Operator Platform: Requirements for Edge Services
Version 2.0
19 February 2026

Security Classification: Non-Confidential

Copyright Notice

Disclaimer

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

The following document details the requirements and use cases for edge computing capabilities using the Operator Platform (OP) architecture defined in GSMA PRD OPG.02 [1]. The document utilises the OP architecture and its interfaces [1] to emphasise the delivery of low-latency, high-performance edge services. The OP enables Operators, service providers, Aggregators and application developers to efficiently utilise edge computing resources while ensuring adherence to regulatory and privacy requirements.

The use cases focus on industries such as automotive, mixed/AR, cloud gaming, and remote control services, demonstrating how the OP facilitates improved user experiences through edge computing. These edge services reduce latency, enhance data handling, and enable seamless mobility between networks, all underpinned by well-defined architectures and interfaces supporting both Northbound (application-facing), Southbound (infrastructure-facing) and East-Westbound (other OP-facing) communication. These features are critical for promoting interoperability across Operators and geographies while ensuring Service Continuity, particularly for mobile users.

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 1.5. 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 1.5. ETSI MEC Phase 3 is taken as the basis.

1.2 Scope and Objectives

This document outlines the requirements for edge services that guide the entire industry ecosystem in defining a unified solution for exposing network capabilities and edge computing resources necessary to support edge-related use cases.

The document addresses the following areas:

- **Exposure of Network and Edge Computing capabilities**: The PRD emphasises the integration of edge computing and network services for Application Providers, whether within enterprises or independent third parties. The goal is to enable a straightforward and universal way to interact with networks and edge computing platforms.
- Requirements to enable an end-to-end delivery chain for edge services. It covers the interactions of all stakeholders involved in application delivery, including:
    - Deployment on Edge Resources
    - Deployment of resources in cloud and network environments

1.3 Definitions

The definitions in the GSMA PRD OPG.02 [1] apply to this document.

Term	Description
Application Backend Part	An architectural part of an Application that is to be deployed on public or private (and central) cloud infrastructure.
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.
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 [4].
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 [4].
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.
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's networks.
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.

1.4 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

Term	Description
API	Application Programming Interface
AR	Augmented Reality
BTS	Base Transceiver Station
CAPIF	Common API Framework
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
DDoS	Distributed Denial of Service
DNN	Data Network Name
DNS	Domain Name System
DoS	Denial of Service
EA	Edge Attribute
EAS	Edge Application Server
ECP	Edge Computing Platform
ECS	Edge Configuration Server
ECSP	Edge Computing Service Provider
EEC	Edge Enabler Client
EES	Edge Enabler Server
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
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
HIDS	Host-based Intrusion Detection System
HPLMN	Home Public Land Mobile Network
HR	Home Routing

Term	Description
HTTP	HyperText Transfer Protocol
IaaS	Infrastructure as a service
ID	IDentifier
IDS	Intrusion Detection System
IETF	Internet Engineering Task Force
I/O	Input/Output
IoT	Internet of Things
IP	Internet Protocol
ISG	Industry Specification Group
ITU	International Telecommunication Union
KPI	Key Performance Indicator
LBO	Local BreakOut
LCM	Life-Cycle Management
MCC	Mobile Country Code
MEC	Multiaccess Edge Computing
MEH	Mobile Edge Host
MEO	Mobile Edge Orchestrator
MEP	Mobile Edge Platform
MNC	Mobile Network Code
MSISDN	Mobile Subscriber Integrated Services Digital Network Number
NAS	Non-Access Stratum
NBI	Northbound Interface
NEF	Network Exposure Function
NFV	Network Functions Virtualisation
NPU	Neural Processing Units
NUMA	Non-Uniform Memory Access
OCI	Open Container Initiative
OP	Operator Platform
OSS	Operation Support System
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
PLS	Private LAN Service
PMIPv6	Proxy Mobile IPv6 (protocol)
PRD	(GSMA) Permanent Reference Document

Term	Description
PrM	Privacy Management
QoE	Quality of Experience
QoS	Quality of Service
RI	Roaming and Interconnect (controls)
RN	Radio Network (operational controls)
RRS	Resource Requirements Specification
SA3	Service and System Aspects WG3 (within 3GPP)
SBI	Southbound Interface
SBI-CR	Southbound Interface – Cloud Resources
SBI-NR	Southbound Interface – Network Resources
SBO	Session BreakOut
SCEF	Service Capability Exposure Function
SDK	Software Development Kit
SDN	Software Defined Network
SFC	Service Function Chain
SH-IoT	Smart Home Internet of Things
SLA	Service Level Agreement
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
TAC	Tracking Area Code
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
UE	User Equipment
UNI	User to Network Interface
UPF	User Plane Function
URL	Uniform Resource Locator
URSP	UE Route Selection Policy
V2X	Vehicle-to-Anything
VIM	Virtualised Infrastructure Manager

Term	Description
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.

1.5 References

Ref	Doc Number	Title
[1]	GSMA PRD OPG.02	Operator Platform: Requirements and Architecture
[2]		Operator Platform Concept – Phase 1: Edge Cloud Computing https://www.gsma.com/futurenetworks/resources/operator-platform-concept-whitepaper/
[3]	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/
[4]		Open Glossary of Edge Computing, Linux Foundation Edge, https://github.com/State-of-the-Edge/glossary/blob/master/edge-glossary.md
[5]	3GPP TS 29.522	5G System; Network Exposure Function Northbound APIs, Stage 3 https://www.3gpp.org/DynaReport/29522.htm
[6]	3GPP TS 29.122	T8 reference point for Northbound APIs https://www.3gpp.org/DynaReport/29122.htm
[7]		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/
[8]		OCI Image Format Specification https://github.com/opencontainers/image-spec
[9]		Open Container Initiative Runtime Specification https://github.com/opencontainers/runtime-spec
[10]	GSMA PRD NG.126	Cloud Infrastructure Reference Model https://www.gsma.com
[11]	3GPP TS 23.501	System architecture for the 5G System (5GS) https://www.3gpp.org/DynaReport/23501.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

Ref	Doc Number	Title
[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]	Ranaweera2021	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
[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 TR33.839	Study on Security Aspects of Enhancement of Support for Edge Computing in 5GC https://www.3gpp.org/DynaReport/33839.htm
[20]	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
[21]	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
[22]	3GPP TS 23.548	5G System Enhancements for Edge Computing; Stage 2 https://www.3gpp.org/DynaReport/23548.htm
[23]	GSMA PRD OPG.04	East-Westbound Interface APIs https://www.gsma.com
[24]	3GPP TS 32.257	Telecommunication management; Charging management; Edge computing domain charging https://www.3gpp.org/DynaReport/32257.htm
[25]	GSMA PRD OPG.03	Southbound Interface Network Resources APIs https://www.gsma.com
[26]	GSMA PRD OPG.05	User-Network Interface APIs https://www.gsma.com
[27]	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
[28]	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

Ref	Doc Number	Title
[29]	3GPP TR 23.958	Edge Application Standards in 3GPP and alignment with External Organizations https://www.3gpp.org/DynaReport/23958.htm
[30]	3GPP TS 33.558	Security aspects of enhancement of support for enabling edge applications https://www.3gpp.org/DynaReport/33558.htm
[31]	3GPP TS 33.501	Security architecture and procedures for 5G System https://www.3gpp.org/DynaReport/33501.htm
[32]	IETF RFC 9113	HTTP/2 https://datatracker.ietf.org/doc/rfc9113
[33]	IETF RFC 9110	HTTP Semantics https://datatracker.ietf.org/doc/rfc9110

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

1.6 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 [28].

2 Use Cases

This section presents a selection of use cases that necessitate deploying an Edge Application closer to the End-User. These use cases were chosen for their wide range of functional coverage, rather than attempting to create a complete list of all use cases that can benefit from federated edge computing, as documented in GSMA WP OPG.01 . Together, these use cases highlight some of the essential edge capabilities that an OP must offer.

2.1 UC1 - Automotive - Advanced Horizon

2.1.1 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.

2.1.2 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 OP that authenticates the parties supplying the data, verifies applications and is involved in their discovery would provide the guarantees required for such a service.

2.2 UC2 - Automotive – Remote Driving

2.2.1 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.

2.2.2 OP Dependency

This use case has similar requirements on trustworthiness and communication to other Operators than the use case discussed in section 2.1.

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 OP for discovery, potentially combined with network slicing that the OP 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 OP would help to ensure that the supporting Edge Application is available on those networks.

2.3 UC3 - Multiplayer Augmented Reality (AR) Game

2.3.1 Description

The following use case is a multiplayer AR 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.

2.3.2 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 OP enabling the sharing of Edge Nodes between Operators would be able to support this.

Some games need specialist compute (e.g. Graphic Processing Unit (GPU)). As indicated in the TEC whitepaper [7], a federated model to deliver an OP 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.

2.4 UC4 - Privacy-preserving Health Assistant

2.4.1 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 Internet of Things (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).

2.4.2 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.

2.5 UC5 - Infrastructure sharing

2.5.1 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.

2.5.2 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.

2.6 UC6 - High-resolution media streaming service

2.6.1 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.

2.6.2 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.

2.7 UC7 – Visual Positioning Service (VPS)

2.7.1 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 Global Positioning System (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.

2.7.2 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 Central Processing Unit (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.

2.8 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	5.1.2.2#1	Y	Y	Y	Y	N	Y	Y
Provide info on UE's **location**	SBI-NR	5.3.2	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	5.1.2.1.2#11 GSMA PRD OPG.02 [1] section 4.3.2.3.2	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	6.3 5.2.1.3.1#3	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	5.1.2.2#2	Y	Y - critical	Y & 'fair'	Y - weak		Y	Y
Establish a chain of **trust** between the elements	UNI & OP	5.4.1.5	Y	Y		Y - critical	Extend over E/WBI		
Security Comms Compute Storage	 UNI OP OP	GSMA PRD OPG.02 [1]	 Y Y .	 Y Y		 Y Y Y			
Inter-OP Security		OPG.02 [1]					E/WBI		
Data sharing (Data is 'open' for use by multiple Application Providers)		missing	Y			Y but highly filtered			Y
Specialist compute	SBI-CR	5.3.2			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.

3 High-level Requirements for OP to Support Edge Services

This section defines the requirements that the OP should fulfil to support edge services. Generic platform related requirements are defined in GSMA PRD OPG.02 [1].

3.1 Functionality Offered to OP Ecosystem Party

3.1.1 Functionality Offered to Application Providers

An OP shall fulfil the following requirements related to the functionality offered to Application Providers:

1. An OP shall be able to effectively isolate each Tenant's applications from the applications of all the other Tenants.
2. The OP 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.

3. 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.

4. An OP shall allow an Application Provider to reserve resources for future Application Instance deployments, ensuring the availability of the booked capacity.
5. An OP shall allow an Edge Application to be deployed within an Operator’s network where it can utilise the optimum resources.
6. The OP 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.
7. The OP shall allow an Edge Application deployed within an Operator’s network to interface securely with the application's back-end infrastructure outside of the Operator’s network.
8. The OP shall allow an Edge Application deployed within an Operator’s network to store data in a manner that is secure and compliant with applicable local regulations.
9. The OP shall enable the utilisation of cloud resources that support deploying applications as Virtual Machines (VMs) or Containers.
10. The OP 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.

3.1.2 Functionality Offered to End-Users/Devices

An OP 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's 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.

3.1.3 Functionality Offered to Operators

An OP shall fulfil the following requirements related to the functionality offered to Operators:

1. The OP shall enable Operators to monitor their Subscribers' usage of Edge Cloud resources (including network) in a visited network.
2. The OP shall enable Operators to establish Edge Interconnection Network (EIN) connections, and monitor their usage by the applications for charging purposes.

3.1.4 Functionality offered to other OPs

An OP shall fulfil the following requirements related to the functionality offered to other OPs:

1. The OP shall enable deploying, operating and managing 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 OPs 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 shall enable a Visited OP to receive applications from Home OPs to serve Subscribers, whether they are Home OP Subscribers or Visited OP Subscribers.

3.1.5 Functionality Offered to Aggregators

Editor's note: Requirements are FFS.

3.2 High-Level Roaming Requirements

1. 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: 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.

3.3 Edge Enabling Requirements

This section defines the requirements that the OP should fulfil for the exposure of edge computing capabilities. Section 3.3.1defines the high-level requirements for exposing these capabilities. Section 3.3.2 goes into the management and reservation of compute resources. Section 3.3.3 defines the requirements for integration with development environments for Edge Applications. Section 3.3.4 provides requirements on edge specific enhancements while section 3.3.5 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 3.3.6and section 3.3.7covers the requirements on what an OP should provide to support Edge Compute capabilities serving Subscribers that may be mobile. Section 3.3.8 defines what an OP should support to manage the interconnection network between Edge Cloudlets and to enable Edge Application access to that interconnection network.

