D2.1: Initial definition of use cases

Deliverable

Document ID	NECOS-D2.1
Status	Final
Version	1.5.2
Editors(s)	Luis M. Contreras (TID), Rafael A. López da Silva (TID)
Due	30/04/2018
Submitted	06/06/2018

Abstract

This deliverable reports on the relevant use cases in the problem space to be addressed by the NECOS project, as well as it introduces a number of scenarios of interest from which the project requirements were derived, guiding the NECOS architecture design. In addition, the document includes a preliminary analysis on the business ecosystem and the stakeholders of a NECOS environment, as support of elaborating further commercial and business insights enabled by the proposed solution.

EUB-01-2017

TABLE OF CONTENTS

EUB-01-2017

Table of Figures

Contributors

Contributor	Institution
Luis M. Contreras	Telefónica Investigación y Desarrollo (TID)
Rafael A. López da Silva	Telefónica Investigación y Desarrollo (TID)
Francesco Tusa	University College London (UCL)
Alex Galis	University College London (UCL)
Kleber Vieira Cardoso	Federal University of Goias (UFG)
Leandro Alexandre Freitas	Federal University of Goias (UFG)
Christian Esteve Rothenberg	University of Campinas (UNICAMP)
David Fernandes Cruz Moura	University of Campinas (UNICAMP)
Marcos Rogério Salvador	University of Campinas (UNICAMP)
Antonio Jorge Gomes Abelem	Federal University of Pará (UFPA)
Billy Anderson Pinheiro	Federal University of Pará (UFPA)
Joan Serrat	Universitat Politècnica de Catalunya (UPC)
Javier Baliosian	Universitat Politècnica de Catalunya (UPC)
Ilias Sakellariou	University of Macedonia (UOM)
Panagiotis Papadimitriou	University of Macedonia (UOM)
Lefteris Mamatas	University of Macedonia (UOM)
Sofia Petridou	University of Macedonia (UOM)
Augusto Neto	Federal University of Rio Grande do Norte (UFRN)
Felipe Dantas	Federal University of Rio Grande do Norte (UFRN)
Marcelo Ribeiro Nascimento	CPqD Telecom Research and Development Center
Douglas Salles Viroel	CPqD Telecom Research and Development Center
Cesar Marcondes	Federal University of São Carlos (UFSCAR)
Paulo Ditarso Maciel Jr.	Federal University of São Carlos (UFSCAR)
Fábio Luciano Verdi	Federal University of São Carlos (UFSCAR)
Rafael Pasquini	Federal University of Uberlândia (UFU)

Reviewers

Reviewer	Institution
Alex Galis	University College London (UCL)
Lefteris Mamatas	University of Macedonia (UOM)
Rafael Pasquini	Federal University of Uberlândia (UFU)

Acronyms

API	Application Programing Interface	NEP	Network Equipment Provider
B2B	Business-to-Business	NFV	Network Function Virtualization
C3	Command, Control and Communications	NFVI	NFV Infrastructure
CAPEX	Capital Expenditure	NFVO	NFV Orchestrator
CD	Content Delivery	NOP	Network Operator
CPU	Central Processing Unit	OPEX	Operating Expenditure
CSC	Communication Service Customer	OTT	Over-the-top media services
CSF	Critical Success Factor	PaaS	Platform as a Service
CSP	Communication Service Provider	PE	Provider Edge
CU	Cloud/Central Unit	PNF	Physical Network Function
DC	Data Centre	PoC	Proof of Concept
DC-GW	Data Centre Gateway	PSC	Public Safety Centre
DCO	Data Centre Operator	PSCmer	Public Slice Customer
DCSP	Data Centre Service Provider	QoE	Quality of Expectation
DIP	Data centre Infrastructure Provider	QoS	Quality of Service
E2E	End-to-End	RAN	Radio Access Network
eMBB	Enhanced Mobile Broadband	RB	Resource Broker
FaaS	Function as a Service	RTT	Round-Trip Time
H/W	Hardware	SC	Slice Customer
HSC	Hybrid Slice Customer	SCD	Smart City Data
ICT	Information and Communication Technologies	SDN	Software Defined Networking
IoTDM	Internet of Things Device Manager	SECaaS	Security as a Service
ISP	Internet Service Provider	SLA	Service Level Agreement
KPI	Key Performance Indicator	SlaaS	Slice as a Service

LSDC	Lightweight Slice Defined Cloud	SLB	Slice Broker
MaaS	Management as a Service	SPST	Service Provider Slice Tenant
MEC	Mobile / Multi-Access Edge Computing	S/W	Software
MIM	Metropolitan Intelligent Monitoring	uRLLC	Ultra-Reliable and Low Latency Communications
mMTC	Massive Machine-Type Communications	VIM	Virtual Infrastructure Manager
MNO	Mobile Network Operator	VISP	Virtualisation Infrastructure Service Provider
MPLS	Multiprotocol Label Switching	VNF	Virtual Network Function
MTC	Metropolitan Tourist Centre	VLSP	Very Lightweight Network & Service Platform
NCP	Network Connectivity Provider	vRAN	Virtual Radio Access Network

Definitions

Term	**Definition**
Customer / Tenant	This actor consumes Information and Communication Technology (ICT) services offered by the Data Centre Operator (DCO). Usually, this entity utilises the ICT services so as to store, retrieve, process and transfer business information under a specific contract and Service Level Agreement (SLA) guarantees provided by the DCO. The objective of a Customer is to receive ICT services with a guaranteed SLA that correspond to a reasonable and fair price, or even favourable, in case of such an opportunity arises. A Customer may also opt to participate in a demand-response curtailment program and receive some cost benefits, while contribute to the implementation of policies opting for green-powered ICT services and Data Centres. Along the document the terms customer and tenant are used as equivalent.
Data Center Operator (DCO)	This actor is responsible for operating a Data Centre (DC) or a group of DCs that are part of the same administrative domain. It is responsible for providing ICT services to customers, adhering to specific Service Level Agreements (SLA). The primary objective of a DCO is to offer ICT services under the promised SLA, maximising the utilisation of the ICT infrastructure, while maintaining the electricity consumption, as a main DC expenditure, within levels that are proportional to the ICT processing work delivered. Furthermore, the DCO would take advantage of any demand-response curtailment incentives that would be available, so as to decrease its operational cost and provide greener ICT services. In order to benefit from the incentives, the DCO might take immediate actions or even schedule them for the near future.
Data Center (DC)	A data centre is a centralized repository physical and / or virtual, for the storage, management, services and dissemination of data and information organized around a particular body of knowledge or pertaining to a particular business. The S/W of the data centres are realised and based on the computing cloud principle and relevant technologies. In general, the hardware (H/W) of the data centres includes redundant server and network nodes capabilities as well as backup power supplies, redundant data communication connections, environmental controls (e.g., air conditioning, fire suppression) and various security devices. Large data centres are industrial scale operations using as much electricity as a small town and sometimes are a significant source of air pollution in the form of diesel exhaust.
Functional Requirement	It is a description of what a system is supposed to do and it defines a function, or a feature of a system, or its components, capable of solving a certain problem or replying to a certain need/request. The set of functional requirements present a *complete* description of how a specific system will function, capturing every aspect of how it should work before it is built, including information handling, computation handling, storage handling and connectivity handling.
System Design	It is a plan for implementing functional requirements.
Non-functional requirement	It is a specification criteria that can be used to judge the operation of a system, rather than specific behaviours; it is a description of how well a system performs its functions; it represents an attribute that a specific system must have. The non-functional requirements are controlled by other aspects of the system.
Business objectives	It is a description in business terms of what must be delivered or accomplished to provide value.

System boundaries / limits	It defines the constraints and freedoms in controlling the system. Limits can be determined by analysing how the behaviour of the system depends on the parameters that drive it. Some limits would lead to unexpected and significant behavioural changes of the system, for example the unpredictable boundaries or changes in the scale of magnitude. Some other limits are determined by non-common interactions between the system components.
Architecture	It is a plan for implementing non-functional and functional requirements within the system limits/boundaries. It is conceptual model that defines the structure, behaviour, and a number of views of a system within the system limits
Use Case (UC)	It is a descriptor of a set of precise problems to be solved. It describes steps and actions between stakeholders and/or actors and a system, which leads the user towards a value added or a useful goal. A UC describes what the system shall do for the actor and/or stakeholder to achieve a particular goal. Use-cases are a system modelling technique that help developers determine which features to implement and how to gracefully resolve errors.
Scenario	It is a narrative of foreseeable interactions of user roles ('actors') and the technical system, which usually includes computer hardware and software. A scenario has a goal, which is usually functional. A scenario describes one way that a system is or is envisaged to be used in the context of activity in a defined time-frame. The time-frame for a scenario could be (for example) a single transaction; a business operation; a day or other period; or the whole operational life of a system. The scope of a scenario could be (for example) a single system or piece of equipment; an equipped team or department; or an entire organization.
DC Governance	It is a framework, which enables operators to describe their goals and objectives, through high-level means and govern their network. Includes the derivation of DC policies from the business goals through the use of semantic techniques.
Actor	It is a person, group or organization with an interest in a specific viewpoint of a system.
Viewpoint	It is a representation of a whole system from the perspective of a related set of concerns.
Accessibility	It represents the degree to which a system, device, service, or environment is available to as many people as possible. Accessibility can be viewed as the "ability to access" and benefit from some system or entity.
Availability	It represents the degree to which a system is in a specified operable and committable state at the start of a task. It is the proportion of time a system is in a functioning condition.
Certification	It refers to the confirmation of certain characteristics of an object, element of system. This confirmation is often, but not always, provided by some form of external review, assessment, or audit.
Configuration	It is a function establishing and maintaining consistency of a system and/or its performance. It is changing system's functional and physical attributes with its non-functional requirements, design, and operational information throughout its life.
Compliance	It represents the conformance to a rule, such as a specification, policy, standard or regulation.
Extensibility	It represents the ability to extend a system and the level of effort and complexity required to realize an extension. Extensions can be through the addition of new functionality, new characteristics or through modification of existing

	functionality/characteristics, while minimizing impact to existing system functions.
Interoperability	It represents the ability of diverse systems and subsystems to work together (inter-operate). It is also a characteristic of a system, whose interfaces are completely understood, to work with other systems, present or future, without any restricted access or implementation.
Maintainability	It is a characteristic of design and installation, expressed as the probability that an element of a system will be retained in or restored to a specified condition within a given period of time, when the maintenance is performed in accordance with prescribed procedures and resources.
Operability	It is the ability to keep a system in a safe and reliable functioning condition, according to pre-defined operational requirements.
Performance	It describes the degree of performance of a system (according to certain predefined metrics, e.g., convergence time)
Privacy	It is the ability of system or actor to seclude itself or information about itself and thereby reveal itself selectively.
Resilience	It is the ability to provide and maintain an acceptable level of service in the face of faults and challenges to normal operations.
Reliability	It is the degree to which a system must work. Specifications for reliability typically refer to stability, availability, accuracy, and maximum acceptable/tolerable bugs.
Robustness	It is the ability of a system to cope with errors during execution or the ability of a system to continue to operate despite abnormalities in input or in environment context.

Executive Summary

NECOS project aims to develop a lightweight system for enabling cloud networking – native integration of cloud computing and advanced networking - slicing capabilities in multi-domain scenarios. Two basic use cases are taken as reference for developing such lightweight system: Multi-access Edge Computing and Telco Cloud. These use cases are currently taking traction in the industry as next step on extending computing capabilities in proximity to the end users, to enable challenging scenarios.

This document describes a number of scenarios on top of those use cases in order to identify requirements to be supported by the NECOS solution. Additionally, critical success factors and KPIs of the system are defined, to guide the architectural development and validation task during the project.

Complementary, this document provides a first insight on the NECOS stakeholders as a way of describing the ecosystem enabled by NECOS.

1 Introduction

The NECOS project aims to address the challenging problem of realising slicing in cloud networking - native integration of cloud computing and advanced networking -environments in a lightweight manner with the following characteristics: being service-agnostic (but adapting the slices to the desired service characteristics), automating the process of optimal cloud configuration in multi-domain federated environments, and providing a uniform management with a high-level of autonomicity.

This deliverable describes in detail the two use cases considered as a starting point of the NECOS project, namely the Telco Cloud and the Multi-Access Edge Computing (MEC). These use cases act as platforms supporting a number of relevant scenarios, controlled and managed through the NECOS platform, leveraging on the concept of slicing as a form of segregating multiple services in the same (federated) cloud computing substrate.

The Telco Cloud and the MEC have emerged as the evolutionary alternatives for the network operators and service providers to extend the capillarity of the existing centralized data centres, enabling new environments of geographically spread cloud capabilities. The availability of computation, storage and network resources as well as the kind of workloads and services to be supported differ from the traditional ones leveraging on the large centralized Data Centres (DCs). This fact imposes the need of new mechanisms to manage and control the computing infrastructure for this new kind of demands.

