    - API type, selected API payload (not mandatory to include and dependent on the service) and API result code
    - Correlation Information
    - Selected Network parameters coming from SBI NR response (not mandatory to include and dependent on the service)
6. The Operator's Charging Engine processes the charging request (rating and charging is done based on provided information in the charging request) and sends a response back to the Operator Platform using the SBI-CHF.

4.9.2 Charging for Data Traffic Usage in Operator Network

This flow describes how the charging concept for data traffic usage will be performed as a result of Service API invocation.

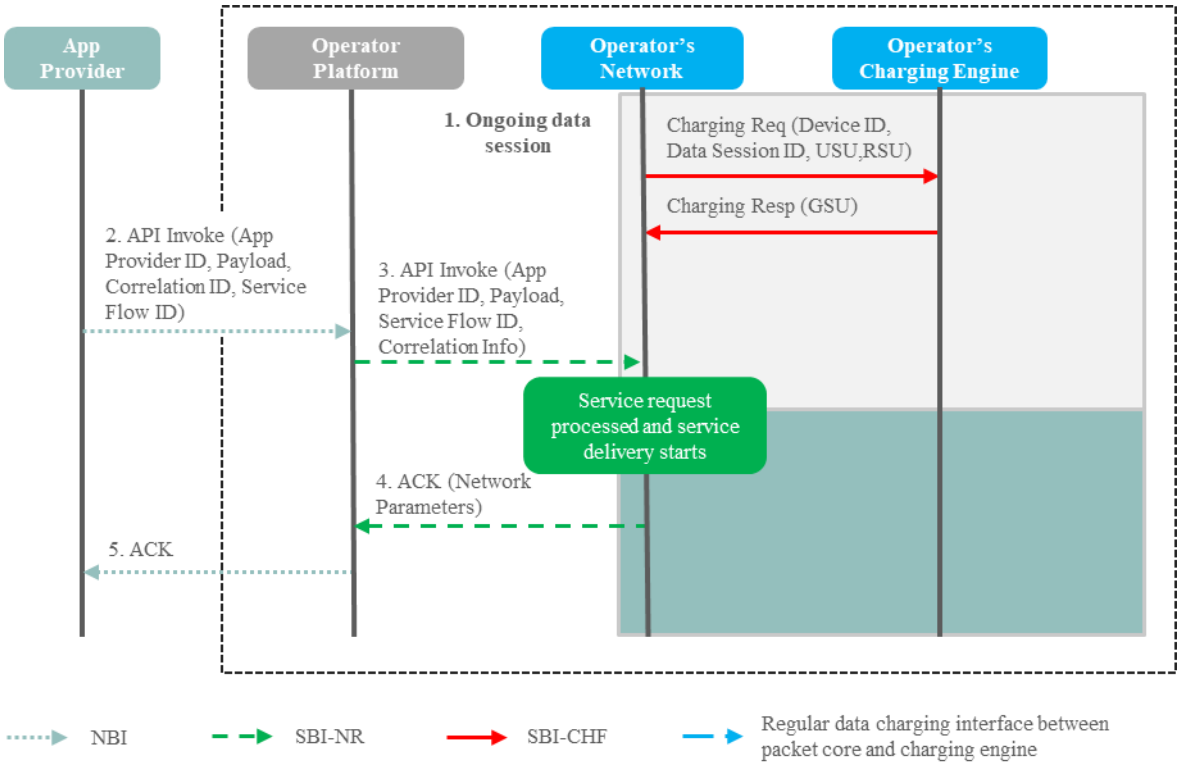


Figure 21: Charging for Data Traffic Usage in Operator Network – Part 1 of 2

1. There is an ongoing data session for a customer. A regular dialogue (session-based charging) is performed between the Operator's data packet core and the Operator's Charging Engine. Online or offline charging could be used depending on the Operator's decision (the online mode shown here as an example) and is out of the scope of this document.
2. An Application Provider invokes an API that influences data traffic usage of a device. Information may include:
    - Party identifiers: App ID, Application Provider ID, Customer device ID.
    - API Payload
    - Correlation ID
    - Data Service flow ID
3. The Operator Platform invokes the corresponding API(s) using the SBI-NR. The Operator Platform provides the required parameters to enable correlation, Correlation Information, between API invocation and the data session in the Operator's Charging Engine.
4. The service request is processed in the Operator's Network and the service is starting to be delivered. An ACK is sent back to the Operator Platform optionally including some additional information which may be relevant for charging.
5. The Operator Platform sends back the response (ACK) to the Application Provider using the NBI

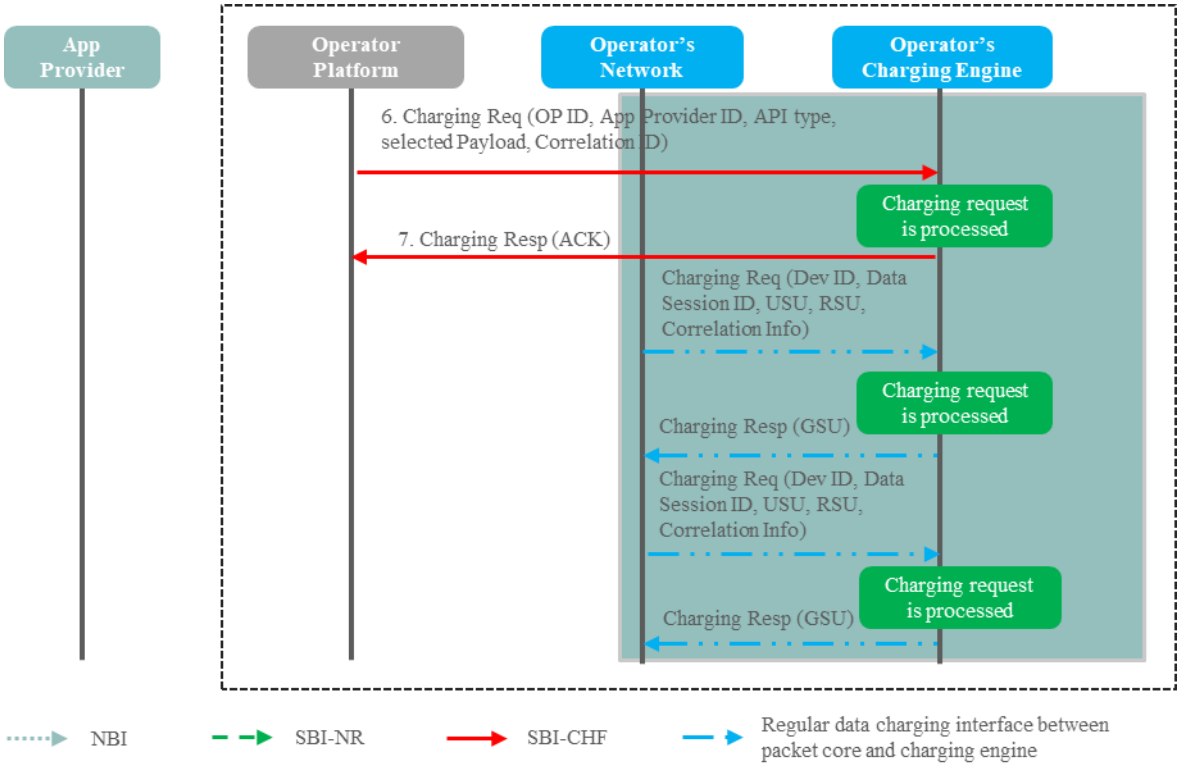


Figure 22: Charging for Data Traffic Usage in Operator Network – Part 2 of 2

6. The Operator Platform sends a charging request to the Operator's Charging Engine using the SBI-CHF to ask for API invocation charging. A charging request may include:
    - Party Identifiers: Leading OP ID, App ID, App Provider ID, Customer device ID
    - API type + selected API payload (optional)
    - Correlation Information
    - DataSessionID
    - Time/duration (optional)
    - Network parameters (optional)
7. The Operator's Charging Engine processes the charging request (rating and charging is done based on provided information in the charging request) and sends a response back to the Operator Platform.

During the service delivery (time/volume where the data traffic is impacted by the API call) the charging dialogue between the Operator's data packet core and the Operator's Charging Engine will continue.

Note: Besides the regular information exchanged for the ongoing charging of the data session, The Packet core will include the Correlation Information required by the Charging Engine to be able to identify the data traffic volume impacted by the Service API Invocation. The contents of this correlation information are for further study.

4.9.3 Charging for Edge Enabling Infrastructure Resource Usage

This flow describes charging for Edge enabling infrastructure resource usage.

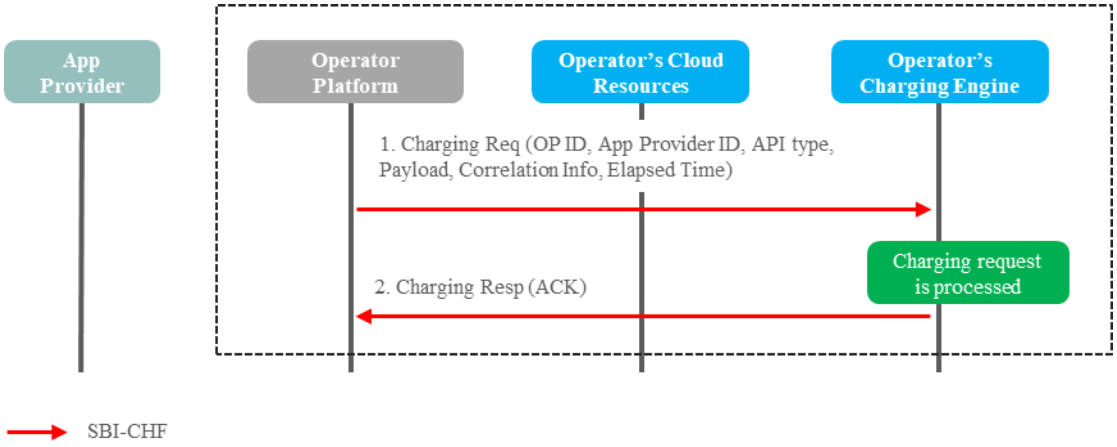


Figure 23: Charging for Edge Enabling Infrastructure Resource Usage

1. The Operator Platform monitors the usage of Edge infrastructure resources and sends a charging request to the Operator's Charging Engine using the SBI-CHF. This can be done periodically based on a configurable timer or one request for the whole period. A charging request may include:
    - Party Identifiers: OP ID, App ID, App Provider ID, Pool ID
    - API type + selected API payload (vCPUs, memory, storage, incoming/outgoing data volume, time period)
    - Correlation Information
2. The Operator's Charging Engine processes the charging request (rating and charging is done based on provided information in the charging request) and sends a response back to the Operator Platform.

4.10 Charging Concepts in Federated Scenarios

The following flows describe charging concepts for Edge computing services in federated scenarios.

Note: Further analysis of other federation services will be included in future versions of this document. This includes any scenarios with intermediate parties.

The E/WBI will be used in communication between the Operators and a pre-requisite is that there is an existing federation agreement in place. In federation scenarios the focus of charging is to ensure that both Leading and Partner Operators have records of the service use as part of their record keeping. This will be the input to settlement and reconciliation between the two Operators later as well as for any wholesale charging between the Leading Operator and the Application Provider.

Please see section 5.1.5 for the requirements related to the SBI-CHF and Annex I.5 for the charging factors and the service categories underpinning the scenarios.

4.10.1 Federated Service API Invocation

This flow describes charging for different type of Service API invocations based on payload in a federated scenario for Edge Computing.

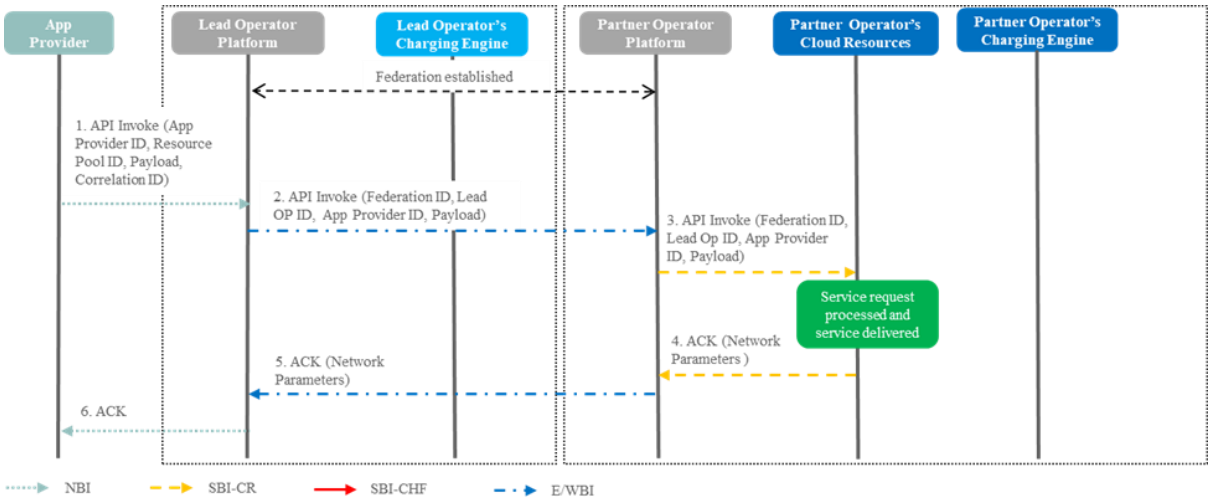


Figure 24: Federated Service API Invocation – Part 1 of 2

1. The Application Provider invokes a Service API
2. The Leading Operator Platform forwards the API request using the E/WBI towards the Partner Operator Platform
3. The Partner Operator platform invokes the corresponding API(s) using the SBI-CR
4. The service request is processed in the Partner Operator's Cloud Resources and the service is delivered. An ACK is sent back to the Partner Operator Platform
5. The Partner Operator Platform sends an ACK back to the Leading Operator Platform
6. The Leading Operator Platform sends back a response (ACK) to the Application Provider using the NBI

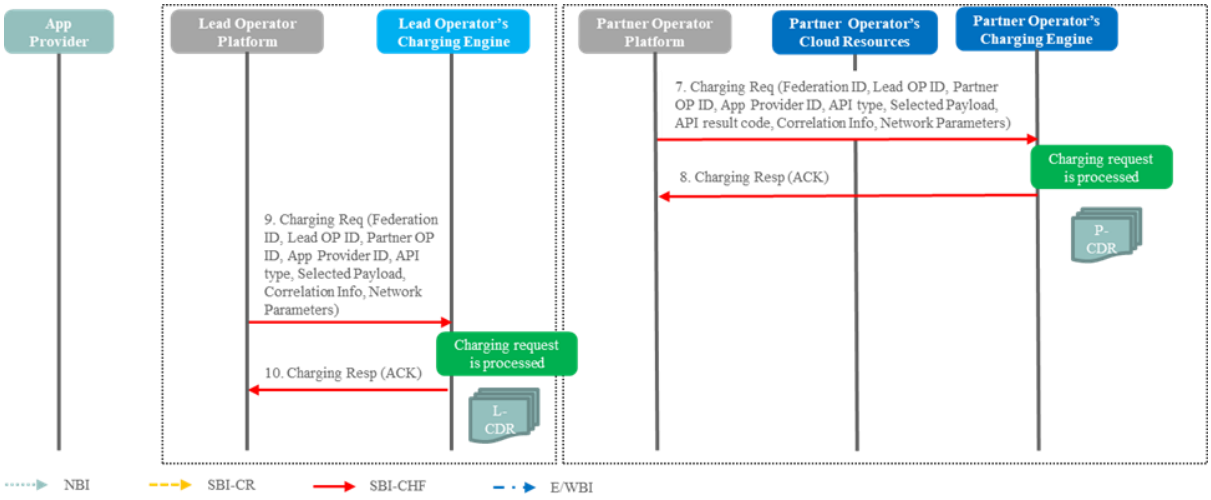


Figure 25: Federated Service API Invocation – Part 2 of 2

7. The Partner Operator Platform sends a charging request to the Partner Operator's Charging Engine using its SBI-CHF. This can be done periodically based on a

configurable timer or be one request for the whole period. A charging request may include:

- Party Identifiers: Federation ID, Leading OP ID, Partner OP ID, App ID, App Provider ID, Pool ID
- API type + selected API payload (not mandatory to include and dependent on the service) + API result code
- Correlation Information

8. The Partner Operator's Charging Engine processes the charging request (rating and charging is done based on provided information in the charging request) and sends a response back to the Operator Platform. The result of charging is stored by the Partner OP as events used for settlement and reconciliation. How and where the events are stored is up to the Operator to decide.
9. The Leading Operator Platform sends a charging request to the Leading Operator's Charging Engine using the SBI-CHF. This is based on the results received from the Partner OP. A charging request includes:

    - Party Identifiers: Federation ID, Leading OP ID, Partner OP ID, App ID, App Provider ID, Pool ID
    - API type + selected API payload (not mandatory to include and dependent on the service) + API result code
    - Correlation Information

10. The Leading Operator's Charging Engine processes charging request (rating and charging is done based on provided information in the charging request) and sends a response back to the Operator Platform. The result of the charging is stored by the Leading OP as events used for settlement and reconciliation. How and where the events are stored is up to the Operator to decide.

Note: The CDRs generated by the Leading and Partner OPs Charging Engines are input to settlement and reconciliation processes outside of charging and hence not in scope.

4.10.2 Federated Edge Enabling Infrastructure Resource Usage

For federated scenarios, it will be possible to periodically exchange information around the effective Edge resource usage over the E/WBI. Consideration needs to be taken to ensure that the resource consumption used for charging on the Partner and Leading Operators are synchronised to reduce risk of reconciliation issues.

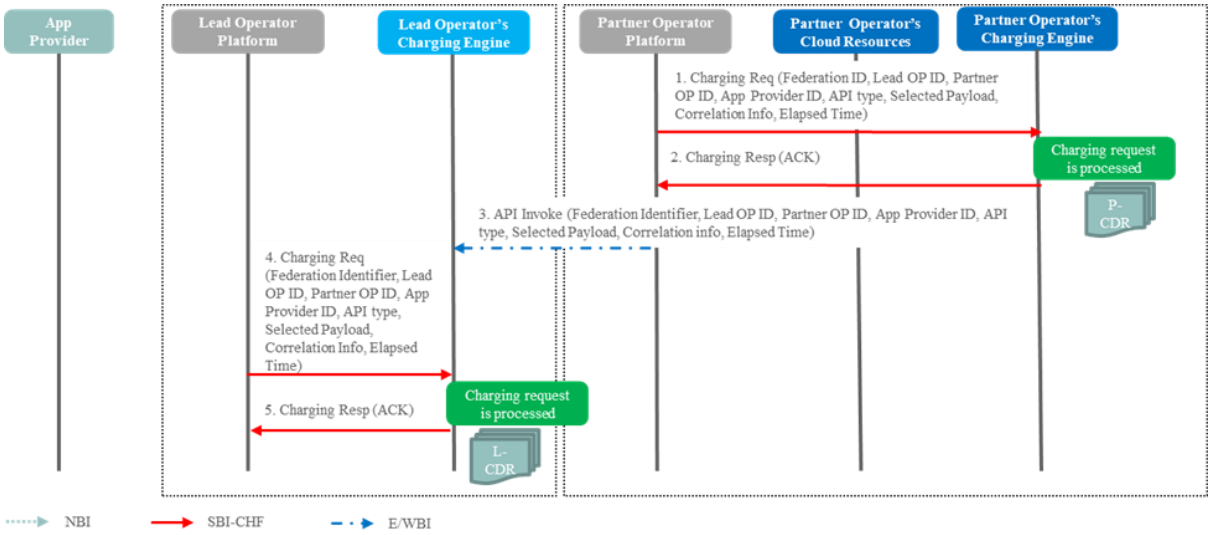


Figure 26: Federated Edge Enabling Infrastructure Resource Usage

1. The Partner Operator Platform monitors the usage of Edge infrastructure resources and sends a charging request to the Partner Operator's Charging Engine using the SBI-CHF. This can be done periodically based on a configurable timer or be one request for the whole period. A charging request may include:

    - Party Identifiers: Federation ID, Leading OP ID, Partner OP ID, App ID, App Provider ID, Pool ID
    - API type + selected API payload (vCPUs, memory, storage, incoming/outgoing data volume, time period)
    - Correlation Information

2. The Partner Operator's Charging Engine processes the charging request (rating and charging is done based on provided information in the charging request) and sends a response back to the Operator Platform. The result of the charging is stored by the Partner OP as events used for settlement and reconciliation. How and where the events are stored is up to the Operator to decide.
3. The Partner Operator Platform invokes the API to exchange the actual resource usage using the E/WBI towards the Leading Operator Platform
4. The Leading Operator Platform sends a charging request to the Leading Operator's Charging Engine using its SBI-CHF. This is based on the actual resource usage received from the Partner OP. A charging request includes:

    - Party Identifiers: Federation ID, Leading OP ID, Partner OP ID, App ID, App Provider ID, Pool ID
    - API type + selected API payload (vCPUs, memory, storage, incoming/outgoing data volume, time period)
    - Correlation Information

5. The Leading Operator's Charging Engine processes the charging request (rating and charging is done based on provided information in the charging request) and sends a response back to the Operator Platform. The result of the charging is stored by the Leading OP as events used for settlement and reconciliation. How and where the events are stored is up to the Operator to decide.

The CDRs generated by the Leading and Partner Ops Charging Engine are input to settlement and reconciliation processes outside of charging and hence not in scope.

4.11 Privacy Management

Depending on the legal basis associated with the Purpose of Data Processing, it might be needed to interact with the end user to obtain authorization for sharing personal data with an Application (e.g., for Consent legal basis – see Annex K). For an OP to trigger the capture of Consent Records by the Privacy Management Function in the CSP domain (records that could be cached in the OP or even shared with a partner OP on federation scenarios provided the local regulations allows it), several procedures could be in place depending on the use case and scenario.

4.11.1 Explicit end user opt-in

In this scenario, an end user, an AP, and a CSP are present, as shown in Figure 27.

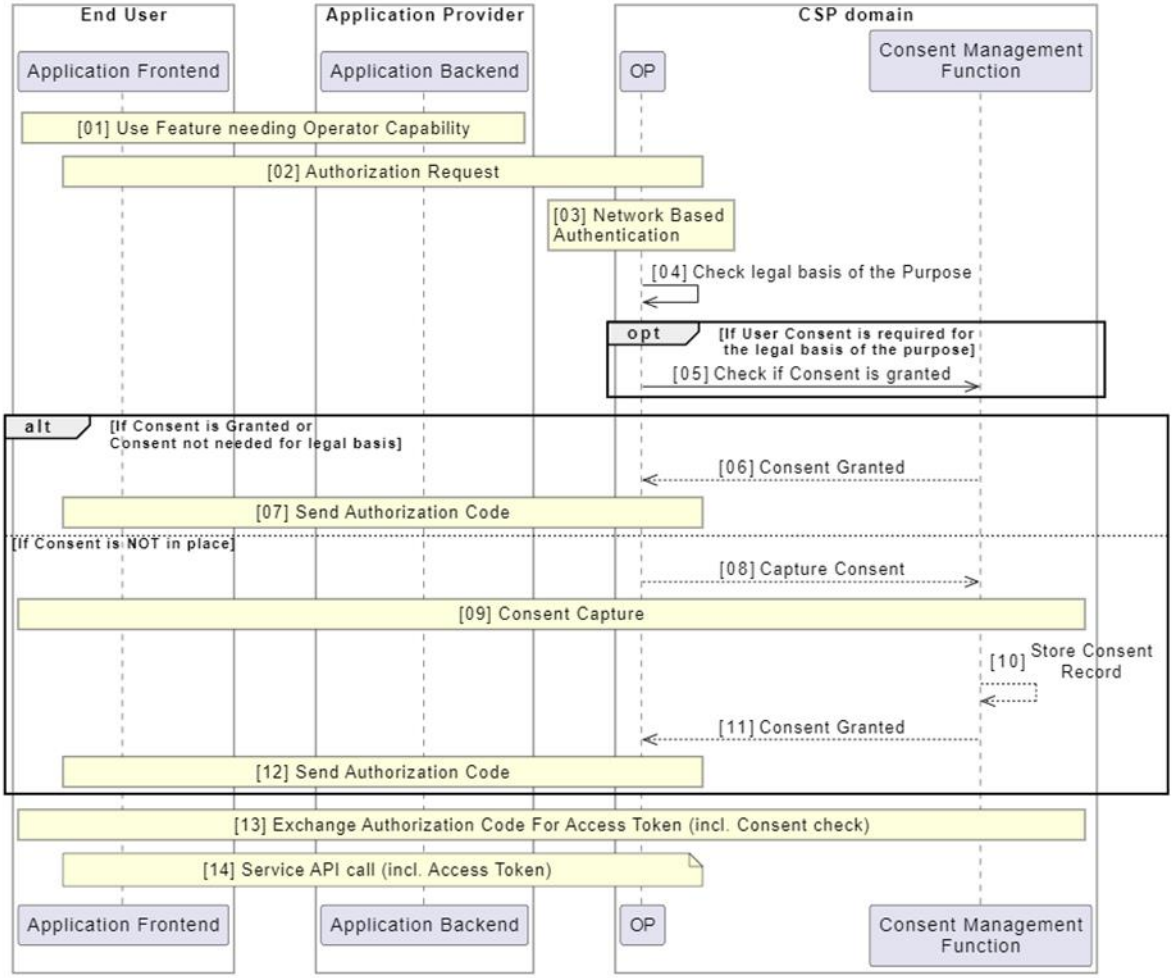


Figure 27: Consent capture from end user

As shown in Figure 27, if some operator capabilities are needed while using an application (e.g., requesting location information in Step 1), an Authorization request is triggered (e.g., by the Application Backend) to the OP carrying (among other information) the Purpose of Data Processing (Step 2). Each Purpose of Data Processing is associated with a legal basis (see Annex K for more details).