3.3.1 Edge Enabling 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 shall allow an application deployed on Cloudlets within an Operator's network to interact with low latency with applications deployed at nearby Operator's network Cloudlets, including those of other Operator's networks in the same area.

3.3.2 OP-enabled Edge Resource Management Requirements

3.3.2.1 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 Edge 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).

- The 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.

3.3.2.2 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.

3.3.2.3 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.

3.3.3 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 (CD) through code development pipelines similar to those provided in a public cloud.
2. Support for Continuous Integration (CI) through staging in edge test sites.

3.3.4 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.

3.3.5 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.

3.3.6 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 Quality of Experience (QoE), geographical store and privacy policies;
3. Create the application backend instance, including autoscaling.

3.3.7 3GPP Connectivity Models to Enable Edge Services

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 [22] 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 Quality of Service (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's OP and Operator deployments may differ but should be compatible with 3GPP specifications.

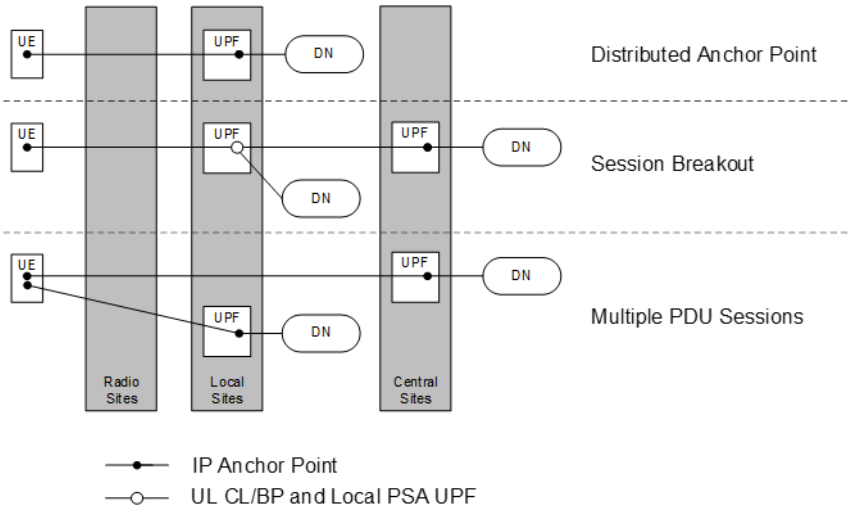

Figure 1: 5GC Connectivity Models for Edge Computing, REF 3GPP TS 23.548 [22]

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:

1. 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 [22].

2. An OP, networks and terminals shall support all Session and Service Continuity (SSC) modes.

3. 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
4. The network shall be able to notify the OP in case of UPF reselection. The Network Exposure Function (NEF) Application Programming Interface (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:

1. The Home OP shall be able to influence the URSP rules sent to the UE related to the applications managed by the OP.

Note: 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.

3.3.7.1 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 6.3.2 provides more details on the instantiation process.

Note: Area restrictions should be bound to Availability Zones

3.3.7.2 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/ Packet Data Network (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 MNO, 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 Virtual Reality (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 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.

3.3.8 Edge Interconnection Network (EIN) Management

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

- Communicating connection information of the peer Edge Clouds
- Enforcing security guidelines and providing security parameters to Edge Clouds
- 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.

3.4 Security Requirements

Generic OP security requirements are defined in Section 2.3 in GSMA PRD OPG.02 [1].

3.5 Edge Platform Model and Views

This model provides a simple high-level profile of the Operator Platform in the context of Edge Services. This can help external groups gain an initial understanding of the OP concepts, Ecosystem Parties and their interactions. The model is presented as a series of views from the perspective of each Ecosystem Party.

Each view uses existing OP terminology and represents certain requirements outlined in section 3, and will therefore omit details not relevant to that view.

Diagram Key:

-  Interface publisher
-  Interface consumer
-  Interface is used but is not defined in Operator Platform

The term ‘Lifecycle’ indicates management operations (CRUD) upon the indicated subject.

3.5.1 Basic Model – overall perspective

This model represents the high-level requirements for an OP to support Edge Services [3]. Note that the functionality offered to aggregators is marked as ‘FFS’ in section 3.1.5 – hence Aggregators/Aggregation Platforms are not explicitly included in the current model.

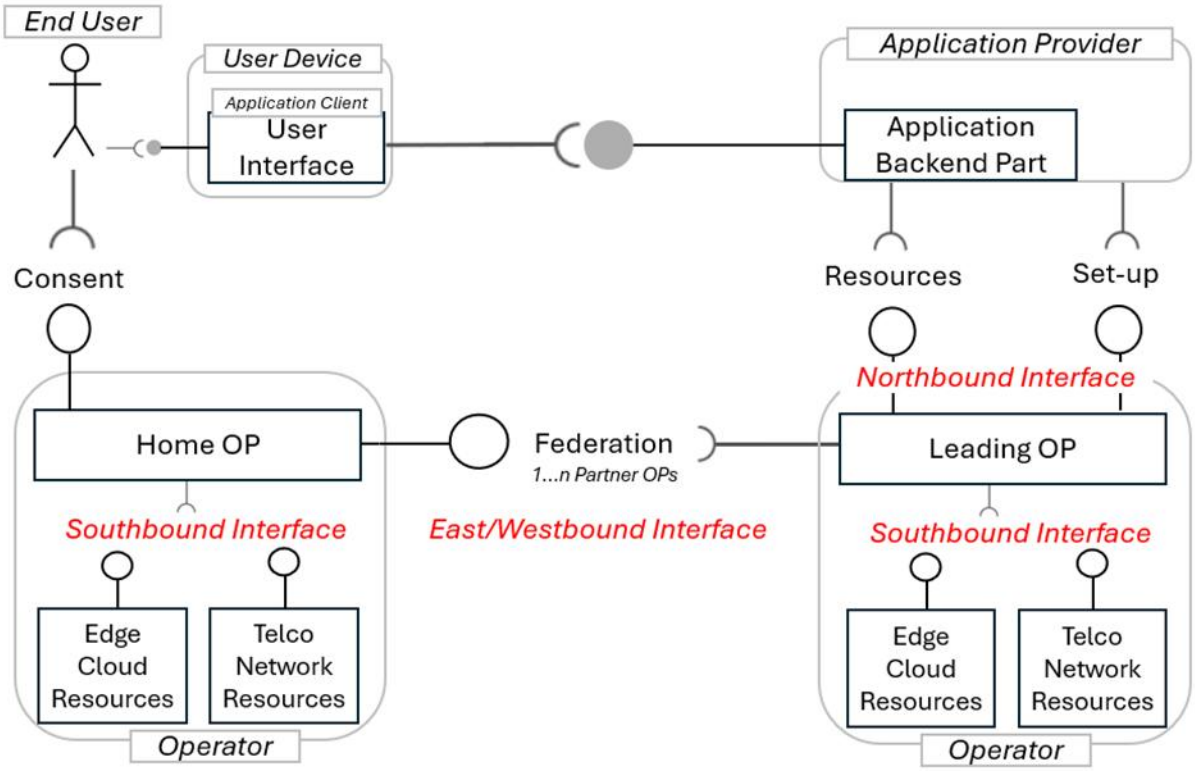

Figure 2: Basic Model showing the Edge Services profile of the OP Architecture

3.5.2 Application Provider perspective

3.5.2.1 Set-up view

This view represents the requirements in section 5.1.2.1 and 7.6.

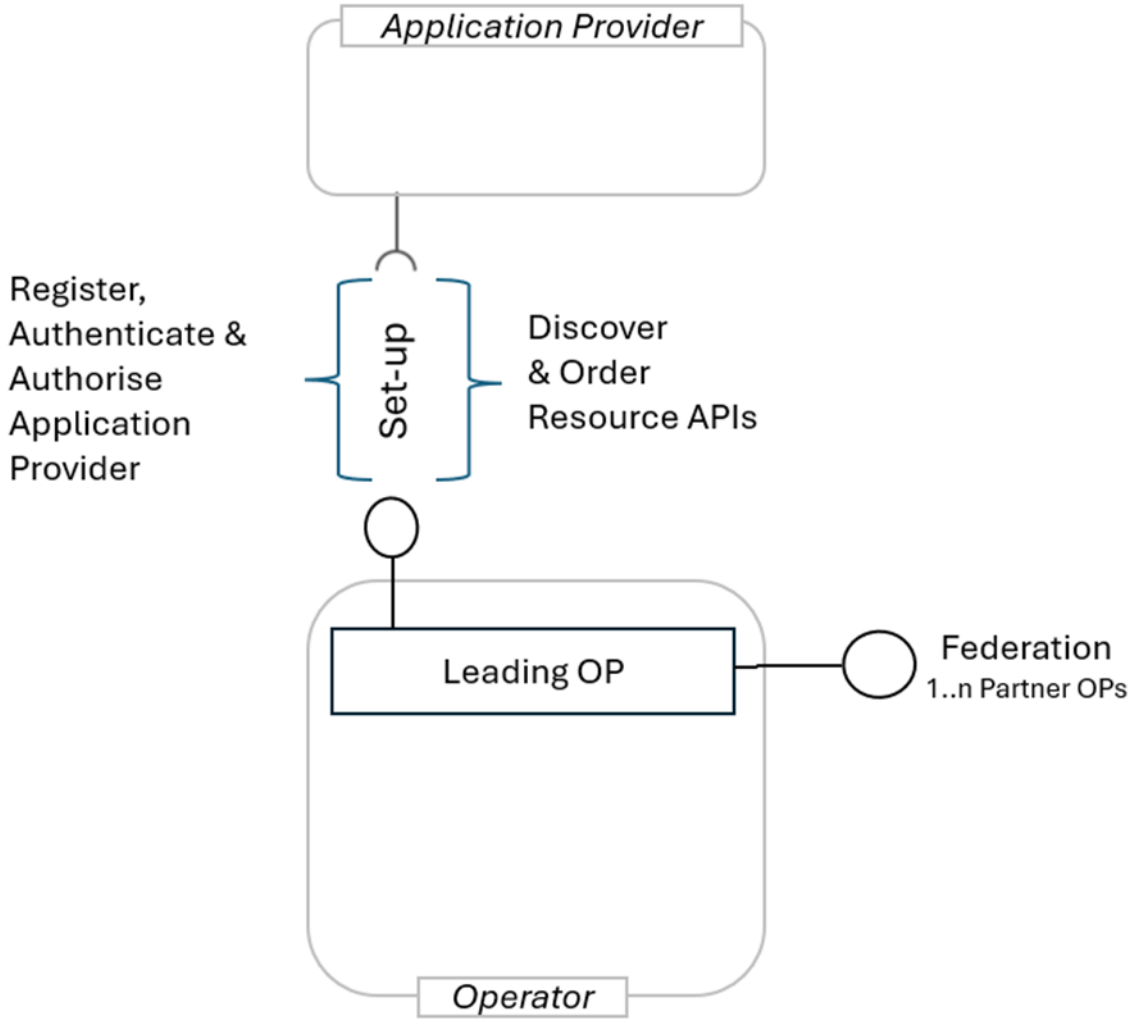

Figure 3: Set-up view – Application Provider perspective

3.5.2.2 Application Functional view

This view represents the requirements 2, 3, 4 and 5 in section 3.1.1 and those in section 5.1.2.