Key on the evolution of the networks and cloud environments is the concept of multi-tenancy, enabling novel network slicing ideas on top of the existing telecom networks. The paradigms of network virtualization, mainly based on the Network Functions Virtualization (NFV) approach, and network programmability through Software Defined Networking (SDN), have tremendously fostered this evolutionary view, appearing as tools leveraging the implementation of slicing. However, these two paradigms are not sufficient for network slicing. Especially for the latter, there is a prevalent NFV-centric view in the industry which is not necessarily satisfying all the cases and scenarios. Overloading the NFV orchestration and management artefacts with slicing mechanisms could lead to inefficient cloud or network partitioning. NECOS intends to address such a gap with a new slicing architecture that is valid for multiple scenarios and a wide-range of requirements, i.e., a first relevant analysis is described here.

The focus of WP2 is to describe the use cases and associated service scenarios, the requirements deriving from them, and their associated business models. This deliverable provides a preliminary insight on the four key NECOS aspects, described below, serving as an input for the rest of work packages in the project, especially for the WP3 defining the NECOS architecture.

The deliverable contributes to the project objectives in the following manners:

- Objective 1 (NECOS Platform) – The deliverable identifies a number of critical success factors and requirements to be satisfied by the NECOS platform.
- Objective 2 (Service Provisioning) – The deliverable presents a number of different service scenarios to be provisioned on top of the NECOS infrastructure enabling the Telco Cloud and MEC environments.
- Objective 3 (Uniform and Efficient Management) – The deliverable reports a number of KPIs to be supported by the NECOS infrastructure through a unified control and management system.
- Objective 4 (Impact Validation) – The deliverable provides alternative service scenarios that highlight the impact of the NECOS approach and enable its validation.

1.1 Structure of this document

The deliverable is structured in a manner that helps the reader to better understand the rationale on the need of a lightweight and service-agnostic cloud slicing architecture and relevant mechanisms.

First, the key NECOS characteristics are summarized in Chapter 2, as an introduction to the project scope and objectives. Chapter 3 surveys existing business models related to the NECOS project which serve as a reference for defining the NECOS actors and roles as well as identifies gaps that can benefit from the project outcomes, i.e., both aspects are manifested as the NECOS business model. A core contribution of the deliverable is reflected in Chapter 4. The chapter elaborates on the Telco Cloud and the MEC use cases and proposes alternative relevant scenarios. These scenarios are reported in a way that key information can be extracted from them, such as: *(i)* the technical enablers required from the scenarios; *(ii)* their critical success factors (CSFs) and key performance indicators (KPIs); *(iii)* their mappings to the four main NECOS characteristics described in the following section; and, *(iv)* their functional and non-functional requirements. This fundamental information will be further processed in the context of the other work packages, especially in the WP3 that covers the definition of the NECOS system architecture. Chapter 5 recaps all the requirements, CSFs and KPIs, facilitating the readiness of the provided information. Finally, Chapter 6 gives a summary of the deliverable and defines the next steps in the project implementation.

1.2 Contribution of this Deliverable to the project and relation with other Deliverables

As discussed above, this deliverable covers a collection of scenarios applicable to the use cases defined in NECOS, i.e., the Telco Cloud and the MEC environments. Such scenarios allowed us to identify a number of requirements that enrich the functional and non-functional capabilities of the NECOS system architecture to be reported in D3.1. Furthermore, the NECOS proof-of-concept demonstrators will be based on these scenarios, documented in D6.1. Finally, the forthcoming D2.2 will be based on the D2.1 outcomes, since it will evolve the detailed use-cases, scenarios, platform requirements, the economic ecosystem and the NECOS stakeholders, through an iterative process that considers but also influences all the parallel work in the other work packages.

2 NECOS key characteristics

An integration of network and cloud computing domains, and their related management operations, will allow for huge savings and an ability for greater flexibility in service provisioning. These operations are mainly carried out separately, so combining them is a goal of NECOS. Maturing paradigms such as SDN and network virtualization, for instance with NFV, when properly designed and deployed, can help in fulfilling the requirement of making such an innovation for services more reliable, faster, and simpler.

Slicing is associated to partition of resources, being able to create and redefine these partitions as needed. A slice is a grouping of physical or virtual (network, compute, storage) resources which can act as a seemingly independent sub-cloud, sub-network and can accommodate service components.

The key characteristics of the Cloud Slicing are, as depicted in Figure 1:

- The concurrent deployment of multiple logical, self-contained and independent, shared or partitioned slices on a common infrastructure platform.
- Dynamic multi-service support, multi-tenancy and the integration means for vertical market players.
- The separation of functions, simplifying the provisioning of services, the manageability of networks, and integration and operational challenges especially for supporting communication services.
- Network operators/ ISP and Cloud infrastructure owners can exploit slicing for: reducing operations expenditure, allowing programmability and innovation necessary to enrich the offered services, for offering tailored services, and allowing network programmability to OTT providers and other market players without changing the physical infrastructure.

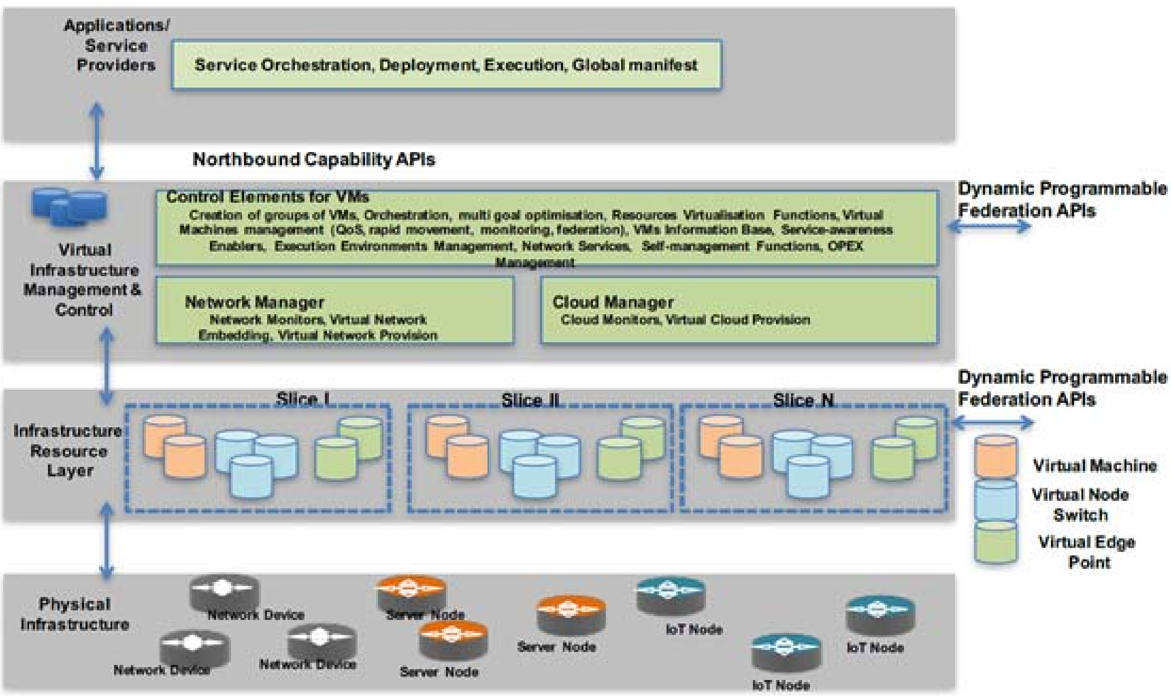

Figure 1 - Cloud slicing main enablers

In NECOS, we propose the Lightweight Slice Defined Cloud (LSDC) architecture to present the concept of Cloud Slicing (*Slice as a Service*), to be built on top of the already available cloud platform features and functions. The Lightweight Slice Defined Cloud (LSDC) represents a novel approach for automating the process of optimal cloud configuration by creating the Cloud Slice concept across all of resources in a set of federated data centres, as well as providing a uniform management of the currently separated computing, connectivity and storage resources. In addition to the above Cloud slicing characteristics, the LSDC adds the following differentiating factors:

1) It empowers a new service model – the Slice as a Service, by dynamically mapping service components to a slice. The enhanced management for the infrastructure creates slices on demand and slice management takes over the control of all the service components, virtualized network functions and system programmability functions assigned to the slice, and (re)configure them as appropriate to provide the end-to-end service.
2) It enables easy reconfiguration and adaptation of logical resources in a cloud networking infrastructure, to better accommodate the QoS demand of the Slice, through using software that can describe and manage various aspects that comprise the cloud environment.
3) It allows each aspect of the cloud environment – from the networking between virtual machines to the SLAs of the hosted applications – to be managed via software. This reduces the complexity related to configuring and operating the infrastructure, which in turn eases the management of the cloud infrastructure.
4) The LSDC platform will offer the ability to a specific cloud provider to federate his own infrastructure with other cloud providers with different configurations in order to realize virtualized services through the use of the Slice as a Service concept. The users of the LSDC APIs and platform will be able to create virtual services that can span the merged cloud infrastructure offered by different cloud providers. This concept is not purely technical, it can also encompass business, cultural, geographical or in any other domain.

Since LSDC platform and APIs main users are meant to be individuals seeking to create a Slice, or other cloud infrastructure providers seeking to form a federated virtual cloud in order to participate in the mechanisms to provide the Slice as a Service, LSDC is a promising solution to a widespread audience, since it may be seen as an enhancer to several scenarios, like those described in this document.

2.1 Barriers overcome by NECOS approach

In current scenarios, different cloud providers offer different interfaces and services. This prevents users from effectively choosing the more economically viable cloud for their needs; locking them to the cloud provider they chose to run their first service. It is a task left to the consumer to make sure that resources required to run their services are present in all the cloud infrastructures, limiting automatic workload distribution across different cloud providers.

NECOS technological advancement in cloud federation will help in offering a unique set of functionality and resource access across different cloud providers, enabling a currently unfeasible scenario for the user like cloud hoping to achieve the best ROI for their expenditures. NECOS will provide advancements in cloud management and orchestration allowing for tightly coupled cloud federation, enabling the automatic deployment of services among several cloud infrastructures.

NECOS orchestration and supporting tools, such as monitoring, allow the cloud operator to achieve high degrees of efficiency and flexibility, resulting in an optimised dynamic resource allocation. All this will provide an advantage to cloud providers using the NECOS models and tools, since they will be able to provide a richer set of functionality than current cloud offerings while keeping their operational costs at a minimum.

The following barriers will be overcome by the NECOS approach and project:

- *Lack of standards:* There is lack of standards for interoperability of software defined infrastructures and services. The ETSI Specification Group for Network Functions Virtualisation (ISG NFV) is developing requirements and architecture specifications to support virtualised functions and guidelines for network and network cloud functions. However, ETSI and other SDOs do not cover the horizontal software defined network clouds interoperability and Slice as a Service. NECOS is intended to actively contribute to OGF (APIs for federation), ITU-T IMT2020 (5G networking and servicing), and IETF (network cloud slicing) to bridge this gap.

- *Network, Compute, and Storage technical complexity:* NECOS concepts, management uniform abstractions, virtualisation, and orchestration over such complex environments would enable simplicity in multi-domain operations.
- *Complexity of service and business coordination:* NECOS will mitigate this risk via its platform proposition, business processes, standard interfaces/SLAs, and open API approach. NECOS empowers a new service model – Slice as a Service, by mapping dynamically service components to a slice.
- *Bootstrapping-adoption:* Technology adoption models indicate that a critical mass is needed for a novel product solution or service paradigm to be accepted in the market. The strength of the NECOS consortium industry members and the size of their respective industry affiliates mitigate this threat thus increasing the potential of NECOS adoption in the market.
- *Barriers to entry-sustainability:* The long-term sustainability of NECOS heavily depends on openness, transparency, healthy competition, rich customer choices, and fair prices. Myopic business strategies to deny participation to potential competitors or new entrants results in barriers to entry and could damage the cloud networks ecosystem. NECOS will adopt an open standards and open APIs approach so that this risk will not be materialised, ensuring fair competition and long-term sustainability of the exchange.

To overcome the aforementioned barriers, NECOS project concept is driven to provide three main features: service provisioning, uniform and efficient management, and validation. Each of them will be briefly stated in the following subsections.

2.2 Service provisioning

This feature is based on the development of the artefacts needed to make the LSDC a service provisioning environment, characterised by the integration of resources within the collection of independent infrastructure providers. A goal of the LSDC approach is to reduce the complexity and timescale for service provisioning and deployment in federated DCs, thus reducing the OPEX for the infrastructure owner.

As measurable results, it brings a demonstration of the LSDC approach to provide a service deployed over multiple clouds (i.e., local, edge, remote and federated cloud) using the Slice as a Service model.

The LSDC is intended to demonstrate a service provisioning approach in the context of the Telco Cloud use-case, in a way that allows the deployment of services as on a cloud computing environment, with an optimal usage of resources according to the actual user demand. Finally, it presents a demonstration of the LSDC approach to support multiple virtual elements (local, edge cloud, remote and federated cloud) in a slice for the Multi-access Edge Computing use case.