Upon reception of the authorization request, a network-based authentication may take place to obtain a network-specific identifier. If the applicable legal basis is Consent (Step 4), then the OP queries the Privacy Management Function over the SBI-PrM interface to check whether the relevant Consent Record is already in place (Step 5).

If the Consent was already granted, then an Authorization Code is sent back (Step 6, 7) and that Authorization Code can be later exchanged for an operator's Access Token (Step 13) that should be carried in the NBI API call in Step 14.

If there is no associated Consent Record, the OP signals the Privacy Management Function in the CSP domain to start the capture of the Consent from the end user (Step 8). Once the Consent is requested (Step 9), the result of that operation is stored (Step 10). If the consent is not granted by the end user (for instance the user does not agree with sharing its information for the signalled Purpose of data processing) the authorization process fails and the access to the personal information resources is denied. If the consent is granted an Authorization Code sent back (Step 11,12) which can be later exchanged for an operator Access Token (Step 13). Upon reception of NBI API call carrying the operator Access Token in Step 14, the OP decrypts the Access Token and forwards the request to receive a response which is sent back.

Note: When an Aggregator is involved, an Operator Resolution is needed (based on an end user identifier like MSISDN or IP address) to forward the Authorization Request (and subsequent NBI API calls) to the right OP.

4.11.2 Consent capture in federated environments

In this scenario, an end user, an Application Provider, a Leading OP and a Partner OP are present, as shown in Figure 28. The AP interacts with the Leading OP via NBI whereas the Leading OP and the Partner OP interact via EWBI. Additionally, the subscription of the end user belongs to the CSP that owns the Partner OP.

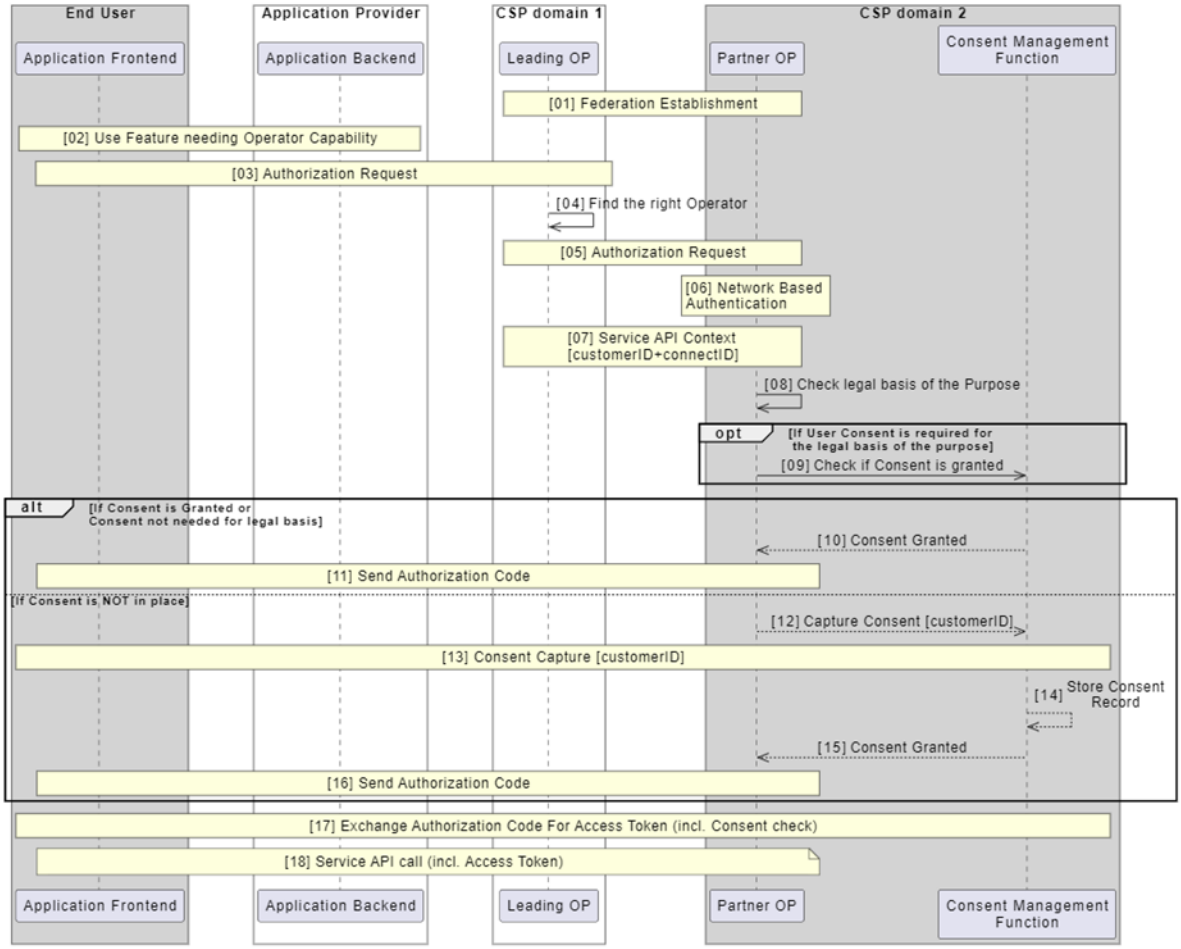


Figure 28: Consent capture in federated environments

As shown in Figure 28, first a federation agreement needs to be established (Step 1) between the two operators [29]. To let the Application Provider to access services and personal information resources hosted in the Partner OP, first an authorization request is sent to the Leading OP (Step 2,3). Based on a user identifier, the Leading OP can find the right operator to forward the authorization request to (Step 4). The Leading OP generates a customerID [29] mapping to the Application Provider, which will be used in the Partner OP to capture Consent (provided it is the applicable legal basis).

The Authorization request is sent to the right Partner OP (Step 5). Upon reception of the authorization request, a network-based authentication may take place in the Partner OP to obtain a network-specific identifier (Step 6). The Service API Context may be created (Step 7) considering the customerID (signalled by the Leading OP) and a connectID (generated at the Partner OP) [29]. If the applicable legal basis is Consent (Step 8), then the OP queries the Privacy Management Function in the CSP domain over the SBI-PrM interface to check whether the relevant Consent Record is already in place (Step 9).

If the Consent was already granted, then an Authorization Code is sent back (Step 10, 11) and that Authorization Code can be later exchanged for an operator's Access Token (Step 17) that should be used for the service API call in Step 18.

If there is no associated Consent Record, the OP signals the Privacy Management Function to start the capture of the Consent from the end user (Step 12). Once the Consent is

requested (Step 13), the result of that operation is stored in the Privacy Management Function (Step 14). If the consent is not granted by the end user (for instance the user does not agree with sharing its information for the signalled Purpose of data processing) the authorization process fails and the access to the personal information resources is denied. If the consent is granted an Authorization Code sent back (Step 15,16) which can be later exchanged for an operator Access Token (Step 17). Upon reception of NBI API call, it is routed to the EWBI (carrying the operator Access Token) in Step 18, the Partner OP decrypts the Access Token and forwards the request in order to receive a response which is sent back.

Even though in Figure 27 and Figure 28 a frontend interaction is needed to capture the Consent from the end user, there could be scenarios in which the Consent is captured through out-of-band mechanisms chosen by an operator (e.g., when no frontend application is available on the device). Additionally, it could happen that the authentication device (providing the Consent) and the “consumption device” (consuming network resources) are different. In those scenarios, the Consent capture shall follow a different mechanism, e.g., OpenID Connect Client-Initiated Backchannel Authentication flow in [54].

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 service, resource and end-user provisioning capabilities through the NBI.
2. The NBI shall offer the Edge Cloud, Network or other Operator Capabilities to Application Providers and Aggregators, 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.
    c) Network Capabilities operators agreed to expose
    d) Other Operator Capabilities

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: Access control to NBI Service and Operate APIs can be based on an access token mechanism.

4. The NBI shall allow to easily identify the target resource (e.g. UE, Subscriber, Availability Zone) to which the request relates.
5. For End-user identity, an OP shall accept MSISDN, IP-address or User Identity Token as identifier and for routing purposes

    a) End-user identification based on an MSISDN requires the Operator to register the MSISDNs using number ranges that can be shared or Number Portability Databases.
    b) End-user identification using IP-address ranges requires the Operator to share its service IP ranges amongst participating operators and Aggregators for routing purposes.
    c) End-user identification using User Identity Token requires the Operator to include an User Identity Token generating and check function as described in section 5.1.10

6. An OP shall expose routing information to Aggregators to find the Home OP of the End-User for all subscriptions served by the OP
7. The Home OP shall be found regardless of whether its OP serves a MVNO or MNO, including sub brands owned by these organizations.
8. The Home OP shall be discoverable when a device is connected to Wi-Fi. Note that in this case the IP address will not point to the home operator.
9. The Operator shall ensure that MSISDNs including non dialable IOT MSISDNs can be used for identification of the Home OP if these are needed for routing purposes.
10. For an API routing lookup method with User Identity Token check, the end-to-end latency including device, network and OP shall aim for sub 1 second and not exceed 4 seconds
11. The lookup mechanisms shall support devices connected without an MSISDN such as game consoles, laptops and IOT devices (e.g. based on IP address or User Identity Token)
12. The service should return the routing information in case of secondary devices owned by the Subscriber.

Note: Device ownership being different from subscription owner (e.g. connecting someone else’s laptop to your UE) is FFS.

5.1.1.2 Onboarding and Deployment Profile

5.1.1.2.1 General

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

- receiving the request,
- authorising/authenticating the Application Provider,

- 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, and
- mapping the Application Provider's request for exposed network capabilities to the available capabilities in the target network(s).

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.

An 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. Subscriber 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. Network Capability requirements, for example, but not limited to

    a) QoS (e.g. Linux Foundation CAMARA project QoD / L4S / URSP /DSCP / etc.)
    b) Connectivity Events
    c) Network-based location
    d) Network statistics
    e) Network analytics

7. 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.

8. Policies that allow the Application Provider to manage circumstances where user conditions do not comply with the deployment criteria.
9. Support for telemetry information from the operator.
10. 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.

11. Service availability in visited networks required/supported.
12. Application lifecycle management policies specifying actions to be taken if the OP cannot provide the requested Service Levels, e.g. terminating the application instance, transport reset, etc.
13. Session Continuity sensitivity indicating the edge application's capabilities to support application session relocations across Cloudlets
14. Alternative QoS References in order of priority that the OP may apply to PDU sessions if the specific QoS as requested by the Application Provider cannot be met for a given application.

5.1.1.3 Management Profile

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

1. An OP shall enable Application Providers to request Edge Cloud in an Availability Zone (within the OP and federated OPs):

    a) On a basis where the Application Provider 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 Provider 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 Providers.
    d) An Application Provider may provide an 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. An 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 Provider 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. An OP shall describe the exposed capabilities of the Leading OP's network(s) and those of the federated target networks
6. An OP shall offer a structured workflow for application deployment and instantiation: CRUD functions.
7. An OP shall allow an Application Provider 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.
8. An OP shall allow an Application Provider to specify whether or not it requires service availability on visited networks (that is, when a UE roams away from its home network operator).
9. An OP shall provide an Application Provider with telemetry information concerning the performance of the Edge Cloud service, including fault reporting.
10. An OP shall allow an Application Provider to request a particular granularity for the telemetry information they receive.

Note: Possibly using a publish and notification approach.

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

11. An OP shall allow an Application Provider to require that outbound access to the internet is prohibited.
12. An 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.
13. An OP shall support Single Sign-on based on login credentials for an Application Provider.
14. An OP shall offer functionality that supports the Application Provider to manage its application instances. For example, to monitor operational performance, get diagnostic logs and help with debugging.
15. An OP shall offer functionality that supports the Application Provider in managing the application development, integration and deployment.
16. An OP shall allow an Application Provider to request to receive application relocation event notifications.
17. An OP shall allow an Application Provider to request to be notified about the abstract Service and Session Continuity modes applied for application sessions.
18. An OP shall allow an Application Provider to request to receive application QoS change notifications if the requested Service Levels drops below a threshold
19. An OP shall allow an Application Provider request to receive application location change event notifications.
20. An OP shall allow an Application Provider to request to receive UE radio access type change event notifications.
21. An OP shall allow an Application Provider to request to receive UE IP address change event notifications.
22. An OP shall allow an Application Provider to request assignment of cloudlet-specific FQDNs for Edge Applications that an Application Client can resolve to an Edge Application's instance IP address.

5.1.1.4 Resource Reservation Profile

5.1.1.4.1 General

When an Application Provider accesses An OP portal or uses an 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 authorised entities.
2. 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 (Exposure Functions) and the Application Provider. The security policies are enforced by the NBI API Gateway.
3. This authentication shall rely on commonly used API authentication mechanisms (e.g. OpenID Connect, Oauth, etc.).
4. 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.

5.1.1.6 Network Communication Service lifecycle and end user profile

An OP shall be able to support the following requirements:

1. An OP shall allow an Application Provider to manage the lifecycle of the Network Communication Service, optionally with specifying geographical region, timeframe, target operators, etc. as defined in section 3.4.16.
2. An OP shall allow an Application Provider to request a Network Communication Service lifecycle state change.
3. An OP shall support the notification for the Network Communication Service lifecycle status change.
4. An Application Provider should be able to assign and switch a Network Communication Service for the end user to access the application.

   Note: For this action, the Application Provider is expected to know related information such as Network Communication Service ID as defined in section 3.4.17.

5. An OP shall notify the Application Provider if end user’s profile data changes.
6. An OP shall enable the Application Provider to request an end user’s profile information.

5.1.1.7 Operator Capabilities Service lifecycle

An OP shall be able to support the following requirements:

1. An OP shall allow an Application Provider to access Operator Capabilities, directly or through an Aggregator. The Operator Capabilities are accessed via APIs, known as Service APIs, which are consumed by the Application Provider's applications.
2. An OP shall expose functionalities for privacy management to access the Operator Capabilities, when those capabilities require the treatment of personal information of the Operator's subscribers.
3. An OP shall enable OAM functionalities for an Application Provider or Aggregator to configure, monitor, measure and control the Operator Capabilities. The OAM functionalities are accessed via APIs, known as Operate APIs. OAM capabilities requirements include:

    a) An OP shall allow an operator to expose the catalogue of products, including the available Operator Capabilities, to an Application Provider or Aggregator.
    b) An OP shall allow an Aggregator to register a new Application provider so that they can access Operator Capabilities.
    c) An OP shall allow an Application Provider or Aggregator to register a new application so that they can access Operator Capabilities.
    d) An OP shall allow an Application Provider or Aggregator to request a specific Operator Capabilities product to be accessed by a specific registered application.
    e) An OP shall allow an Application Provider or Aggregator to access the usage reporting of Operator Capabilities from their registered applications.
    f) An OP shall allow an Application Provider or Aggregator to supervise the status of the Operator Capabilities accessed by their registered applications.
    g) An OP shall allow an Application Provider or Aggregator to report platform fault or issues of the Operator Capabilities accessed by their registered applications.

4. An OP shall provide routing information to the Application Provider or Aggregator for identifying which operator is responsible for handling the required Operator Capability, as related to the operator's subscriber.

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.
2. An OP shall be able to identify the UCs among OP instances.
3. An OP shall be able to identify the Application Providers among OP instances.
4. An OP shall be able to identify the applications among OP instances.
5. The E/WBI shall allow to easily identify the target resource (e.g. UE, Subscriber, Availability Zone) to which the request relates.
6. The E/WBI shall allow to update Partner OPs on changes related to the resource identifiers for which the OP can offer services (e.g. IP address ranges for UEs, Availability Zones offering Edge Resources).

5.1.2.2 Security Requirements

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

OPs 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) Edge Resources shall be published as “edge resources” entities, referred to a specific Availability Zone.
    b) Specific edge node information shall not be shared.

2. An OP shall only expose the resources to its Partner OPs previously agreed with each specific Partner.
3. The E/WBI shall provide an authentication mechanism to enable access only to authenticated and authorised entities. Therefore, mutual authentication shall be provided between the instances of the OP.
4. The E/WBI shall provide a security mechanism to safeguard the confidentiality, integrity and authenticity of the exchanged data
5. 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.

5.1.2.3 Application Management

The E/WBI needs to replicate the behaviour and functions available on the NBI to transmit the workload, requirements, mobility decisions, privacy considerations (if already in place on the originating OP side), 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.
    b) If privacy considerations are already in place in the Leading OP side, e.g., Consent records (when Consent is the applicable legal basis), the E/WBI shall support forwarding the relevant Consent-related configuration to any partner OP.
    c) If the Leading OP avails relevant Consent-related configuration (i.e., Consent records), the E/WBI shall allow notifications (to any partner OP) on changes of that configuration on the Leading OP-side.

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.
e) An OP receiving an instantiation request through its E/WBI shall be responsible of the Network Communication Service management for the service and the (to be) connected subscribers.
f) An OP receiving an instantiation request through its E/WBI along with Consent records (when Consent is the applicable legal basis), should cache this information for Authorizing future invocations and be able to process any future notification on changes of the Consent records.

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 forward the network events for use by the Leading OP and the Application that may have requested those from the Leading OP.
6. The E/WBI shall be employed for managing the service continuity on visited networks.
7. The E/WBI shall forward the network and analytics information to be shared with the Leading OP of the Application Provider.
8. The E/WBI shall forward the Network Communication Service lifecycle management and related subscriber information to be shared with the Leading OP of the Application Provider.
9. The E/WBI shall forward the Capture Consent message (whenever Consent is the applicable legal basis) to the Leading OP, which in turn should forward it to the associated Privacy Management Function to capture the consent from the end user.

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.

An Operator Platform shall be able to access the cloud resources of its 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 Resource Orchestration Function

A cloud provider/operator may want to expose the cloud resources through an orchestration function. However, this integration may not expose the whole set of functionalities that an Operator Platform may need to provide. In this case, a serverless computing approach would be available where the provider’s orchestration function performs the instantiation of the application based on the request from the OP.

With this cloud resource orchestration function integration, an OP shall be able to provide access to:

- Application onboarding/instantiation on specific edge/cloud site (Cloudlet);
- Image management;
- Application lifecycle management;
- Routing / Forwarding resources management;
- Retrieval of limited resource usage statistics for settlement.

The capabilities exposed by the cloud resource orchestration function may not allow an OP to enlarge or reduce the resources reserved for edge application purposes. Furthermore, the information provided does not enable the OP to ensure an application's instantiation until the orchestration function performs the internal infrastructure procedures. These limitations endorse the serverless computing approach of this integration.

The cloud resource management and orchestration capabilities and the statistics that an cloud resource orchestration function offers to an OP are restricted to the cloud resources used and the assigned orchestrator's tenant's scope.

OP SBI-CR integration shall allow adopting industry references for cloud resource orchestration function integration.

5.1.3.1.2 Integration with Infrastructure Manager

If direct integration with a cloud resource manager is done, e.g., directly using the Virtualised Infrastructure Manager (VIM) or Container Infrastructure Service Manager (CISM), an OP may offer access to additional functions beyond those offered based on integration with a Cloud Resource Orchestration function. These functions include, for example,

- transforming / mapping resource management requests,
- transforming / mapping reservation requests,
- returning transformed / detailed statistics,
- offering resource catalogue and
- load reporting.

Access to these additional functions may result in the OP offering cloud resource management function 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 offer access to:

- Application onboarding/instantiation on a specific edge/cloud site (Cloudlet);
- Image management;
- Application lifecycle management;
- Routing / Forwarding Resources management requests (e.g., based on the outcome from Transformation Functions in Figure 4);
- 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 e.g., Openstack or Kubernetes(see Figure 6).

5.1.3.1.3 Integration with Hyperscalers

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

An OP shall be able to access edge capabilities via the SBI-CR and re-expose themto Application Providers through the NBI. The OP shall do this in a manner that provides the complete set of needed functionalities, restricted to the 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 authorised entities. Therefore, mutual authentication shall be provided between the OP and the cloud resource provider / orchestrator / management function(s).
2. The SBI-CR shall support the use of authorisation mechanisms by its endpoints that 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 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.
5. The SBI-CR shall support the security mechanisms that the cloud resources and their interconnection should provide to protect the live migration of Edge Application services between Edge Nodes.
6. The SBI-CR shall safeguard the protection and integrity of parameters and controls for steering user traffic to the application instances.

5.1.4 Southbound Interface to Network Resources

5.1.4.1 General

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

An OP shall be able to access network capabilities that the Operator has chosen to expose 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 network capabilities that the operator has chosen to expose, e.g. QoS, Network Events/Statistics.
- 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.
- The OP shall be able to retrieve network analytics information (when available) in a standardised way: load level information, network performance, service experience, etc.
- The OP shall be able to retrieve resource analytics information (when available) in a standardised way.
- The OP shall be able to access an end user's data profile.
- The OP shall be able to retrieve end-user's roaming access details (e.g. status, network connected).

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

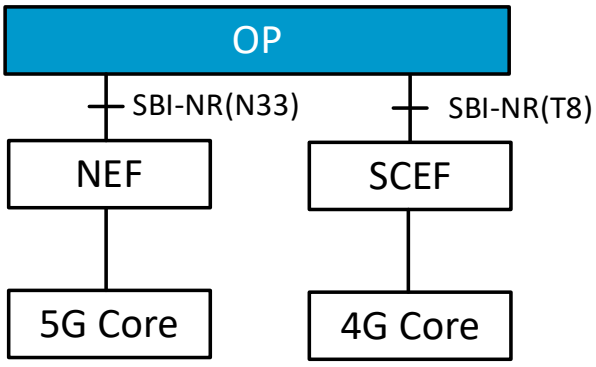


Figure 29: Functional mapping between OP and NEF/SCEF

A combined SCEF+NEF may be deployed by the MNO to hide the specific network technology from applications for user devices having capabilities for 4G and 5G access. In those scenarios it is expected that an OP should be able to support the communication with the combined SCEF+NEF on the SBI-NR interface.

Different sets of APIs can be supported by the two network types i.e., EPC and 5GC. From this perspective an OP should be able to discover the API capability differences while interacting over the SBI-NR with mobile core network.

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 request to be informed on 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 a UE mobility analytics report.

10. An OP's SBI-NR may be able to retrieve a UE communication pattern report (e.g. UL/DL volume per application).
11. An OP's SBI-NR may be able to retrieve a network performance report (e.g. gNB active ratio, gNB computing resource usage).
12. An OP's SBI-NR may be able to report QoS change statistics in a specific area.
13. An OP's SBI-NR may be able to retrieve UE status reports (e.g. location information, reachability, roaming status).
14. An OP's SBI-NR may be able to control the transfer of data in the background for UCs.
15. 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).
16. An OP's SBI-NR may be able to configure the Alternative QoS References applicable to different access technologies for cases where the specific QoS target requested by the Application Provider cannot be met.
17. An OP's SBI-NR may be able to receive QoS relevant notifications based on UE connection statistics.
18. An OP's SBI-NR may be able to configure the charging party of the UE data sessions.
19. An OP's SBI-NR may be able to configure service-specific parameters for UCs (e.g. network slice).
20. 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).
21. An OP's SBI-NR shall be able to influence the URSP rules sent to the UE to provide the mapping of applications to the DNN/NSSAI applicable to the serving network.
22. An OP's SBI-NR may be able to influence the 5G mobile core network to establish a user plane for PDU sessions requiring access to edge services based on OP-provided criteria.
23. An OP's SBI-NR may be able to report access type change notifications for UCs due to user mobility.
24. For the APIs that are common to EPC and 5GC, an OP's SBI-NR shall be able to support operations to be informed on their availability or the expected level of support.
25. An OP's SBI-NR shall be able to work with the Common API Framework (CAPIF) when available.

Note: An OP's SBI-NR can work without CAPIF. If CAPIF is not supported, the SBI-NR API will provide an alternate means of providing these functions.

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 authorised entities. Therefore, mutual authentication shall be provided between the OP and the NR.
2. The SBI-NR shall support the use of authorisation mechanisms by its endpoints that 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 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.
5. The SBI-NR shall support security mechanisms to protect the network functions discovery procedure of the NEF/SCEF by an OP.