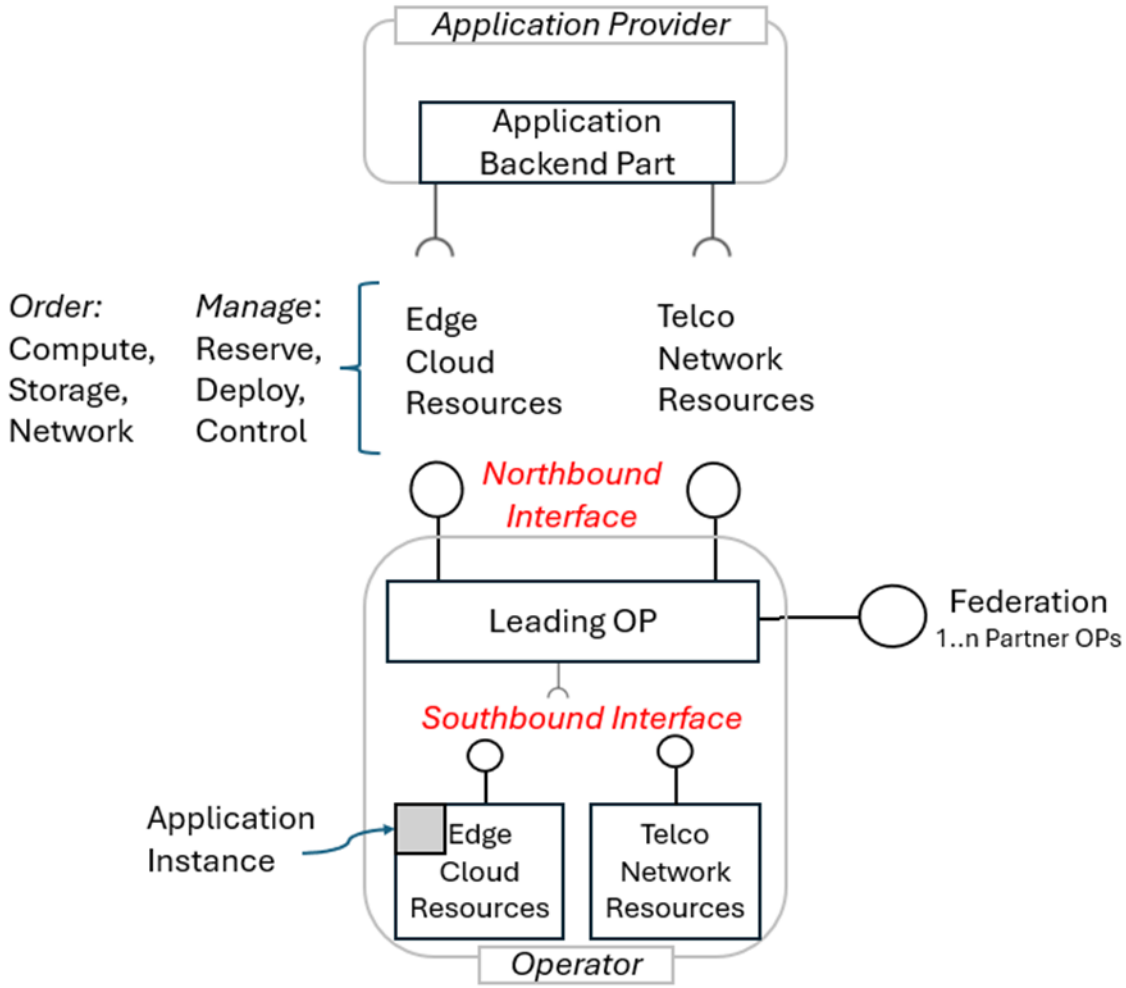

Figure 4: Application Functional view – Application Provider perspective

3.5.2.3 Security view

This view represents requirements 1, 7 and 8 in section 3.1.1 and those in section 3.3.5.

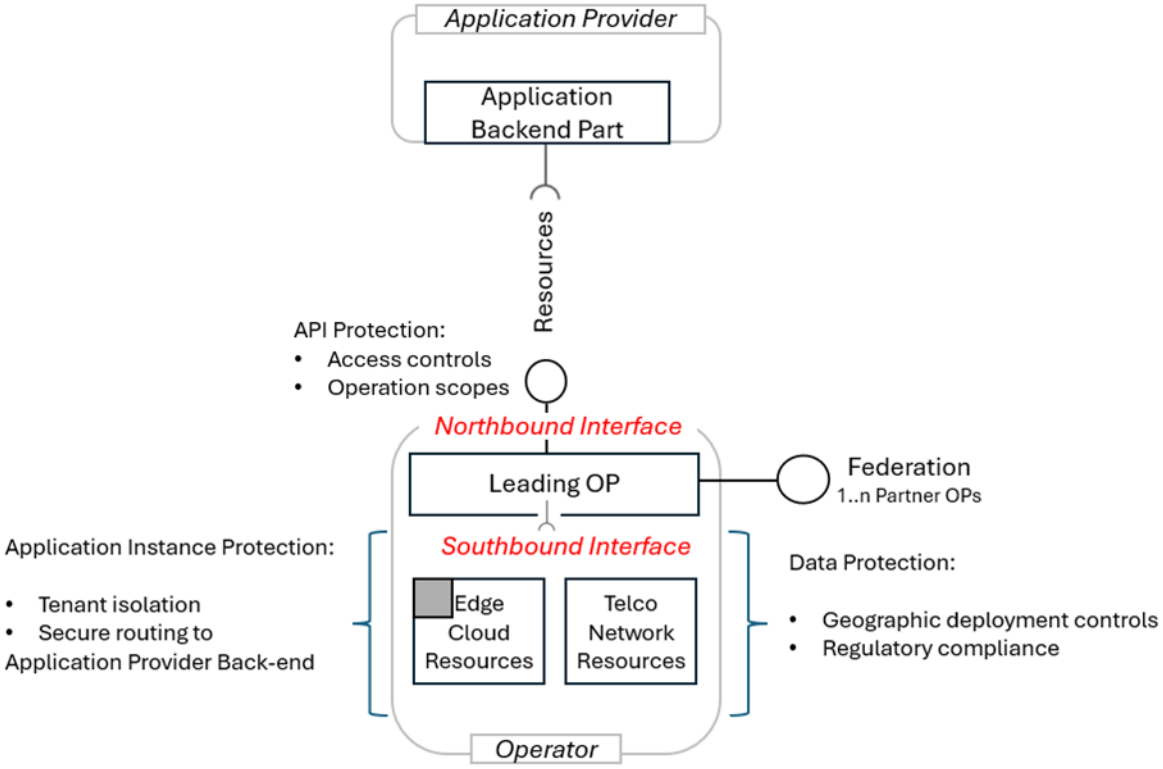

Figure 5: Security view – Application Provider perspective

3.5.3 End User perspective

3.5.3.1 Functional view

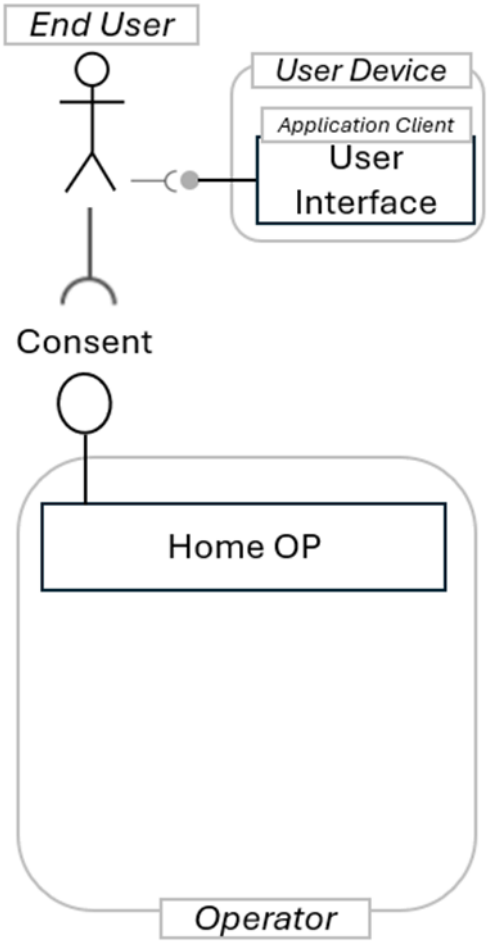

Figure 6: Functional view – End-User perspective

4 Deployment Scenarios Using OP Architecture

The OP architecture, its interfaces and functions are defined in GSMA PRD OPG.02 Section 3 [1].

4.1 Resource Catalogue Synchronisation and Discovery

Operators shall 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. 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 shall be supported over the E/WBI to achieve the above.

4.2 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 section 3.1 of GSMA PRD OPG.02 [1]) 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 section 3.1 in GSMA PRD OPG.02 [1].

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).

4.3 Service Availability on Visited Networks Management

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

Note: It is highly recommended that when entering into a federation agreement, MNOs also agree to enable Local BreakOut (LBO)/ Session BreakOut (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 User Client 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 Uniform Resource Locator (URL). e.g. http://register.op.mnc.mcc.pub.3gppnetwork.org.

During User Client registration, to support the Edge service discovery procedure for the User Client in the Visited OP, the Home OP shall identify that the User Client is in a visited network and provide the User Client with the discovery URL of the Visited OP to redirect the User Client 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 User Client 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/ Service Capability Exposure Function (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 User Client tries to access a service when on visited networks, the Visited OP authorises the User Client using the authorisation information received via the E/WBI from the Home OP of the User Client 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 User Client 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.

4.4 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.

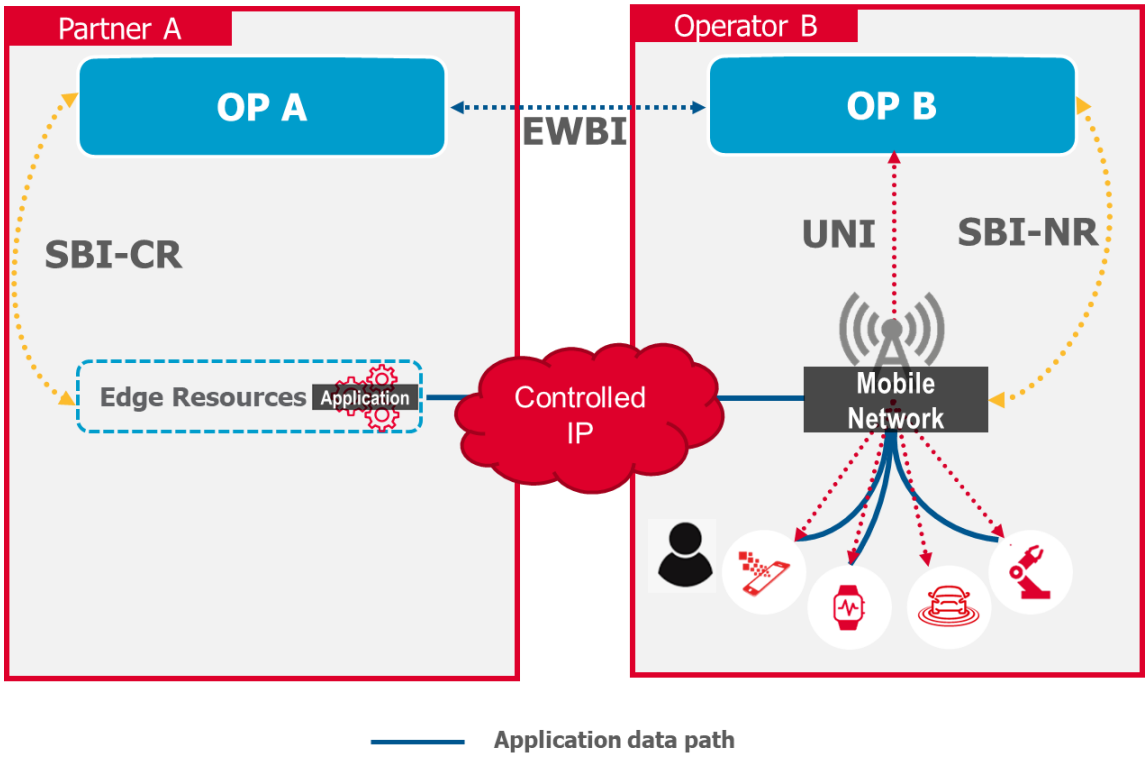

Figure 7: Edge Node Sharing

Figure 7 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 User Clients in their own radio network either by their own edge deployments or via the Partner edge deployments closest to the User Clients.

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/OP 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. Fully Qualified Domain Name (FQDN)) on the Cloudlet at which the Subscriber can connect to the application.

Note: The 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.

4.5 Edge Cloud Resource Monitoring

1. 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.
2. Usage data shall be enabled by default.
3. Event and Performance Metric data shall 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.
4. 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.
5. The monitoring of any resources required for the functioning of the Operator’s charging engine shall be enabled.
6. Data collection shall be subject to the security requirements of GSMA PRD OPG.02 [1] section 2.3 and Annex B.

4.6 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.

There are a few essential scenarios considered for automation:

- starting new Application Instances
- the reconfiguration of resources and network to maintain Service Level Agreements (SLAs)

- the execution of application policies
- the reservation and release of resources

4.7 Low Latency Interaction Between User Clients and Applications in Different Networks

The end-to-end latency between an User Client 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 Vehicle-to-Anything (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 7. 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.

4.8 Containers

4.8.1 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

4.8.2 Container Image and Repository format

An OP shall support the Open Container Initiative (OCI) Image-spec [8], specifying how container images are bundled.

4.8.3 Container runtimes

An OP shall support the Open Container Initiative (OCI) Runtime-spec [9] for container applications on Cloudlets. The Runtime Specification outlines how to run an “OCI Image bundle” unpacked on a disk.

4.8.4 Cloudlet Host OS

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

4.8.5 Supported Architectures

A Cloudlet shall support x86_64 CPU architectures to run containers.

4.9 Virtual Machines

4.9.1 Description

As indicated in section 3.1.1, 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 VM 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: Non-Uniform Memory Access (NUMA), CPU Pinning, use of dedicated core, Affinity/non-affinity, etc.

4.9.2 Guest OS support

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

4.9.3 CPU Architecture support

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

4.10 Serverless

4.10.1 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.