2.3 Uniform and efficient management

This feature is mandatory, given the huge need to develop the service and resource orchestration, as well as the management methods for LSDC infrastructure resources that are located within and at the edge of the network. This service orchestration and management approach includes the automatic re-allocation of resources and services across distributed and geographically separated computing, data storage, and network infrastructures within separate slices.

It will be accomplished by targeting three goals: at first, a detailed design and demonstration of the APIs to allow the LSDC approach to federate cloud service providers using Slice as a Service. Afterwards, it shall bring a detailed design of a cloud management approach aimed at mapping the orchestration decisions into the allocation of resources to the substrate infrastructure for multiple domains of a cloud federation, using slices. Finally, it addresses a detailed design of a monitoring architecture of the substrate cloud infrastructure, including its monitoring policies and monitoring approach.

RNP
REDE NACIONAL DE ENSINO E PESQUISA

2.4 Validation through Proofs-of-Concept

Given its experimental approach, the full impact of the NECOS solutions will be validated by proofs-of-concept (PoC) developments based on the scenarios for each use case. The envisioned metrics to evaluate feature accomplishment are stated as follows:

- Detailed design and demonstration of the operation of the algorithms and capabilities of the NECOS platform through the demonstrations of the presented project use cases;
- A number of scenarios that the consortium will be able to set up to demonstrate the increased capacity and competitiveness that NECOS adopters will have.
- Demonstration of the NECOS platform to show that it is supporting Multi-access Edge Computing and Telco Cloud.
- Demonstration of the NECOS platform capability to allow for federation of geographically distributed cloud providers, using the slicing mechanisms.

2.5 Conclusion

In this section, we have discussed NECOS key characteristics, presenting LSDC architecture as an enabler of a novel paradigm on cloud slicing, stated as "*Slice as a Service*", to be built on top of the already available cloud platform features and functions. We postulate LSDC is a promising solution to a widespread audience, since it may be seen as an enhancer to several scenarios, as depicted in Section 2.1, given its three main features: service provisioning (Sect. 2.2), uniform and efficient management (Sect. 2.3), and validation through PoC use cases (Sect. 2.4).

Based on that, Section 3 presents NECOS business model and ecosystem, as well as an analysis on related references from organisms like 3GPP, NGNM, and TM Forum.

3 Ecosystem and stakeholders

Cloud and network slicing constitutes an extremely promising approach for the provision of Cloud network services and to support new wholesale offerings. The feasibility of creating logical, full-functional partitions of the infrastructure will allow service providers to support and operate different kind of services with very distinct requirements onto the same infrastructure.

Slice as a Service will change the current ecosystem and business models of communication service providers, considering how the slices are going to be provided and consumed. In this context, different scenarios can be considered in terms of slice management and control capabilities, and how much of these the service provider hands over to the tenant. The service providers can have their own internal slices and can also provide Slice as a Service to tenants in two ways: multiple tenants share a managed slice, where the provider keeps the control of the slice; and one unmanaged slice per tenant, where the provider gives the control of the slice to the tenant. Figure 2 presents these different options described above.

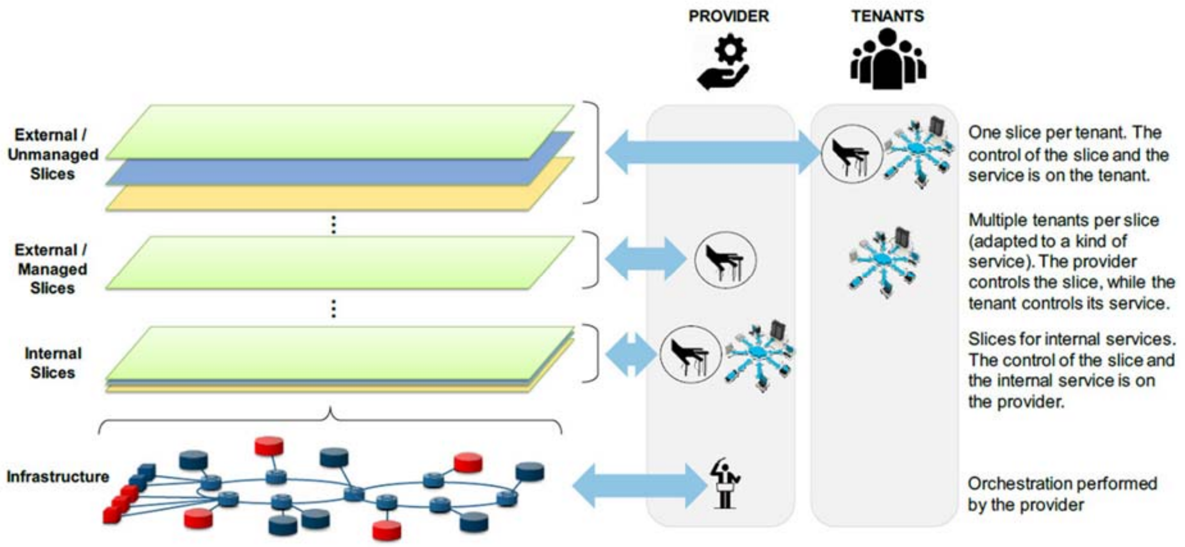

Figure 2 - Types of considered slices and control responsibilities. Source: [10]

3.1 Related business models

The following subsections describe the NECOS business model and ecosystem based on the analysis on the related business models in the 3GPP, NGNM, and TM Forum standardization bodies.

3.1.1 3GPP business model

3GPP is working on the definition of a high-level functional model of business roles for 5G network services, which is described on the specification TR 28.801 - Study on management and orchestration of network slicing for next generation network [1].

3GPP defined the following high-level business roles in the context of next generation networks:

- **Communication Service Customer** (CSC) consumes communication services.
- **Communication Service Provider** (CSP) provides communication services with the responsibilities of designing, building and operating its communication services.
- **Network Operator** (NOP) provides network services with the responsibilities of designing, building and operating its networks to offer such services.
- **Virtualization Infrastructure Service Provider** (VISP) provides virtualized infrastructure services with the responsibilities of designing, building and operating its virtualization infrastructure(s). VISPs may also offer their services to other types of customers without going through the NOP.

EUB-01-2017

- **Data Centre Service Provider** (DCSP) provides data centre services with the responsibilities of designing, building and operating its data centres.
- **Network Equipment Provider** (NEP) supplies network equipment, including Virtual Network Functions (VNFs).
- **NFVI Supplier** supplies network function virtualization infrastructure to its customers.
- **Hardware Supplier** supplies hardware.

For 3GPP, an organization can play one or several roles defined above depending on the scenario of deployment. In this sense, the stakeholders are not clearly defined for the 5G ecosystem. The Figure 3 depicts the relationship among the roles described above.

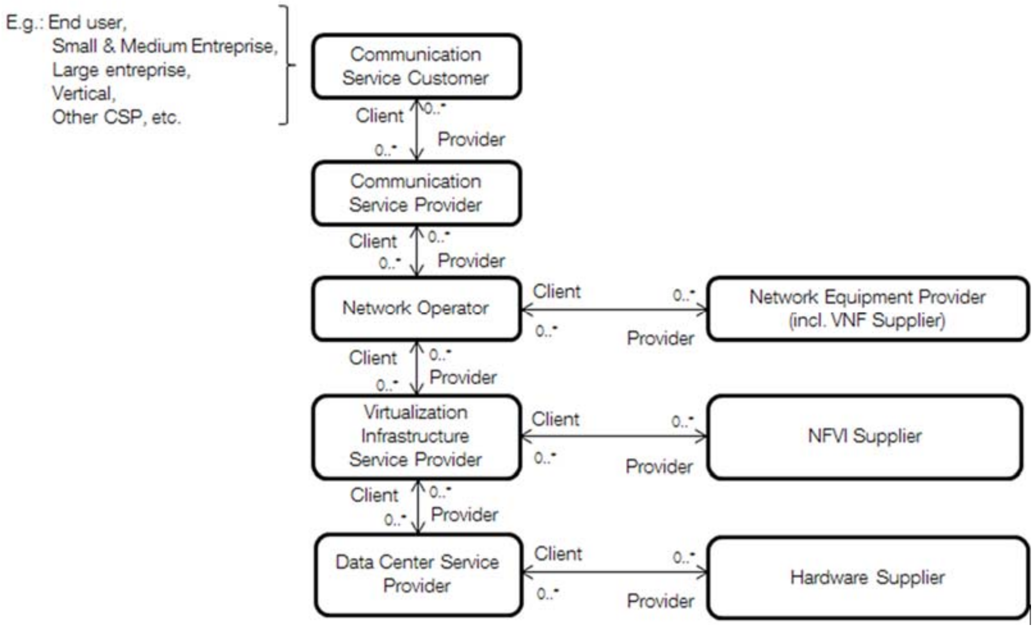

Figure 3 – Roles defined in 3GPP. Source: [1]

3.1.2 NGMN business model

The 5G White Paper [2] from the NGMN Alliance describes a business model for the next generation mobile networks, where 5G expands the current business model to support different types of customers and partnerships. Figure 4 shows examples of models that have to be supported by 5G. The same models are being elaborated below:

- *Asset Provider* is an infrastructure provider, where the assets can be different parts of a network infrastructure that are operated for or on behalf of third parties resulting in a service proposition. Another dimension of asset provisioning is real-time network sharing that refers to an operator's ability to integrate 3rd party networks in the MNO network and vice versa, based on a dynamic and context dependent policies.
- *Connectivity Provider* provides best effort IP connectivity for retail and wholesale customers, which includes QoS and differentiated feature sets (e.g. zero rating, latency, mobility) as well. Furthermore, (self-) configuration options for the customer or the third party will enrich this proposition.
- *Partner Service Provider* can directly address the end customers by providing integrated service offerings based on operator capabilities (connectivity, context, identity etc.) enriched by partner (3rd party / OTT) content and specific applications. This role can also empower partners (3rd parties / OTTs) to directly make offers to the end customers enriched by the operator network or other value creation capabilities.

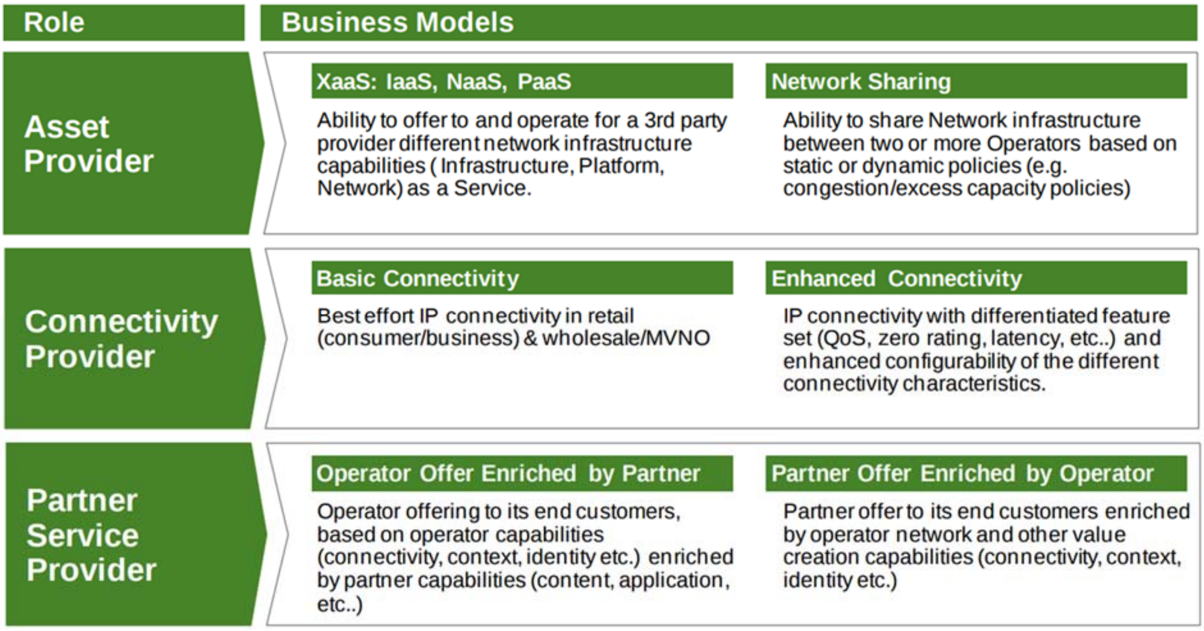

Figure 4 - NGMN business models. Source: [2]

As a reflection of the above business context, the pricing models will also evolve and adapt to represent different types of services and customer profiles, for example,

- Evolved usage-based pricing, which reflects the throughput, latency, data consumption and device movement.
- Event based / real-time charging which may cover, e.g., bandwidth consuming services.
- Tiered offers based on differentiated customer profiles and services.

3.1.3 TM Forum business model

TM Forum, a global industry association that drives collaboration and collective problem-solving to maximize the business success of communication and digital service providers and their ecosystem of suppliers, released the exploratory report IG1152 - Dynamic Network Slices Management and Business Models [3] that describes a business model related to network slicing.