5.1.5 Southbound Interface to Charging Function

An OP shall provide a set of capabilities that will enable the charging and billing for the usage of the Operator's services exposed to third party providers. Although these services and capabilities are quite heterogeneous and in constant evolution, they can be classified into a set of categories that share common characteristics from a charging perspective and that are described in Annex I of this document.

5.1.5.1 General Charging Integration Requirements

1. An OP shall be able to integrate with the CCS (Converged Charging System) deployed in the Operator's network through the SBI-CHF interface. This integration will allow doing the rating and charging for the usage of the services and capabilities exposed by the OP.
2. Considering that there could be different CCS instances deployed in the Operator's network (e.g. dedicated instances for a particular service/customer segment, geo-redundant deployment, etc.) the OP will be able to select the CCS instance that will be used to do the rating and charging of the service.

Note: The criteria used for this CCS instance selection (e.g. CCS discovery mechanism, OP local configuration, etc.) are for further study but as a general approach an OP will provide mechanisms to configure the target CCS instance depending on a combination of different parameters (e.g. type of service used, application provider identifier, etc.)

3. An OP shall support different charging integration models with the CCS. The charging integration models to be supported will be the ones standardised by 3GPP and defined in 3GPP TS 32.240 [35]. As a reference, the following charging models shall be supported:

    - Event Based Charging:

      This charging model is based on a request/response pattern, where an OP would trigger a charging request when an event occurs (e.g. an API invocation) including all the information relevant for rating and charging for the CCS.
      The CCS would use the information provided in the charging request to do the rating and charging for that event and will send the response to the OP with the result.
      The following charging model, defined by 3GPP, will be supported by the OP:

      - PEC (Post Event Charging): a charging request is sent after the service is delivered. (e.g.. an OP receives an API call, makes several API calls through the SBI to deliver the service and a Charging request is sent after the OP makes these API calls through the SBI)

Note: Although 3GPP also defines IEC (Immediate Event Charging) charging model, where a charging request Is sent before the service that is associated to the event is delivered, the support for this charging model in the OP is not mandatory and is left for further analysis as it has dependencies on the evolution of 3GPP standards in the context of 5G SA charging.

- Session Based Charging:

Note: the usage of this charging model in the context of the OP requirements is for further study. The description of this charging model will be expanded in next releases of this document).

The charging model to be used by an OP in the integration with the CCS will depend on the particular service to be charged.

4. An OP shall provide mechanisms that will allow doing the charging for the services in the case of unavailability of the connection with a CCS through the SBI-CHF interface. These mechanisms are for further study but as a reference the following approaches could be used:

   - Usage of a primary/secondary/pool of CCS instances as the result of the CCS instance selection procedure, so that in case the primary instance is not available a secondary one could be used.
   - Ability to log/store charging requests when no CCS instances are available so that this information could be used to do the rating and charging when communication is re-established.

5.1.5.2 Services and capabilities exposure charging requirements

1. An OP shall support rating and charging for the following service categories described in Annex I of this document:

   a) Category 1: Network capabilities exposure services with no impact on the device's data usage.
   b) Category 2: Network capabilities exposure services with impact on the device's data usage.
   c) Category 3: Network provisioning services.
   d) Category 4: Edge application management services.

2. An OP shall support the following charging factors/events for triggering charging for the services included in Category 1:

   a) Service activation charging.
   b) Charging per service API invocation (and related notifications):

      An OP shall be able to send a charging request to the Operator's CCS through the SBI-CHF following the PEC charging model defined by 3GPP. This charging request will be sent once the service is delivered upon confirmation from the Network.

Note: As a reference, the integration between an OP and the Operator's CCS for this charging factor is shown in section 4.9.1 of this document.

3. An OP shall support the following charging factors/events for triggering charging for the services included in Category 2:

    a) Service activation
    b) Charging per service API invocation (and related notifications):

        The same requirements as in requirement2.b) (charging per service API invocation for category 1 services) of this section are applicable for this case.

    c) Charging based on data traffic consumption in the Operator's Network as a result of a previous service API invocation.

        An OP shall be responsible for providing the Operator's Network (through the SBI-NR) with the information that allows the correlation between a service API invocation and a data traffic flow from a device in the Operator's Network.

Note: A charging dialogue will take place between the Operator' Network and the Operator's Charging engine following the regular procedure used in the Operator to do the data sessions charging (out of the scope of this document). The Operator's Network will include the correlation information provided by the OP in the charging requests sent to the CCS to indicate the API Invocation's impact on charging.

Note: As a reference, the integration between an OP, the Operator's Network and the Operator's CCS for this charging factor is shown in section 4.9.2 of this document.

4. An OP shall support the following charging factors/events for triggering charging for the services included in Category 3:

    a) Service activation
    b) Charging per service API invocation (and related notifications):

        The same requirements as in 3.b) above (charging per service API invocation for category 2 services) are applicable for this case.

    c) Charging per service API invocation (service lifecycle modification charging):

        An OP shall be able to send a charging request to the Operator's CCS through the SBI-CHF following PEC charging model defined by 3GPP after a Service lifecycle modification API request is received from the NBI and the service is delivered.
        For the specific case of Network Slice Instance management services, the charging models defined in 3GPP TS 28.202 [37] will be supported. In other scenarios, e.g. realisation through QoS services or dedicated APNs where no specific definition is available, the operator can fall back to traditional ways of charging per event (e.g. QoS) or lifecycle management of a needed APN.

d) Charging based on data traffic consumption in the Operator's Network as a result of a previous service API invocation:

The same requirements as in 3.c) above (charging based on data traffic consumption for category 2 services) are applicable for this case.

5. An OP shall support the following charging factors/events for triggering charging for the services included in Category 4:

a) Service activation
b) Charging per service API invocation (application lifecycle management operations):

An OP shall be able to send a charging request to the Operator's CCS through the SBI-CHF following PEC charging model defined by 3GPP after an Application lifecycle management API request is received from the NBI or from the E/WBI (for the case of federated scenarios) and the service is delivered.

Note: As a reference, the integration between an OP and the Operator's CCS for this charging factor is shown in sections 4.9.1 and 4.10.1 (for federated scenarios) of this document.

c) Charging per service API invocation (charging for edge enabling infrastructure resources usage based on subscribed capacity in API request):

An OP shall be able to send a charging request to the Operator's CCS through the SBI-CHF following PEC charging model defined by 3GPP after an API request is received from the NBI or from the E/WBI (for the case of federated scenarios) requesting for capacity reservation in the Operator's Cloud Resources and the API request is processed.

Note: As a reference, the integration between an OP and the Operator's CCS for this charging factor is shown in sections 4.9.1 and 4.10.1 (for federated scenarios) of this document.

d) Charging based on edge enabling infrastructure resources usage:

An OP shall be able to send a charging request to the Operator's CCS through the SBI-CHF including the information about the effective resources' usage in the Operator's CR over a period of time.
Charging models defined by 3GPP in 3GPP TS 32.257 [39] will be supported for this purpose.

Note: An OP will periodically retrieve the actual resource usage information from the Operator's CR based on the agreed data collection interval from the SBI-CR or from the E/WBI in the case of federated scenarios. This information will be the one included in the charging request sent to the Operator's CCS through the SBI-CHF. The definition of this mechanism is out of the scope of this section.

Reference to diagram flow in section 4.9.3 and in section 4.10.2 (for federated scenarios) of this document is provided for clarifications.

e) Charging based on data traffic consumption in the Operator's Network as a result of a previous service API invocation (in non-federated scenarios):

The OP shall be responsible for providing the Operator's Network (through the SBI-NR) with the information that allows to do the correlation between a service API invocation and a data traffic flow from a device in the Operator's Network that is accessing an application.

Note: A charging dialogue will take place between the Operator' Network and the Operator's Charging engine following the regular procedure used in the Operator to do the data sessions charging (out of the scope of this document). The Operator's Network will include the correlation information provided by the OP in the charging requests sent to the CCS.

Reference to diagram flow in section 4.9.2 of this document is provided for clarifications.

6. An OP shall allow the configuration of charging factor/factors to be used for the services applicable to each category on a per service basis. It will also be possible to configure different charging factor/factors per service depending on the scenario: federated / non-federated.

   In the case of federated scenarios, this configurability is applicable both to the Leading and Partner Ops.

Note: It is an Operator's decision to decide which charging factors - from the ones that are applicable to a category – will be configured/used to do the charging and billing for the usage of a service/capability.

This decision will be dependent on the commercial model chosen by the Operator for the commercialisation of that service (and out of the scope of this document).

The configuration in the OP will need to be aligned with this decision.

7. An OP shall maintain security and data/topology privacy requirements when reporting consumptions to the Operator's Charging Engine (both in federated and non-federated scenarios.

   In case that - for technical reasons - an OP needs to report information disclosing privacy sensitive data/topology to the Operator's Charging Engine (e.g., to allow correlation in scenarios where data traffic consumption in the Operator's Network needs to be correlated with a service API invocation) the responsibility to guarantee these security/privacy requirements will be on the Operator's Charging engine side.

Note: the specific information to be provided for correlation and the mechanisms used in the Operator's Charging Engine are for further study.

5.1.5.3 Charging information

1. The charging requests sent by an OP to the Operator's Charging Engine through the SBI-CHF shall include any information usable by the Operator's CCS to address the rating and charging of the services and enable the final billing process in the Operator. An OP creating or sharing charging data shall guarantee the security, integrity, availability, and non-repudiation of charging data.
2. Charging information to be provided by an OP in the charging requests shall include the identification of the different parties that are involved in the transaction, from the Application Provider to the UC. These identifiers could be used for different purposes by the Operator's CCS (e.g., to determine the chargeable parties, to have end-to-end traceability of the transaction, etc.).
3. An OP shall at least include the following identifiers in the charging requests sent to the Operator's CCS through the SBI-CHF interface:

    a) Party identifiers involved in the transaction: Application ID, Application Provider ID, Customer Device ID.
    b) Operator Platform ID
    c) Partner Operator Platform ID (only applicable in the case of federated scenarios)

4. An OP shall include a correlation identifier of the NBI service API invocation in the charging requests sent to the Operator's CCS. This correlation identifier is a unique identifier for that particular transaction. This ID will allow end-to-end traceability and will assist in the correlation required to enable charging factors where data traffic charging in the Operator's Network needs to be correlated with the service API invocation by the Operator's CCS.

Note: The mechanism used to generate this correlation identifier is out of the scope of this section and for further study.

5. An OP shall include specific information that will depend on the service category and the charging factor in use in the charging requests sent to the Operator's CCS.

Note: The information to be collected is described in the next requirements.

A summary table showing the list of potential charging factors per service category is shown in Annex I of this document.

6. For the services included in categories 1, 2, 3 and in case the charging factor chosen by the Operator is the one based on API invocations or on service lifecycle modification operations received, the OP shall be able to include the following information in the charging requests:

    a) Mandatory information:

        i. API type (identification of the service API that was invoked through the NBI e.g., device location)

    b) Optional information:

i. A subset of the parameters included in the service API invocation. The list of parameters to be included (if any) will be configurable per service.
ii. A subset of parameters retrieved from the Network (e.g., device ID in the Operator's Network) after the service is delivered. The list of parameters to be included (if any) will be configurable per service.
iii. API result code

7. An OP shall expose the most relevant parameters associated to the Network Communication Service (e.g., time, latency, jitter, reliability, coverage area, etc.) in the charging request as part of the optional information.

   In the specific case where the network communication service is realised with slicing in a 5G network, the charging for the network slice management operations described in section H.5 of this document the charging models and charging information defined in 3GPP TS 28.202 [37] will be used.
   The concrete list of mandatory/optional parameters is for further specification but as a general approach any of the parameters included in the GST (Generic Network Slice Template) could be included by an OP based on configuration.
   The operator needs to fallback to traditional ways of charging when the network communication service is realised in a 4G network with QoS or a dedicated APN.

8. In the case of service categories 2 and 3, and if the charging factor chosen by the Operator is based on API invocations to enable simple time-based charging models (charging per unit of time the service is delivered where this time is not measured in the Operator's Network), the OP shall be able to include the time parameter in the charging requests to be used for charging purposes.

   The procedure used in OP to measure this service delivery time is out of the scope of this section.

9. In the case of service categories 2 and 3, and if the charging factor chosen by the Operator is based on data traffic consumption in the Operator's Network, the OP shall include the following information in the charging requests sent through the SBI-CHF:

   a) Mandatory information:

      i. API type (identification of the service API that was invoked through the NBI e.g., QoS influence)
      ii. Correlation information: this information will allow the CCS to correlate the charging requests associated to the devices data traffic consumption received from the Operator's Network with the service API invocation (to distinguish this traffic from the regular data traffic navigation of a customer).

         The OP will also be responsible for providing this correlation information to the Operator's Network through the SBI-NR. The Operator should have in place the mechanisms to guarantee that this correlation information is provided to the CCS in the charging requests sent from the Operator's Network. This mechanism is out of the scope of this document.

The list of parameters to be included (if any) will be configurable per service and is left for further specification.

b) Optional information:

i. A subset of the parameters included in the service API invocation. The list of parameters to be included (if any) will be configurable per service.
ii. A subset of parameters retrieved from the Network (e.g., service flow id in the Operator's Network) after the service is delivered. The list of parameters to be included (if any) will be configurable per service.

10. In the case of services in categories 3 and 4 and if the charging per service API invocation is chosen by the Operator to enable lifecycle management API requests charging, the OP shall include the following information in the charging requests sent through the SBI-CHF:

a) Mandatory information:

i. API type and operation (identification of the service API that was invoked through the NBI: e.g., application instantiation)

b) Optional information:

i. A subset of the parameters included in the service API invocation. The list of parameters to be included (if any) will be configurable per service.
ii. API result code

11. In the case of services in category 4 and if the charging factor chosen by the Operator is for edge enabling infrastructure resources usage based on subscribed capacity in API request, the OP shall include the following information in the charging requests sent through the SBI-CHF:

a) Mandatory information:

i. API type and operation

b) Optional information:

i. A subset of the parameters included in the service API invocation that include the detailed information about the resources to be reserved (independent from the effective usage):
    1) Subscribed compute capacity:
        a. vCPU
        b. Memory
        c. Network Resource Location
        d. Availability zone
    2) Subscribed storage capacity:
        a. Storage
        b. Type
        c. Network Resource Location
        d. Availability zone

3) Subscribed Network capacity:
    a. Input
    b. Output
    c. Label (internet traffic, intra-cluster, inter edge cloud traffic …)
4) Subscribed accelerators capacity:
    a. Accelerator name (Example: GPU)
    b. Type
    c. Network Resource Location
    d. Availability zone

ii. A reservation time period

12. In the case of services in category 4 and if the charging factor chosen by the Operator is for edge enabling infrastructure resources usage (information about effective consumption), the OP shall include the following information in the charging requests sent through the SBI-CHF:

a) Mandatory information:

i. API type and operation

b) Optional information:

i. A subset of the parameters included in the service API invocation that include the detailed information about the resources to be reserved (independent from the effective usage):
    1) Effective compute usage:
        a. vCPU
        b. Memory
        c. Network Resource Location
        d. Availability zone
    2) Effective storage usage:
        a. Storage
        b. Type
        c. Network Resource Location
        d. Availability zone
    3) Effective Network usage:
        a. Input
        b. Output
        c. Label (internet traffic, intra-cluster, inter edge cloud traffic …)
    4) Effective accelerators usage:
        a. Accelerator name (Example: GPU)
        b. Type
        c. Network Resource Location
        d. Availability zone

ii. Covered usage time period.

5.1.5.4 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 authorised entities. Therefore, mutual authentication shall be provided between the OP and the Charging Engine element.
2. The SBI-CHF shall provide an authorisation mechanism to grant access to only the necessary services to which previous authorisation has been granted.
3. The SBI-CHF shall support the use of security mechanisms by its endpoints that safeguard the exchanged data's confidentiality, integrity, and authenticity.
4. 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.
5. An OP shall maintain security and data/topology privacy requirements when reporting federated consumption.

5.1.6 Southbound Interface to Edge Interconnection Network

5.1.6.1 High-Level Requirements

1. An OP shall provide the interface for control/management of the EIN between two ECs.
2. An OP will help enable the EIN, but not keep track of the interface management further.
3. An OP shall help establish the EIN between two ECs, and optionally provide security guidelines.

Note: EIN connection setup and management among different operators is out of scope for this version.

5.1.6.2 Security requirements

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

1. The SBI-EIN shall provide an authentication mechanism to enable access only to authenticated and authorised entities. Therefore, mutual authentication shall be provided between the OP and the EIN management.
2. The SBI-EIN shall support the use of authorisation mechanisms by its endpoints that grant access to only the necessary authorised services and data.
3. The SBI-EIN shall provide security mechanisms to safeguard the confidentiality, integrity and authenticity of the exchanged data.
4. The SBI-EIN shall support security mechanisms for the SDN control plane.

5.1.7 User to Network Interface

5.1.7.1 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 UC (typically located in the UE) and an 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 or non-SIM UE accesses via an untrusted network (such as public Wi-Fi), the OP needs to undertake its own AAA services or rely on the mechanism recommended in section 5.1.7.2.3.

5.1.7.2 User First Attachment

5.1.7.2.1 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.

5.1.7.2.2 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 subscriber 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 UC can request an Edge Cloud service.
3. A UNI UC request shall include identity information and parameters:

   a) For a UE,

      i. UE ID, e.g. MSISDN, GPSI;
      ii. Application ID;
      iii. Location, e.g. cell-ID, TAI. The UNI request does not need to include this information if the OP knows the UE’s location.

   b) For a Non-SIM UE,

      i. UE ID, e.g. UUID (or equivalent)
      ii. Application ID;
      iii. Preferred network ID (For the preferred OP)
      iv. Location, e.g. City, Latitude/Longitude (If possible). The UNI request does not need to include this information if the Non-SIM UE does not support a location feature. In that case, the OP needs to identify the Non-SIM UE's location using the Non-SIM UE's network information.

5.1.7.2.3 First-time registration for non-SIM UE

UCs for non-SIM UEs shall do a first-time(bootstrap) registration with an OP on the very first connection to the OP.

1. A Non-SIM UE shall send a first-time registration request with location information and Non-SIM UE details.

Note: First-time registration authentication and security details for Non-SIM UEs are out of the scope of this document.

2. An OP shall generate a unique ID for the registering Non-SIM UE. The OP can follow approaches like UUID generation or other proprietary mechanisms to identify the non-SIM UE.
3. An OP shall perform the location identification of a Non-SIM UE using the network information based on the public IP address of the registration request.
4. An OP shall register the non-SIM UE using ID, Location and other device information shared as part of the registration process.

    a) Information shall be stored at the OP for use on subsequent connections.
    b) The OP may generate and share authentication/authorisation information for the non-SIM UE and communicate that information in the response message.

5. On successful registration, the UC shall set the status locally as registered and store exchanged ID and authentication details securely on the Non-SIM UE.
6. The Non-SIM UE shall use the exchanged information for identification, authentication and authorisation on subsequent connections.

5.1.7.2.4 UE Authentication and Authorisation

An 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 a UE-based UC.
3. For Non-SIM UEs, an OP shall authenticate using ID and other security parameters exchanged at the first-time registration of the Non-SIM UE (see section 5.1.7.2.3).
4. In addition, the OP shall provide a mechanism to allow efficient authorisation of the UE for subsequent interactions.

5.1.7.2.5 Cloudlet selection

An 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. For Non-SIM UEs, at the time of first-time registration (bootstrap), location information will be identified by UC or OP. The Non-SIM UE and OP shall store this location information and refer to it for cloudlet selection.

Note: Mobility of Non-SIM UEs may be covered in future versions of this document.

3. 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 Provider has previously determined where its Application Clients would be,
   b) satisfies the Application Provider's statement about the requirements for data privacy,
   c) meets the Application Provider's input on requirements for QoS, and the UC'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.

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

5.1.7.2.6 Service Provisioning

An 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. 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),

   a) The OP shall ensure that the Edge Application can be reached by all applicable connectivity services (e.g. best effort, latency optimised and bandwidth optimised) and prioritise.

2. The UE shall be able to test the connectivity characteristic towards the selected Cloudlet.
3. An OP shall be able to inform Application Clients about QoS changes
4. An OP shall be able to inform Application Clients about Edge Application Relocation events.
5. An OP shall be able to inform Application Clients about the new communication endpoints of the relocated edge Application Instance.
6. An Application Client may be able to provide the observed QoS reports to the OP over the UNI.

5.1.7.3 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 authorised UCs and OPs. Therefore, mutual authentication shall be provided between the UC and the OP.
2. The UNI shall provide secure communication between the UC and the OP, assuring integrity protection, replay protection and confidentiality protection.
3. 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.

5.1.8 Southbound Interface to OAM

5.1.8.1 General

The integration with the operation and management APIs on the SBI-OAM allows an OP to expose them to the Application Provider. Depending on the service offerings and the deployment options, the Operator may impose limits on the management capabilities exposed on the SBI-OAM interface.

The OP integration to the operation and management systems should allow:

- The OP to retrieve network slice lifecycle notification in a standardised way
- The OP to manage network slice lifecycle status in a standardised and secure way.
- The OP to manage and retrieve 4G network characteristics in a standardised way.

An OP's SBI-OAM shall be able to interact with the 5G / 4G system via a slice management function (e.g. NSMF or Exposure Governance Management Function (EGMF)) to access these management capabilities.

Note: The APIs used by an OP to interact with a slice management function can either be TM Forum or 3GPP APIs.

5.1.8.2 NSaaS Lifecycle management

Network Communication Services in 5G networks can be realized with NSaaS, in which case the following requirements apply:

1. An OP's SBI-OAM shall be able to interact with the Operator's systems to manage the network slice.
2. An OP's SBI-OAM should be able to retrieve information on a NSI.
3. An OP's SBI-OAM shall be able to collect a network slice lifecycle status for the network slices associated with the services provided by OP.
4. An OP's SBI-OAM should be able to modify the network slice lifecycle status.

5.1.8.3 Security requirements

The following security requirements shall be considered:

1. The SBI-OAM shall provide an authentication mechanism to enable access only to authenticated and authorised entities. Therefore, mutual authentication shall be provided between the OP and the OAM system.
2. The SBI-OAM shall support the use of authorisation mechanisms by its endpoints that grant access to only the necessary authorised services and data.
3. The SBI-OAM shall provide security mechanisms to safeguard the confidentiality, integrity and authenticity of the exchanged data.
4. The SBI-OAM 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.

5.1.8.4 Network Communication Service management in pre 5G networks

The OP shall be able to manage Network Communication Service characteristics in the networks where NSaaS is not available, or the operator chooses to use traditional QoS mechanisms.

1. An OP's SBI-OAM shall be able to interact with the Operator's systems to manage the network.
2. If a dedicated APN/DNN is provided, the end user's traffic should be treated according to the AP's requirements (e.g. specific QoS). The OP shall provision the Policy and Charging Rules Function (PCRF)/Policy Control Function (PCF) with the rules needed to satisfy the AP request for the APN/DNN.
3. If a dedicated APN/DNN is not available, end user's traffic detection is required in the network, and the end user's traffic should be treated according to the AP's requirements (e.g. specific QoS). The OP shall provision the PCRF/PCF with the rules needed to satisfy the AP request.
4. An OP's SBI-OAM shall be able to collect performance metrics for the APN/DNN associated with the services provided by OP.
5. An OP's SBI-OAM should be able to modify the network characteristics related to QoS/APN/DNN.

5.1.9 Southbound Interface for Privacy Management

5.1.9.1 General

The integration with the Privacy Management Function in the CSP domain enables an OP to verify whether a suitable legal basis allows sharing personal data with an Application (owned by an Application Provider) within an Authentication and Authorization context. Depending on the legal basis associated with the Purpose of Data Processing, an explicit interaction with the end user (whom personal data belongs) to grant access to the protected resources (for instance for Consent legal basis) might be needed.

5.1.9.2 High-Level Requirements

The OP integration to Privacy Management Function is valid only when e.g., Consent is the applicable legal basis and shall allow:

- The OP to retrieve whether a Consent Record for a specific API call is already in place in the Privacy Management Function,

- Cache an existing Consent Record and request to receive notifications from the Privacy Management Function related to the change of information in the Consent Record,
- To trigger the capture of the Consent if there is no Consent Record in the local cache or in the Privacy Management Function,
- When applicable, trigger an update of the Consent Records hosted in the Privacy Management Function to enable the exercise of an end user privacy right (see Annex K).

5.1.9.3 Security requirements

The following security requirements shall be considered:

- The SBI-PrM shall be confidentiality and integrity protected.
- The SBI-PrM shall support mutual authentication between the OP and the Privacy Management Function within the CSP domain.
- The SBI-PrM shall support the use of authorisation mechanisms by its endpoints that grant access to only the necessary authorised services and data.
- The SBI-PrM shall support the adoption of strong security mechanisms that guarantee forward secrecy and prevent intervening attacks such as replay, relay and man-in-the-middle attacks.

5.1.9.4 Managing Consent

The OP shall be able to:

- trigger the capture of the Consent when dictated by the legal basis, and when no valid Consent record is found in either the local cache or the Privacy Management Function.
- receive notifications when a subscriber revokes Consent via NBI (and forward that notification over the SBI-PrM and probably EWBI in federation scenarios.)
- receive notifications via SBI-PrM when a subscriber revokes Consent (and forward that notification over the NBI and probably EWBI in federation scenarios).
- keep records of attempts to capture Consent from the subscribers through Logging, Tracing and Auditing functions.