4.10.2 Serverless Computing

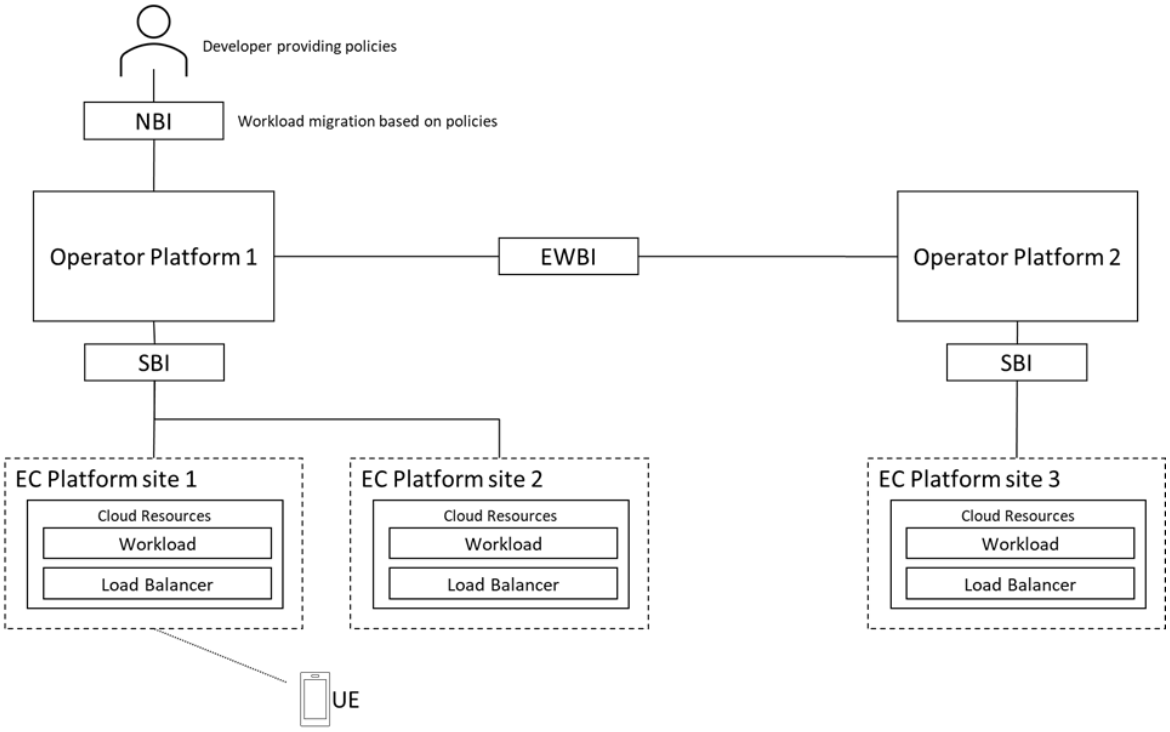

Figure 8: Serverless Computing

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 4.10.4.4.
- 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 User Client. 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.

4.10.3 Serverless Computing Lifecycle

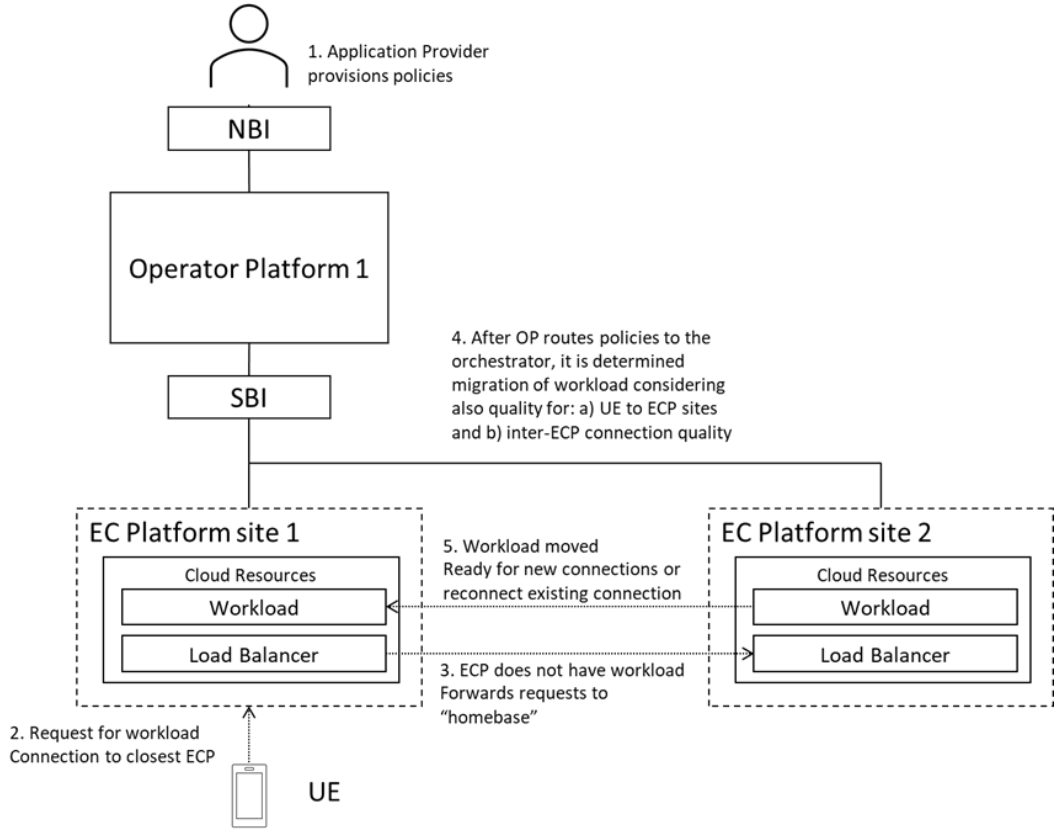

Figure 9: example sequence of a serverless lifecycle

An example sequence of a serverless lifecycle:

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

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

4.10.4 Architectural Components & Considerations

4.10.4.1 Application Packaging

Serverless applications shall be packaged as containers according to the container definition in section 4.8.

4.10.4.2 Serverless Event

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

4.10.4.3 Orchestrator

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

4.10.4.4 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:

1. The number of concurrent connections per Application Instance. Informed through connections request on the Load Balancer.
2. The number of concurrent sessions on an ECP point of presence (as seen by the Load Balancer proxy).

5 Interface Requirements for Enabling Edge Services

5.1 Northbound Interface (NBI) Requirements

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 Continuous Integration / Continuous Development and Deployment (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. The Application Provider may not to be an active participant in orchestration and migration decisions (apart from delivering the intent).

5.1.1 High-level Requirements

1. The NBI shall offer the Edge Cloud, Network or other Operator Capabilities to Application Providers and Aggregators. These capabilities are as follows:
    - a low latency service (and perhaps other application QoS metrics) in a geographical Region;
    - Edge Cloud capabilities are offered whatever Operator the UE is attached to.

5.1.2 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. An OP shall fulfil the following requirements:

1. To allow an Application Provider to “write once, deploy anywhere”, the NBI shall be 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 shall allow 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).

   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.

7. The NBI shall allow 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.
8. The NBI shall allow an OP to accept or reject the request but not to negotiate.
9. The NBI shall allow an Application Provider to upload its application image to the OP.
10. The NBI shall enable an Application Provider to delete its application image.
11. The NBI shall allow an Application Provider to request that their application is instantiated.
12. The NBI shall enable an Application Provider to request that instances of their application are Created, Read, Updated and Deleted (CRUD).
13. The NBI shall allow an Application Provider to specify that their Edge Applications are restricted to a particular geographical area, corresponding to data privacy (General Data Protection Regulation (GDPR)) restrictions.
14. The NBI shall allow 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.
15. The NBI shall allow 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.
16. The NBI shall allow the OP to report telemetry information about the performance of the Edge Cloud service to an Application Provider.
17. 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.
18. The NBI shall provide an Application Provider with information about faults that (may) affect its Edge Cloud service.
19. 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.

20. 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 VMs (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.

21. The other processes of lifecycle management of Edge Applications should follow a similar pattern.
22. 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.

23. 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.
24. The NBI shall allow requesting those network capabilities either as part of the Application Manifest or dynamically.
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.
27. 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.

5.1.2.1 Onboarding and Deployment Profile

5.1.2.1.1 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 4.8 and 4.9, respectively.

5.1.2.1.2 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 [10] 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, 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 User Client's request).
    b) Dynamic: when a User Client 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.
11. 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.

12. Service availability in visited networks required/supported.
13. 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.
14. Session Continuity sensitivity indicating the Edge Application's capabilities to support application session relocations across Cloudlets
15. 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.

5.1.2.2 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.

5.1.2.3 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 Input/Output (I/O) Virtualisation (SR/IOV).

5.1.2.4 Resource Reservation Profile

5.1.2.4.1 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.

5.1.2.4.2 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.

5.1.2.5 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 3.1.
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.

5.1.2.6 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 [10] 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:

a) Computing Resource
b) Storage Resource
c) Network Resource
d) 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 QoE for the End-User at the UE.
The QoS requirements include the following properties:

a) **Bandwidth**, bidirectional data rate between UE and backend application, measured end-to-end with “loopback” application;
b) **Latency**, the round trip delay between UE and backend application, measured end-to-end with “loopback” application;
c) **Jitter**, Variance of round-trip delay between UE and backend application, measured end-to-end with “loopback” application.
d) **Alternative QoS References,** refers to the QoS descriptions that an Application Provider can optionally provide along with the specific QoS.

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:

a) **ID**, a unique identifier of that specific capability to ensure using the same capability over different networks;
b) **Service Level Objectives**, the application requirements for the SLIs of that network capability;
c) **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:

   a) Application mobility allowed/restricted
   b) 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's 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:

   a) Data privacy (GDPR) restriction on the geographical area
   b) 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:

    a) **ID**, a unique identifier of that specific capability to ensure using the same capability over different networks;
    b) **Service Level Objectives**, the application requirements for the SLIs of that network capability;
    c) **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.

d) **Granularity scope**, the definition of granularity of capability requested, depending on the type (e.g. event/notification based).

5.1.2.7 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.

5.1.2.8 Image Management

An Application Provider deploys the application by providing an image for containers (per section 4.8) or VMs (per section 4.9. 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.

5.1.2.9 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 Key Performance Indicators (KPIs) monitoring and rollback.
- Support of additional services like GitOps, for facilitating Application Provider CI/CD integration.

5.1.2.10 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 Infrastructure as a service (IaaS).

5.1.2.11 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.

5.2 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 section 3.1 of GSMA PRD OPG.02 [1].

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

5.2.1.1 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.

5.2.1.2 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).

5.2.1.3 Application and Resources Management

5.2.1.3.1 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 support instantiation requests for applications depending on Containers and VMs that comply with the format criteria specified in sections 4.8 and 4.9, respectively.
2. An OP receiving an instantiation request through its E/WBI shall get in charge of the management of the application:

    a) An OP receiving an instantiation request through its E/WBI shall be responsible for the Operator deployment criteria management.

b) 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.
c) An OP receiving an instantiation request through its E/WBI shall be in charge of the application mobility management.

3. The E/WBI shall forward the application mobility notifications and procedures towards the Leading OP for management with the Application Provider.

5.2.1.3.2 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 4.8 or VMs per section 4.9) 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 5.2.1.3.4).

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).

5.2.1.3.3 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.

5.2.1.3.4 Edge Node Sharing Service

Edge node sharing is a scenario wherein an OP, when serving the User to Network Interface (UNI) requests originating from (its own) User Clients, decides to provide the application

from the Edge nodes of a Partner OP (where the application is available). Like the scenario discussed in section 4.4, 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 User Clients;
- In a non-roaming scenario where an OP needs to allow its own User Clients, 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". Then, the OP serving the Subscriber can use that information for sharing connection parameters with the User Client (e.g. application IP address or access token).

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

5.2.1.3.5 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.

5.2.1.4 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.

5.2.1.5 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 User Clients.

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

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

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

5.3 Southbound Interface

5.3.1 Southbound Interface – Charging Function (SBI-CHF) Requirements

Editor's note: This is a placeholder for future requirements.

5.3.2 Southbound Interface – Cloud Resource (SBI-CR) Requirements

5.3.2.1 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 GSMA PRD OPG.02 [1]. 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.