In order to identify actors and attached roles, the TM Forum describes different high level deployment scenarios as the one illustrated by Figure 5. In this scenario, the Slice Provider is partnering with administrative/operational domains in order to build and deliver E2E Services. Each domain provides a slice as 'a service' to the uppers layer with an associated SLA and is responsible for Management & Orchestration of the 'resources' it delivers. The upper layer then consumes this 'slice' to complement its own slice ‘segment’ in order to build a full E2E Network Slice and support E2E Services.

In this scenario, the actors and attached roles are the following:

- The *Service Provider* offers its product, E2E Network Slices, to the Customers that can be End Users or Enterprise. The service provider is responsible for the design, provisioning, delivery and operation of the E2E Network Slice. To achieve this objective, it relies on the Administrative/Operational domains as Slice Consumer of their resources. The Service Provider also plays the role of Network Slice Manager and Orchestrator, thanks to Slice Orchestration Policies and Governance tool-chain.
- In the illustrated scenario, the *Slice Producer* is composed of three Producers that owns, administrates and operates their Platforms. They provide 'segments' as parts of the Service Provider's Network Slice 'segment' that combined, by the Service Provider, form a Network Slice back-end in order to delivery E2E Service Instances.
- The *Tenant* can be seen as a specific 'Platform' that manages and orchestrates E2E Network Slices and offers E2E Services to customers. The owner of this Platform can be either the Service Provider or a third party who will offer E2E Services to customers. The owner of the

Platform manages, operates and delivers the E2E Service Instances and performs usage collection, charging, billing and invoicing to Customer's order.

- The *Customer* interacts with the Service Provider to order Network Slices and supported Service Instances that it wants to consume. The Customer can also monitor the status of its Service Instances through a dashboard offered by the Service Provider.

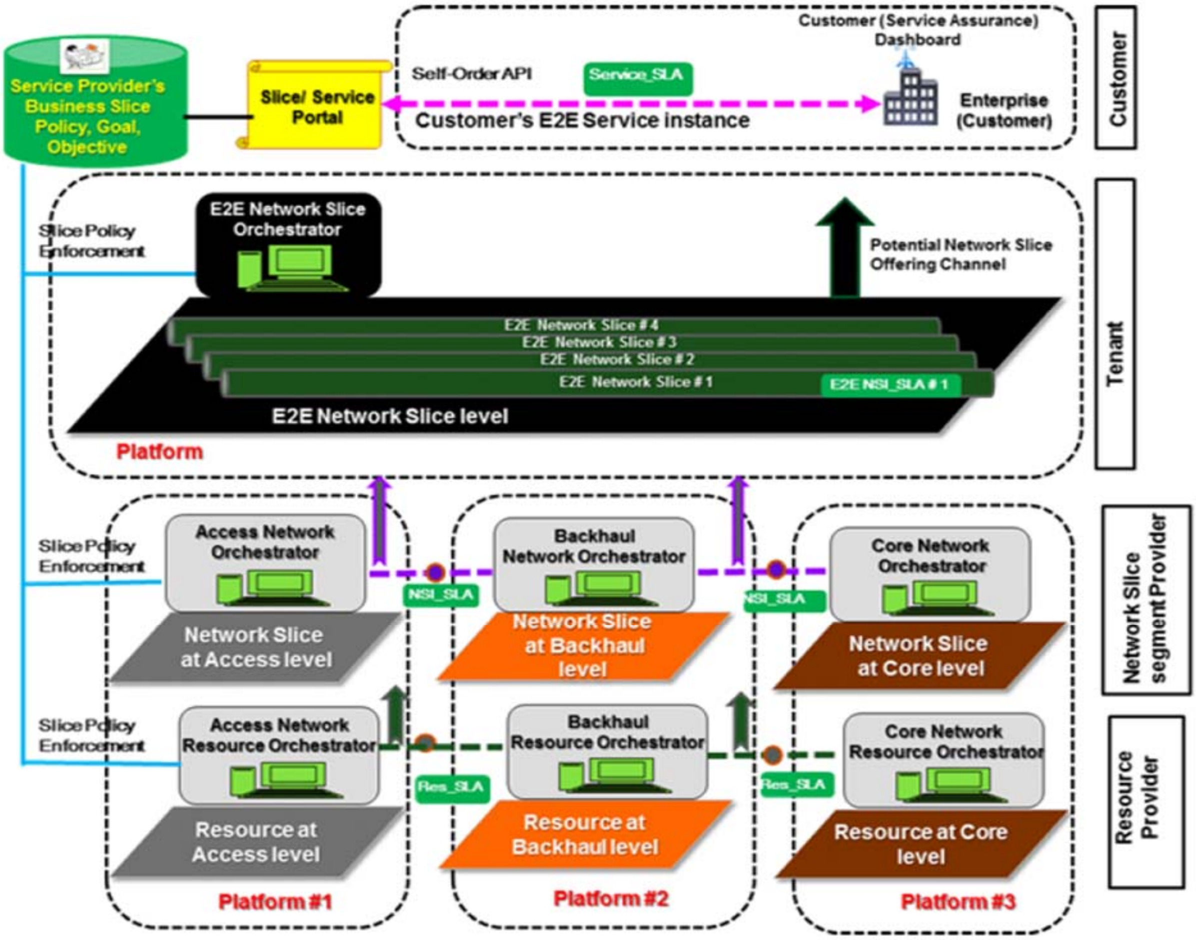

Figure 5 - 5G Slicing Orchestration layering model in Multi-domain use case. Source: [3]

The TM Forum also illustrates the case of a Slice Provider partnering with other Slice Producers, in order to provide E2E Services. These Slice Partners can be specialized (Resource, Network, Service) or offer slice capabilities in all three slicing layers.

3.1.4 Business model analysis

In general, the business models that have been proposed by standardization organizations, as described above, can be summarized considering three main groups of roles: infrastructure/resource provider, service provider and customer. Each of these groups can be composed of more specialized subgroups, where a single party can play multiple roles.

The slicing concept proposed by NECOS defines a slice as a composition of a partition of connectivity, compute and storage resources within services. In this sense, the business model for NECOS must include new roles related to the stakeholders that can provide services. In addition, a broker could play an important role by consolidating the request for resources and service from multiple providers through a marketplace. These new roles are described in the next section.

3.2 Initial NECOS stakeholders / actors / business model

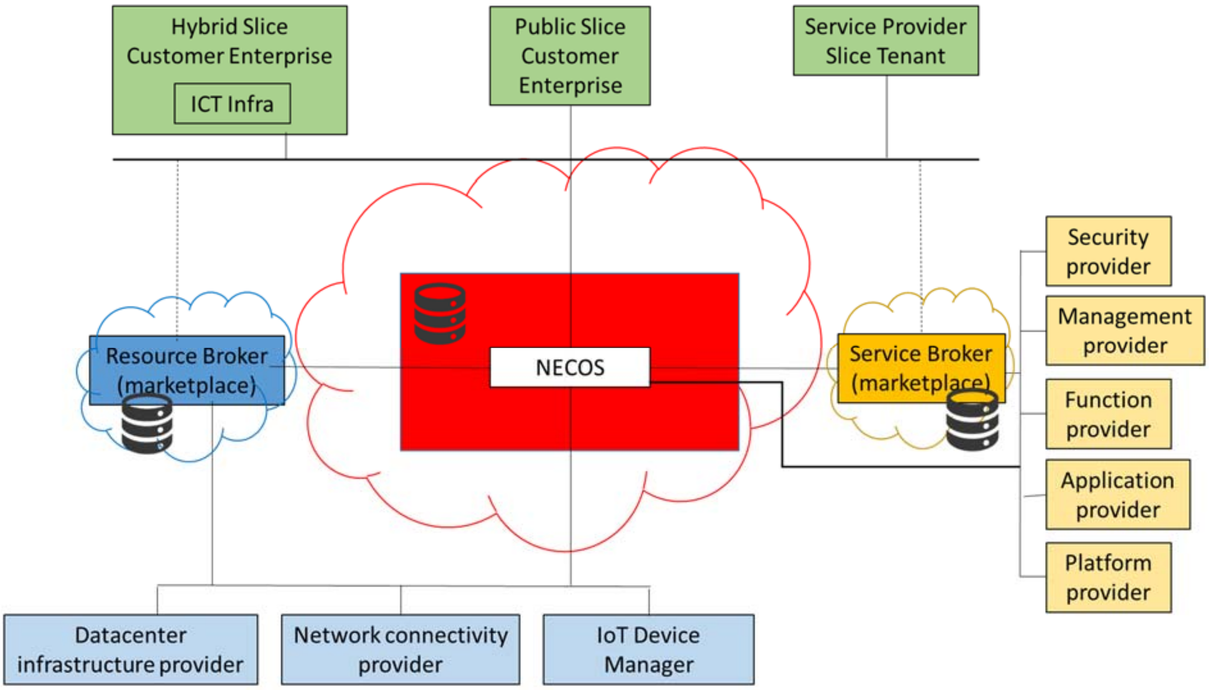

Figure 6 - NECOS ecosystem stakeholders

Figure 6 depicts the NECOS ecosystem and its stakeholders. It is possible to classify the ecosystem stakeholders into four big categories:

- *Slice Provision*: consists of the entities that offer slice services to customers and tenants (red box in Figure 6), i.e., it is basically the NECOS orchestrator and its internal elements. Elements in this category only exist if both elements in the other categories exist and collaboration or commercial agreements among them are established;
- *Infrastructure Resource Provision*: consists of the entities that provide physical and virtual resources to slices (blue box in Figure 6). Elements in this category have no dependency on elements in the other categories. That said, they only make sense to exist to serve customers and tenants, the exception being infrastructure resource owners (not providers) offer idle capacity;
- *Service Provision*: consists of the entities that provide services to slices (yellow box in Figure 6). Elements in this category have no dependency on elements in the other categories. That said, they only make sense to exist to serve customers and tenants;
- *Slice Consumption*: consists of the customers and tenants that consume slices (green box in Figure 6). Elements in this category do not necessarily depend on the Slice Provision services to exist, but have their needs greatly facilitated by the automation provided by those services.

The Slice Provision category consists of the following stakeholders:

- *NECOS*: Entity that provides slice services to customers upon requests. To serve its customers, NECOS interacts with Resources Brokers and Services Brokers and orchestrates actions to deliver to customers whatever was requested.

The Infrastructure Resource Provision category consists of the following stakeholders:

- *Resources Broker*: Entity that provides a resource directory service to NECOS and to Slice Customers to support slice requests or increase/reduction of resources in an existing slice. The Resources Broker offers an API to allow for resources providers to register/unregister the resources they want to offer;
- *Datacenter Infrastructure Provider*: Entity that provides data centre resources (e.g., compute, storage, network) to support building of slices. The Datacenter Infrastructure Provider registers the resources for slices with the Resources Broker;

EUB-01-2017

- *Network Connectivity Provider*: Entity that provides wired and wireless connectivity resources to support building of slices. The Network Connectivity Provider registers the resources for slices with the Resources Broker;
- *IoT Device Manager*: Entity that provides shared or exclusive access to IoT devices to support building of slices. The IoT Devices Manager registers the IoT devices for slices with the Resources Broker.

The Service Provision category consists of the following stakeholders:

- *Service Broker*: Entity that provides a services directory to the NECOS and to Slice Customer to add to slices (e.g., Security as a Service) or on top of slices (e.g., Management as a Service). The SB offers an API to allow for service providers to register the services they want to offer;
- *Security as a Service*: Entity that provides security services (e.g., Intrusion Detection) to slices as a service. The Security as a Service provider registers the security services available for slices with the Services Broker;
- *Management as a Service*: Entity that provides management services (e.g., monitoring) to slices as a service. The Management as a Service provider registers the management services available for slices with the Services Broker;
- *Function as a Service*: Entity that provides functions (e.g., VNFs) to slices as a service. The Function as a Service provider registers the functions available for slices with the Services Broker;
- *Software as a Service*: Entity that provides software applications to slices as a service. The Software as a Service provider registers the applications available for slices with the Services Broker;
- *Platform as a Service*: Entity that provides platforms (e.g., SDN Controller) to slices as a service. The Platform as a Service provider registers the platforms available for slices with the Services Broker.

The Slice Consumption category consists of the following stakeholders:

- *Public Slice Customer*: Entity that requests to NECOS a slice using external resources and services (i.e., registered at the Resource Broker and at the Slice Broker);
- *Hybrid Slice Customer*: Entity that requests to NECOS a slice using both internal resources and services (i.e., owned by the customer) and external resources and services (i.e., registered at the Resource Broker and at the Slice Broker);
- *Service Provider Slice Tenant*: Entity that requests to NECOS a slice using external resources and services. The tenant may provide to customers various services (e.g., Slices, Cloud Services) and play various roles (e.g., Cloud Service Provider, Virtual Network Operator) on top of the requested slice.