5.1.10 Southbound interface for User Identity Token check

Network API invocation must be secured in a way to prevent intruders from taking the identity of applications and / or devices to get hold of network and device specific information. Most use cases deal with applications that are used from a device and that leverage Network APIs that target the very same device. The mechanism for API invocation must therefore support a means of authenticating the End-User and the Application via some User Identity Token that is supplied by OP or retrieved this elsewhere in the operator domain (e.g. entitlement server…) and that is used throughout the API invocation chain. The integration with the User Identity Token Manager in the operator domain enables an OP to verify whether the application and user inside the token is correct. Integration with the User Identity Token Manager using the SBI-PrM (Privacy Management) reference point shall be supported.

5.1.10.1 High level requirements

The OP must support using an User Identity Token for End-User and Application authentication, which has been provided by the User Identity Token Manager function to the Application running on the device and passed through the API invocation chain to the OP platform:

1. The OP shall support application registration with globally used ids (e.g. ID of the device application client as known in the device vendor app store)
2. The OP shall provide the Operator ID to the User Identity Token Manager function for inclusion in the User Identity Token for call routing purposes.
3. The OP may support the User Identity Token as login hint on the different variants of authorization (e.g. Oauth 2.0, OIDC, Mobile Connect) protocol
4. The OP shall decompose the User Identity Token , extract the network subscription ID (e.g. MSISDN), Operator ID, and application ID and authenticate these against the registered data
5. The OP shall be able to map the device application client ID against the backend application invoking the API to ensure that these belong to each other.
6. Secondary devices like smartwatch share the same external MSISDN but have an internal technical MSISDN which is not revealed to the subscriber. An OP shall be able to use this technical MSISDN for API resolution.

Note: Dual SIM situations are for FFS.

7. The User Identity Token shall contain identifiers with operator ID in the token to allow an Application Provider or an Aggregator to route the request toHwards the home OP
8. The identity of the issuing User Identity Token Manager may be obfuscated in the token for privacy reasons, in which case trusted Aggregators and APs should be able to deobfuscate the routing information for routing purposes.

5.1.10.2 Security requirements

The following security requirements shall be considered:

1. The SBI-PrM shall be confidentiality and integrity protected.
2. The SBI-PrM shall support the use of authorisation mechanisms by its endpoints that grant access to only the necessary authorised services and data.
3. The SBI-PrM shall support the adoption of strong security mechanisms that guarantee forward secrecy and prevent intervening attacks such as replay, relay and man-in-the-middle attacks.
4. The User Identity Token shall be confidentiality and integrity protected.
5. The User Identity Token shall be End-user and application specific.
6. The User Identity Token shall contain an encrypted section for End-user identifiers as well as application identity and a potentially an unencrypted section indicating the domain of the home operator.

    a) There shall be the option to encrypt the User Identity Token section indicating the domain of the home operator.

7. The User Identity Token may be used through the whole API invocation chain, even if there are multiple partners between Application Backend and the OP.

5.1.10.3 Checking the User Identity Token

The OP shall be able to:

1. Interact with the User Identity Token Manager function for token validation over the SBI-PrM.
2. Optionally trigger User Identity Token invalidation in case validation has failed
3. Keep records of User Identity Tokens used on API invocations through Logging, Tracing and Auditing functions.
4. OP shall be able to interact with the User Identity Token Manager function over the SBI-PrM to retrieve the appropriate device/End-user identifier (such as MSISDN) and application identifier from the supplied User Identity Token.

5.2 Functional Elements

5.2.1 Exposure Functions

5.2.1.1 High-level requirements

The Exposure Functions serve as intermediary layer between the Application Provider and the Leading OP and transitively to those OPs federated with the Leading OP. To carry out this function, it shall satisfy the requirements listed below.

Note: In some cases, a requirement associated with the Exposure Functions specifically applies to its endpoint to the Application Provider, i.e. the NBI. In those cases, the requirement will be specified for the NBI.

1. The Exposure Functions shall present a data model to the Application Provider that is consistent with the Common Data Model of section 3.4.
2. The Exposure Functions shall support an Application Manifest model consistent with the Edge Application Data Model of Table 3.
3. The Exposure Functions shall present a QoS Profile model to the Application Provider that is consistent with Table 21
4. The Exposure Functions shall present a Cloudlet data model to the Application Provider consistent with Table 4 for scenarios in which Cloud Resource information is collected and inventoried.
5. The Exposure Functions shall present an Availability Zone data model to the Application Provider consistent with Table 7.
6. The Exposure Functions shall present a set of Availability Zones to the Application Provider that is representationally consistent with the Availability zones of the OPs that the Application Provider can reach and internally consistent. This means that the Application Provider does not need to re-build or re-link applications because of inconsistencies in the specification of Availability Zones. Differences in Availability Zone representations that can be accommodated in an Application Manifest/Metadata or similar means is acceptable.
7. The Exposure Functions shall present a data model, as shown in Table 13, consistent among Leading and Partner OPs.
8. The Exposure Functions shall present an information model to the Application Provider that is consistent among the Leading OP and the Partner OPs federated with it.

9. The Exposure Functions shall allow the Application Provider to present a workload profile with a common specification to the OP and enable the common specification to apply to the Leading and federated Partner OPs. The common workload specification shall be consistent with the QoS information profile of Table 21.
10. The Exposure Functions shall support Application Life Cycle scenarios consistent with Table 1.
11. The Exposure Functions shall support a secure means of authentication and authorisation, operating over the NBI.
12. The Exposure Functions shall support a common model for telemetry data (i.e., data arising from resource monitoring) and a means of configuring telemetry data collection, as described in section 3.3.7.
13. The telemetry system should be consistent with the SBI-CHF interface of section 3.5.2.3.
14. The Exposure Functions should support default values for all configurable parameters in manifests, profiles, and other data structures to allow for an “easy” default deployment of an application.
15. An Application Provider may request deployment of an application by specifying parameters in an Application Manifest. The Leading OP shall try to satisfy the manifest, potentially in a Partner OP, but need not guarantee that it will be satisfied. The response of the Exposure Functions to the Application Provider, both for a successful or an unsuccessful request, shall be consistent.
16. The Exposure Functions shall present NSaaS data models that are consistent with the NSaaS Data Model in Table 15 and Table 17 in Section 3.4.

5.2.1.2 Security Requirements

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

1. The Exposure Functions shall provide an authorisation mechanism to grant access to only the necessary authorised services and data. The security enforcement point is the NBI API Gateway.
2. The Exposure Functions shall provide a fine-grained authorisation mechanism to grant authenticated entities selective access to the NBI exposed services and functionalities.
3. The Exposure Functions shall use profile-based access control to provide appropriate restrictions on the amount of functionality that the OP offers to a particular Application Provider, according to their operational profile and the type of access requested.
4. When defining and assigning the authorisation profiles, the Exposure Functions shall apply the principle of least privilege, ensuring that any entity should have only the minimum profile roles necessary to perform its function.
5. Given the external exposure of the NBI, the Exposure Functions shall provide security mechanisms to counteract/prevent attacks aimed to undermine the availability of the NBI, such as DoS and DDoS attacks, reconnaissance attacks (attempts to identify service or API vulnerabilities) and brute force attacks.
6. The Exposure Functions should provide isolation between resources of different Application Providers (e.g. when providing telemetry data or when accessing and managing Edge Applications configuration data).

7. The Exposure Functions should provide security mechanisms to protect accounting and guarantee safe logging (e.g. integrity, non-repudiation, etc.) of the activity over the NBI.

5.2.2 Federation Functions

5.2.2.1 Federation and Platform Interconnection

5.2.2.1.1 General

One of the Operator Platform's primary purposes is offering Application Providers an extended operator footprint and capabilities through interconnecting with other operators' resources and subscribers. 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.

5.2.2.1.2 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.

5.2.2.2 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.

These services will be provided over the SBI-CHF where the CDRs generated by the Leading and Partner Ops Charging Engine are input to settlement and reconciliation

processes outside of charging and hence not in scope. Reference to diagram flows in section 4.10 of this document are provided for clarifications.

5.2.2.3 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. An OP shall be able to share the exposed network capabilities.

5.2.2.4 Security Requirements

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

1. The Federation Functions shall provide an authorisation mechanism to grant access only to the necessary authorised services and data for a Partner OP. The security enforcement point is the EWBI API Gateway.
2. The Federation Functions shall provide security mechanisms to counteract attacks aimed to prevent the availability of the E/WBI, such as Denial of Service (DoS) attacks
3. The Federation Functions should provide security mechanisms to protect accounting and guarantee safe logging (e.g., integrity, non-repudiation, etc.) of the activity over the E/WBI.

5.2.2.5 Routing of Requests

When having relationships with one or more Partner OPs the following requirements apply

1. The OP shall be able to determine whether it needs the support of a Partner OP to provide a service.
2. When needing the support of a Partner OP, the API Federation Management Function (depicted in Figure 4) shall be able to determine what E/WBI the OP needs to use to reach that Partner OP.
3. For cases where routing depends on the Subscriber or UE to which the service request relates, the API Federation Management Function shall be able to identify the appropriate Partner OP and E/WBI based on

- a public IP address through which the UE or Subscriber is identified,
- the MSISDN associated to the subscription, and
- A network-specific token that identifies the Subscriber (e.g. an external GPSI including a domain).

4. For cases where routing depends on network or cloud resources to which the service request relates, the API Federation Management Function shall be able to identify the appropriate Partner OP and E/WBI based on

   - Identifiers for the Availability Zone where the resources would be located.

5. For this identification, the API Federation Management Function shall take into account the information provided by the Partner OPs over the E/WBI that the OP has with them (see section 5.1.2.1).
6. An OP's API Federation Management Function shall update Partner OPs on changes in the resource identifiers for which the OP can offer services (e.g. IP address ranges for Subscribers or UEs, Availability Zones offering Edge Resources).

5.2.3 Transformation Functions

No general requirement have been identified so far for the Transformation Functions in Figure 4.

Note: Future general requirements for the Transformation Functions (e.g., identified in other groups), may be transposed into this document.

5.2.4 Integration Functions

5.2.4.1 Network/Operator Criteria for edge selection

When several edge nodes meet the Application Provider criteria and support operator policies, the platform shall consider the following criteria to enable selection of the edge where the application shall be deployed:

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.

5.2.4.2 Instantiation Strategy for Edge Applications

An OP shall be able to map / forward / route instantiation requests for 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 map / forward / route requests for static instantiation of the application on a specific edge node.
2. An OP shall be able to map / forward / route requests for 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 map / forward / route requests for the dynamic instantiation of an Edge Application based on a user's request.
5. If a dynamic instantiation is considered necessary, the OP shall trigger the deployment the application image and create an instance on the selected Cloudlet.

5.2.4.3 Edge Application Relocation

5.2.4.3.1 General principles for Edge Application Relocation

In the context of this document, Edge Application Relocation deals with the transfer 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 IP address change.

- Due to UE mobility, or or the OP's measurements, 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 functions(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.).

- An OP should be aware 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 relocation 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, Edge Application Relocationincludes support for a change of operator and OP.

5.2.4.3.2 Relocation 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. Relocation 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: the latter seems less likely, as it is hard for an OP to measure application-level performance accurately, but some simple measures such as packet loss may be possible.

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

2. Relocation 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;

Note: 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).

Note: The application should not over-report relocation triggers.

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

5.2.4.3.3 Application relocation Conditions/Restrictions

An OP shall be able to consider the application-specific requirements for managing relocation 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 relocationfor all the edge services associated with each UC.
3. An OP shall consider the relocationsensitiveness of the applications.
4. An OP shall take into account the active Edge Application on the UC for considering the relocation.

    a) An OP shall ensure that all the active Edge Applications are relocatedcorrectly when network change is required.
    b) An OP shall not perform a network relocation if an active application does not support relocation.
    c) An OP shall not perform application relocationto 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 relocation.

5.2.4.3.4 Application relocation (Server-Side)

An 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.
4. An OP shall ensure that the information is available to configure an edge application's traffic flows towards the selected edge node statically or dynamically via the SBI-NR, e.g. via the Operation and Management plane.

5.2.4.3.5 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 addresses cases where the Application Provider has requested, as part of the initial policy phase, to be notified prior to any change of the edge compute resource hosting 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 context from the old edge compute resource to the new one.
3. The application indicates to the OP when it is ready to be relocated to the new edge compute resource. This approach means that the application is generally in charge of the timing of the relocation (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 relocation. Then, the OP shall be able to cancel the relocationprocedure. 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 relocation.

5. The application shall confirm the completion of the relocationof the Edge Application onto the new Cloudlet to the OP.
6. Relocationof the UE may require that the operator changes the IP address used by the application client.
7. An OP shall support the capability for Edge Applications to request to be informed on application clients' IP address change events and shall be able to notify the applications when events are reported over the SBI-NR.

5.2.4.3.6 Relocation 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 any relocation procedures.
6. Network GW location may not be needed in case of service degradation due to an edge node saturation.

5.2.4.4 Service Availability on Visited Networks

5.2.4.4.1 General

Service availability on visited networks shall be considered to allow the users to use an 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.

5.2.4.4.2 Requirements

1. When a device first attaches to a visited network, there shall be messaging between the UC, Home OP and Visited OP. The messaging's purpose is for the Home OP to

authenticate the UC and authorise it to use the Edge Cloud and Network Capabilities 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 UCs of the Home OP. It is not the intention to define a granular charging /settlement mechanism ("granular" meaning, for example, per UC or per application instance).

2. User plane LBO/SBO shall be available for the UC in the visited network.

   a) If no LBO/SBO is available or there is no service availability agreement among operators, the UC 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 UC.
5. The Visited OP shall match the Application Provider's requirements on Network Capabilities to the exposed capabilities in the visited operator network.
6. 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).
7. The Visited OP shall be able to provide the abstract application Service and Session Continuity capabilities over the E/WBI for roaming users to their Home OP
8. The Visited OP shall be able to provide the application mobility relocation monitoring events information to an application's Leading OP over the E/WBI.

   Note: UC mobility management is handled with existing mobility management mechanisms.

5.2.4.5 Application Operation and Management

An OP shall expose to an Application Provider a set of management capabilities including:

1. The capability to

   a) Create Cloudlets within an Availability Zone
   b) 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. The capability to manage the exposed network capabilities
6. Capabilities for the operator to monitor Cloudlet usage in terms of compute, memory, storage and bandwidth ingress and egress
7. The capability to monitor the above metrics per tenant.
8. 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 request to be notified about that event, receive it. Services that requested notifications, update their corresponding data. For a failed transaction event, the service that requested notifications 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 OP shall allow for access to 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 OP shall release reserved resources after the reservation expires (in case of reservation).

9. The capability to monitor Cloudlet events, alarms logs
10. The capability to monitor Cloudlet performance metrics
11. The capability to offer operator interfaces to federated partners to monitor usage across Cloudlets
12. The capabilities for Edge Applications FQDN management

    a) Management of DNS subdomain(s)
    b) Management of FQDN allocation to Edge Applications
    c) Synchronisation of DNS records (i.e., FQDN to IP address(es) mapping) updates with the DNS service.

5.2.4.6 Seamless Application Service and Session Continuity

5.2.4.6.1 General principles for application session continuity

A mobile user actively engaged with an edge application instance hosted in a Cloudlet may, during their movement from one place to another, not always get the desired quality of experience. This is due to various network access factors like poor radio connectivity, network congestion, etc.

The quality of experience from an application's perspective is affected by different aspects impacted by these network access factors, e.g. uninterrupted transport-level session continuity for a TCP session. For some categories of edge applications (e.g. video streaming), the client and server applications may be able to maintain a seamless user experience despite interruptions in connectivity through application domain-specific algorithms. For other categories, e.g. gaming applications, such interruptions may affect the user experience significantly.

An OP shall be able to provide a consistent user experience during the mobility of the user device.

As general principles, the following are essential requirements to provide application Session Continuity in the OP architectural model:

- An OP shall rely upon the 5G core network capabilities for supporting Service and Session Continuity in mobile networks to deliver the OP's application Session Continuity services.
- An OP, based on the network capabilities for supporting Session Continuity, shall expose abstract Session Continuity models towards the Application Provider or an Aggregator over the NBI interface
- An OP shall interact with the mobile network and the 3GPP-defined standard services over the SBI-NR interface to synchronise with the 5G core network procedures to support application Session Continuity.
- When required, an OP shall inform UCs over the UNI interface about the prior indications of application IP address change events, post notifications of application IP address changes, and the new location of the Application Instances after application session relocation.

Note: The abstract service and session continuity modes corresponding to 3GPP defined SSC modes 1, 2 and 3 are typically described as “IP Preservation”, “Break-Before-Make”, and “Make-Before-Break" respectively.

5.2.4.6.2 Access technologies support for application session continuity

The SSC capabilities in a mobile network depend considerably on the type of the radio network, i.e. 4G, 5G, Wi-Fi etc. and on the support for Session Continuity defined for these networks in standards like the 3GPP's. It also depends on whether the operator has deployed such services for their subscribers.

Depending on their access hardware and software capabilities, UEs may attach to mobile networks following the access policies configured for the subscription and network capabilities deployed and operated by the mobile service providers.

The UE may perform its network attachment to the radio networks available in the UE's location. Those networks could be broadly segregated into 3GPP or non-3GPP (trusted or untrusted) access technologies. As part of the SIM configuration, an Operator can configure their preference for the selection of access technologies to the UE. The network to which a UE is currently attached would also determine the level of support available for session continuity in that network what an application can expect.

Handovers and associated SSCprocedures may be triggered by the mobility of UEs within the mobile network coverage area. These procedures or capabilities are defined for devices attached to a mobile network using 3GPP's 5G radio technologies. Table 27 describes the SSCthat an OP shall support in the current version of this document when 5G capable UEs attached to a 5G radio network are served by the 5G core network (i.e. 5G Standalone (SA)).

	Support in Home NW	Support in Visited NW
5G to/from 5G	Supported	Supported
5G to/from 4G	Supported	Supported
5G to/from non-3GPP trusted access	Not Supported	Not Supported
5G to/from non-3GPP untrusted access	Not Supported	Not Supported

Table 27: Access Technologies Supported In OP Architecture For Application Session Continuity

Note: For the above scenarios where an OP supports application SSC, the cases involving mobility from one Operator network to another Operator’s network is for future study.

Note: For non-3GPP access technologies, the SSCcapabilities continue to evolve and, therefore, are not supported.

5.2.4.6.3 Network and OP responsibilities for application session continuity

Assuming a subscriber actively engaged with an edge application starts moving in a network operated by their home Operator, this may result in network procedures to reselect a network attachment point for the UE to maintain agreed QoS levels.

As described in section 2.2.7.3, "Requirements for Application Session Continuity”, the mobile core network may activate SSC mode (starting with 3GPP Release 15 for 5G's Standalone Architecture (SA)) specific procedures based on the user's subscription and the network policies defined by the Operator.

Due to the SSC mode procedures execution in the core network, the following events may occur that require external entities to take application-specific actions such as triggeringapplication session context relocation to a new target Cloudlet:

1. For SSC mode 1, which could be named as “IP preservation mode”, in which the network may assign a different attachment point while keeping the IP address for the UE unchanged:

- The mobile network may assign SSC mode 1 to a PDU session considering factors such as user subscription information, operator configured local policy, an indication from authorised Application Functions (AF), e.g. an OP, if a PDU session involving access to edge application cannot be relocated (application relocation indication) and the address should be preserved
- In this situation, the mobile core network may be unable to provide the desired QoS needed by the application (as defined in Application Manifest).
- In such cases, an OP should have access to information related to the network attachment point change (user plane reconfiguration) event and the QoS that the network can provide for the UE PDU session
- A UC application may need to adapt its behaviour according to the QoS that the network can deliver end-to-end. If the mobile network cannot maintain the requested QoS during the mobility period, then based on QoS change notifications from the NEF, an OP can timely notify the QoS change events to UCs over the UNI interface. Application Clients can gracefully adapt their behaviour using such notifications, e.g. switching to a lower frame rate for video streaming. An OP may also allow edge applications to request to be notified about this kind of event, allowing them to take appropriate actions to provide consistent quality of experience to their users.
- An OP shall also publish over NBI the monitoring information regarding the change in QoS for the application sessions

2. For SSC mode 2, which could be named “Break-Before-Make” mode, the network may change the existing user plane and assign an optimum user plane in the new location of the UE, which would cause the IP address of the UE to change. It may be possible for the mobile network to provide the desired QoS as needed by the application without preserving the session continuity

    - An OP should have access to information related to the user plane change preparation event for the UE PDU session in the mobile network via notifications related to user plane change events requested over the SBI-NR interface
    - An OP could use these events to notify the UCs to be prepared for a possible connectivity break over the UNI interface.
    - An OP shall also provide notification services on the NBI interface to edge applications for these events to enable edge applications and UCs to prepare for a possible application session context relocation.
    - An OP on receiving a network event on the SBI-NR interface for a possible session connectivity interruption for an application session shall perform the application session state/context relocation function to minimise the connectivity disruption time. The OP may use the following information to select an adequate target edge cloud to host the new application instance.

        - Application provider criteria (see section 3.5.1.4)
        - Application data privacy policies
        - Operator defined policies, e.g. cost functions associated with edge clouds
        - Location information on the UE received through the SBI-NR interface
        - Edge sites and available resources at the UE's location as received through the SBI-NR interface
        - The Application Session Continuity mode of the UE PDU session

- Based on these criteria, the OP shall attempt to select a Cloudlet where a new Application Instance for the session can be launched or an existing Application Instance of the application can be assigned.
- The OP shall launch the Application Instance at the selected Cloudlet in the new location of the UE. As per the network configuration, the OP shall also generate the traffic steering rules to route the application traffic from the UC PDU session to the new UE Cloudlet where the application instance is created.
- If an Application Instance is already available, the OP may use that instance's information to generate the traffic steering rules for the UC PDU session in the selected edge site.
- The OP shall interact with the cloudlet over the SBI-CR interface to perform the required functions, e.g. application instance creation, and shall record the status of operations performed
- An OP shall provide capabilities over the NBI interface for Application Providers or Aggregators to perform the application session/context relocation functions
- An OP shall indicate the completion of the application session state/context relocation procedures to the core network via the SBI-NR interface and the NEF (as per 3GPP NEF specified procedures).
- On receiving the UE user plane change progress indication over the SBI-NR interface, the OP, in response to the network, shall provide the SBI-NR API parameters, e.g. description of the traffic steering rules for the application traffic, QoS reference, a period of time or a traffic volume, etc. to the mobile core network over the SBI-NR interface to steer the UC traffic towards the new Application Instance
- The OP shall provide the new Application Instance communication endpoints to the UC over the UNI interface

Note: It is important to note that 3GPP specifications do not put any time constraints for external AFs to respond to the core network notifications and acknowledge the application's readiness for the session/context relocation. Therefore, any OP implementation shall follow the behaviour described in 3GPP specifications and treat the acknowledgements towards the core network independent of any specific OP procedures, e.g. session/context relocation, application instantiation etc.

Note: It is important to note that due to user mobility in mobile networks, events like a user plane change may result in a UE IP address change managed by the core network. Similarly, circumstances outside the mobile network (e.g., edge application relocation to a new target Cloudlet) could change application endpoints, i.e., an IP address change managed by the OP. Any implementation of an OP that supports application SSC will need to consider such aspects from both the application's and UC's perspective.

3. For SSC mode 3, which could be named “Make-Before-Break” mode, the network may, similarly to SSC mode 2, assign a different user plane to UE due to its mobility. This user plane change would cause a modification of the UE's IP address later. However, in this mode, UC application traffic can still reach the previous application instance over the existing connection in the meantime.

- It may be possible for the mobile network to provide the desired QoS as needed by the application and more time for the OP to create new edge Application Instances in the target Cloudlet and synchronise any application states for stateful applications. An OP shall have the mechanisms to minimise the time simultaneous sessions with old and new Application Instances remain active to optimise the network and compute resources.
- An OP shall indicate completion of all the application relocations tasks to the mobile network over the SBI-NR interface, allowing the network to reclaim the network resources of the previous session and start steering the UE traffic towards the new instance.