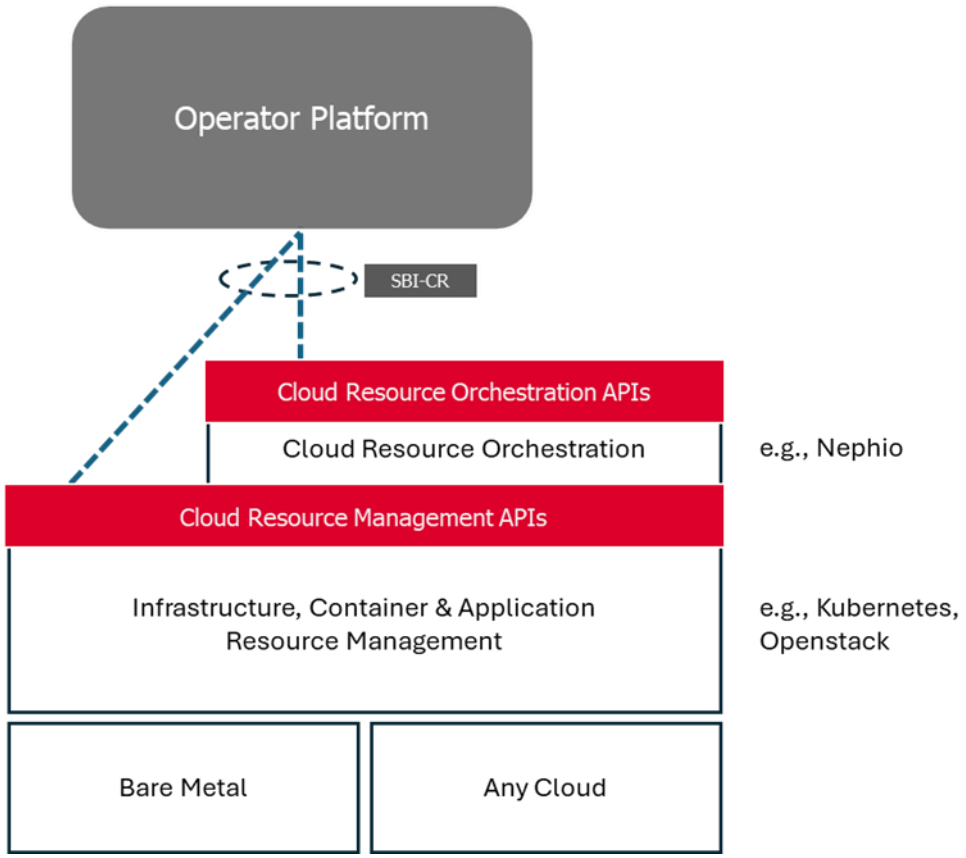

Figure 10: Possible SBI CR integrations

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

5.3.2.2 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.

5.3.2.3 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.

5.3.2.3.1 Cloud Resources Management

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

An OP 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.

5.3.2.3.1.1 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 OP 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.

5.3.2.3.1.2 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/ Container Infrastructure Service Manager (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 section 3.1 of GSMA PRD OPG.02 [1]);
- 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 10).

5.3.2.3.1.3 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.

5.3.2.4 Security Requirements

Based on the attack surface analysis provided in Annex B, 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.

5.3.3 Southbound Interface – Edge Interworking Network (SBI-EIN) Requirements

5.3.3.1 High-Level Requirements

1. An OP shall provide the interface for control/management of the EIN between two Edge Clouds.
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 Edge Clouds, and optionally provide security guidelines.

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

5.3.3.2 Security requirements

Based on the attack surface analysis provided in Annex B, 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 Software Defined Network (SDN) control plane.

5.3.4 Southbound Interface – Network Resources (SBI-NR)

Editor's note: This is a placeholder for future requirements.

5.3.5 Southbound Interface – Operation & Maintenance (SBI-OAM) Requirements

Editor's note: This is a placeholder for future requirements.

5.3.6 Southbound Interface – Privacy Management (SBI-PrM) Requirements

Editor's note: This is a placeholder for future requirements.

5.4 User to Network Interface

5.4.1.1 High-Level Requirements

1. The UNI shall be universal, meaning that the Application Provider does not have to modify its applications for different Operators or OPs.
2. The UNI between the User Client (typically located in the UE) and an OP 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 5.4.1.3.3.

5.4.1.2 General Requirements

1. The User Client should be implemented on EU software, e.g. through an Software Development Kit (SDK) or OS add-on.
2. The UNI shall allow the User Client to discover the existence of an Edge Cloud service.
3. An OP's UNI shall allow the User Client registration process with the OP, which entails the following:

    a) It enables the End-User device to establish an encrypted communication channel with the OP.
    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 Authentication, Authorisation and Accounting (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 User Client's usage tracking. For example, to support integration with the network Operator's billing infrastructure.

4. An OP's UNI shall allow the User Client to trigger the selection of a Cloudlet through the OP.
5. An OP's UNI shall allow the User Client to trigger the instantiation of an Application Instance on the selected Cloudlet.
6. An OP shall measure network performance metrics for tracking the average latency characteristics of the edge network.
7. Based on metrics and location information, the User Client 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 5.1) allowing to provide OP services to Application Clients on UEs that do not support the UNI.

5.4.1.3 User First Attachment

5.4.1.3.1 General

When a User Client 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. User Client connectivity should be available to allow initiating this request. User Client connectivity is out of the scope of this document.

5.4.1.3.2 Edge Cloud Service Discovery

The User Client 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's network.
2. An OP shall offer a general URL that can be constructed based on Operator information available to the UE, e.g. MCC/MCN, to which a User Client can request an Edge Cloud service.
3. A UNI User Client request shall include identity information and parameters:

   a) For a UE,

      i. UE ID, e.g. MSISDN, Generic Public Subscription Identifier (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.

5.4.1.3.3 First-time Registration for Non-SIM UE

1. User Clients for non-SIM UEs shall do a first-time(bootstrap) registration with an OP on the very first connection to the OP.
2. 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.

3. 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.
4. 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.
5. 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.

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

5.4.1.3.4 UE Authentication and Authorisation

An OP shall authenticate the User Client 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 User Client.
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 5.4.1.3.3).
4. In addition, the OP shall provide a mechanism to allow efficient authorisation of the UE for subsequent interactions.

5.4.1.3.5 Cloudlet Selection

An OP processes all the information from the User Client, 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 User Client 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 User Client's selection of QoS (including bandwidth and latency),
   d) Takes account of the capacity and usage of the Cloud Resources (e.g. CPU and memory) at the various Cloudlets and the Network Resources (e.g. congestion),
   e) The choice of Cloudlet may result in the UE needing to be redirected to a different UPF /PGW.

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

5.4.1.3.6 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.

5.4.1.4 Security Requirements

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

1. The UNI shall provide an authentication mechanism to enable access only by authenticated and authorised User Clients and OPs. Therefore, mutual authentication shall be provided between the User Client and the OP.
2. The UNI shall provide secure communication between the User Client 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.

5.4.1.5 Establishing Chain-of-Trust Between Architectural Elements

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

- the UE and the OP;
- the User Client and the OP;
- the Application Client and the Edge Application;
- the Operator's Network and the Edge Application;
- 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 User Client and establish a chain of trust between the User Client 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's Network and the Edge Application are implementation-dependent.

The mechanism shall use a registration procedure from the User Client 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 an User Client. 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.

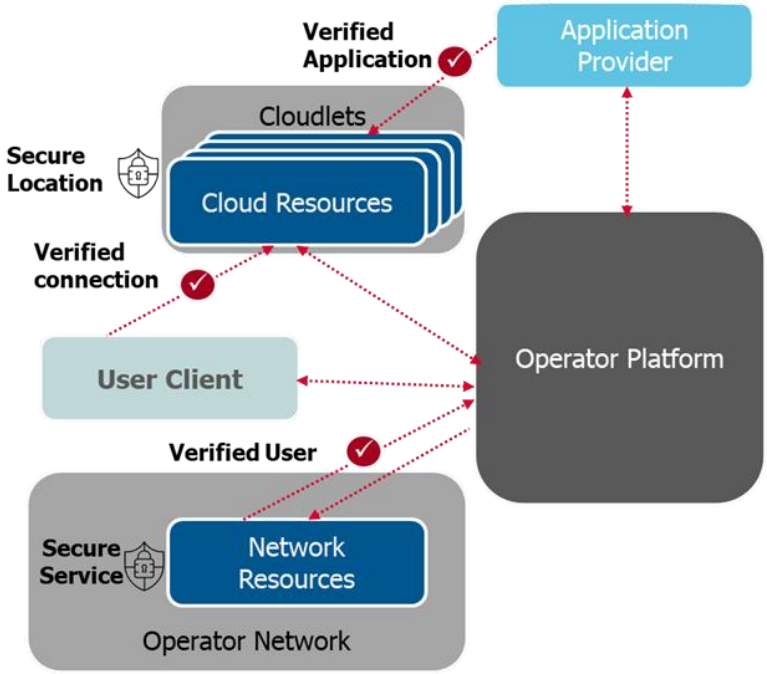

Figure 11: Chain of Trust: High-level Diagram

6 Requirements on Functional Elements

This section defines the requirements of the OP’s functional elements to enable edge services.

6.1 Exposure Functions

1. 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..
2. The Exposure Functions shall support Application Life Cycle scenarios.
3. 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.
4. 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.

6.2 Federation Functions

6.2.1 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 needed for the edge services 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.

6.3 Integration Functions

6.3.1 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:

- Edge node resources and load.
- Network resources and load.
- Network usage forecast.
- Edge usage forecast.
- Application availability (already deployed/onboarded on Edge Node).
- Reserved resources availability.
- UE mobility supported.
- Network mobility supported (integration with data packet core).
- Specific constraints/barring for users, application or Edge Nodes selection.
- Specific considerations to abide by commercial agreements between involved parties.

6.3.2 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.

6.3.3 Edge Application Relocation

6.3.3.1 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 (User Mobility) (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, see GSMA PRD OPG.02 [1] section 5.4.

Note: the application cannot reject or delay the IP address change.

- Due to User Mobility, 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 session(s).

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 Relocation includes support for a change of Operator and OP.

6.3.3.2 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.

- Application Session Relocation triggers from the core network relayed by the OP:
    - Related to User Mobility that causes a change in session anchor (PGW/UPF) network point, see GSMA PRD OPG.02 [1] section 5.4;
    - Related to User Mobility that causes a change in the serving network (i.e. PLMN change);
    - Related to User Mobility that causes a change in the Application Client's IP address, see GSMA PRD OPG.02 [1] section 5.4;
    - Related to User Mobility (for instance, for each Edge Cloud location, the Operator identifies the set of base stations that it most naturally supports);
    - 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);
    - Related to usage forecasts about its edge compute resource and network;
    - 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 TS 23.501 section 6.3.3.3 [11].

- Relocation triggers from the Application:
    - Related to its measurement of QoS parameters (such as latency, jitter and bandwidth);
    - 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, Application Edge Part, Application parts in a central cloud).

6.3.3.3 Application relocation Conditions/Restrictions

1. An OP shall be able to consider the Application-specific requirements for managing relocation over different Edge Nodes.

2. 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.
3. An OP shall manage the Application Session Relocation for all the edge services associated with each User Client.
4. An OP shall consider the relocation sensitiveness of the Applications.
5. An OP shall take into account the active Application Edge Part for considering the relocation.
6. An OP shall ensure that all the active Application Edge Parts are relocated correctly when network change is required.
7. An OP shall not perform Application Session Relocation onto another Operator's network domain if an active Application does not support roaming.
8. An OP shall perform a network relocation if an Application requires mandatory relocation.

6.3.3.4 Application Session Relocation (Server-Side)

An OP needs to manage the reconfiguration of the Edge Application environment, selecting a new Edge Instance to have the Application session available.

1. An OP shall be able to ensure that the selected Edge Instance has enough capacity.
2. An OP shall be able to request the instantiation of the Application Edge Part on a 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 Instance.
4. An OP shall ensure that the information is available to configure an Edge Application's traffic flows towards the selected Edge Instance statically or dynamically via the SBI-NR, e.g. via the Operation and Management plane.

6.3.3.5 Session Mobility (User Side)

Application session mobility is mandatory for maintaining the session continuity on stateful applications, where the Subscriber's Edge Application traffic 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 for the session to be relocated to the new Edge Instance. 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 relocation procedure. Note that the service may be degraded or even lost. Note also that, as part of the initial policy phase, the Application may give a permanent indication that it cannot handle relocation.

5. The Application shall confirm the completion of the relocation of the Edge Application session onto the new Edge Instance to the OP.
6. Relocation of 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.

6.3.3.6 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 Session Relocation based on Application requirements and different information, e.g. network and physical location or Edge Resources usage.
    a) 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.
    b) When a new Application Instance is required to host the Application session, the OP may use the following information to select an adequate target Edge Cloud to host the Application Instance.
        i. Application Provider criteria
        ii. Application data privacy policies
        iii. Operator defined policies, e.g. cost functions associated with Edge Clouds
        iv. Location information on the UE received through the SBI-NR interface
        v. Edge Sites and available resources at the UE's location as received through the SBI-NR interface
        vi. The Application Session Continuity mode of the UE PDU session
    c) 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.
    d) 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 UE’s PDU session to the new UE Cloudlet where the Application Instance is created.
    e) If an Application Instance is already available, the OP may use that instance's information to generate the traffic steering rules for the UE’s PDU session in the selected Edge Site.
    f) 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.
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.

5. The OP shall coordinate the different procedures with the Edge Application.

a) An OP shall provide capabilities over the NBI interface for Application Providers or Aggregators to perform the application session/context relocation functions
b) 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: 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 Performance Profile 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., UE IP address change event.