In this architecture, a Slice Customer requests a slice to NECOS. NECOS attempts to meet the request by first verifying its internal database for available slices ready to use. If no slice meets the requirements, then the NECOS checks with the Resouce Broker for resources and with the Service Broker for services to build the slice. The slice to be delivered to the Slice Costumer may consist of resources only or both resources and services. For elasticity purposes, a Slice Costumer may request to NECOS or directly to the Resource Broker to add resources to or remove resources from the slice. Likewise, a Slice Costumer may request to NECOS or directly to the Service Broker to add services to or remove services from the slice (e.g., VNF) or from the top (e.g., MaaS) of the slice.

This initial proposal of interaction among stakeholders in NECOS will be revisited and defined in forthcoming deliverables (mainly D2.2) according to the progress on the architectural work and the development of the project. It is included here just as initial reference of the work on the topic.

3.3 Business Actors with respect to NECOS

The proposed architecture and ecosystem provides for various benefits and business opportunities:

- *Resource Provider*: it is an infrastructure provider or an enterprise customer with capacity available on the premises (and registered with Resources Broker) to support building of slices. The ability to provide resources to multi-domain slices opens us opportunities to serve more slice customers and, hence, increase revenues;
- *Service Provider*: a service provider offers services in the marketplace to support building of slices. The ability to provide services to or on top of multi-domain slices opens us opportunities to serve more slice customers and, hence, increase revenues. Examples of services include Function as a Service, Management as a Service, Mobile Virtual Network Operator;
- *NECOS*: by means of interacting with Resources Broker(s) and Services Broker(s) NECOS has the knowledge of resources and services of various types and purposes and from multiple administrative domains. This knowledge allows for NECOS to play with various (e.g., cost, price, performance) strategies and, hence, increase revenues and operations margins;
- *Public Slice Customer Enterprise*: a slice customer with no resources or services on the premises to build the slice on. The ability to get and use slices on demand comprising of various types of resources, services and functions provided by others with little upfront investment provides for cost reduction and business agility;
- *Hybrid Slice Enterprise Customer*: a slice customer that also has resources and possibly services on the premises that should be used to build slices, but that are not enough to meet the customer´s requirements. The ability to build a slice that uses its own private infrastructure as part of the slice provides for elasticity and possibly to serverless computing, thus reducing costs compared to static slices;
- *Service Provider Slice Tenant*: it is both a slice customer and a retail seller of slices with added-value to others customers (e.g., Mobile Virtual Network Operator). The ability to deliver slices comprising various types of resources, services and functions provided by others with little upfront investment and sell these slices to customers from multiple administrative domains opens us business opportunities that would otherwise not be possible;

Business models to capitalize on the business opportunities vary depending on the role in the ecosystem:

- *Infrastructure Providers* can establish swap agreements among themselves to increase market reach and penetration without increasing costs to do so;
- *Slice Providers* may sell different slice services, such as blank slices (i.e., Infrastructure as a Service), blank slices with building blocks to build value-add services, turn-key slice solutions. The higher up in the stack, the higher the recurring revenues. Business models are similar to those of Cloud Service Providers, although elasticity is not always possible in networks;
- *Service Providers* may sell services indirectly via the Service Broker or directly to the Slice Customer or the Slice Provider (i.e., NECOS). Services that will be part of the slice can be sold directly or indirectly, but for services that will run on top of the slice (e.g., Slice Management as a Service) it makes more sense to sell it directly. Recurring revenues tend to be higher selling directly to the Slice Customer or the Slice Provider (i.e., NECOS);
- The *Resources Broker* has the global knowledge of costs and prices and, consequently, negotiation power over resources providers. Hence, the Resources Broker can apply various pricing strategies to increase the revenue and/or the operational margin;
- The *Services Broker* has the global knowledge of costs and prices and, consequently, negotiation power over service providers. Hence, the Services Broker can apply various pricing strategies to increase the revenue and/or the operational margin.

In the following section, we elaborate on the use-cases and scenarios defined in the work of WP2.

4 Use cases

The overarching challenge for the NECOS project is to research and develop the technologies to enable Cloud Slicing infrastructure providers to provide value for themselves and to their customers – typically, application suppliers. To this end, NECOS picked two use cases from its industrial and academic partners for the following three reasons: *(i)* to derive the requirements for the design of the NECOS architecture; *(ii)* to test the developed systems; and *(iii)* to demonstrate the validity of the NECOS solution. These two use-cases are described below.

4.1 Overview of NECOS use cases

Conventional centralized IT data centre architectures for cloud computing, as shown in Figure 7, usually employ a leaf (also called Top-of-Rack, ToR) and spine switching fabric to connect the servers hosted in the data centre. Connections from the data centre to external networks are achieved by deploying another equipment, called the Data Centre Gateway (DC-GW), in a redundant configuration that connects the cloud data centre infrastructure to the network operator equipment, usually an MPLS Provider Edge (PE) node. This back-to-back connection of these 2 pieces of equipment (DC-GW and PE) allows having an administrative demarcation point between the cloud provider and the network provider. On the other hand, the overhead implied by this DC-GW hardware equipment is shared by the high number of servers and applications (running on top of them) present in a central cloud location.

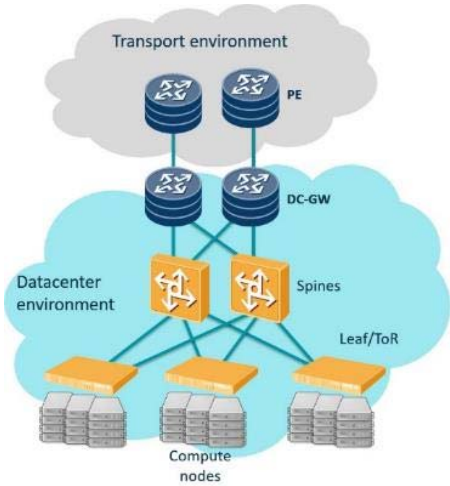

Figure 7 - Central Cloud location network architecture

The existence of these centralized data centres, with such infrastructure overhead, seems to be however insufficient to satisfy new communication trends, which will require higher levels of capillarity in terms of edge cloud locations.

Mobile devices can often create situations where hundreds of devices are present in a few square meters, transmitting and receiving data to be processed and transformed to appropriate stimulus to their owners, in order for the latter to seamlessly manage their digital lives. Such mobile devices can be smartphones that have become the main hub of citizens' digital life, wearable devices that aim the healthcare of elderly people, workers in adverse conditions like in oil plants, mines, etc. Although the nature of the data, the type of processing and the range of applications is significantly broad, the user devices share some common characteristics mainly due to their mobile nature, i.e., rather low processing capability, limited storage capacity, restricted energy supply and they follow the movement of their users.

Consequently, intense data processing is not feasible or desirable on resource-constrained devices, but ideally has to be carried out remotely and the results fed back to the devices. In a similar manner, the large amounts of data accumulated by the intensive use of such devices, has to be done at a remote location, possibly in some high level summarized form. In situations like this, where we can have high-density distributions of devices in wide geographical areas, a scalable approach consists of extending the capabilities of mobile devices using the massive resources offered by cloud data centres. However, instead of the classical cloud servers situated on the Internet or central cloud environments, the edge cloud can serve as a computation and storage resource near the user. Such edge cloud environment is enabled by the **Multi-access Edge Computing (MEC)** paradigm [4]. Although an edge cloud has limited capacity compared to the classical (core) cloud, its proximity to end-users with dense geographical distribution can lead to reduced latency, improved QoS and better mobility support for the above-mentioned applications and services.

This configuration results in multiple levels of processing/storage (local, edge cloud, remote and federated cloud). The criteria to decide where the former has to be done would be a combination of several factors, among which we can mention the minimization of energy consumption, the network load balancing, as well as the cloud load balancing, and the fulfilment of the relevant regulations of personal data protection. Such decisions should be transparent to the user, who would use the service, as if it was hosted on its own device.

Complementary to MEC, an additional innovation is emerging around the transformation of the conventional Central Offices towards cloudification. The traditional Telco business is evolving towards offering a richer set of services beyond basic connectivity services. Future telco networks are expected to support the needs of a hyper-connected society, which is continuously demanding very high data rate access, independency from the technology of attachment to the network, and an increasing number of almost permanently connected devices. Conventional ways of engineering services are not valid anymore, i.e., based on monolithic devices statically located in the network. Evolution in time, location, and requirements of the workloads generated by the end-users advocates a flexible infrastructure able to allocate resources that can be instantiated and removed, scaled-up and down, and being made closer to the user, according to the real needs of the overall services, in real time.

A versatile execution environment is required, capable of running different workloads. This is only economically sustainable when using commodity hardware, and defining simple and automated lifecycle processes associated to both the infrastructure and the workloads. Cloud computing and virtualization is the key technology to allow for this change. However, traditional cloud computing could not be enough to address this challenge due to a number of essential differences with respect to a telco cloud (e.g., performance bound to CPU vs. performance bound to I/O; node-centric vs. network-centric; many and small VMs vs. few and large VMs; etc.). Then, the evolution of the existing Central Offices by means of the integration of both virtualized and non-virtualized equipment is creating the concept of **Telco Cloud** environment. One example of this evolution to a Telco Cloud is the CORD (Central Office Rearchitected as a Data centre) project [5] of the ON.Lab initiative, led by AT&T. Another example of the move to the Telco Cloud is the ongoing work of the Broadband Forum (BBF) to define the Cloud Central Office (Cloud-CO) that has released the architectural framework TR-384 [6].

In terms of network layout, as shown in Figure 8, the different Telco Cloud proposals from the industry share a target network reference architecture that just includes the leaf and spine switching fabric, general purpose servers and some Physical Network Functions (PNFs) hosting access-facing and network-facing I/O cards connected directly to the switching fabric. There is no hardware specifically devoted to the DC-GW and PE functions as in a traditional data centre architecture, to reduce the overhead and to become self-contained, being the necessary functions to interact with external networks distributed across the other elements in the architecture (fabric, servers and I/O cards). This is the final target of a transition to a Telco Cloud for traditional network operators to become edge cloud providers.

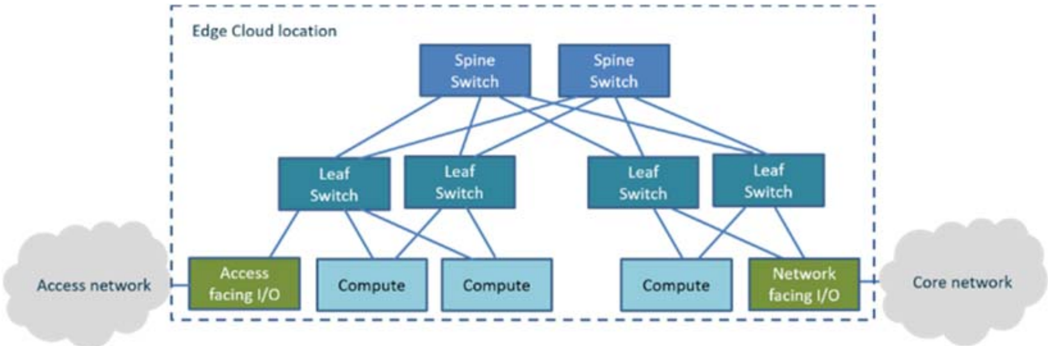

Figure 8 - Telco Cloud architecture

4.2 Impact of MEC use case

By using the Multi-access Edge Computing technology, a mobile operator can efficiently deploy new services for specific customers or classes of customers. The technology also reduces the load in the core network and can host applications and services in a less costly way. It also collects data about storage, network bandwidth, CPU utilization, etc., for each application or service deployed by a third party, which allows a better network planning and management. Application developers and content providers can take advantage of close proximity to cellular subscribers and real-time RAN information.

Multi-access Edge Computing is an emerging computing approach, which can be viewed as an important enabler for Mobile/Internet of Things (IoT) applications and services. Positive side effects with 5G technology are also observable in the sense that the latter can be an enabler for the former. Finally, outsourcing the demand to the federated cloud has been proposed as a means to overcome the local demand spikes without having to rely on over dimensioning or just rejecting service requests at peak demand times. Nevertheless, the decision of which and how much data has to be outsourced is not trivial and it directly impacts in the revenues of all service providers of the federated ecosystem. The existence of a platform adopting autonomous decisions on that respect widens the market of cloud providers and stimulates new entrants to enter in the game.

4.3 Impact of Telco Cloud use case

The key idea of this use case is to elaborate on a general-purpose infrastructure used for many different purposes, covering internal IT applications, through to the offering of B2B services (i.e., Cloud services), and to the own realization of the network itself (i.e., NFV).

To make it easy and convenient to address these different infrastructure usages, an appropriate control and management environment running on top of the distributed infrastructure is needed. Such environment is responsible for allowing: *(i)* the common infrastructure to be used by multiple isolated workloads; and *(ii)* the sharing or the partitioning of resources.

The Telco vision is virtualizing the network from end-to-end, from customer premises to the inner network infrastructure. Virtualization brings the opportunity to build mouldable networks with software based network functions deployed over general-purpose hardware.