SSC Modes	Key Characteristics	Capability Name	Network Capability Description	Key Mobility Events Handling in OP
1	UE IP Preserved	IP Preservation	Preserve UE IP agnostic to user location change for active sessions	•Request notifications on UE Mobility events over SBI-NR •Monitor application session QoS •Enforce Application Provider policies and Operator defined policies
2	UE IP Not Preserved	Break-Before-Make	PDU session modification with a new PDU Session Anchor(PSA) and IP connectivity disruption	•Request notifications on UE mobility events over SBI-NR •Coordinate OP activities, e.g., application session relocation in synchronism with mobile network •Enforce Application Provider policies and operator-defined policies •Notify session continuity events to NBI and UNI interface •Assist edge apps to prepare for and handle short disruption in session continuity via service APIs
3	UE IP Not Preserved, Concurrent Sessions	Make-Before-Break	PDU session modification with a new PDU Session Anchor(PSA) and with simultaneous connectivity with the previous	•Request notifications on UE mobility events over SBI-NR •Indication of simultaneous connectivity temporarily maintained for the source and target PSA based on app criteria

SSC Modes	Key Characteristics	Capability Name	Network Capability Description	Key Mobility Events Handling in OP
			session anchor	•Coordinate OP activities, e.g., application session relocation in synchronism with mobile network •Enforce Application Provider policies and Operator defined policies •Notify session continuity events over the NBI and UNI interfaces •Assist edge applications to prepare for and handle concurrent sessions with edge Application Instances via service APIs

Table 28: Summary of OP responsibilities for supporting 3GPP-defined SSC modes

Note: Edge applications should be able to communicate with external applications over the internet. An Application Provider might use this to coordinate or synchronise edge application states. An OP, in such cases, will need to provide the capabilities like controlled access to the internet for edge applications and managing and automating the corresponding functions, e.g., application traffic routing and QoS control etc.

Note: As a possible approach, an Application Provider can also choose to deploy application instances statically and use the OP provided network services to replicate application state information or use another application hosted outside of the OP for this purpose. An OP would need to offer services to edge applications to receive events, e.g., UC IP address change event.

Note: It is expected that to support application Session Continuity in 5G mobile networks, the Operator would need to support features like UL CL (Uplink Classifier) or IPv6 multi-homing as defined by 3GPP for the UPF

Note: Based on some of the events on the SBI-NR interface, e.g. location monitoring events, QoS status notification events etc., an OP may determine the level of QoS provided by the mobile network to application sessions against the QoS level requested by the application. In such cases, the OP may initiate the user plane relocation (e.g., by using Traffic Influence APIs) services on the SBI-NR interface. Possibly this may result in the triggering of session mobility procedures in the mobile network

5.2.4.6.4 5G Core Network managed informational elements required by OP

To support application Session Continuity for edge applications, an OP shall support various procedures defined by 3GPP for an external application function (AF). An OP in the role of AF would need to manage network events and notifications over the SBI-NR interface (NEF

APIs) and enable orchestration of edge Application Instances in target Cloudlets and synchronisation of the associated application states to provide application Session Continuity.

An OP will need access to network location information associated with the UEs typically managed by the mobile network. Network location information will enable the OP to correlate network events with the edge deployment topology and enabling functions like target edge cloud selection, generating traffic steering rules, applying data privacy rules for information protection etc.

To facilitate access to the managing function of the Cloudlet deployment topology, an OP should use some of the following UE network location information that the 5G mobile core network uses to track the UEs in the mobile network coverage area (not an exhaustive list):

- Cell-IDs,
- Tracking Area Codes(TACs),
- Registration Area (RA),
- Geo Location (Latitude/Longitude),
- Data Network Access Identifiers (DNAIs),
- Data Network Name (DNN),
- Single – Network Slice Selection Assistance Information (S-NSSAI).

An OP should be able to correlate the current location of the UE received over the SBI-NR (NEF API Notifications, e.g. event monitoring, User Plane change events etc.) with the Cloudlets in the UE's location to enable selection of an adequate Cloudlet to serve the UE by using this network topology information associated to the Cloudlets.

An OP shall also use the Application Provider's criteria for determining the adequate Cloudlet for the dynamic selection of a target Cloudlet to serve UEs in motion. The OP shall ensure that the agreed level of QoS for edge application sessions with the UCs are maintained irrespective of the device mobility.

Note: The information mentioned above is indicative and has been taken from 3GPP specifications on the NEF APIs as a possible approach to relate network resources with edge clouds located outside of the core network.

5.2.4.6.5 Edge Applications responsibility in Session Continuity Process

UE mobility may trigger the mobile network to initiate the user plane change process. It may also result in the OP starting a synchronised application relocation process for edge applications.

While an OP prepares for the possible application relocation process based on the network events received over the SBI-NR interface on a particular PDU session, the edge application may also require access to some information for performing application-specific functions to support relocations. Some of the information that an OP can expose to edge applications can be

- Target Application Instance information
- Old and new IP address of the UEs in case of User Plane reselection
- Application communication endpoint (IP, Port, Protocol) on the target edge node

- Requested and achieved QoS level information
- Current access network and access network change events
- UE Location events based on UE privacy permission

Note: It is expected that the UCs should be able to detect the change of the IP address assigned by the 4G/5G core network to the user device due to the mobility events using application-level logic e.g., connection reset events on existing application sessions in client applications or in UE APIs etc.

Note: The use of Network Address Translation (NAT) by the MNO in mobile networks may result into a mapping of the UEs' private addresses to a different set of public IP/port combinations that are visible to external applications. This may pose additional complexities to OP functionalities. Any consideration for NAT is for further study in a future version of the PRD.

5.2.4.6.6 Application Session Continuity Support for Roaming Users

An OP shall support the application session continuity for roaming users when they roam into locations served by a Partner OP.

To provide session continuity services for roaming users, the Partner OP shall provide the following information to the Leading OP over E/WBI interface (Not an exhaustive list),

- Supported Abstract Session Continuity Modes (as described in section 5.2.4.6.1)
- LBO Capability
- Supported Service APIs
- Relocation Failure Events
    - Application relocations denied by OP
    - Application relocation execution failures and causes

Note: Some network capabilities and applications relocation event monitoring information shared by a Partner OP over E/WBI can be published over the NBI to inform application providers on the Partner OP capabilities before deploying the applications. This information can be helpful if the applications are sensitive to session continuity capabilities supported by the Partner OP.

5.2.4.6.7 Application Session Continuity Support for handovers between 4G and 5G

An OP shall support the application session continuity for application clients and edge applications when the user devices support both 4G and 5G capabilities. The mobile network may provide the interoperability between 4G and 5G for UEs that support both 5GC NAS and EPC NAS and may also offer the network capability exposure APIs based on combined SCEF+NEF via CAPIF (see section 5.1.4.2.1).

An OP shall request notifications on the SBI-NR to be informed about the expected level of support for network services or network capability exposure APIs. Based on the UE's serving network, the OP shall use these APIs as per the level of support available.

For devices attached to the 5GC with SSC mode 1 or in the EPC with an IP preservation session, an OP shall request notifications on the SBI-NR for the Core Network (CN) type

(EPC, 5GC) change events for the PDU sessions used by applications that are mobility sensitive. An OP shall interact with the mobile network to monitor the QoS level provided by the mobile network for a given PDU session.

Depending on the monitored QoS level notified over the SBI-NR, an OP may provide the Application Provider requested Alternative QoS References to the mobile network over the SBI-NR. These Alternative QoS References are defined in relation to a CN type. An OP shall determine the set of QoS references according to the CN type that the UE is attached to. Also, based on the OP receiving notifications related to QoS level change events for a PDU session, QoS level information as received over the SBI-NR shall be made available over the NBI to the Application Providers.

In the scenario when a user with an application session in a 5G network with SSC mode 2 or 3, is handed over from the 5G to a 4G network it may not be possible for an OP to ensure seamless session continuity. To support these scenarios, the SBI-NR should provide early notifications during the 5GC to EPC handover initiation process. The OP shall use these notifications to inform applications that requested those notifications of the upcoming handover allowing those to take appropriate application-level actions to ensure the most optimal user experience. Also, an OP may request over the SBI-EIN interface to configure the connectivity between the application instances on source and target cloudlets for synchronizing session states.

Note: The various 3GPP access technologies and core network together provide QoS models that may not be harmonised in an abstract QoS model to interface with a mobile core network on the SBI-NR interface in an access network-agnostic manner.

5.2.4.7 Network slice provisioning for an end user

An OP shall offer a centralised management plane for the Operator to manage end user's profile data and to map it to the corresponding AP ID and (Edge) Application ID.

Note: S-NSSAI, DNN list and NSI are network slice related information associated with the end user managed by the OP in the case of a 5G network.

5.2.4.8 Security Requirements

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

1. The OP shall provide security mechanisms to counteract attacks on the OP's Southbound Interfaces (i.e. the SBI-CR, the SBI-NR, the SBI-CHF, the SBI-EIN and the SBI-OAM) aiming to prevent data availability, such as DoS attacks.
2. The OP shall protect Personally identifiable information (PII) of subscribers while in storage.
3. Privacy and tracking protection: Information originating in the UE should be protected for integrity, privacy, confidentiality, nonrepudiation.
4. The OP shall provide an authorisation mechanism for the UNI requests to grant access to only the previously authorised services. The authorisation mechanism shall ensure that the EC is authorised to access the provisioned services and that the UE can access the edge data network.

5. Given the external exposure of the UNI, the OP shall provide security mechanisms to counteract attacks on the OPs UNI aimed to prevent the availability of the interface, such as DoS or DDoS attacks.

5.2.5 User Client

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

6 Realisation of the OP

A consistent set of standards is required to realise Operator Platform (OP) services supporting federation among operators. These standards must be well-supported by Standards Development Organisations (SDOs) and cover the requirements identified and documented in this current document. Next to those, Open-Source communities (OSCs) exist with API specifications and software blueprints that may approach the OP requirements. Details on which specifications are relevant for the realisation of the OP's requirements, are provided in the PRDs detailing the APIs used for the realisation of the different OP interfaces (e.g. GSMA PRDs OPG.03 for the SBI-NR, OPG.04 for the E/WBI, etc.).

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. All the documents are available for the public at the ETSI site https://www.etsi.org/committee/1425-mec.

A.1.2 ETSI ISG MEC specifications relevant for the architecture and support of mobility

- ETSI ISG MEC 003: The framework and reference architecture describing application placement on an edge compute resource.
- 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.
- ETSI ISG MEC 012: Radio network information API provides specifications related to radio network events and fetching them.
- ETSI ISG MEC 021: Specification provides application mobility service API

A.1.3 ETSI ISG MEC specification defining interaction with the UE

- ETSI ISG MEC 016: UE Application Interface

A.1.4 ETSI ISG MEC specifications relevant for Network Capability Exposure

- ETSI ISG MEC 014: UE Identity API
- ETSI ISG MEC 009: General principles for MEC service APIs
- ETSI ISG MEC 015: Bandwidth management API
- ETSI ISG MEC 013: Specification describes the location API
- ETSI ISG MEC 029: Specification provides fixed access information API
- ETSI ISG MEC 044: Study providing potential requirements and enhancements to the MEC system needed to support MEC Application Slices.
- ETSI ISG MEC 045: Specification providing QoS Measurement API (including predictive QoS provided by AI/ML components if available)

A.1.5 ETSI ISG MEC activities relevant for federation

ETSI ISG MEC provides various specifications to enable inter-MEC communication. In particular, the MEC architecture defined in MEC003 [18] supports inter-MEC communication, either directly via the Mp3 reference point or via MEC federators. ETSI ISG MEC 040 [42] defines the APIs to support MEC federation.

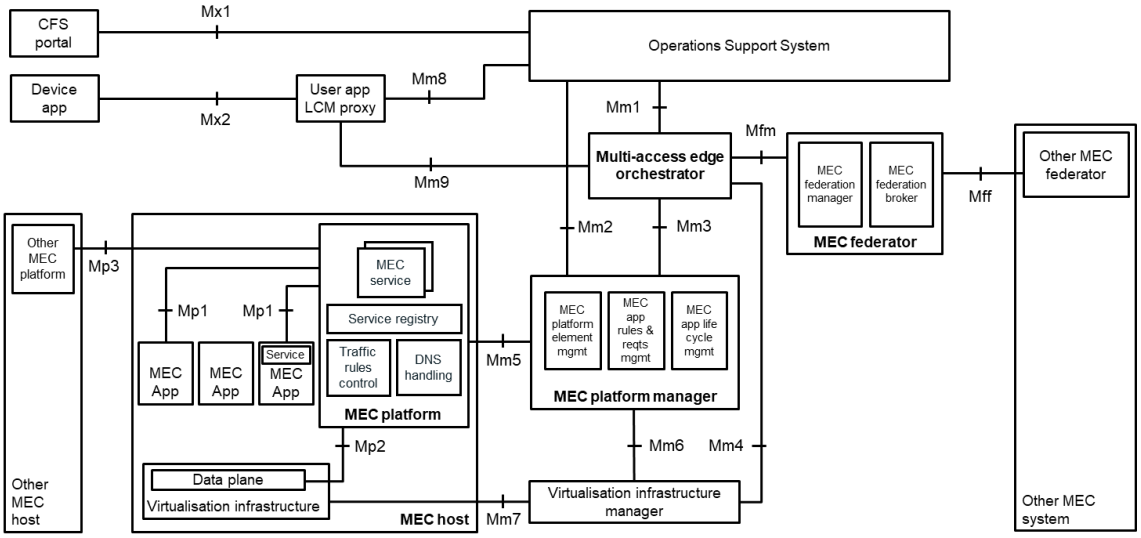


Figure 30: Multi-access edge system reference architecture variant for MEC federation in ETSI MEC003 [18]

A.1.6 ETSI ISG MEC activities relevant for cloudlet interconnection

ETSI ISG MEC 003 [18]: The MEC framework and reference architecture mentions the mp3 interface for the Inter-MEC communication and lists the requirements.

A.1.7 ETSI ISG MEC activities relevant for application LCM

- ETSI ISG MEC 010: Information flows for lifecycle management of MEC applications, and describes interfaces over the reference points to support application LCM
- ETSI ISG MEC 037: specifies the structure and format of a MEC application package and data models of the MEC application descriptors.

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 in TS 23.558 [17]. In clause 6.2 of TR 23.958 [44], 3GPP provides a functional mapping between 3GPP EDGEAPP and GSMA OP.

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 PCF/ 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 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.

GSMA PRDs OPG.03 [40] and OPG.05 [41] provide a detailed mapping of the APIs required to realise the OP's SBI-NR and UNI interfaces to the APIs exposed by the SCEF/NEF and the ECS and EES.

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. The EDGE-1 and EDGE-4 reference points can support similar function(s) as the OP's UNI, providing the EEC (corresponding to Edge/User Client in OP) with the information required to access the edge services [44].
- 3GPP SA6 defines the EDGE-2 and EDGE-8 interfaces for the interactions from the edge platform to the network. 3GPP SA5 also defines more details on the cloudlet management aspects. The EDGE-2 and EDGE-8 reference points can support similar function(s) as the OP's SBI-NR, through which the edge enabler layer (corresponding to the OP) accesses the 3GPP network capabilities and services (e.g. SCEF/NEF).
- 3GPP SA6 defines the EDGE-3 interface for the cloudlets to communicate with the edge platform. The EDGE-3 reference point can support similar function(s) as the OP's NBI, exposing the capabilities of the EES to the EASs hosted on the edge.
- To support the OP's EWBI, 3GPP SA6 defines some interactions over the EDGE-9 and EDGE-10 reference pointsfor the Operator Platforms to communicate with each other. Additionally, some Provisioning Management Services are foreseen to enable the interactions between the ECSP management system of the Leading OP and Partner OP.

Note: The EDGE-9 reference point in the EDGEAPP architecture can be used to discover an EAS from the EES of the partner ECSP for edge node sharing scenarios.

- 3GPP SA5 defines the Nchf interface for charging as specified in [39].
- According to 3GPP SA5, the ECSP management system caters to the requirements of OP's SBI-CR interface.
- 3GPP SA3 defines the security details of all the EDGEAPP interfaces in [45].

A.2.3 3GPP Exposure Interfaces

3GPP SA2 defines the interfaces N33 and T8 for 5G and 4G, respectively, enabling the following APIs:

- 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 an 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 is 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 resources 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, personalised 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 centralised 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 optimisation. 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/Application 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 #10 5.1.4.2.2 #20	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	5.2.4.4 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		5.2.2.1.2					E/WBI		
Data sharing (Data is ‘open’ for use by multiple application providers)		missing	Y			Y but highly filtered			Y
Specialist compute	SBI-CR	5.2.4.3			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 an Operator Platform.

C.1 Relationship with OP and Operator

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

In Figure 31, 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.

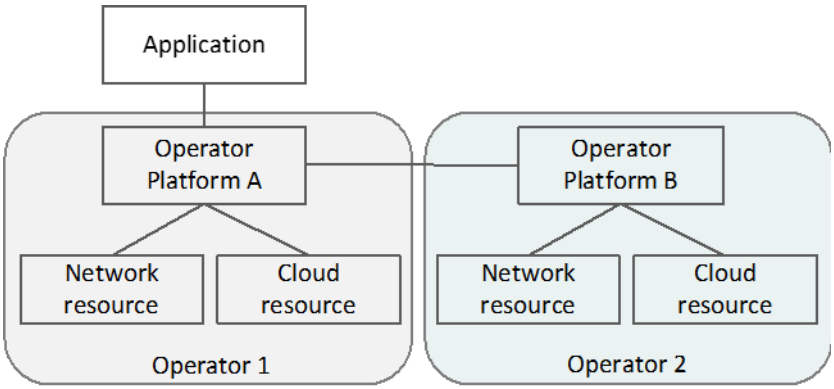


Figure 31: Each operator has an own Operator Platform

In Figure 32, 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.

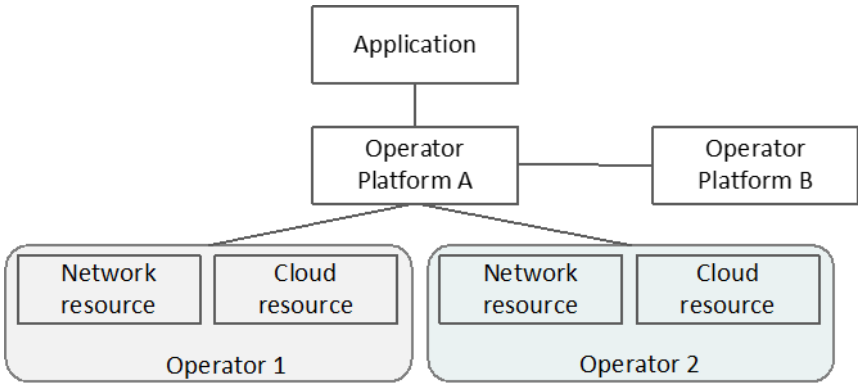


Figure 32: Multiple operators share the same OP

C.2 Relationship with hyperscalers from a single Operator perspective

An Operator can have their own cloud resource and collaborate with a hyperscaler simultaneously. An 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 an OP. First, hyperscalers can be considered enterprise customers to the 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 enterprise customers (i.e. Application Provider 2). In addition, Hyperscaler resources can be available for OP A to offer its enterprise customers (i.e. Application Provider 1).

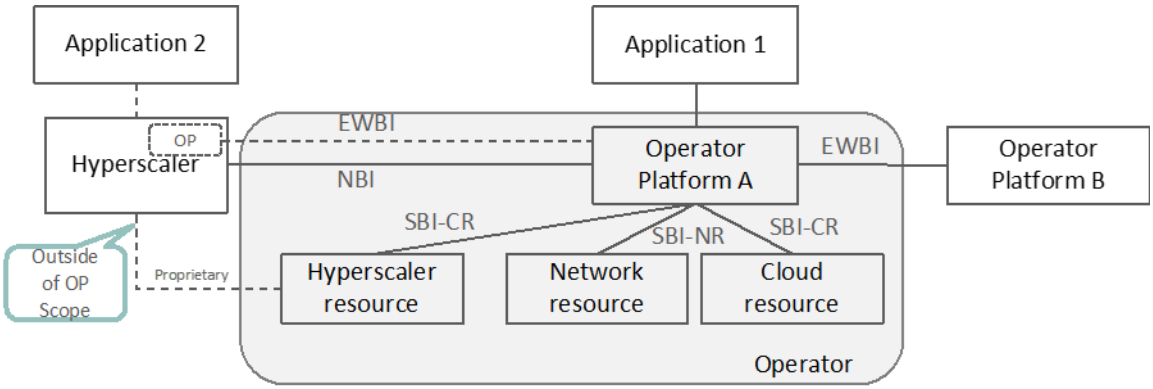


Figure 33: Relationship with hyperscalers

Annex D Aggregation / Marketplace Platform

An Aggregation/Marketplace Platform is a store the services exposed by the Operators via an OP. In addition, there may be OPs that offer additional 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 Aggregation/Marketplace Platform would be to enable Application Providers to discover such additional services and possibly buy them.

The following are potential functionalities supported by Aggregation/Marketplace Platform:

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

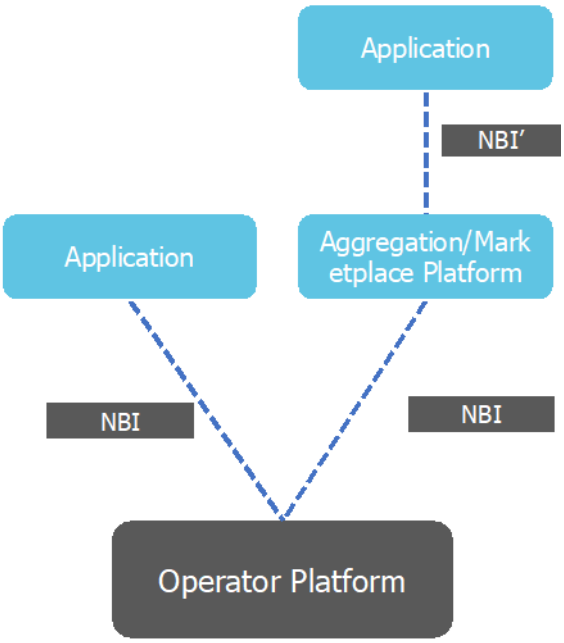


Figure 34: Operator platform with Marketplace

Annex E Operator Platform Security

E.1 Introduction

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

This Annex contains informative text that supplements and supports 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. As not all threats can be mitigated through the OP's architecture and interface definitions, section E.5 of this Annex provides guidance for the implementation, deployment and operation of an OP and the edge resources that it exposes.

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 characterisation. 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 characterisation 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 characterisation 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 36, 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 virtualisation, 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 characterise 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 characterisation 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 characterisation.
- 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 has studied the security aspects of edge computing support in the 5G Core (e.g., [20], [16]) and has specified the main security aspects in [45]. 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 – GSMA OP and 3GPP EDGEAPP – GSMA OP mapping (see Annex A) to associate the threat vectors to the OP architecture directly.
- Create tables of countermeasures for each of the threat vector identifiers appearing in Figure 36. 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.

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 characterises 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 characterisation provides an initial attack surface characterisation 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 categorises 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.