6. The OP shall coordinate the different procedures with the Edge Application, from the original node to the target.
   a) When SSC mode 2 applies, the OP shall assist Edge Applications to prepare for and handle short disruption in Session Continuity via Service APIs
   b) When SSC mode 3 applies, the OP shall assist Edge Applications to prepare for and handle concurrent sessions with Application Instances via Service APIs
7. The OP shall coordinate the different procedures with the Application Client on the User Client.
8. 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.

9. An OP shall ensure that the User Client is forced to apply any relocation procedures.
10. Network location may not be needed in case of service degradation due to an Edge Node saturation.

6.3.3.7 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 QoS Performance Profile for Edge Application sessions with the UEs 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.

6.3.3.8 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 Performance Profile information
- Current access network and access network change events
- UE Location events based on UE privacy permission

Note: It is expected that the User Clients 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.

6.3.4 Service Availability on Visited Networks

1. Service availability on visited networks shall be considered to allow the users to use an Edge Application outside of their Operator's 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 Application would be delivered from home network resources with the inherent latency and service degradation.

2. 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.
3. Based on the information received from Home OP and the internal policies, the Visited OP shall instantiate the Edge Application on a Cloudlet for use by the User Client.
4. The Visited OP shall be in charge of selecting the Cloudlet within the Visited OP best placed to host the Application Edge Part (including when the user device moves within the Visited OP).
5. The Visited OP shall be able to provide the Application Session Relocation monitoring events information to an Application's Leading OP over the E/WBI.

6.3.5 Application Operation and Management

An OP shall expose to an Application Provider a set of management capabilities. These capabilities are as follows:

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 Domain Name System (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.

7 Service Flows

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

7.1 User Client/UE Registration to the OP using UNI

7.1.1 User Client Registration to the OP - Home OP

This procedure describes the registration between the User Client and an OP, allowing the User Client to be authenticated and authorised to access the service.

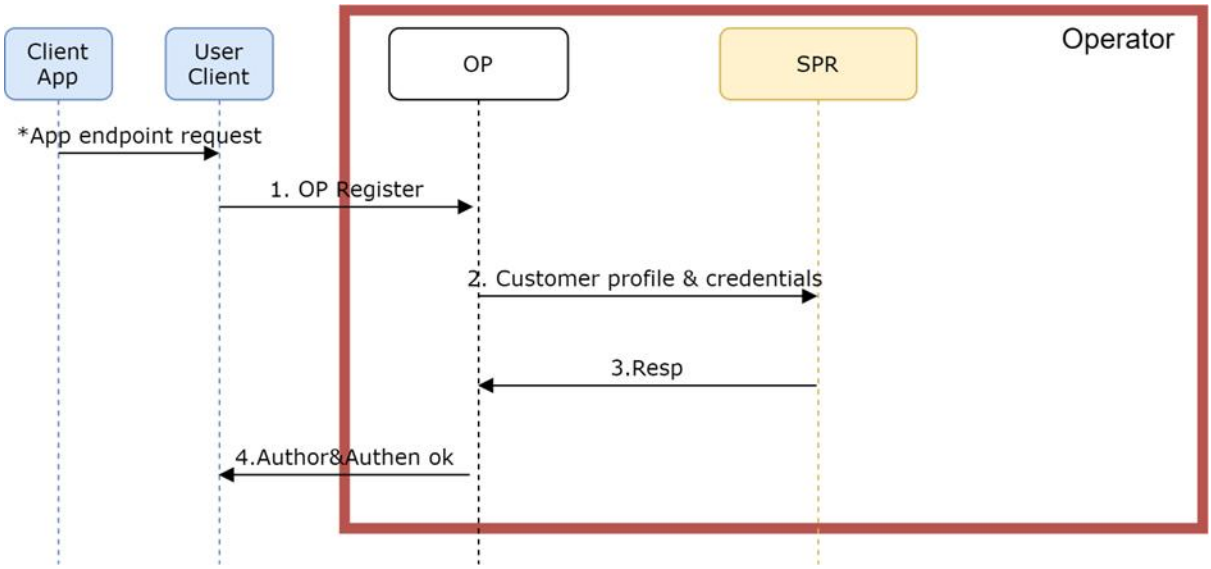

Figure 12: User Client Registration to the OP - Home OP

1. An User Client 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 User Client in the registration request.
4. The User Client receives the authentication validation and is authorised to request the OP services from that moment onwards (e.g. Cloudlet discoveries).

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

7.1.2 User Client Registration to the OP - Visited OP

This procedure describes the User Client registration with an OP while accessing the service from a visited network. For such cellular roaming, two models exist as defined in section 4.3:

- **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 7.1.1. Figure 13 shows the relations between the networks in this

case. This scenario comes with limitations on application availability due to increased latency (see section 7.5).

- **Local breakout**, to access Edge Nodes available in the visited network. This model is preferred because the Edge Cloud service is provided closer to the User Client then.

  In this case, the Home OP 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 14 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 7.5) 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.

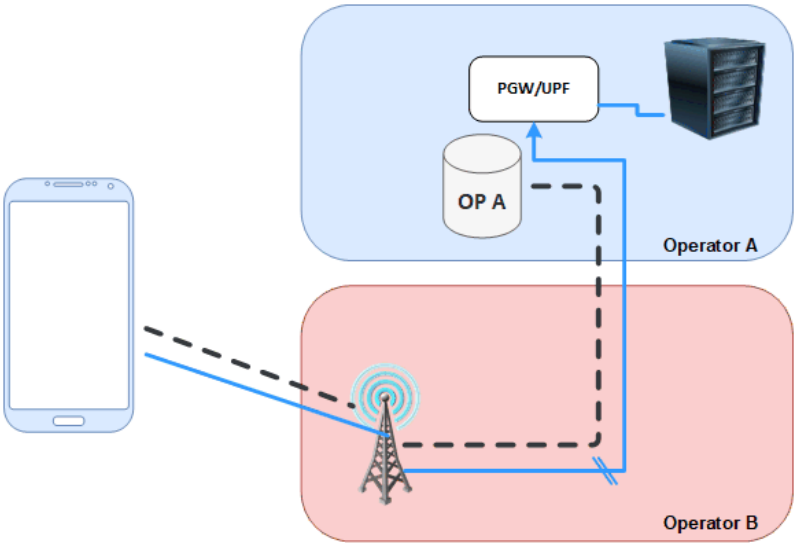

Figure 13: Roaming access to OP and Edge Resource - home routing

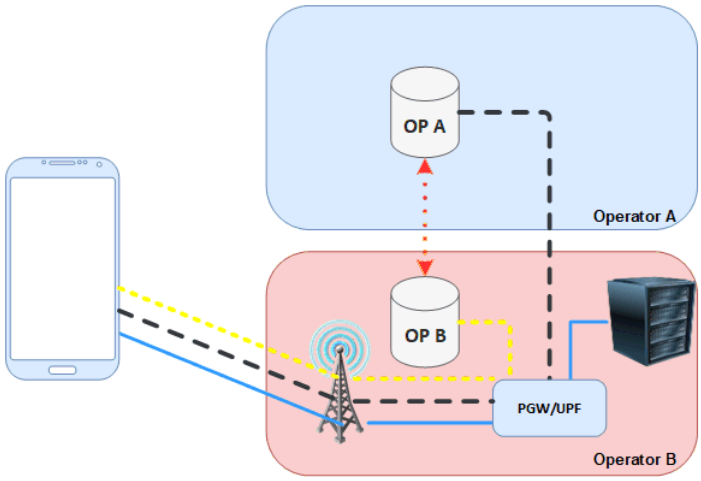

Figure 14: Roaming access to OP and Edge Resource – local breakout

7.2 Service delivery by the OP without UNI

The procedure for service delivery by the OP when UNI is not present, is documented in GSMA PRD OPG.02 [1].

7.3 Edge discovery in the home network

This procedure describes the edge discovery by an User Client 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 GSMA PRD OPG.02 [1].

7.4 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.

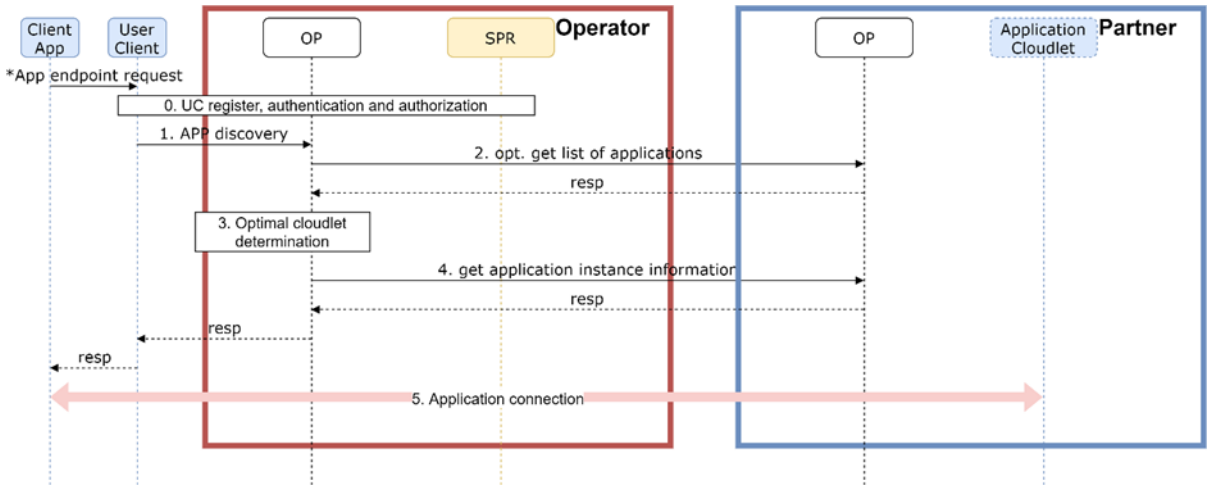

Figure 15: Edge discovery in an edge-sharing Partner network

1. An User Client on a UE requests a discovery query for a particular application. The User Client previously registered with the OP as in the procedure described in section 7.1.

2. Optional. Operator’s OP (Home OP) may trigger a discovery request for the applications available on the Partner's resources.

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

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

7.5 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 section 7.1) also apply to Edge Discovery. When using home-routing, the discovery is similar to the case described in section 7.3. 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 when UNI is not available would use the flow provided in GSMA PRD OPG.02 [1].

7.6 Application deployment In the Home Operator Domain

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

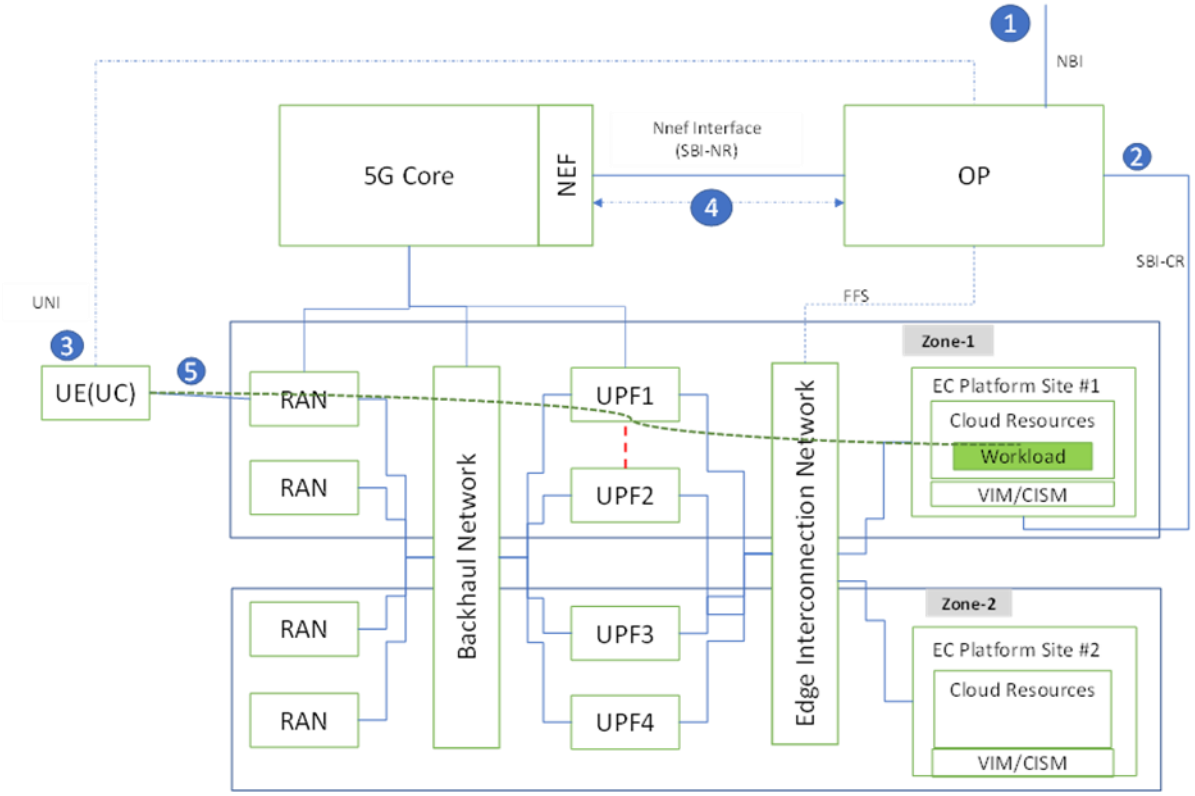

Figure 16: OP Application Deployment flow with 5G core network interaction

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 16, 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 “Edge Cloud 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 “Edge Cloud Platform Site #1”.
5. The User Client connects through UPF1 with the instance on Edge Cloud Platform Site #1 using the communication endpoints returned in step 3.