In terms of CAPEX, this improves capacity in a flexible and efficient way, leading to simpler networks, while avoiding vendor lock-in, since it will allow using and operating a common infrastructure, independently of the final purpose of the application or network function running on top of it. From the point of view of generation of new income, this Telco cloud reduces Time-To-Market and general delivery of innovative and new services. Finally, from an OPEX perspective, the Telco cloud, through the integrated control and management provided by NECOS, allows for automation and cost sharing (e.g., energy or space), as well smooth interaction with the transport network.

Last but not least, the final end-user benefits through a tailored offering of services accommodated to the timely needs and circumstances of the user due to the flexibility mentioned before regarding the time, location and requirements of each specific demand.

4.4 NECOS platform as enabler of these innovative environments

NECOS targets the combination of cloud and networking resources in Slices, in order to make the most appropriate resource provisioning to applications and services and demonstrates the aforementioned characteristics.

NECOS will have to provide an ecosystem of lightweight distributed clouds (at different distribution levels as enabled by MEC and Telco Cloud) to be in support of applications, forming a homogeneous cloud environment constituted by the corresponding virtual machines and their controlling functions. For the realization of this scenario, NECOS will be facing the following key challenges: QoS and optimal resource allocation, scarcity of resources, lack of appropriate connectivity and users' mobility.

NECOS project will address the provision of a distributed but integrated distributed cloud architecture for the hosting and deployment of virtualized components enabling those new services in an automated fashion acting on the resources offered by different kind of Data Centres around the network. In this context, services consist of the concatenation of Service Components, which in turn are running on different Virtual Machines (VMs) or Containers on top of different physical resources that altogether constitute a Cloud Slice. From the point of view of the control and management of all the above, the service orchestration will deal with the creation of end-to-end services by composing different service components, and the topology management of the network service instances.

In the following section a number of potential scenarios making use of MEC and / or Telco Cloud are described. These scenarios will be taken by the project as a reference for the definition and identification of functional and non-functional requirements of the NECOS platform.

4.5 NECOS scenarios

This section introduces a number of scenarios built on top of previous use cases. These scenarios are taken as reference for deriving requirements to be satisfied by the NECOS platform.

4.5.1 5G Networks

A primary set of scenarios is focused on mobile telco networks.

4.5.1.1 5G infrastructure (virtual RAN)

Generically speaking, the virtual RAN (vRAN) scenario take advantage of the virtualization trends starting to be considered for mobile networks, specifically in the radio access components.

4.5.1.1.1 Description

Mobile Network Operators (MNOs), including the ones leveraging on the infrastructure of other operators, known as Mobile Virtual Network Operators (MVNOs), are adopting virtualization as the technological paradigm for the deployment of network and services in a general way. While the virtualized approach is already quite evolved in the core network, this trend is also being extended to the Radio Access Network (RAN) for a number of reasons:

- To open the industrial ecosystem by decoupling H/W and S/W for RAN nodes.
- To reduce costs, by means of sharing infrastructure resources, more interestingly at the remote locations where the scarcity of resources can be larger.
- To improve network performance in general terms.
- To provide flexibility to adapt to standard evolutions and traffic demands.

The infrastructure deployed for providing network functions associated to the RAN can also be extended to support different services providers. In this context, vRAN presents potential benefits for multiple stakeholders. For example, Network Connectivity Providers can make available near to the edge resources to attend applications that demand for ultra-Reliable and Low Latency Communications (uRLLC); and tenants have an infrastructure to provide localized or contextualized services. However, this scenario involves sharing of a common infrastructure among tenants with different needs. Additionally, the owner of the infrastructure is interested on optimizing the usage of the resources, what must happen transparently to the tenants.

In Figure 9, we present an example to illustrate the main problem faced in this scenario. The figure shows where two tenants have users and so where they demand for coverage. Tenant 1 offers an enhanced Mobile Broadband (eMBB) service to its users, while Tenant 2 users rely on an uRLLC application. Additionally, the MNO also has its own traditional mobile users. To cover the area, the MNO employs a combination of vRAN and MEC infrastructure with two Cloud/Central Unit (CU), i.e., CU1 and CU2. The MNO needs to assure the performance required by each tenant but also wants to optimize the resource usage and attend its own users.

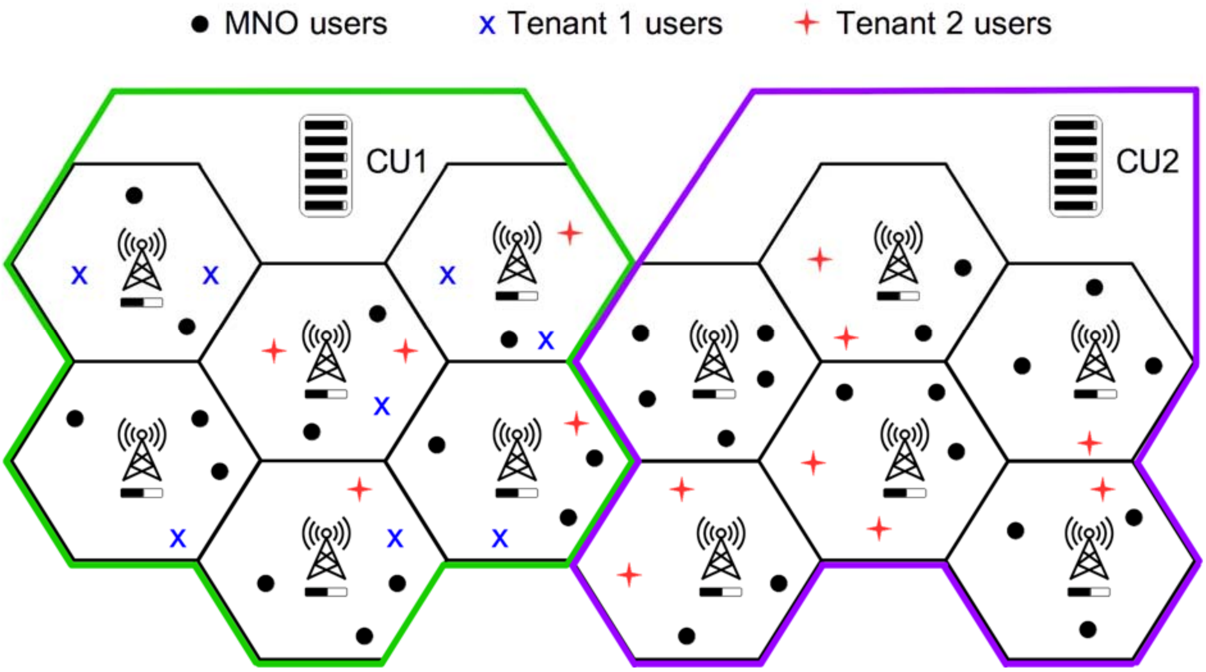

Figure 9 - vRAN scenario with two distinct tenants

4.5.1.1.2 Technical enablers

In order to attend this vRAN scenario, NECOS introduces cloud slicing which encompass several technical enablers. Figure 10 illustrates the NECOS approach in which three slices are created to address the issues raised previously. There is one slice for each tenant and an additional slice for the MNO.

The slicing system must be able to create an isolated set of resources (i.e., a slice) for each tenant. These resources include network, computing, and storage. All resources in the same slice must be connected and must be identified by the tenant as a unique infrastructure. A tenant must be able to monitor the performance metrics concerted in its SLA. A tenant must be allowed to request management and orchestration capabilities for its slice. Naturally, due to the isolation, the operation performed by a tenant must not affect any other tenant. The slicing system must also facilitate the service deployment whatever the software technology employed by a tenant.

The MNO must be able to monitor, manage, orchestrate, and optimize the whole infrastructure, but these actions must not disturb the SLA negotiated with any tenant. As illustrated by this scenario, an MNO may need to create a slice in order to support its own users. However, an MNO may need a slice due to other reasons, e.g., to offer communication services for tenants in a transparent manner.

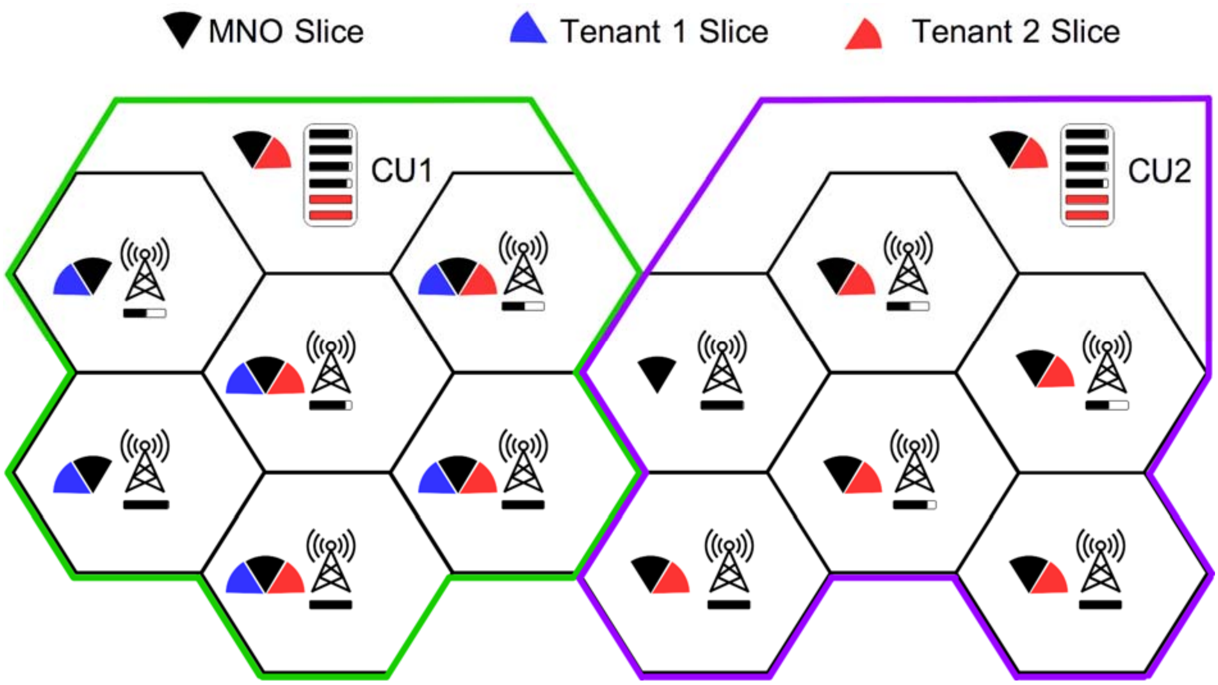

Figure 10 - Slices to support the vRAN scenario

4.5.1.1.3 Critical success factors and KPIs

The following factors are identified as critical to success of this scenario:

- F1 - High degree of scalability (variable and dynamic number of users), with a short-time response for the deployment on the field;
- F2 - Fast service deployment;
- F3 - Real-time monitoring;
- F4 - Support to ultra-reliable and low latency communications;
- F5 - Support to enhanced mobile broadband.

These critical success factors are mapped to the following KPIs:

- K1 - Accomplishment of operations such as creation, enlargement, shrink of slice. Metrics - Provisioning time, and decommission time;
- K2 - Provision of the slice to the tenants considering a service that defines the time for this provisioning. Metrics - Service provisioning time;
- K3 - Provision of monitored data to the operator and the tenant. Metrics - Monitoring-data availability;
- K4 - Accomplishment of operations such as creating, enlarging, reducing and deactivating a slice within acceptable time periods. Metrics - end-to-end delay;
- K5 - Provision of data transfer rate in order to meet the tenants. Metrics - throughput.

4.5.1.1.4 Mapping to NECOS key characteristics

The following NECOS characteristics are identified as key enablers of this scenario:

- C1: Slice as a Service model dynamically allocates, modify, or deallocates slices on-demand;
- C2: Adaptations and reconfigurations are done at a per slice level, keeping the proper isolation;
- C3: Each aspect that comprises the cloud environment - from the networking between virtual machines, to the SLAs of the hosted applications - is managed via software;
- C4: Lightweight management and virtualization systems deployable on large number of small servers at the network edge.

4.5.1.1.5 Requirements

The following functional requirements were identified:

- RF.vRAN.1: *Service Level Agreement*. The system must assure all the performance metrics (bandwidth, latency, CPU) negotiated for each slice.
- RF.vRAN.2: *Accountability*. The system must offer monitoring and accounting in a per slice basis.
- RF.vRAN.3: *On-demand slice provisioning*. The system must allow the operator to create or adapt the slices on demand.

The following non-functional requirements were identified:

- RN.vRAN.1: *Isolation of slice resources.* Tenants must be protected from each other, i.e., it must be avoided any information leakage (e.g., monitoring measurements) among the slices.
- RN.vRAN.2: *Fairness*. Operator must be able to optimize the resource usage without negative impact on any tenant.
- RN.vRAN.3: *Fault detection.* Any service should operate continuously for a long time, so failures in the slice level should be automatically handled without impacting the service.