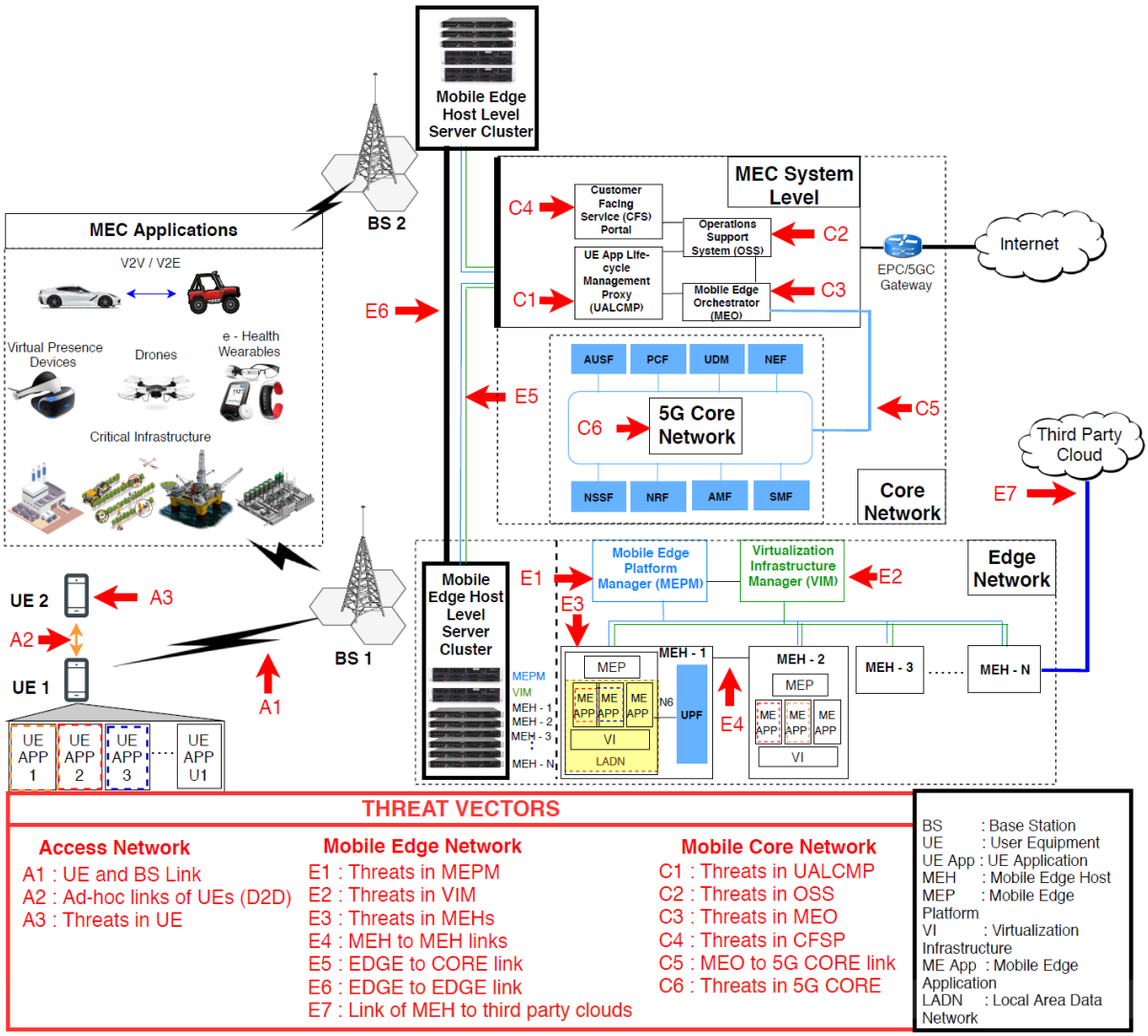


Figure 35: ETSI ISG MEC Access, Edge and Core Threat Vectors (from [15])

Table 29 summarises 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 summarised 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 [15]
A2	Ad-hoc connectivity between UE [15]
A3	UE vulnerabilities [15]
AR1	Network Slicing (NS) [15]
AR2	Traffic Steering [15]

Threat Vector (TV) ID	Description
AR3	Service Migration [15]
AR4	Mobility Management [15]
C1	User Application lifecycle management (LCM) Proxy (UALCMP) [15]
C2	Operation Support System (OSS) [15]
C3	Mobile Edge Orchestrator (MEO) [15]
C4	Customer Facing Service Portal (CFSP) [15]
C5	Connectivity of MEO and 5G Core Network [15]
C6	5G Core Network [15]
E1	Mobile Edge Platform Manager (MEPM) [15]
E2	Virtualisation Infrastructure Manager (VIM) [15]
E3	Mobile Edge Host (MEH) [15]
E4	Connectivity between MEHs [15]
E5	MEC platform connectivity between Edge and Core [15]
E6	Connectivity between MEC apps operated under hosts at different BTSs [15]
E7	Link of MEH to third party clouds [15]
MEC1	Required signalling for secure inter-MEC systems [22]
MEC2	MEC system discovery supporting authentication, authorisation, identity management, etc. [22]
MEC3	MEC platform discovery supporting authentication, authorisation, identity management, etc. [22]
OTV1	Charging and billing for MEC subscriptions [15]
OTV2	Service impeding/delaying threats [15]
OTV3	Mobile offloading [15]
OTV4	Virtualisation and orchestration of the edge [15]
Privacy	Privacy-related threats [15]
3GPP1	Authentication and Authorisation between EEC and EES – EDGE-1 [20]
3GPP2	Authentication and Authorisation between EEC and ECS – EDGE-4 [20]
3GPP3	Authentication and Authorisation between EES and ECS – EDGE-6 [20]
3GPP4	Edge Data Network authentication and authorisation [20]
3GPP5	Edge Data Network user identifier and credential protection [20]
3GPP6	Transport security for the EDGE-1-10 interfaces [20]
3GPP7	Security of network information provisioning to local applications with low-latency exposure [20]
3GPP8	Authentication and authorisation in EES capability exposure – SCEF/NEF northbound APIs [20]
3GPP9	Security of EAS discovery procedure [20]
3GPP10	Authorisation during edge data network change [20]

Table 29: Threat Vector Descriptions (adapted from [15], [20], [22])

E.2.2 Threat Vectors Identified by 3GPP SA3

3GPP Service and System Aspects (SA) Working Group 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 34.

3GPP 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 30 summarises 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 35 and summarised in Table 29 (a composite table of all threat vectors identified from all sources).

Threat Vector (TV) ID	Description	Location
3GPP1	Authentication and Authorisation between EEC and EES	EDGE-1
3GPP2	Authentication and Authorisation between EEC and ECS	EDGE-4
3GPP3	Authentication and Authorisation between EES and ECS	EDGE-6
3GPP4	Edge Data Network Authentication and Authorisation	edge data network
3GPP5	Edge Data Network User Identifier and Credential Protection	edge data network
3GPP6	Transport security for the EDGE-1-10 Interfaces	EDGE-1 through EDGE-10
3GPP7	Security of Network Information Provisioning to Local Applications with low latency exposure	UPF, AF, NEF
3GPP8	Authentication and authorisation in EES capability exposure	SCEF/NEF northbound APIs, CAPIF
3GPP9	Security of EAS discovery procedure	EAS
3GPP10	Authorisation during Edge Data Network Change	edge data network

Table 30: 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 31 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 [18] network element to provide secure signalling
MEC2	MEC system discovery supporting authentication, authorisation, identity management, etc.	Definition of a new reference point (Mff-fed) to support secure interaction between Federation Managers [18]
MEC3	MEC platform discovery supporting authentication, authorisation, identity management, etc.	Support of authentication, authorisation, identity, etc., to be supported at application level. Possibly different keys, certificates, CAs, from those for MEC system discovery.

Table 31: 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 32 lists vulnerabilities for different domains (RN: Radio Network Operation, RI: Roaming and Interconnection, CN: Core Network, EC: Edge Computing, and CC: Container Controls) in the “threat vector” summary form. This table nor Table 34 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
FS3	Unspecified intrusion into or disruption of network	RN-003
FS4	Unauthorised 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	Unauthorised access to interconnect network elements	RI-002
FS8	Need for roaming log information	RI-003

Threat Vector (TV) ID	Description	[14] reference
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 subscriber identity on network	CN-004
FS13	Unsolicited messaging traffic to subscriber	CN-005
FS14	Unconsistent 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 virtualisation 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 32: Threat vectors identified in [14]

E.3 OP Threat Vectors and Countermeasures

Figure 36 depicts the threat vectors identified in the OP architecture

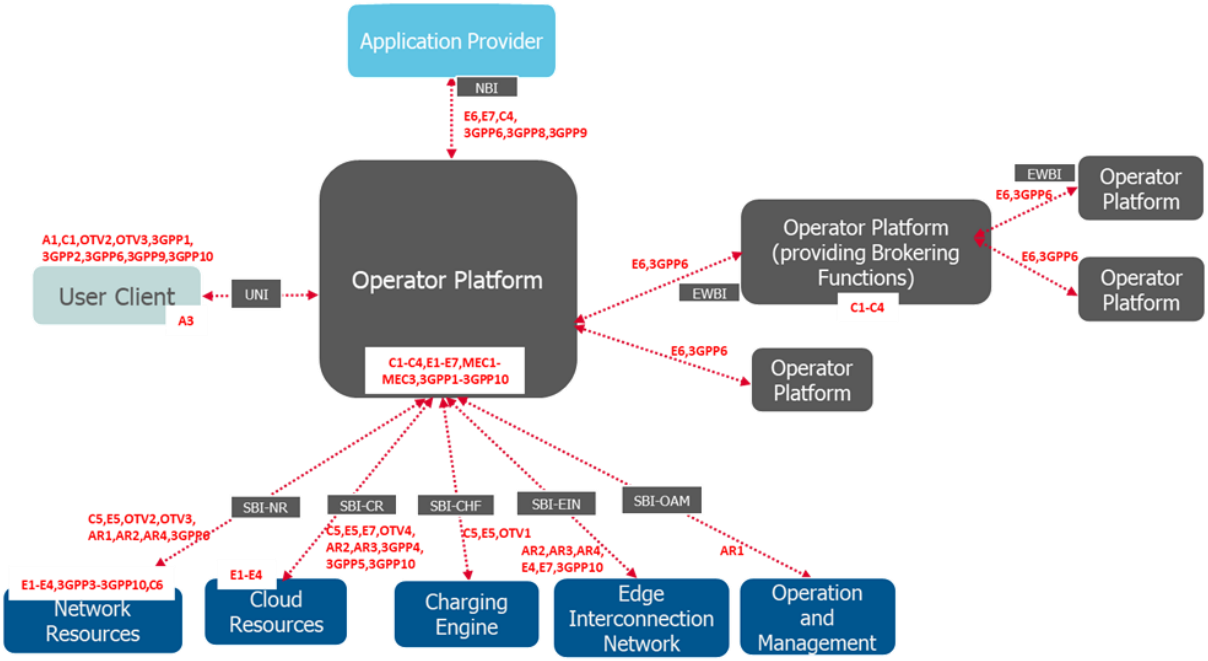


Figure 36: OP Threat Vectors

E.3.1 Access Threat Vectors

According to Figure 36, 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 application 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 application and UC, 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
A3	Security and Privacy Enhanced (SPE) framework for UEs and intent-based validation policy

Table 33: 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 35 (from [15]). Instead, they were added in Figure 36.

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. 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 categorised as architectural and so appear in this table. Because of the large number of items identified in [14], they are summarised by their identifiers in Table 32.

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
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

Threat Vector (TV) ID	Countermeasure Recommendation
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	Virtualisation Controls, FS-30 – FS-33
AR1	NS: Network Services Controls, [14] 2.2.8

Table 34: 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 may map onto the functional levels depicted in Figure 4, 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	Countermeasure Recommendation
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 virtualised 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 35: 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 E/WBI.

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 36: 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 35. Instead, they are provided in Figure 36.

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 (see clause E.4 for more details).

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)
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

Threat Vector (TV) ID	Countermeasure Recommendation
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 37: 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	Authorised and Curious Adversaries
P5	Computational Offloading privacy threats
P6	Service Migration privacy threats.

Table 38: Privacy Threat Vectors (derived from [15])

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

Privacy Objectives	Recommendations
O7	Optimal utilisation of UE resources with embedded privacy-enhancing mechanisms
O8	Comply with GDPR legislation.

Table 39: Privacy Objectives and Recommendations (derived from [15])

Some general privacy considerations are provided in Annex K.

Some privacy-preserving solutions specific for ETSI 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 Recommendations from 3GPP

3GPP has specified the security aspects for supporting edge computing in [45]. Considering the functional mapping presented clause 6.2 of TR 23.958 [44], it is possible to list the following recommendations:

- For the OP's UNI interface (provided it is mapped to the 3GPP EDGE-1/4 reference points), it should support the use of HTTP/2 with "https" URIs as specified in RFC 9113 [47] and RFC 9110 [48]. Authentication between the UC and the OP shall be done during the execution of the TLS handshake protocol. Server side certificate-based TLS authentication shall be supported [45]. Authorization based on OAuth 2.0 shall be supported [45].
- For the OP's SBI-NR (provided it is mapped to the 3GPP EDGE-2/8 reference points), if NEF APIs are meant to be used, then security aspects including the protection of the NEF-AF interface (and support of CAPIF) as defined in clause 12 of TS 33.501 [46] shall be considered: mutual authentication based on client and server certificates using TLS. TLS shall be used to provide integrity protection, replay protection and confidentiality protection for the interface between the NEF and the AF. The support of TLS is mandatory. The NEF shall authorize the requests from an AF using the OAuth-based authorization mechanism.
- For the OP's NBI interface (provided it is mapped to the 3GPP EDGE-3 reference point), it should support the use of HTTP/2 with "https" URIs as specified in RFC 9113 [47] and RFC 9110 [48]. Mutual authentication shall be supported over the NBI (EDGE-3). TLS shall be used.

- For the OP's EWBI interface (provided it is mapped to the 3GPP EDGE-9/10 reference points), mutual authentication (among OPs) should be supported. TLS shall be used. Authorization among OPs shall be based on local authorization policies. It should support the use of HTTP/2 with "https" URIs as specified in RFC 9113 [47] and RFC 9110 [48].

It should be notice that TLS is recommended to be used to provide integrity protection, replay protection, and confidentiality protection for OP interfaces.

E.5 Guidance for the implementation, deployment and operation

Some threats identified in this Annex cannot be mitigated through the OP's architecture and interface definitions. Therefore, this section provides guidance for the implementation, deployment and operation of an OP and the edge resources that it exposes. The following guidance is to be taken into account at a high-level:

1. The implementation and deployment of an OP needs to use 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.
2. An OP implementation needs to 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.
3. An OP implementation is recommended to support telemetry for intrusion detection.
4. An OP deployment and its operation are recommended to follow best practices for DevSecOps (i.e., the practice of introducing security practices into DevOps), as described in GSMA FS.31 [14].
5. An OP implementation needs to employ telemetry and analytics to detect and report application security policy violations at runtime to localise and isolate malicious application behaviour.
6. An OP implementation needs to 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.
7. An OP implementation is recommended to 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.

8. An OP implementation is recommended to 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.

9. An OP implementation is recommended to enable resource isolation, sharing authorisation, and residual data clean-up to protect shared network resources/slices from tampering and data theft.

10. An OP implementation is recommended to employ message filtering of HTTP control plane signalling and firewall configurations to protect network resources from spoofing attacks from roaming interconnections.
11. An OP deployment is recommended to enable security audits on the access privilege management to avoid identity theft or fraud.
12. An OP implementation is recommended to employ secure storage of account credentials to avoid identity theft or fraud.
13. An OP implementation needs to employ secure initialisation and secure configuration data storage to avoid the exploitation of network configuration data weaknesses.
14. An OP deployment should provide hardware root-of-trust based tools to guard network configuration status.
15. An OP deployment is recommended to support centralised and unified log management to protect from any tampering, whether malicious or inadvertent,
16. An OP implementation is recommended to support the automation of security operations.
17. An OP implementation needs to provide secure tracing and logging of charging and billing data requests.

The following guidance is to be considered for the edge resources:

1. Services, processes, and tenants running in containers and virtual machines, and their data, need to be protected.

Note: Approaches to protecting them include process isolation via name-spacing or hypervisor controls and trusted enclaves.

2. The Cloud Resources need to provide security mechanisms to prevent attacks from containers or VMs, of which Docker or VM Escape attacks are examples.
3. The Cloud Resources need to provide security mechanisms to counteract attacks on the SBI-CR aiming to prevent data availability, such as DoS attacks.

Annex F 5G Core Network Application Session Continuity Enabler Services

Native support for enabling edge computing in 3GPP based networks is specified starting with the release 15 of the 3GPP specifications. 3GPP has introduced requirements for various network capabilities to support application session continuity.

As per the 3GPP standards, various APIs (network capabilities) expose essential network capabilities to external AFs via the NEF. These capabilities can be used to support application Session Continuity in the OP. The OP will require close coordination with 5G core network procedures and will use different services exposed by NEF to achieve that.

Some of the key services (or network capabilities), as specified in 3GPP standards, that can be used to support Session Continuity are,

- Event Reporting: Provides support for event exposure
    - NEF Service: Nnef_EventExposure

  - Allows for configuring the specific events, the event detection, and the event reporting to the requested parties
  - Events may include, e.g. loss of connectivity, Location reporting, Roaming status, etc.
  - Location reporting events may help authorised external AF (e.g., an OP in the role of AF) to confirm the UE location and influence the mobile core network over the SBI-NR to trigger a User Plane change when needed

- AS session with QoS: Requests the network to provide a specific QoS for an AS session

  - NEF Service : Nnef_AFsessionWithQoS
  - Input parameters include a description of the application flows, a QoS reference, an applicable period or a traffic volume for the requested QoS. These can be included in the request to NEF
  - The QoS reference refers to pre-defined QoS profiles which have been configured by the Operator in the core network and which can be used by an external AF to request a specified QoS for application sessions
  - An OP can also infer from QoS status notifications from NEF if the requested QoS requirements provided by the Application Provider are not being met. In that case, the OP may initiate a user plane relocation (Traffic Influence APIs) via the NEF APIs (and the SBI-NR) to request the 5G Core to start the user plane reselection process. Possibly this may result in the triggering of session mobility event in mobile network
  - The end-to-end QoS requirements for an edge application are expected to be known by the Application Providers. They should be able to select the QoS profiles offered by OP, which can provide a good quality of experience (QoE) to the users of the applications. At the same time, subscribers consuming the edge applications could have a subscription plan with their home operator, defining the QoS entitlement that they may expect from the network.

- Traffic Influence: Provide the ability to influence traffic routing

  - NEF Service: Nnef_TrafficInfluence
  - The request to the NEF may include parameters e.g.

    - The IP address of the UE, if available, GPSI, DNN, traffic filtering information, a list of DNAI(s), N6 traffic routing information
    - Indication of application relocation possibility, AF acknowledgement to be expected, Early and late notifications about UP path management events

  - External Application Functions, e.g. OP, in the role of AF, need to provide various parameters as indicated above in SBI-NR interface APIs to the mobile core network
  - Some of the parameters, e.g. DNAIs, DNN etc., may need to be configured by the MNO to OP for setting up the network topology information

- Chargeable party: Requests to become the chargeable party for a data session of a UE

- NEF Service: Nnef_ChargeableParty
- External entities like an OP in the role of an AF can initiate requests towards the 5G core network via the NEF containing parameters, e.g. UE address, description of the application flows, Sponsor Information, Sponsoring Status, etc.
- QoS notifications containing information about application session quality may be used by an OP to derive an average QoS level offered for an application session that may have experienced multiple session relocations across Cloudlets due to device mobility
- As the QoS for the application sessions at different Cloudlets may not always have the same level, the QoS data collected via these notifications can be used by an OP to profile the QoS distributions across Cloudlets

Note: Insights collected from the QoS distribution profiles can potentially be used for different purposes, e.g. optimizing the application placement decisions.

Annex G Client-side mechanisms to control QoS

G.1 Introduction

Edge applications often need low latency and jitter to function properly. For this QoS can be applied from the network by calling QoS APIs for each IP flow where the application needs the traffic to be treated with priority. On top of the network-initiated QoS, additional mechanisms that allow an application to ask for priority from the client side exist, such as URSP traffic categories, L4S tagging or DSCP tagging. This can be done by the application supporting URSP traffic categories, L4S tagging or DSCP tagging. Applications can tag priority IP flows using these mechanisms and depending on the network, support and operator settings might benefit from a better treatment then.

The Application Provider is informed about the supported mechanisms so that the client-server connection between the Application Client and the Application Edge part can select a mechanism that is supported.

G.2 URSP traffic categories

The Operator can provide the end user (UE/device) with the URSP configuration. A UE application can indicate traffic categories for selected flows that will be mapped by the UE to active URSP configurations. It is up to the UE OS implementation how a network slice is selected or access to the network is done. It is important to associate standardised categories of application traffic with the specific connectivity defined by the Operator.

GSMA PRD NG.135 [33] has listed a range of traffic categories that can be applicable to OP services that require specific traffic treatment for specialised (edge) services. The categories would allow to separate normal internet traffic from edge specific traffic. Figure 37 shows an application client that utilises 2 different traffic categories, PDU session A and PDU session B are treated separately with different characteristics according to the traffic category applied. Here e.g. Session A could be for normal priority and session B for real-time interactive traffic flows.

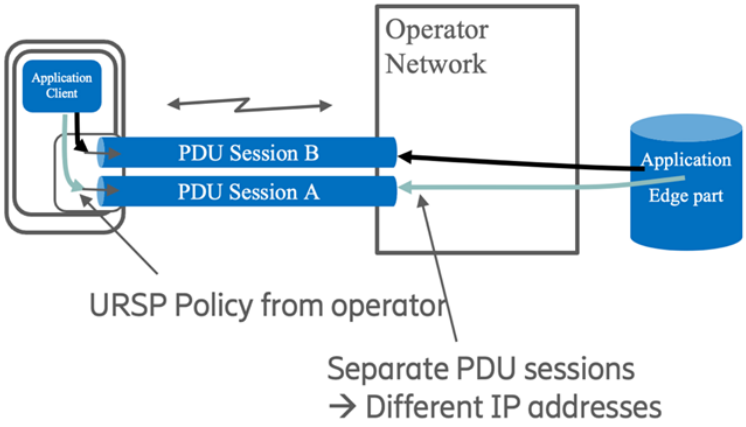


Figure 37: URSP mapping to operator services

G.3 L4S

L4S is a new technology based on an Internet Engineering Task Force (IETF) standardisation, which provides high throughput and low latency for IP traffic, resulting in improved, fast rate adaption management, and reduced network congestion, queuing and packet loss.

L4S relies on ECN (Explicit Congestion Notification) in the IP header to indicate queue build-up in the radio access network to the application. The congestion signals are then managed at the sender and receiver side thanks to scalable congestion control algorithms. In turn, the technology signals to the application server to adjust the application bit rate to meet the capacity of the established communication link. As a result, L4S is effective in delivering a seamless user experience even with variable traffic load and radio conditions. The application tags the communication handle (socket) according to desired QoS treatment (L4S or no tag)

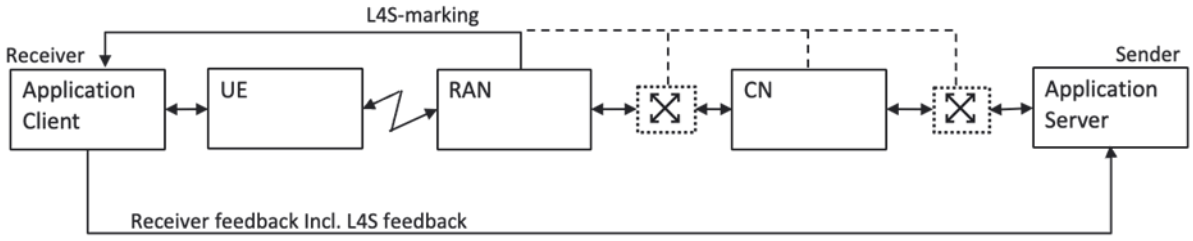


Figure 38: L4S handling

The operator determines how L4S traffic is handled. When the subscriber is allowed to use this mechanism from their subscription, the operator can support L4S using packet filters and prioritise L4S-tagged traffic; see also Figure 39 where the black optimised queue handles only L4S-tagged traffic. The application is expected to respect the ECN bits that indicate congestion and take measures when congestion is detected. This can e.g. by reducing bit-rates for streaming or other measures.

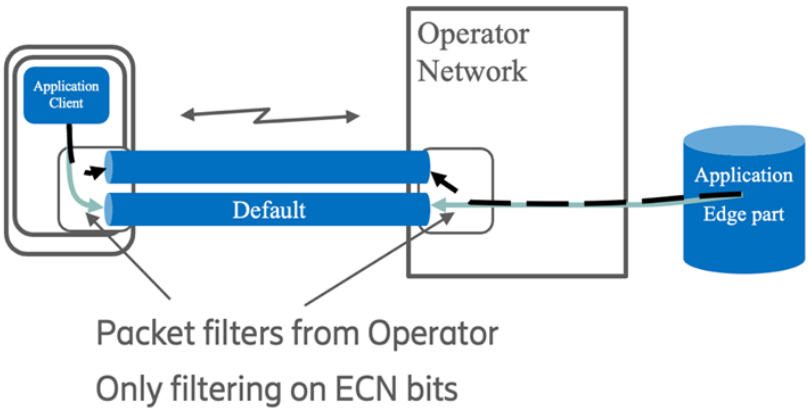


Figure 39: L4S based queue selection

G.4 Other mechanisms

Note: Other client-side QoS mechanisms are for further study. These could include DSCP or other relevant mechanisms.

Annex H Network Communication Service as a Service (NCaaS) realised with NSaaS

This annex explains the relation between the Network Communication Services and Network Slice as a Service (NSaaS).

From the OP point of view, the Application Provider acts as a Communication Service Customer (CSC), while the Operator takes the role of Communication Service Provider (CSP).

Depending on actual scenarios, the Operators and the Application Providers can play one or several roles simultaneously, as depicted in Figure 40.

Note: The OP is assumed to be in the Operator's domain.

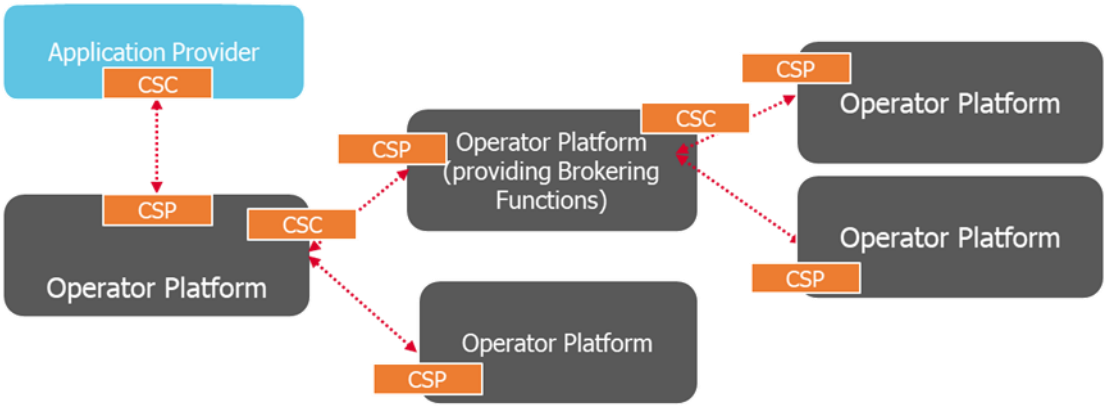


Figure 40: OP and AP roles

A CSP may offer a Network Communication Service to its CSC(s) where they can ask for special treatment of traffic for devices. This service can be realised with the CSP's network

via different mechanisms as long as the special treatment can be applied for CSC's device traffic.