7.7 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.

7.8 Application Service and Session Continuity in the home network

Figure 17 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 “Edge Cloud Platform Site #1” belongs to Zone-1, and “Edge Cloud 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 17 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 7.6).

Figure 17 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.

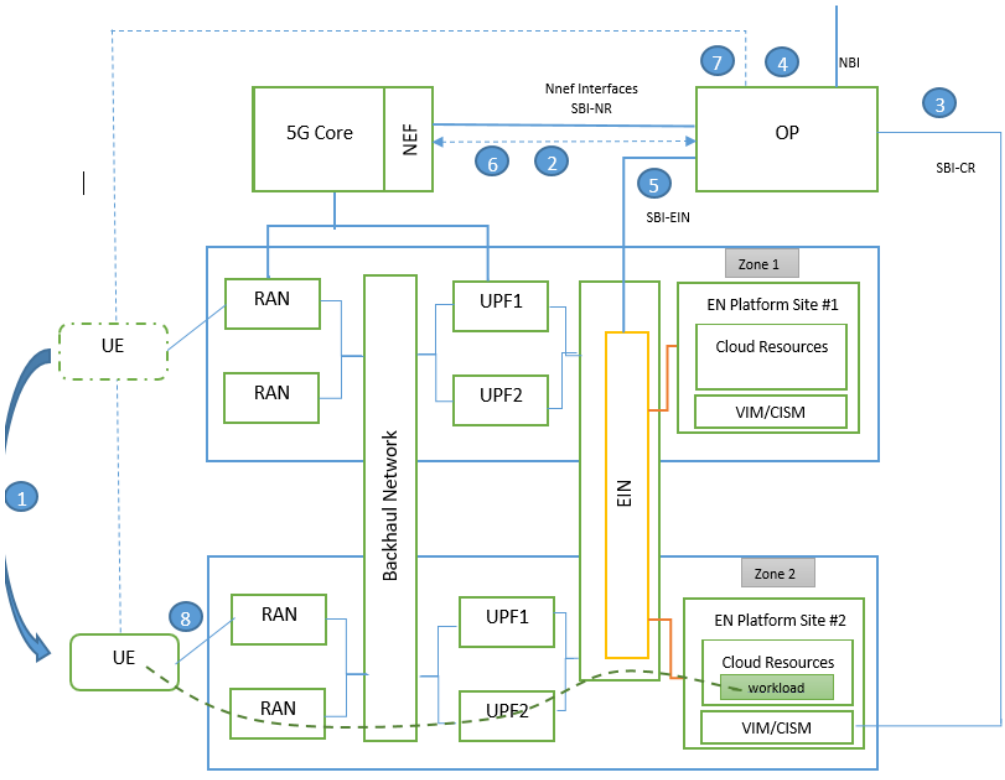

Figure 17: OP Application Relocation flow

Following are the details of the various steps marked in numbers in Figure 17, 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 3.3.7.2. Figure 17 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. “Edge Cloud 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 “Edge Cloud Platform Site #1” and “Edge Cloud Platform Site #2 ”. If not, then the OP will request and establish the EIN connection between the two Edge Clouds. 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 User Client of the new end points for the application while the User Client 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 User Client 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 “Edge Cloud 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 the User Client over the UNI to enable the Application to prepare for any application-level relocation tasks

Note: The EIN and interface with an OP are for future studies but can be used by the OP or the Edge Applications for application state management across the Cloudlets.

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 [27] defines the APIs to support MEC federation.

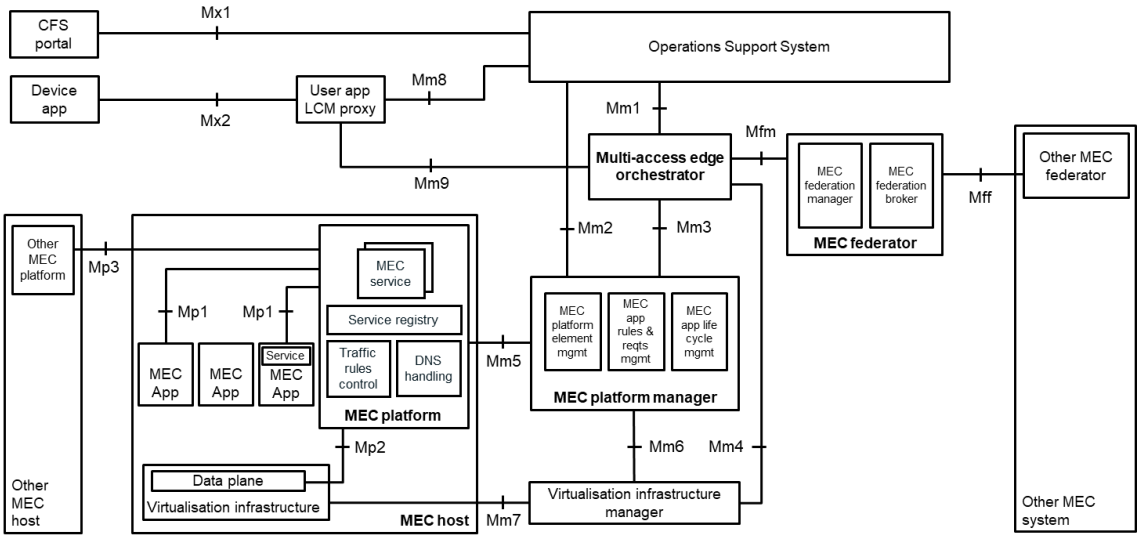

Figure 18: Multi-access edge system reference architecture variant for MEC federation in ETSI MEC003 [18]

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 Life-Cycle Management (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 3GPP TS 23.558 [17]. In clause 6.2 of Rel18 3GPP TR 23.958 [29], 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 Policy Control Function (PCF)/ Policy and Charging Rules Function (PCRF), application traffic configuration APIs, and service APIs exposed by SCEF/NEF.

Note: The EEC(s) may be provisioned with the ECS address(es) information also by the Session Management Function (SMF) at 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 [25] and OPG.05 [26] 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 [29].
- 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 points for the OPs 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 [24].
- 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 [30].

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 [5].
- 3GPP ReportingNetworkStatus NEF API [5] and SCEF API [6].
- 3GPP Monitoring NEF API [5] or SCEF API [6].

- 3GPP AsSessionWithQoS NEF API [5] or SCEF API [6].
- 3GPP ChargeableParty NEF API [5] or SCEF API [6].
- 3GPP DeviceTriggering NEF API [5] or SCEF API [6].
- 3GPP ServiceParameter NEF API [5].

Annex B OP Security

B.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 OP 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 B.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 B.3 is an initial mapping of the threat vectors affecting the OP 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." [21].

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

In section B.3, the threat vectors are mapped to the OP architecture. The mapping is shown in Figure 20, labelled by the identifiers provided for the threat vectors of section B.2.

The threat vectors are in various categories, and each category is covered in a separate subsection of B.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 B.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.

B.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., [19], [16]) and has specified the main security aspects in [30]. 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 [20].

B.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 20. These tables are provided in section B.3
- Use the tables of section B.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.

B.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.

B.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.

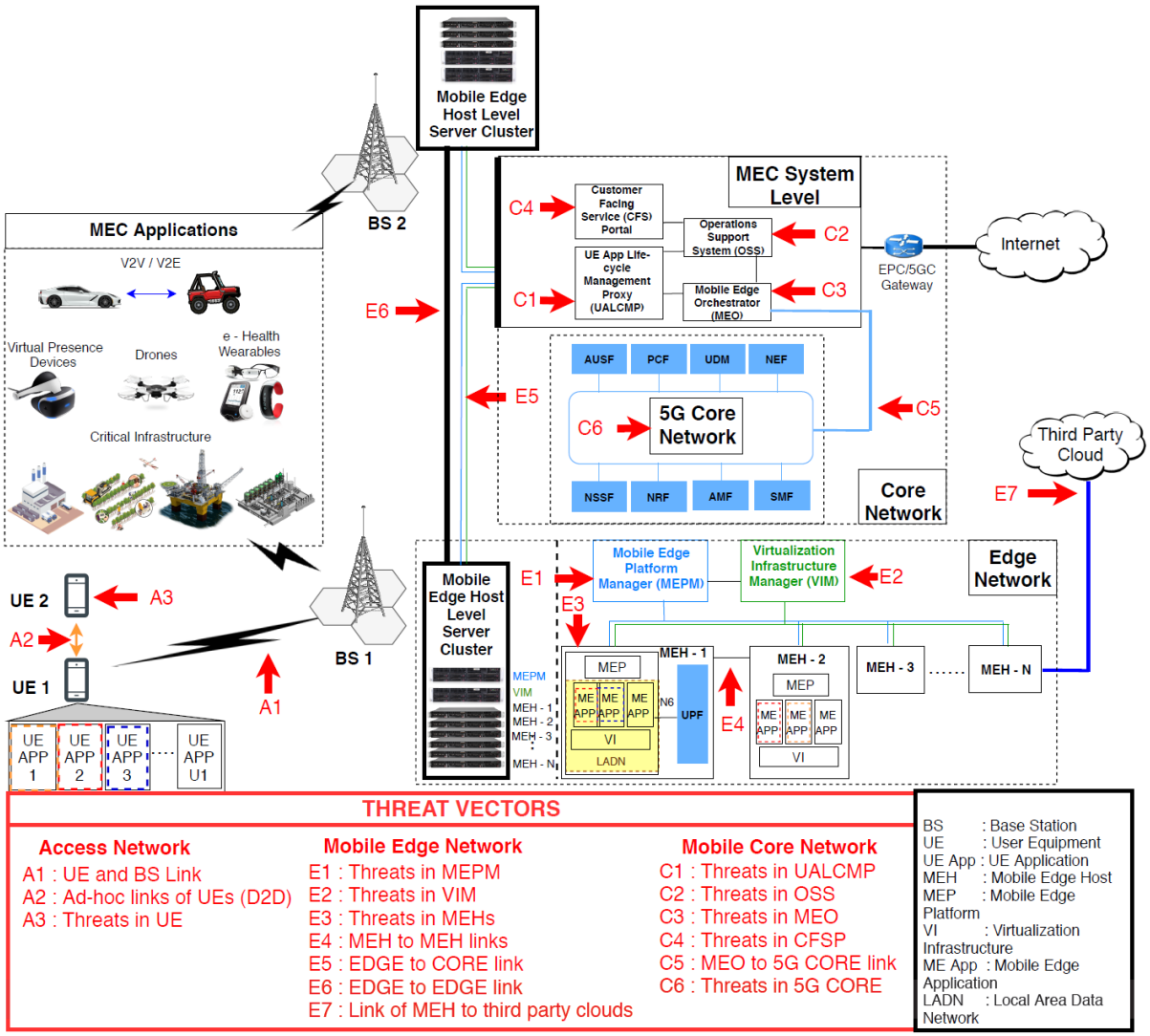

Figure 19: ETSI ISG MEC Access, Edge and Core Threat Vectors (from [15])

Table 1 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 [19] (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 Base Transceiver Station (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 [20]
MEC2	MEC system discovery supporting authentication, authorisation, identity management, etc. [20]
MEC3	MEC platform discovery supporting authentication, authorisation, identity management, etc. [20]
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 [19]
3GPP2	Authentication and Authorisation between EEC and ECS – EDGE-4 [19]
3GPP3	Authentication and Authorisation between EES and ECS – EDGE-6 [19]
3GPP4	Edge Data Network authentication and authorisation [19]
3GPP5	Edge Data Network user identifier and credential protection [19]
3GPP6	Transport security for the EDGE-1-10 interfaces [19]
3GPP7	Security of network information provisioning to local applications with low-latency exposure [19]
3GPP8	Authentication and authorisation in EES capability exposure – SCEF/NEF northbound APIs [19]
3GPP9	Security of EAS discovery procedure [19]
3GPP10	Authorisation during edge data network change [19]

Table 1: Threat Vector Descriptions (adapted from [15], [19], [20])

B.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 1.5. 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 OP 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 6.