4.5.1.2 5G services

Network operators are facing now the need of adapting their existing networks in order to be able of providing forthcoming 5G services. Three main types of services have been identified so far [7]: enhanced Mobile Broadband (eMBB), massive Machine-Type Communications (mMTC) and ultra-Reliable and Low Latency Communications (uRLLC). The eMBB service type encompasses the challenge of providing an unprecedented volume of data delivery, associated with e.g., high-definition video sharing. The mMTC focuses on applications where a large number of IoT devices, such as sensors, collectively creating a significant data volume passing through the network. Moreover, these data are highly localized and are often associated with requirements like privacy, data ownership, etc. Finally, the uRLLC type refers to services in the need for extremely low end-to-end latency, like Tactile Internet, Interactive Gaming, Virtual Reality, Automotive, Industry and Automation. Figure 11 summarizes some characteristics of each type of service.

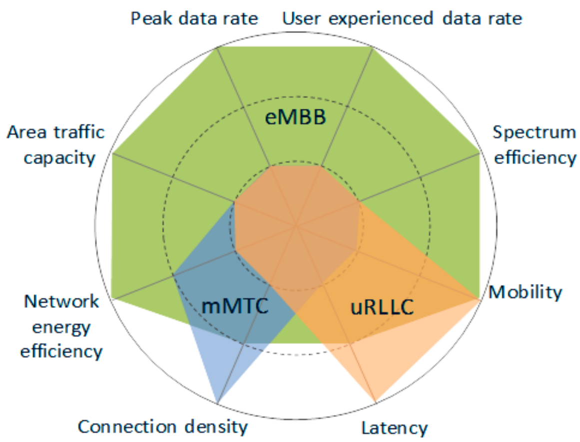

Figure 11 - Characteristics of different kinds of 5G services

3GPP have identified a preliminary set of performance requirements for different scenarios expected to be supported by 5G networks [8]. Table 1 collects requirements for high data rate and traffic density scenarios, while Table 2 shows a characterization of low-latency and high-reliability scenarios.

Table 1 - Performance requirements for high data rate and traffic density scenarios. Source: [8]

	Scenario	Experienced data rate (DL)	Experienced data rate (UL)	Area traffic capacity (DL)	Area traffic capacity (UL)	Overall user density	Activity factor	UE speed	Coverage
1	Urban macro	50 Mbps	25 Mbps	100 Gbps/km^2 (note 4)	50 Gbps/km^2 (note 4)	10 000/km^2	20%	Pedestrians and users in vehicles (up to 120 km/h	Full network (note 1)
2	Rural macro	50 Mbps	25 Mbps	1 Gbps/km^2 (note 4)	500 Mbps/km^2 (note 4)	100/km^2	20%	Pedestrians and users in vehicles (up to 120 km/h	Full network (note 1)
3	Indoor hotspot	1 Gbps	500 Mbps	15 Tbps/km^2	2 Tbps/km^2	250 000/km^2	note 2	Pedestrians	Office and residential (note 2) (note 3)
4	Broadband access in a crowd	25 Mbps	50 Mbps	[3,75] Tbps/km^2	[7,5] Tbps/km^2	[500 000]/km^2	30%	Pedestrians	Confined area
5	Dense urban	300 Mbps	50 Mbps	750 Gbps/km^2 (note 4)	125 Gbps/km^2 (note 4)	25 000/km^2	10%	Pedestrians and users in vehicles (up to 60 km/h)	Downtown (note 1)
6	Broadcast-like services	Maximum 200 Mbps (per TV channel)	N/A or modest (e.g., 500 kbps per user)	N/A	N/A	[15] TV channels of [20 Mbps] on one carrier	N/A	Stationary users, pedestrians and users in vehicles (up to 500 km/h)	Full network (note 1)
7	High-speed train	50 Mbps	25 Mbps	15 Gbps/train	7,5 Gbps/train	1 000/train	30%	Users in trains (up to 500 km/h)	Along railways (note 1)
8	High-speed vehicle	50 Mbps	25 Mbps	[100] Gbps/km^2	[50] Gbps/km^2	4 000/km^2	50%	Users in vehicles (up to 250 km/h)	Along roads (note 1)
9	Airplanes connectivity	15 Mbps	7,5 Mbps	1,2 Gbps/plane	600 Mbps/plane	400/plane	20%	Users in airplanes (up to 1 000 km/h)	(note 1)

NOTE 1: For users in vehicles, the UE can be connected to the network directly, or via an on-board moving base station.
NOTE 2: A certain traffic mix is assumed; only some users use services that require the highest data rates.
NOTE 3: For interactive audio and video services, for example, virtual meetings, the required two-way end-to-end latency (UL and DL) is 2-4 ms while the corresponding experienced data rate needs to be up to 8K 3D video [300 Mbps] in uplink and downlink.
NOTE 4: These values are derived based on overall user density.
NOTE 5: All the values in this table are targeted values and not strict requirements.

Table 2 - Performance requirements for low-latency and high-reliability scenarios. Source: [8]

Scenario	End-to-end latency (note 3)	Jitter	Survival time	Communication service availability (note 4)	Reliability (note 4)	User experienced data rate	Payload size (note 5)	Traffic density (note 6)	Connection density (note 7)	Service area dimension (note 8)
Discrete automation – motion control (note 1)	1 ms	1 µs	0 ms	99,9999%	99,9999%	1 Mbps up to 10 Mbps	Small	1 Tbps/km^2	100 000/km^2	100 x 100 x 30 m
Discrete automation	10 ms	100 µs	0 ms	99,99%	99,99%	10 Mbps	Small to big	1 Tbps/km^2	100 000/km^2	1000 x 1000 x 30 m
Process automation – remote control	50 ms	20 ms	100 ms	99,9999%	99,9999%	1 Mbps up to 100 Mbps	Small to big	100 Gbps/km^2	1 000/km^2	300 x 300 x 50 m
Process automation – monitoring	50 ms	20 ms	100 ms	99,9%	99,9%	1 Mbps	Small	10 Gbps/km^2	10 000/km^2	300 x 300 x 50
Electricity distribution – medium voltage	25 ms	25 ms	25 ms	99,9%	99,9%	10 Mbps	Small to big	10 Gbps/km^2	1 000/km^2	100 km along power line
Electricity distribution – high voltage (note 2)	5 ms	1 ms	10 ms	99,9999%	99,9999%	10 Mbps	Small	100 Gbps/km^2	1 000/km^2 (note 9)	200 km along power line
Intelligent transport systems – infrastructure backhaul	10 ms	20 ms	100 ms	99,9999%	99,9999%	10 Mbps	Small to big	10 Gbps/km^2	1 000/km^2	2 km along a road
Tactile interaction (note 1)	0,5 ms	TBC	TBC	[99,999%]	[99,999%]	[Low]	[Small]	[Low]	[Low]	TBC
Remote control	[5 ms]	TBC	TBC	[99,999%]	[99,999%]	[From low to 10 Mbps]	[Small to big]	[Low]	[Low]	TBC

NOTE 1: Traffic prioritization and hosting services close to the end-user may be helpful in reaching the lowest latency values.
NOTE 2: Currently realised via wired communication lines.
NOTE 3: This is the end-to-end latency required for the 5G system to deliver the service in the case the end-to-end latency is completely allocated to the 5G system from the UE to the Interface to Data Network.
NOTE 4: Communication service availability relates to the service interfaces, reliability relates to a given node. One or more retransmission over the radio interface may take place in order to satisfy the reliability requirement.
NOTE 5: Small: payload typically ≤ 256 bytes
NOTE 6: Based on the assumption that all connected applications within the service volume require the user experienced data rate.
NOTE 7: Under the assumption of 100% 5G penetration.
NOTE 8: Estimates of maximum dimensions; the last figure is the vertical dimension.
NOTE 9: In dense urban areas.
NOTE 10: All the values in this table are targeted values and not strict requirements. Deployment configurations should be taken into account when considering service offerings that meet the targets.

Looking at the stringent requirements in terms of latency, bandwidth or supported number of connections (as main characteristic of each type of service), it can be concluded that the 5G use scenarios will require some processing of data and/or proximity at the edge of the Radio Access Network (RAN). Furthermore, mixing services like eMBB, uRLLC and mMTC, altogether on the same network, makes quite difficult to define common architecture and engineering patterns capable of keeping simultaneously the requirements of each of them in an ordered and structured form. It seems to be much more convenient to segregate those services on specialized network partitions, designed and optimized for each of the types of services to be provided.

Last but not least, from the business perspective, 5G will open a new ecosystem facilitating the offering of such kind of slices to very different kind of vertical industries. The convergence of the telecom and IT industries provides a common very high capacity 5G infrastructure, with service invocation capabilities. 5G by design enriches the customer-facing services (the so-called “vertical industries” or simply “verticals”) with new capabilities and quality features. It is expected to revolutionize service provisioning and end-user experience over multiple service domains, as foreseen by the 5G-PPP [9] for the most prominent vertical services in the areas of Media and Entertainment, eHealth, Energy, Automotive, and Manufacturing-Factories of the Future. Then, the verticals will act as tenants of the shared infrastructure provided by the 5G networks, possibly with different levels of control and management on the allocated resources forming the slice on top of which the vertical service is implemented [10].

4.5.1.2.1 Description

Figure 12 presents the scenario considered here. The purpose is to segregate the eMBB, uRLLC and mMTC services in different slices on top of the same infrastructure, then creating network partitions tailored and adapted to them. In fact, it could be even possible to create specific slices per each of the scenarios described in Table 1 and Table 2, as mean of providing finer granularity and better fit to a particular service scenario.

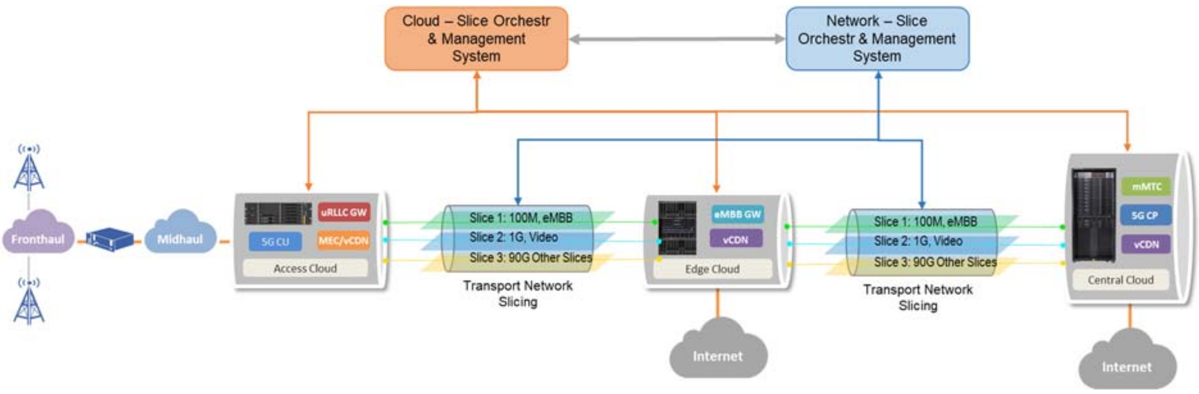

Figure 12 - NECOS 5G scenario

It is worthy to note that in 5G networks the slicing capabilities should extend also to the transport network. Then some mechanisms should be in place to jointly provide the aforementioned network partitions, including both IT and network resources. Consequently, there is a need for interactions with other systems, in this case transport-related orchestrator and management systems.

Finally, sufficient mechanisms for control and management have to be facilitated for external tenants to manage their own services running on the slice, including the allocated resources.

4.5.1.2.2 Technical enablers

When applies to 5G, the slicing system must be able to create slices tailored to the specific requirements dictated by the kind of scenario to be supported. Parameters like latency, bandwidth or number of sessions should be interpreted for allocating necessary resources and for implementing smart decisions with respect to the kind of resources and the location of them in the network.

Since the usage of distributed resources can be the norm, the slicing system should interact with control and management systems devoted to configure the transport network in such a way that the service continuity is ensured. The possibility of enforcing the chaining of service functions, soliciting paths with specific characteristics, etc., should be considered in the design.

Finally, since the tenants in this scenario can be assumed to be vertical industries, compliance to SLAs is a mandatory task. This implies the need of reporting sufficient performance indicators in such a way that SLAs can be checked periodically, reacting in case of any violation of the agreed terms.

4.5.1.2.3 Critical success factors and Key Performance Indicators

The following factors are identified as critical to success of this scenario:

- F1 - High degree of scalability (variable and dynamic number of users), with a short-time response for the deployment on the field;
- F2 - Fast service deployment;
- F3 - Real-time monitoring;
- F4 - Support to ultra-reliable and low latency communications;
- F5 - Support to enhanced mobile broadband;
- F6 - Support of massive communications (high density of sessions);
- F7 - Flexible mechanisms for supporting service requests from different vertical industries;
- F8 - Means of verifying compliance of agreed SLAs.