In a 5G network, this can be realised using 3GPP network slice mechanism which can be offered in the form of a service [34]. This can also be realized via specific QoS flows in both 5G / 4G network. In 4G networks, a network slice is typically not available in which case the operator can revert to other mechanisms such as dedicated APNs combined with QoS flow mechanisms. In the latter case also charging, lifecycle management, etc will need to fallback to traditional mechanisms.

A network slice is defined as a logical network that provides specific network capabilities and characteristics [10]. 3GPP has defined standardised network Slice /Service Types (SSTs) in Section 5.15.2.2 of 3GPP TS 23.501 [10].

The CSC communicates its needs by specifying the Network Communication Service that they need. If that needs to be mapped to a network slice in the CSP's south bound, the OP will make use of Network Slice as a Service (NSaaS).

Using NSaaS, Network slice characteristics and capabilities created by CSP are tailored to satisfy the agreed Network Communication Service needs i.e., the network slice is still offered in terms of a Network Communication Service which satisfy certain network characteristics as needed by CSC. Note that the performance requirements of the network slice are based on characteristics of the network i.e.

- Radio access technology
- Location
- Bandwidth
- End to end latency
- Reliability
- QOS
- Security, etc.

There are many ways of creation and management of the network slices using NSaaS. One option is to leverage GSMA-defined Generic network Slice Template (GST) and several NEtwork Slice Types (NESTs) in GSMA PRD NG.116 [32]. GST contains a list of attributes that can be used to characterise a type of network slice/service. A NEST is a selection of GST attributes filled with values. A network slice can be tailored to provide a specific service. Figure 41 below shows GST and NEST in the context of the network slice lifecycle. A CSP can leverage NEST for the SBI realisation of a Network Communication Service. That way they can hide the technical values and complexity of a network slice from the CSC.

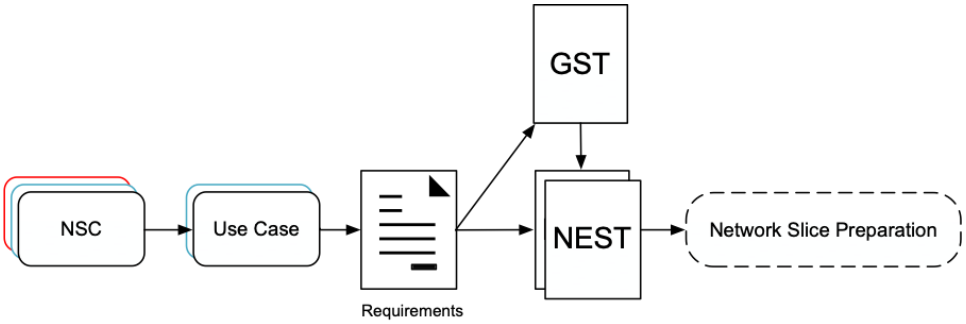


Figure 41: GST and NEST in context of the network slice lifecycle [33]

A network slice could span across multiple domains – access network (or RAN), core network and transport network. A variety of Network Communication Services can be supported by a network slice (see Figure 42 below). The Network Communication Services using a network slice may include 5G eMBB service, AR/VR service, V2X services and many others.

Figure 42 below shows that there can be various combinations of how a Network Slice Instance (NSI) can be deployed in a network. How the NSIs are deployed will depend on the Operator's decision and the service level requirements of each Application Provider. It is also possible for operator to make use of APN/DNN to realize a communication service.

- An NSI can share (or not at all) a certain level of infrastructure resources with other NSIs. Depending on the service level requirements to be delivered by the NSI, it is an operator decision how and where to allocate (and/or dedicate) resources for the network functions serving the NSI.
- An NSI can be shared by different Application Providers. In that case, all the Application Providers using the same NSI will experience the same service level requirements for their services. A single S-NSSAI will be used by all of them.

In both cases, the OP may be able to trigger various operations from commissioning to decommissioning, together with the other lifecycle operations in between (e.g., configuration, activation, modification, and deactivation). It is an operator decision how to land the service level requirements from an Application Provider to a concrete NSI, for example if it needs to deploy a dedicated NSI with a certain service level requirements or it can reuse an existing NSI because the service level requirements from the Application Provider can be delivered with an existing NSI.

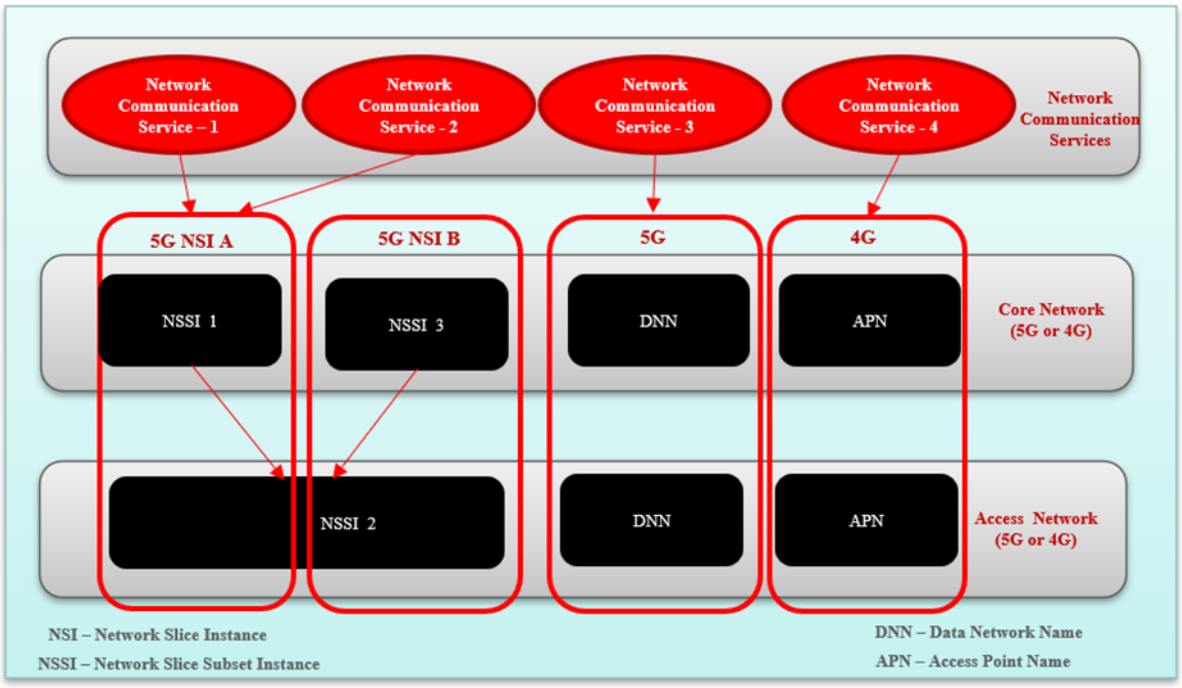


Figure 42: A variety of services provided by multiple network slices [34]

A Network Slice Instance (NSI) is a set of Network Function (NFs) instances and the required resources (e.g. compute, storage and networking resources) which form a deployed network slice.

An S-NSSAI identifies a network slice/service [10] and comprises of:

- **Slice/Service type (SST)**, which refers to the expected network slice behaviour in terms of features and services;
- **Slice Differentiator (SD)** which is optional information that complements the SST(s) to differentiate amongst multiple network slices of the same SST.

H.1 Network slice lifecycle management

A Network Slice, like any other network, has various cycles of its life. The Operator needs to understand the Application Provider's requirements to fulfil them correctly (e.g. using a GST as described above). Once the requirements are well defined and captured in a NEST a new network slice can be created, or an existing slice will be updated whenever there is a requirement for a new Network Communication Service in form of a network slice from a CSC.

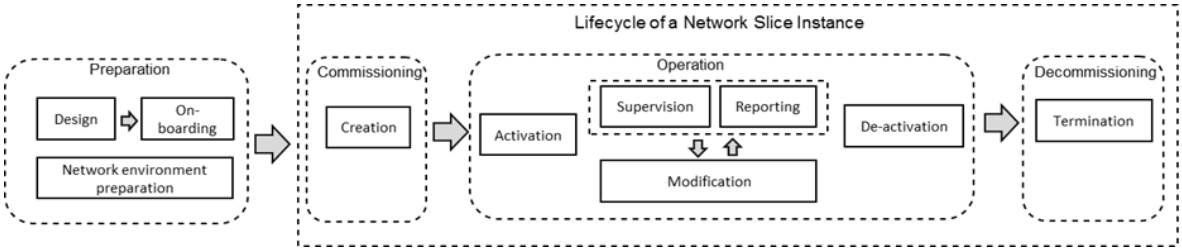


Figure 43: Management aspects of network slicing [34]

There are multiple ways an operator can manage the preparation and operation of network slices. 3GPP has defined four phases of the network slice lifecycle in TS 28.530 Section 4.3 [34] , wheras TM Forum has documented similar in NaaS Transformation IG1224 V13 [49]. It is depending on the Operator's deployment architecture what level of integration is best suited.

- **Preparation, also known as Service Qualification and Service feasibility in TM Forum**: This phase includes network slice design, network slice capacity planning, on-boarding and evaluation of the network functions, list of Network Communication Services supported by the network slice, preparing the network environment and other necessary preparations required to be done before the creation of an NSI. In this phase, the NSI does not exist.

  The Operator will decide on what elements to use in each domain for a particular network slice. It could be possible that two different network slices share some elements (radio, transport, part of the core) but also have other elements that are dedicated to this network slice to meet the requirements of that network slice (as shown in Figure 42).

- **Commissioning or Service Provisioning**: In this phase, the NSI is created. During NSI creation, all needed resources are allocated and configured to satisfy the network slice requirements.
- **Operation**: This phase includes the activation, supervision, performance reporting (e.g. for KPI monitoring), resource capacity planning, testing, problem management, modification, and de-activation of an NSI.
- **Decommissioning**: includes decommissioning of non-shared constituents if required and removing the NSI specific configuration from the shared constituents. After this phase the NSI, does not exist anymore.

Depending on the service offering, the Operator may impose limits on the NSaaS management capabilities exposed to the Application Provider. There may be various levels of the NSaaS management capabilities, from managing only specific characteristics (e.g. bandwidth, end-to-end latency, QCI) to managing the network slice lifecycle (e.g. activation, decommissioning).

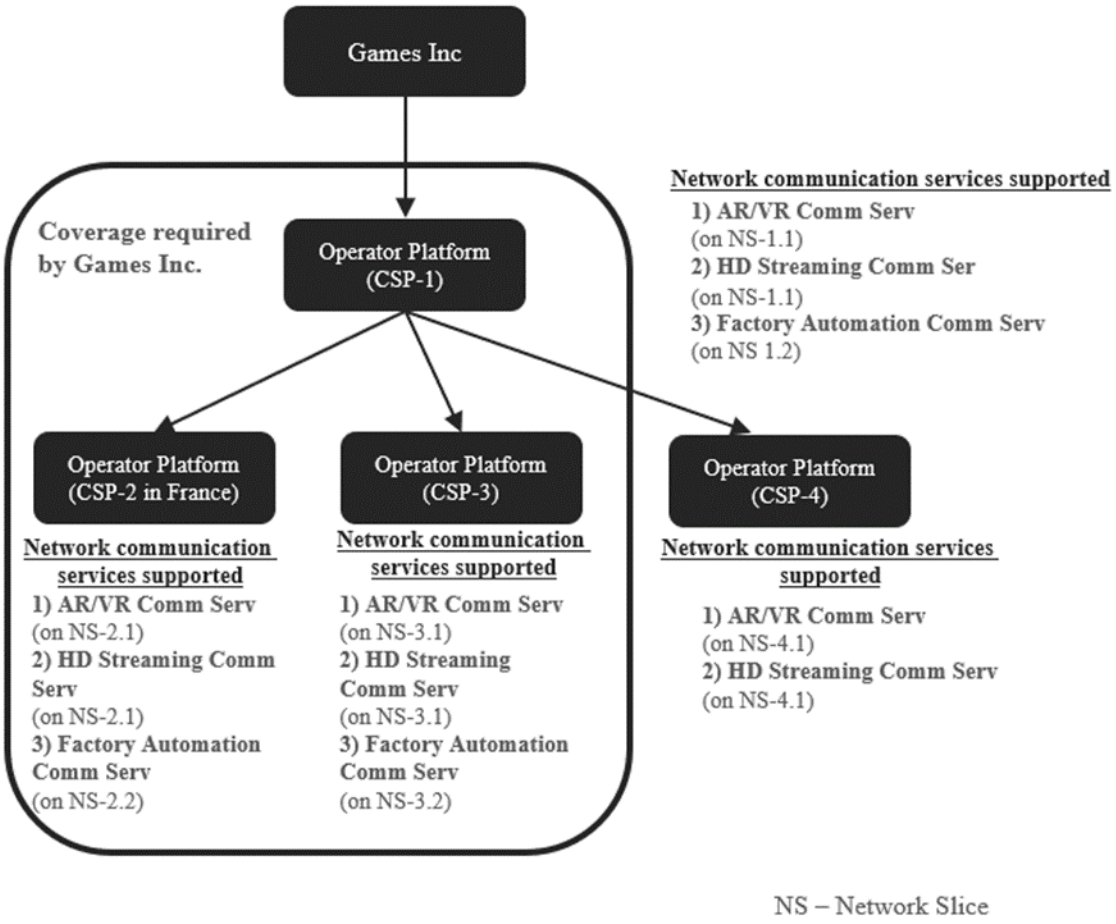
H.2 Roaming

Note: Roaming is for further study.

H.3 Federation

For the federation scenario, one of the CSPs acts as Leading CSP/OP, providing a single point of contact to the Application Provider for interaction with all involved CSPs. Thus, the Application Providers do not need to have a direct relation with all the CSPs. Whenever they need to have access to a Network Communication Service, they specify that Network Communication Service's needs and the location needs (e.g., coverage in France, or coverage in a stadium) to their Leading CSP. The Leading CSP will identify the target CSPs and distribute the Network Communication Service needs provided by the Application Provider to them. The figure below provides an example use case depending on this

approach.



In this use case, Games Inc has an AR/VR game which requires low latency connectivity (Network Communication Service) to provide a good experience to their premium game end-users. Their game end-users have subscriptions that span across multiple CSPs (CSP-1, CSP-2, CSP-3), but not CSP-4 which serves a different geographical area that is not covered by Games Inc. They intend to operate only in France which is covered by CSP-1, CSP-2 and CSP-3 whereas. CSP-4 may be operating in the UK for example.

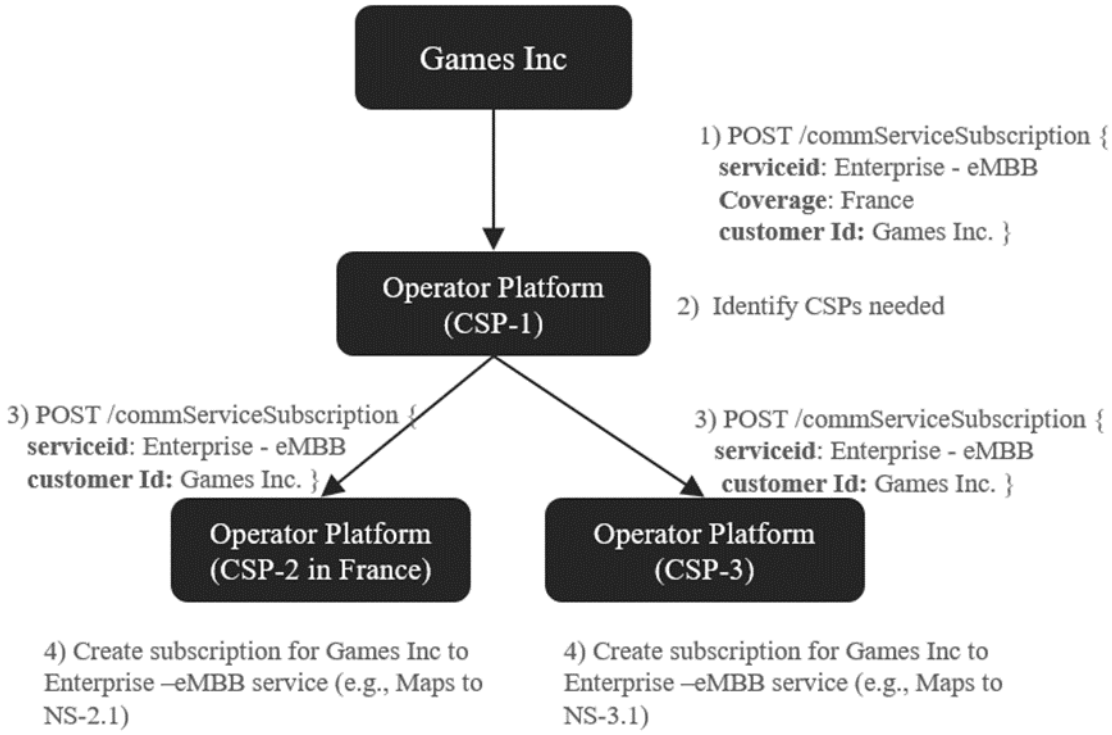
To have this special low latency connectivity for their gaming end users, Games Inc. need all the CSPs in France to support a low latency Network Communication Service. They sign-up with CSP-1 as their primary contact (Leading OP) to have the nation-wide Network Communication Service setup in France. They put a request to CSP-1 for such Network Communication Service to be setup passing these key parameters (ignoring all commercial aspects for now):

- **Network Communication Service characteristics**: LOW_LATENCY (Traffic Category)
- **Network Communication Service Coverage area**: France

As a simplification, the solution proposed leverages the concept of an agreed set of Network Communication Services (with identifiers) among all the CSPs. In this case, the Network Communication Services and their characteristics are agreed with all the CSPs and then published to all the Application Providers. Application Providers do not need to discover any

specific Network Communication Service based on some network characteristics. Instead, they can leverage the agreed set of Network Communication Services provided by the involved CSPs' Operator Platforms.

Whenever an Application Providers needs to get the Network Communication Services access, they send out the request to the Leading OP as shown in the figure below (using the Games Inc use case introduced earlier).



Continuing the earlier use case, Games Inc triggers the Network Communication Service subscription in France by giving a POST request passing the Network Communication Service id which is required (Enterprise - eMBB) and coverage area (France).

CSP-1's OP validates the requests and identifies the required target CSPs (in France, CSP-2, CSP-3). Note that as explained earlier CSP-4 is not required since the request relates to France only. CSP1's OP forwards the request to all the required CSPs' OPs. All CSPs' OPs validate the request and create a subscription for Games Inc. to this shared Network Communication Service. They may just forward the request to each of the target CSPs using an E/WBI API similar to the NBI API that the Leading OP exposes to Games Inc. In this example also CSP-1 checks if they offer this Network Communication Service (given that they are based in France themselves). If available, they create / add Games Inc. to that Network Communication Service.

Note: It is possible for each CSP to decide how they want to realize a Network Communication Service as long as the Network Communication Service characteristics can be assured.

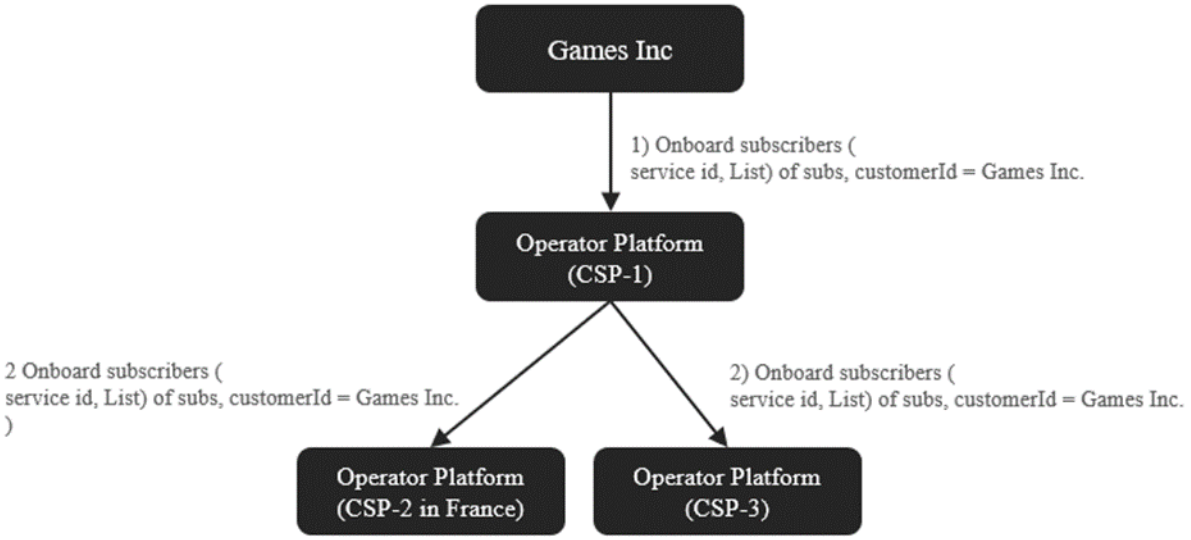
In this simplified solution, all the CSPs are aware of the Network Communication Service needed by Games Inc. So, there is no need for any Shared Catalog / inventory between the

CSPs. It might still happen that some CSPs cannot offer that Network Communication Service for any reason. In that scenario they will respond with an error message to the Leading OP. The Leading OP can report that error back to the Application Provider (Games Inc) so that they can act accordingly.

Once the Network Communication Services are setup, the Application Provider (Games Inc) can trigger onboarding of their devices (or PDU sessions etc.) on to the Network Communication Service.

Note: This is just one of the options with another one being that the subscriber might trigger getting onboarded to the Network Communication Service themselves, for example by buying add-on subscriptions.

The figure below shows what the onboarding flow triggered by Application Provider looks like:



Note: Simplistic view shown for easy understanding

As shown in the figure, once the Network Communication Service is setup, Games Inc can send the list of their subscribers/devices to be onboarded to this Network Communication Service (identified by a service id) to CSP-1.

Note: Rather than a list, Games Inc. might provide also a single subscriber/device in the request if they would loop around all their subscribers.

CSP-1's OP identifies the CSP for each of subscribers provided by Games Inc and prepares a sub-list of subscribers for CSP-2 and CSP-3. Next to triggering the onboarding requests to CSP-2 and CSP-3 with these sub-lists of subscribers, CSP-1 triggers onboarding of their own subscribers from the list provided by Games Inc to the Network Communication Service that was set up.

The Leading OP is responsible to distribute this request to the downstream CSPs with a filtered list of devices. Each CSP then decides how to onboard the devices to the Network Communication Service offered on their network (e.g. by subscription in network, URSP provisioning, QoS flows etc.).

H.4 Security

NSaaS provides on-demand requirements based on the needs of the Application Provider for specific network slices. The OP and the Application Provider use appropriate security control policies to be able to protect against unauthorised access and inappropriate use of the E2E network slice. Moreover, the capabilities for authentication of management service requests for allocating, deallocating, or modifying an NSI are expected to be supported by the OP or Application Provider either explicitly (directly) or implicitly(indirectly), as per 28.533 section 4.9 [38].

Authorised services allow a CSP to provide management capabilities and grant provisioning permission to the CSC. Once the CSC has provisioning permission, it will manage the network slice instance lifecycle (allocating, deallocating, or modifying an NSI) and its services.

Authorised services can either be explicit or implicit. Explicit authorisation of a token is obtained from a CSP so that the CSC can interact with a CSP. A CSP can enforce access control and verify the access token. Implicit authorisation is when the CSP enforces access control based on local policies and synchronises the policies across a centralised authorisation service.

H.5 Charging

3GPP has produced a set of technical specifications that define the architecture and protocols that enable Network Slice charging using the Operator's Converged Charging System (CCS). In the context of this Annex, the most relevant ones are:

- TS 28.202 "Charging management; Network Slice management Charging in the 5G System (5GS);Stage 2" [37]
- TS 28.201 "Charging management; Network slice performance and analytics charging in the 5G System (5GS); Stage 2" [36]

As explained in section H.1 of this Annex, 3GPP has defined the network slice lifecycle, which optionally can be managed by the CSC/Application Provider if supported by the MNO/CSP/OP. If this capability is allowed, 3GPP has defined in 3GPP TS 28.202 [37] the protocol that allows doing the rating and charging associated with the following operations related to the Network Slice lifecycle management:

- Network Slice Instance creation.
- Network Slice Instance modification.
- Network Slice Instance termination

When a CSC/Application Provider invokes one of the lifecycle management operations included above, and the operation is successfully completed, the OP triggers a charging request to the Converged Charging System (using the SBI-CHF) to ask for the rating and charging associated with that particular operation.

It is important to note that to enable the rating and charging process, a provisioning in the Operator's BSS and CCS may also be needed.