3GPP SA3 has recently engaged in studying edge computing security aspects in the 5G core network in [19]. They identified security gaps, locations, and solutions, in an approach similar to that of [15]. Table 2 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 19 and summarised in Table 1 (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 2: Threat Vectors derived from [19] with a location indication

B.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 [20]. 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 3 is extracted informally from [20] to align the security gaps and solutions with the threat vector/name/countermeasure approach of other sources. The threat vector tags are

applied to figures depicting threat vectors, and the countermeasures are adapted from the proposed solutions.

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

Threat Vector (TV) ID	Description	Solution
MEC1	Required signalling for secure inter-MEC systems	Creation of Federation Manager network element to provide secure signalling
MEC2	MEC system discovery supporting authentication, authorisation, identity management, etc.	Definition of a new reference point (Mff-fed) to support secure interaction between Federation Managers
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 3: Derived Threat Vectors and Solutions from [20]

B.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, Network Functions Virtualisation (NFV), edge computing, and other related areas.

They identified numerous vulnerabilities and countermeasures in [14]. Table 4 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 6 include 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 4: Threat vectors identified in [14]

B.3 OP Threat Vectors and Countermeasures

Figure 20 depicts the threat vectors identified in the OP architecture.

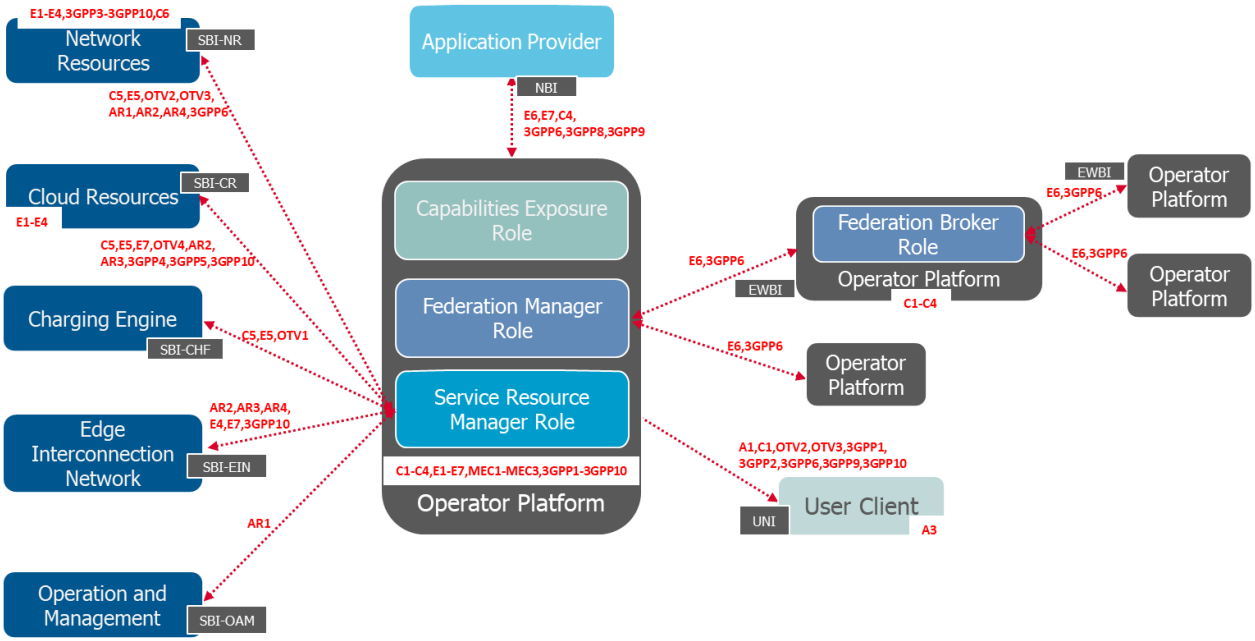

Figure 20: OP Threat Vectors

B.3.1 Access Threat Vectors

According to Figure 20, 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/ E-UTRAN Node B (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 the User Client, are also relevant.

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

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

Table 5: Access Threat Vectors and Countermeasure Recommendations (from [15])

B.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 19 (from [15]). Instead, they were added in Figure 20.

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 4.

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	Service Function Chain (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 Proxy Mobile IPv6 (protocol) (PMIPv6)
AR4	PMIPv6 based security protocol for Smart Home Internet of Things (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 6: Access Threat Vectors and Countermeasure Recommendations (from [15], [14])

B.3.3 Core Threat Vectors

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

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

Threat Vector (TV) ID	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 7: Core Threat Vectors and Countermeasures (from [15])

B.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 Host-based Intrusion Detection System (HIDS)
E1, E2, E3	Policy based VM Intrusion Detection System (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 8: Edge Threat Vectors and Countermeasures (from [15])

B.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 19. Instead, they are provided in Figure 20.

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 section B.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 (Distributed Denial of Service) attacks
OTV2, OTV3	Genetic Algorithms

Threat Vector (TV) ID	Countermeasure Recommendation
OTV2	Leveraging edge computing to mitigate IoT-DDoS attacks
OTV2	Hardening resource management (elaborated from paper)
OTV2	Anomaly detection on QoE requests (elaborated from paper)
OTV3	Private LAN Service (PLS) model for multi-user multi-carrier MEC channels
OTV3	Secure UE (modified from Unmanned Aerial Vehicle (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 9: Core Threat Vectors and Countermeasures (from [15])

B.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 10: 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

Privacy Objectives	Recommendations
O6	Declaring minimum specification requisites of UE for subscribing Mobile Edge Services
O7	Optimal utilisation of UE resources with embedded privacy-enhancing mechanisms
O8	Comply with GDPR legislation.

Table 11: Privacy Objectives and Recommendations (derived from [15])

Some general privacy considerations are provided in GSMA PRD OPG.02 [1].

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.

B.4 Recommendations from 3GPP

3GPP has specified the security aspects for supporting edge computing in [30]. Considering the functional mapping presented clause 6.2 of TR 23.958 [29], it is possible to list the following recommendations:

- For the OP's SBI-NR (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 [31] shall be considered: mutual authentication based on client and server certificates using TLS. The support of TLS is mandatory with at least version 1.3. The NEF shall authorize the requests from an AF using the OAuth-based authorization mechanism.
- For the OP's NBI interface (mapped to the 3GPP EDGE-3 reference point), it should support the use of HTTP/2 with "https" URIs as specified in RFC 9113 [32] and RFC 9110 [33]. Mutual authentication shall be supported over the NBI (EDGE-3). TLS shall be used with at least version 1.3.
- For the OP's EWBI interface (mapped to the 3GPP EDGE-9/10 reference points), X.509 certificates shall be used for authentication. The identities in the end-entity certificates shall be used for authentication and policy checks. Identities in the end-entity certificate shall be based on endpoint information (e.g., URI, FQDN, IP address etc.) as described in TS 23.558. Mutual authentication (among OPs) should be supported. TLS shall be used with at least version 1.3. 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 [32] and RFC 9110 [33].

It should be noted that TLS is recommended to be used to provide integrity protection, replay protection, and confidentiality protection for OP interfaces.

Note: EWBI Authorization using a local authorization policy is FFS in the current version of this document.

B.5 Guidance for the implementation, deployment and operation

The following guidance is to be considered for the Edge Resources:

- Services, processes, and Tenants running in containers and VMs, and their data, need to be protected.
- Note: Approaches to protecting them include process isolation via name-spacing or hypervisor controls and trusted enclaves.
- The Cloud Resources need to provide security mechanisms to prevent attacks from containers or VMs, of which Docker or VM Escape attacks are examples.
- The Cloud Resources need to provide security mechanisms to counteract attacks on the SBI-CR aiming to prevent data availability, such as Denial of Service (DoS) attacks.

Annex C OP Managed DNS Service related to Edge Applications

C.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.

C.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).

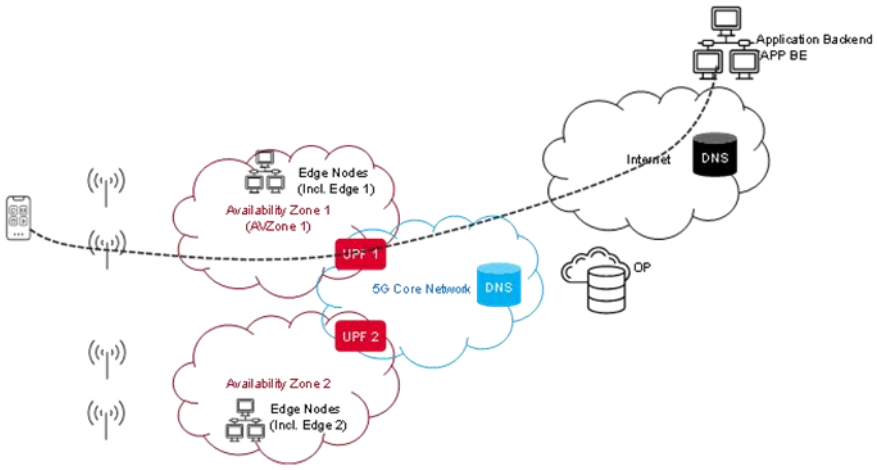

Figure 21: A reference deployment of DNS service with OP

C.3 Role of the OP

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.

C.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.

C.5 Implementation Guidelines

An OP can 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 D Local interface on an End-User device

Using edge computing through an OP should be as easy as possible from an Application Provider's perspective. As envisioned in the OP architecture, the UNI interface between the User Clients 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 22).

Note: By nature, such a common implementation would be device platform-specific; see section D.1.3 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 D.1.1). 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 User Client to consume the 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 User Client 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.

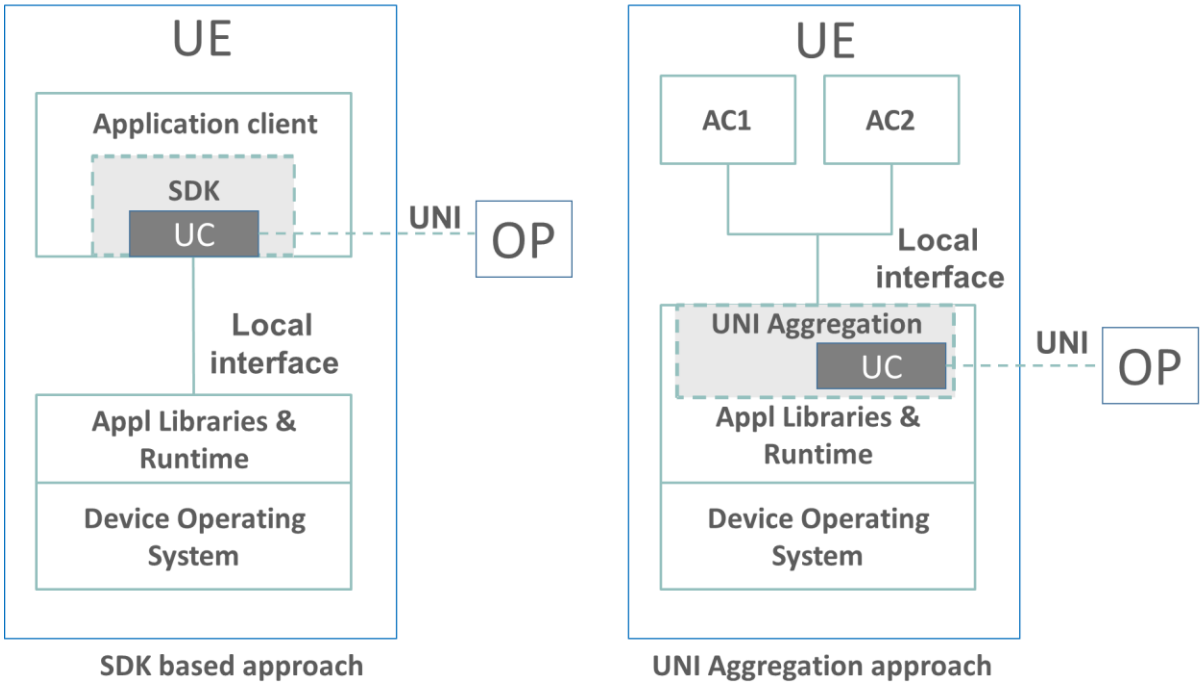

Figure 22: OP Device side architecture (local interface)

Note: As per two of the possible approaches for building UNI support for Application Clients, i.e. SDK and the UNI aggregation, Figure 22 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.

D.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.

D.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.
  - Mobile Subscriber Integrated Services Digital Network Number (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.

D.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 [22] 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.

D.1.3 Key considerations for architectural requirements on the local interface

The client applications or the User Clients 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.

Annex E Document Management

E.1 Document History

Version	Date	Brief Description of Change	Approval Authority	Editor / Company
1.0	09 May 2025	New PRD OPG.11 based on Requirements in PRD OPG.02 v8.0	ISAG	Sandra Ondrusova / CK Hutchison
2.0	19 Feb 2026	Update implementing OPG.11 CR1002	ISAG	Tom Van Pelt / GSMA

E.2 Other Information

Type	Description
Document Owner	Operator Platform Group
Editor / Company	Cristina Santana Casillas / Telefonica

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.