3GPP has defined different potential architectures for this charging integration, where the charging requests could be triggered:

- Directly from the element managing the lifecycle management operation (embedded charging trigger function -CTF- in 3GPP terminology).
- By the CEF (Charging Enablement Function), an element defined by 3GPP that gets the notifications about an operation's completion and triggers the charging request to the CCS.

Note: The architecture to be used will be dependent on the capabilities available in the Operator and is for further study.

The Charging dialogue with the CCS is based on a Request/Response pattern where:

- The charging requests will include all the information elements that could be relevant for the CCS to calculate the appropriate tariff and do the charging (the commercial model used in the Operator is out of the scope of this document). As a reference, the following elements should be included in the charging request:
    - CSC/Application Provider/Operator Platform identifiers that are invoking the operations
    - Operation type invoked (creation/update/termination)
    - Network Slice Instance identifiers
    - Set of parameters related to the service profile associated with that Network Slice instance (e.g. latency, resource sharing level, Jitter, reliability …)
- The Charging request responses from the CCS will inform about the outcome of the charging operation.

As a result of this charging operation, the CCS will generate a CDR (Call Details Record) including all the details about the charging operation (parameters used in the charging request, price, balances after doing the charging, etc.).

Note: The potential use of this CDR is for further study.

Besides the charging scenarios associated with network slice management operations, 3GPP has also defined the following charging integration scenarios that are for further study [36], [37] :

- Charging based on a device's 5G data usage using a particular Network Slice Instance. This charging scenario is enabled by the information provided by the 5G Packet core in the charging requests sent to the CCS that are associated with the device's data usage.
- Charging based on the performance of a particular Network Slice (based on the SLA commitment for a particular Network Slice Instance).

  In this particular case, the charging integration is defined in 3GPP TS 28.201 [36] and - as in the case of network slice management charging scenarios - the charging

dialogue is also based on a request/response pattern where charging requests will include information about the analytics for that Network Slice that are periodically collected by the Operator (in the NWDAF). Potential examples of these analytics are the Network Slice load, the device observed service experience, etc.

H.6 Provisioning for end user

Note: It is for further study if the subscription will be managed by the OP directly or if the Application Provider needs to communicate with the Operator's BSS directly.

To enable a UE to select the appropriate Network Communication Service, it is essential to provide the UE with the correct network details. This is done during the provisioning when the end user profile is updated with the service and network slice or APN/DNN information. The end user profile includes but is not limited to S-NSSAI, DNN, APN, URSP rules, and location information.

Annex I Service and capability exposure charging concepts

As described in section 2.1.4, the Operator Platform architecture needs to allow Operators to charge for the services and capabilities that are exposed by that Operator to Application Providers, subscribers, and other Operator Partners.

Any decision relating to charging and/or billing for the usage of the services as described in this Annex is for an individual Operator to decide.

A set of technical requirements are necessary to enable these charging and billing capabilities. These technical requirements will support potential commercial models defined by Operators –for federation and towards end customers/developers.

Note: The definition of commercial models is out of the scope of this document.

An Operator Platform exposes different Operator's services and capabilities to third parties. Although this set of services and capabilities is quite heterogeneous and is in constant evolution, it is possible to establish a classification of these services/capabilities from a charging perspective. The following service categories can be considered:

- Network capabilities exposure services with no impact on the device's data usage.
- Network capabilities exposure services with impact on the device's data usage.
- Network provisioning services.
- Edge application management services.

A detailed description of these categories together with examples of potential charging factors used for services/capabilities will be provided in the next sections of this Annex.

In addition to the categories listed above, there is one more that can be considered that groups "General purpose services" into its own category. This category would include the set of services/capabilities that are exposed by the Operator as "enabler" services (e.g., to manage the connection from the Application Provider to the OP, to manage permissions/consents, etc.). This category may require generation of file records (e.g., XDRs) that could be used by the Operators for charging and/or reporting purposes.

I.1 Network capabilities exposure services: with no impact on device's data usage

This category includes the group of services that are consumed by the Application Providers to access the capabilities exposed by the Operator's Network and that have no impacts on the device's data traffic usage as a result of the service invocation. These services are normally used to get information from the Operator's Network and some potential examples are:

- Network information retrieval related services: for example, to get or verify the location of a device that is registered in the Operator's Network, to get or check the device's registration status, to be informed about a device's location changes etc.
- Services to receive notifications related to analytics information provided by the Operator's Network.

From a technical perspective, the following potential charging factors, supporting charging and billing, could be used:

- Service activation charging:

  A one-time fee or a periodical fee (e.g., a monthly fee) could be charged by an Operator to third parties. This fee would enable the access to a particular service (different fee per service/group of services). This fee will not be dependent on the service usage.

- Charging per API invocation received:

  In this case charging would be based on the service API that is invoked by the Application Provider. Depending on the Operator's decision, this charging factor would allow the operator to charge based on:

  - The particular API (operation) that is invoked by the Application Provider, without considering the parameters in the payload included in the service API invocation.
  - The particular API (operation) that is invoked by the Application Provider and considering some parameters included in the service API invocation (selected API payload).

    Note that in this case only a subset of parameters, that will be dependent on the service, would be considered (e.g., in a device location service request, the precision included in the API payload could be used to use that level of precision as a potential parameter to consider in the rating and charging).
    The reason for considering only a subset of the parameters is to avoid unnecessary complexity and potential latency/dimensioning issues.

  This charging factor would allow the Operators to have the possibility to do the charging and billing based on:

  - The number of API invocation requests for Network information retrieval received (e.g., Charging per device location query request received)

  - The number of API invocation requests for a notification service received (e.g., Charging for requests to receive notifications from an analytics information service during a period of time)

- Charging per notification sent to the Application Provider (as a result of a request for such notifications):

  In this case charging would be based on the type of notification that is sent (e.g., Charging per analytics information notification delivered to the Application Provider)

The list of charging factors are the potential ones that the Operator can choose to support the commercial models for the services included in this category.

The related technical requirements that need to be supported by the Operator Platform for these charging factors are described in section 5.1.5 of this document.

I.2 Network capabilities exposure services: with impact on device's data usage

This category includes the group of services that are consumed by the Application Providers to access the capabilities exposed by the Operator's Network and that have an impact on the device's data traffic usage. Some potential examples of these services are:

- Services that influence the device's QoS (e.g., to request a specific QoS – 'High' QoS – to be delivered to a specific PDU data traffic session of a device)
- Services that allow sponsorship of data traffic usage (e.g., A particular PDU data traffic session of a device is sponsored by an Application Provided)
- Services that influence how the data traffic of a device is steered in the Operator's Network.

From a technical perspective, the following potential charging factors, supporting charging and billing, could be used:

- Service activation charging:

  A one-time fee or a periodical fee (e.g., a monthly fee) could be charged by an Operator to third parties for enabling the access to a particular service (different fee per service/group of services). This fee would not be dependent on the service usage.

- Charging per API invocation received:

  In this case charging could be based on the service API that is invoked by the Application Provider.
  As in the previous category, depending on the Operator's decision, charging can be based on the operation that is invoked (API type) or on a combination of the operation invoked and a subset of parameters included in the API invocation payload.
  Through this charging factor, the Operators would have the possibility to use time-based charging models to do the charging and billing of a service (e.g., charging per unit of time that a particular QoS is provided to a device/PDU session)

- Charging based on data traffic usage as a result of a previous service invocation:

  In this case charging could be based on the data traffic consumption of a device in the Operator's Network as a result of a previous service API invocation (e.g., charging per each unit of traffic volume that is carried over an active QoS session) Using this charging factor, it would be possible to enable volume-based charging models to do the charging and billing of a service.
  The feasibility of using this charging factor depends on the ability to correlate the data traffic in the Operator's Network that is impacted by a particular API call with the API invocation.

The list of charging factors described above are the potential options that an Operator could use to support the commercial models that an Operator chooses to carry out the charging and billing for the services included in this category.

The technical requirements that need to be implemented by an Operator Platform to support each of these charging factors are described in section 5.1.5 of this document.

I.3 Network provisioning services

This category includes the group of services that are consumed by the Application Providers to manage different aspects of Network Services Provisioning in the Operator's Network.

In this category, the Application Providers are also accessing services and capabilities provided by the Operator with impact on the devices data traffic. The main difference compared to the previous category (*Network capabilities exposure services: with impact on device's data usage*), is that the exposition of these services requires previous provisioning activities in the Operator's Network (e.g., to provision a particular APN or Network Slice Instance in the Operator's Network).

Note: The Operator's BSS/OSS should be involved during these services provisioning flows. How this is done is Operator-dependent and out of the scope of this document.

From a technical perspective, the following potential charging factors, supporting charging and billing, could be used:

- Service activation charging:

  A one-time fee or a periodical fee (e.g., a monthly fee) could be charged by the Operator's to third parties for enabling the access to a particular service (different fee per service/group of services). This fee would not be dependent on the service usage.

- Charging per API invocation received (service lifecycle modification):

  In this case charging would be based on the service API that has been invoked by the Application Provider and, depending on the Operator's decision, charging could be based just on the operation that is invoked or a combination of the operation invoked, and a subset of parameters included in the API invocation payload.
  Some potential examples of this charging factor include:

  - Charging per service allocation/deallocation operation received (e.g., Charging per operation received to assign an Application Provider to an existing NSI)
  - Charging based on the number of devices that are provisioned with a particular Network service (e.g., charging per number of devices that are using a particular APN based on the number of subscription/unsubscription operations received)

- Charging based on data traffic usage:

  In this case charging would be based on the device's data traffic usage using a particular Network Service (e.g., an APN or an NSI).
  Depending on the Operator's decision, the following charging models could be used:

  - Time based charging: charging per each unit of time that the devices are using a particular Network Service (e.g., charging per each unit of time devices are using an APN)
  - Volume based charging: charging per each unit of traffic volume that devices are consuming using a particular Network Service (e.g., charging per each unit of traffic volume that is carried over an APN)

  The feasibility of enabling this charging factor depends on the ability to correlate the data traffic in the Operator's Network that is impacted by a particular API call with the API invocation.

The technical requirements that need to be implemented by an Operator Platform to support each of these charging factors are described in section 5.1.5 of this document.

I.4 Edge application management services

This category includes the group of services that are consumed by the Application Providers to manage their applications. Some potential examples are:

- Resources reservation services.
- Application onboarding and lifecycle management services.
- Application resources monitoring services.

From a technical perspective, the following potential charging factors, supporting charging and billing, could be used:

- Service activation charging:

  A one-time fee or a periodical fee (e.g., a monthly fee) could be charged by the Operator's to third parties for enabling the access. This fee would not be dependent on the service usage.

- Charging based on Edge enabling infrastructure resources usage:

  Depending on the Operator's decision the following possibilities could be considered:

  - Charging based on the subscribed capacity requested by the Application Provider in the API request (a subset of the parameters included by the Application Provider in the API request payload).

As an example, the following resources could be considered to do the rating and charging: number of vCPUs, memory, storage, incoming/outgoing data volume, etc.

- Charging based on real resources usage (based on periodical information retrieved from the Operator's Network)

- Charging based on Application lifecycle management API requests:

  In this case charging could be based on the type of operation that is invoked (instantiation/upgrade/termination) and potentially, depending on Operator's decision, a subset of the parameters received in the API payload (e.g., availability zone).

- Charging based on data traffic usage in the Operator's Network:

  In this case charging would be based on the data traffic consumption of the devices in the Operator's Network accessing an Edge Application.
  The feasibility of enabling this charging factor depends on the ability to correlate the data traffic in the Operator's Network that is impacted by a particular API call with the API invocation.

The technical requirements that need to be implemented by an Operator Platform to support each of these charging factors are described in section 5.1.5 of this document.

I.5 Charging factors summary

The table below summarises the list of potential events/charging factors that could be used by the Operator to carry out the charging and billing, depending on the service category exposed. The factors marked with "YES" are the ones potentially applicable for the service category.

Potential Events/ Triggers for Charging	Service Categories				- Technical Complexity +
	Network capabilities exposure: no impact on data traffic	Network capabilities exposure: with impact on data traffic	Network services provisioning	Edge capabilities management	
Service activation	YES	YES	YES	YES	
Service API invocation (and related notifications)	YES (API + payload)	YES (API+payload)	YES (API+payload) Service lifecycle management	YES (API+payload) Reserved Infra resources App lifecycle management	
Data traffic usage in the Operator's Network	NO	YES Only if volume-based charging (info provided by the Network) (*)	YES Only if volume-based charging (info provided by the Network) (*)	YES Only if volume-based charging (info provided by the Network) (*)	
Edge enabling infrastructure resources usage	NO	NO	NO	YES Only in case charging based on effective use infra resources (**)	

Table 40: Charging factors summary

(*): although information for charging is provided by the Operator's Network it will have implications for the Operator Platform that will have to provide the Network with information – via the SBI-NR – that can be used by the Operator's Charging engine to correlate API invocations with Data Traffic usage.

(**): Event-based charging model to be used for this purpose.

As already mentioned in this Annex, the Operator will be responsible for selecting the charging factors to use for a particular service depending on the selected commercial model for that service. Therefore, any decision relating to charging and/or billing for the usage of the services as described in this table is for an individual Operator to decide.

The table also shows the level of technical complexity that would be required for the implementation of the different charging triggers/factors, where the first rows have lower level of complexity than the rows at the end. Each charging trigger/factor is independent from one another.

Annex J OP Managed DNS Service

J.1 Introduction

DNS provides the lookup service or the name (FQDN) resolution process to translate an FQDN to the corresponding IP address(es). A DNS server in the role of an authoritative server may be contacted to resolve a given FQDN that is derived from a Top-Level Domain (TLD).

An OP can use FQDNs to refer to the IP address(es) for the Edge Application instances on OP-managed cloudlets. It can enable Application Providers to use these FQDNs with Application Clients to discover the IP address(es) for communicating with an Edge Application.

Cloudlets hosting Edge Applications e.g., containers, VMs etc can be assigned the FQDNs for the IP addresses that external clients can use to connect with the Edge Application's instances.

An OP can use external DNS services for managing the DNS records i.e., the mapping of FQDNs to IP address(es) for edge resources. This DNS service can act as an authoritative DNS server that can be reached from the telco network to resolve the DNS queries for the OP managed subdomains.

J.2 A Use case for Edge Application IP Address Discovery

An Application Client on a UE can use the cloudlet-specific FQDN of the Edge Application for the discovery of the application instance IP address(es).

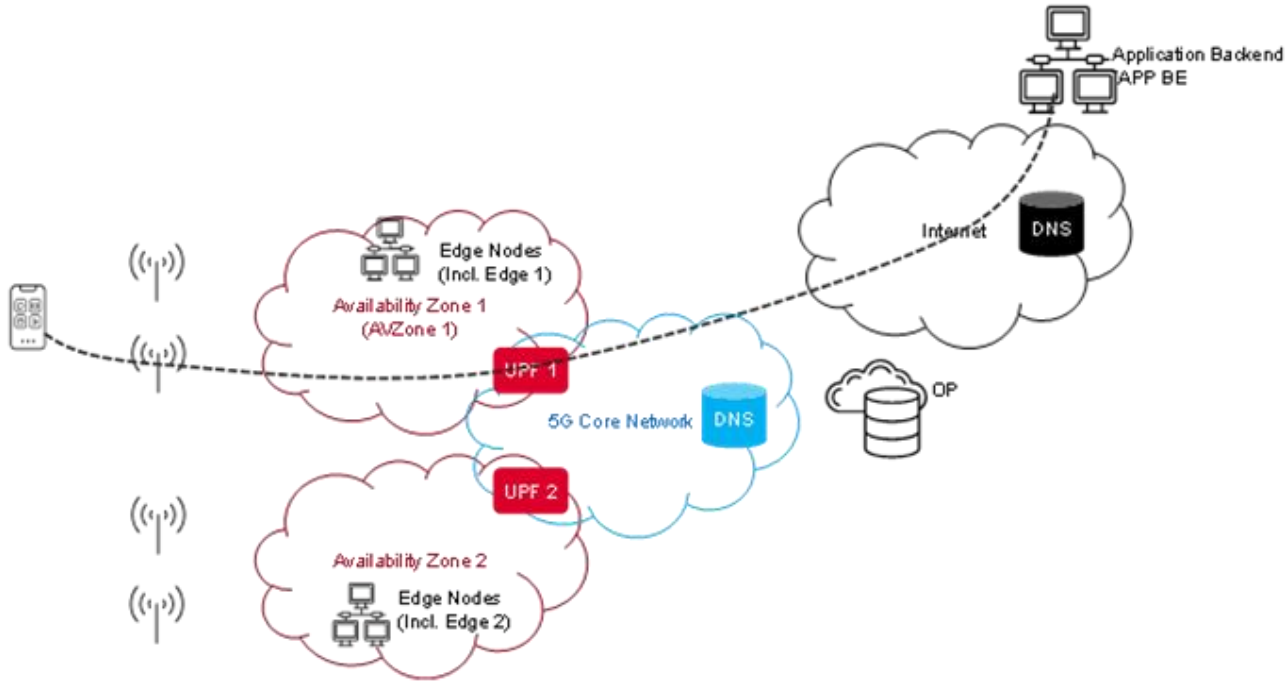


Figure 44: A reference deployment of DNS service with OP

J.3 Role of the Operator Platform

For supporting the use case for Edge Applications that can be referenced by a FQDN, an OP can support a capability to manage and assign cloudlet-specific FQDNs for Edge Applications which could belong to different Availability Zones.

An OP can use one or more DNS subdomains derived from TLDs (e.g., .com, .io etc) for the allocation of the FQDNs to the Edge Resources. These subdomains can be obtained from the public domain registrars and the OP can use them to derive the FQDNs for Edge Resources.

J.4 Role of the Application Providers

Application Providers can configure or deliver the FQDNs to the Application Clients using mechanism outside of the OP scope. For example, Applications Clients can use the Linux Foundation CAMARA Simple Edge Discovery API to retrieve the FQDNs for the edge resources and use the FQDNs in a DNS query to resolve to their IP address(es).

If there are multiple Application Instances in an Availability Zone, then the Application Providers can determine their own mechanism for the use of the IP address(es) to connect with Edge Resources.

J.5 Implementation Guidelines

An OP may use one or more DNS subdomains that it can use to assign to the Edge Resources. The FQDNs for Edge Resources derived from the OP-managed subdomains would be valid for the OP-defined Availability Zone and any DNS query from different Availability Zones may not result into a valid response.

The Operator’s mobile network infrastructure should be able to route the DNS queries for these FQDNs towards the DNS service that is an authoritative server for the OP-managed subdomains. This DNS service could be an external service outside of the OP architecture.

The blocks of IP Addresses associated with the Edge Resources as used with the FQDNs can be from the Operator’s private subnets which are accessible only from the specific locations associated with an Availability Zone. Or they could be from be from public subnets that may be accessible from a wider set of locations or Availability Zones.

An OP can have the policies to select the nature of the IP addresses to be used with FQDNs in an Availability Wone.

Note: DNS support under user mobility is FFS in the current version of this document.

Note: A guidance to use the optimum Time to Live (TTL) with the OP is FFS

Note: For clients using any kind of VPN service to connect with the Edge Resources by using FQDNs, the OP may not be able to ensure the DNS query resolution time.

Note: The Application Providers should be aware of the nature of the IP address(es) while using them for connecting to Edge Resources or applications in various cloudlets.

Annex K Privacy Management considerations

K.1 General

From an OP perspective, data processing is limited to sharing data to an Application Provider (potentially through an Aggregator), so an Application (owned by the Application Provider) can perform any further processing on the shared personal data. To declare what an Application Provider wants to do with a set of personal information resources, a Purpose of Data Processing must be declared. Each Purpose of Data Processing is associated with a legal basis which must be compliant with local regulations. Only pre-defined Purposes of Data Processing can be used by the AP and the use of personal data cannot go beyond that Purpose of Data Processing. Whenever the legal basis dictates direct interaction with end users (e.g., Consent legal basis), the Application signals the Purpose of Data Processing to the end-user through the OP (an Aggregator could be involved) and Privacy Management Function in the CSP domain. In turn, the end-user must opt-in and the Privacy Management Function must capture the result of that operation. It is expected that the data processing in the Application takes place exclusively under the indicated Purpose of Data Processing.

Several legal bases are well-established as indicated e.g., in [51]:

- Consent

  Note: Throughout this document, the terms “Consent” and “Application-related Consent” are interchangeable

- Context of a contract (to which the end user is party)
- Compliance with a legal obligation (to which the controller is subject)
- Protect vital interest (of the data subject or of another natural person)
- Performance of a task carried out in the public interest
- Legitimate interests (pursued by the controller or by a third party)

Although most of the technical interest on legal bases revolves around the Consent for processing personal data (e.g., there is dedicated 3GPP study to deal with Consent for accessing 3GPP services [52]), other legal bases for processing personal data could be used. Nevertheless, it is noteworthy that some definitions related to other legal bases are rather relative to local regulations, e.g., what public interest or vital means could vary around the globe, so having universal mechanisms to fulfil local regulations is challenging.

If Consent is the applicable legal basis for processing, users must actively agree through an affirmative action (opt in). How Consent can be captured depends on the concrete use case and on the laws of the jurisdictions which govern the use case. Even though Consent can be obtained through a variety of methods and techniques (e.g., ticking a box on a website or writing/accepting a letter confirming the grant for processing personal data), having the Consent captured during runtime is a well-established approach for some scenarios in which the so-called “resource owner grants or denies the client’s access request” (see OAuth 2.0 Authorization Code Grant [53]).

There is not a universal solution for Consent Management. It depends for instance in the controllership of the device (e.g., only one application should have control on the device), or the type of service provided by an application running on a generic device. For the former

case, a device identifier could be considered (along with other information) for granting access to personal data, whereas for the latter case, depending on the type of service, a server-side IP address could be considered (along with other information) for granting the application access to personal information.

Additionally, the Consent can be granted for one device or several at the same time. For the latter case, many mechanisms could be in place, e.g., providing a list of devices upfront while signing a contract, or via API calls through a portal.

K.2 Requirements for supporting relevant end user rights

General requirements to OP can be derived from analysis of a subset of rights that an end user could be entitled to. Direct interactions among end users and the Application Provider, as well as direct interactions among subscribers and the CSP domain are considered out-of-the-scope of this document. Table 41 presents technical requirements for the OP, SBI-PrM, NBI and EWBI to support a subset of privacy-related rights in which OP plays a relevant role.

		Requirements on:		
		Privacy Management Function (SBI-PrM)	**OP**	**Application Provider (NBI)**
End user privacy-related rights	Information (to understand what will be done with their data)	• Upon indication from the OP, capture the Consent from the end user signaling the Purpose of Data Processing for an Application ID • Store the Consent information in the Privacy Management Function • Create a subscription to get the OP notified about any possible change on the Consent information	• Ensure that a suitable legal basis (compliant with local regulations) supports sharing personal data with an AP. • If the applicable legal basis is Consent: 1. Trigger the Consent capture carrying the intended Purpose of Data Processing, AP ID and Application ID 2. Cache Consent Records (if allowed by local regulations) 3. Get notified about changes on Consent information	• -Signal the Purpose of Data Processing (associated to a legal basis) while onboarding and obtaining authorization process
	Access (to get a confirmation whether their personal data is being processed)		• Logging, Tracing and Auditing functions	• Logging, Tracing and Auditing functions

		Requirements on:		
		Privacy Management Function (SBI-PrM)	**OP**	**Application Provider (NBI)**
	• Restricting of processing (to request ceasing of all processing of their data) • Object	• Upon notification, update relevant Privacy Management Function entries • Log, trace and audit any further attempt to capture Consent • Keep the Consent-related configuration (i.e., Consent records) stored during the time indicated by the local regulation	• Upon notification, stop sharing associated personal information and (potentially) notify other entities: 1. If applicable legal basis is Consent, notify Privacy Management Function 2. Notify a Partner OP • Log, trace and audit any further invocation involving personal data	• Upon end user request, stop any associated processing of personal data and notify OP or Leading OP

Table 41: Requirements for supporting end user privacy-related rights

Annex L Document Management

L.1 Document History

Version	Date	Brief Description of Change	Approval Authority	Editor / Company
1.0	30 Jun 2021	New PRD OPG.02, based on requirements proposed in OPG.01.	ISAG	Tom Van Pelt / GSMA
2.0	14 Apr 2022	Update implementing OPG.02 CR1002	ISAG	Tom Van Pelt / GSMA
3.0	03 Oct 2022	Update implementing OPG.02 CR1003	ISAG	Tom Van Pelt / GSMA
4.0	29 Mar 2023	Update implementing OPG.02 CR1004	ISAG	Tom Van Pelt / GSMA
5.0	26 Jul 2023	Update implementing OPG.02 CR1005	ISAG	Tom Van Pelt / GSMA
6.0	16 Feb 2024	Update implementing OPG.02 CR1006	ISAG	Tom Van Pelt / GSMA
7.0	20 Sep 2024	Update implementing OPG.02 CR1007	ISAG	Tom Van Pelt / GSMA
8.0	28 Feb 2025	Update implementing OPG.02 CR1008	ISAG	Tom Van Pelt / GSMA

L.2 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.