These critical success factors are mapped to the following KPIs:

- K1 - Accomplishment of operations such as creation, enlargement, shrink of slice. Metrics - Provisioning time, and decommission time;
- K2 - Provision of the slice to the tenants considering a service that defines the time for this provisioning. Metrics - Service provisioning time;
- K3 - Provision of monitored data to the operator and the tenant. Metrics - Monitoring-data availability;
- K4 - Accomplishment of operations such as creating, enlarging, reducing and deactivating a slice within acceptable time periods. Metrics - end-to-end delay;
- K5 - Provision of data transfer rate in order to meet the tenants' needs in case of eMBB kind of services. Metrics – throughput;
- K6 – Provisioning of delay measurements to meet the tenants' needs in case of uRLLC kind of services. Metrics – latency;
- K7 – Provisioning of computing power to meet the tenants' needs in case of mMTC kind of services. Metrics – CPUs.

4.5.1.2.4 Mapping to NECOS key characteristics

This scenario considers the applicability of NECOS as solution for creating the slices necessary to support differentiated services in 5G. The following NECOS characteristics are identified as key enablers of this scenario:

- C1: Slice as a Service model, considering very distinct kind of slices to fit the different nature of the 5G services.
- C2: Smart allocation and placement of resources for ensuring SLAs as requested by the vertical customers (the tenants).
- C3: Lightweight management, especially towards the access, where the cloud resources will be limited.

4.5.1.2.5 Requirements

The following functional requirements were identified:

- RF.5G.1: *Service Level Agreement*. The system must assure all the performance metrics (bandwidth, latency, CPU) negotiated for each slice.

- RF.5G.2: *Accountability*. The system must offer monitoring and accounting in a per slice basis.
- RF.5G.3: *On-demand slice provisioning*. The system must allow the operator to create or adapt the slices on demand.
- RF.5G.4: *External control and management of the offered slices*. The system must allow the control and management of the resources allocated to the tenant if it is demanded by the tenant in the service request.

The following non-functional requirements were identified:

- RN.5G.1: *Isolation of slice resources.* Tenants must be protected from each other, i.e., it must be avoided any information leakage (e.g., monitoring measurements) among the slices.
- RN.5G.2: *Fairness*. Operator must be able to optimize the resource usage without negative impact on any tenant.
- RN.5G.3: *Fault detection.* Any service should operate continuously for a long time, so failures in the slice level should be automatically handled without impacting the service.

4.5.2 vCPE

Fixed telco environments are also target for the usage of cloud capabilities where locating total or partial communications functions, such as a virtual Customer Premises Equipment (vCPE).

4.5.2.1 Description

Traditional CPE deployments have network functions installed on customer site to provide local NAT, local DHCP, IGMP proxy-routing, PPP sessions, routing and etc., along with remote site connectivity. This model has the following characteristics:

- Heterogeneous installed park;
- Many legacy devices;
- Unequal services portfolio;
- Expensive operations and support.

Virtual CPE claims to address many issues related to the current model by turning the CPE into a very simple standard device while moving the network functions to a cloud infrastructure, where they run over virtualization technologies. The goal is to simplify the deployment, support and maintenance procedures of CPEs for network service providers, allowing them to delivery dynamic services to their subscribers, as depicted in the Figure 13.

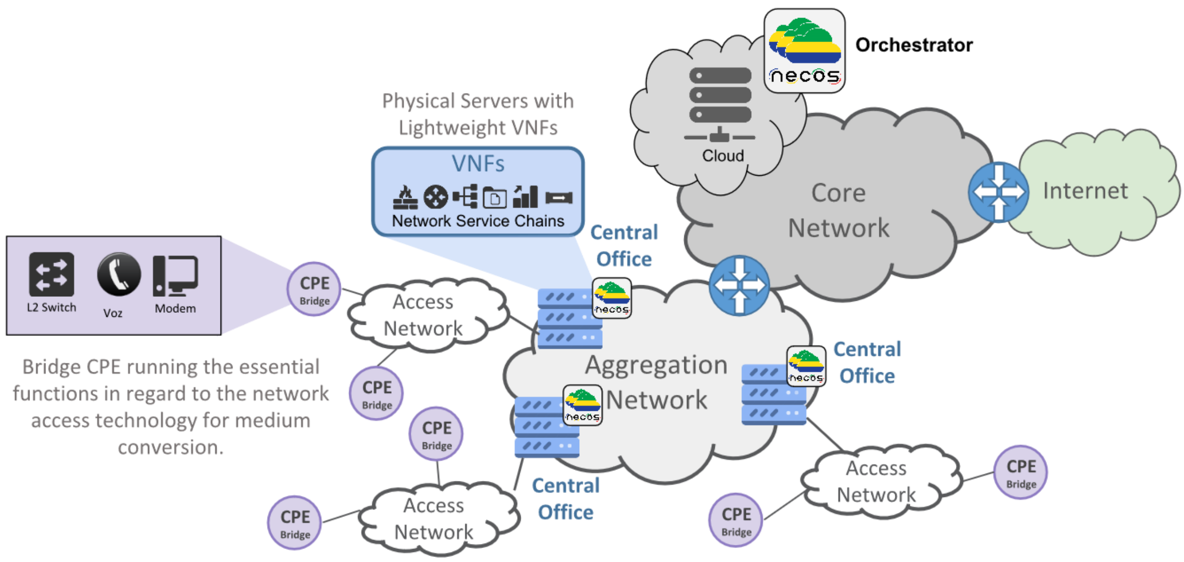

Figure 13 - vCPE deployment scenario

The virtual CPE enables the evolution of Central Offices by applying the key concept of Telco Clouds: the integration of both non-virtualized equipment (bridge CPE) and virtualized elements (virtual network functions), enabling new services to be deployed in an automated fashion. The vCPE services can also be deployed at the Edge Cloud, near to the end-user, in order to reduce latency and improve the user experience.

Although vCPE can bring many benefits, the implementations have faced the challenge of providing a high scalable platform without increasing latency and network bottlenecks. In this context, NECOS is an optimized solution that brings this scenario into reality by providing a deep integration of DCs and networking systems.

Decoupling the owner of the physical infrastructure (infrastructure provider) from who runs and manages the service (service provider) can lead to innovation, allowing new business models and a reduction in the complexity of running services. Since the world of computing has a similar experience, a relevant model can be applied to networking by using the cloud infrastructure to provide virtualized network services.

The infrastructure provider, which owns the cloud and network infrastructure, will be responsible for the management of the life cycle of the network slices and to guarantee the individual service-level agreement (quality, availability, responsibilities) for each slice. The service provider, which owns the slice, can use the slice to provide Internet access for its customers through virtual network functions (VNFs). Multiple service providers can share the same cloud infrastructure in order to meet the demand for network services in different niches, such as residential access, smart factory, agribusiness, smart cities and so on. The Figure 14 describes the scenario where multiple Virtual Internet Service Providers (Virtual ISPs) share the same infrastructure through fully isolated slices in order to provide Virtual CPE service.

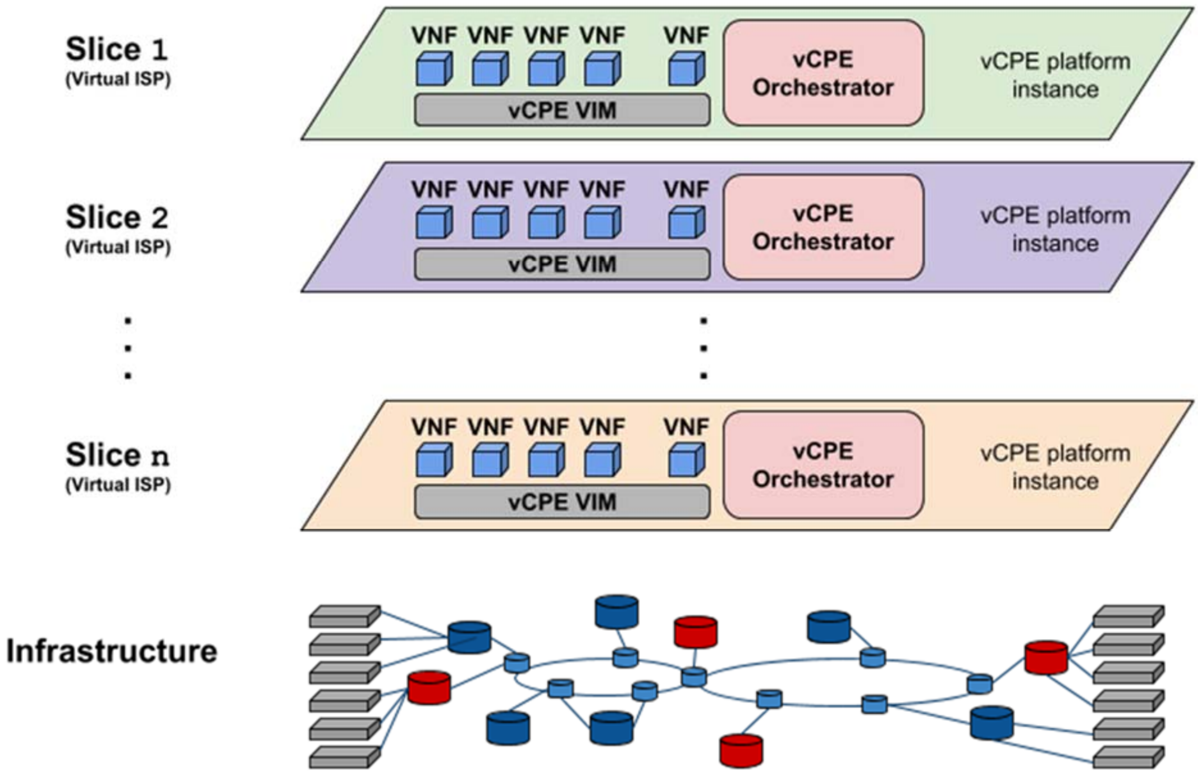

Figure 14 - vCPE deployment scenario with multiple instances

A network operator can play both roles, such as: *(a)* provide network services to the end user through internal slices within its infrastructure (e.g., Virtual CPE); and *(b)* offer unused resources as external slices for third-party companies providing network services, such as Virtual ISPs, e.g., as shown in Figure 14.

Different slicing models can be applied, as represented in Figure 15. For example, the external slice can be managed by the network operator and the latter can delegate the service management to the Virtual ISPs. So, the virtual ISPs can leverage the geographically distributed infrastructure from the network operator, instead of building its own infrastructure. Besides, the Virtual ISP does not have to deal with the complexity of managing the infrastructure and the slice.

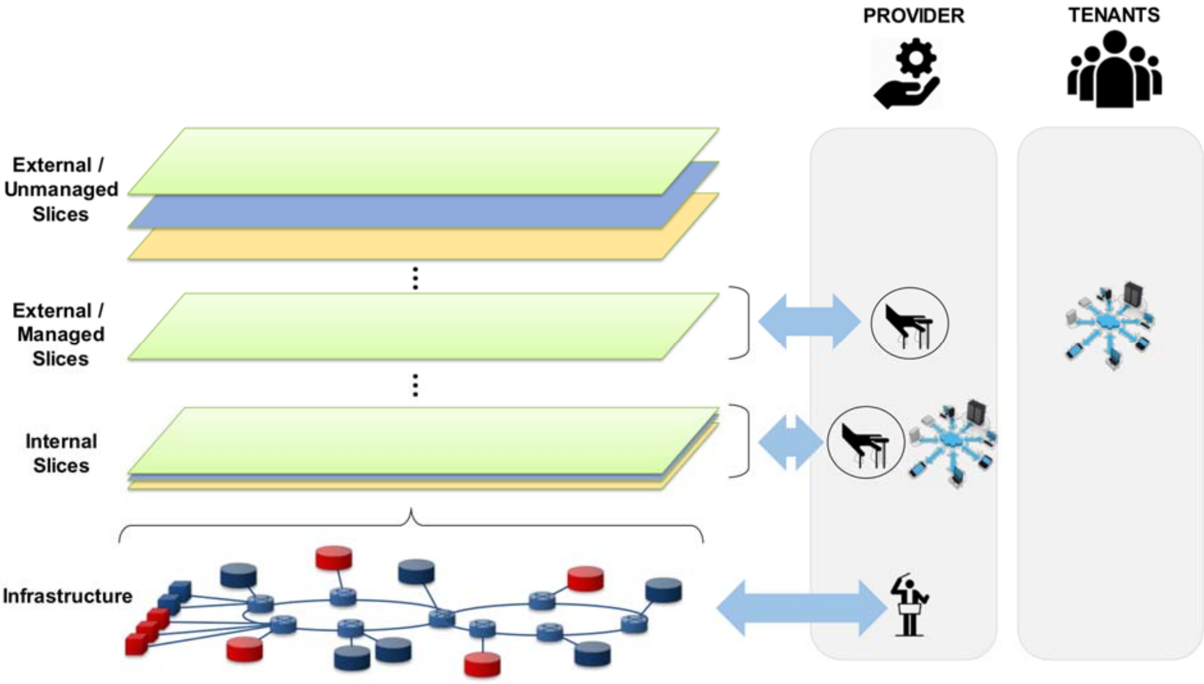

Figure 15 - Slice management in vCPE scenario

NECOS will be deployed by federated network operators, which can offer communication services in two ways: (a) provide Virtual CPE to their customers through internal slices; and (b) provide Virtual CPE as a service to ISPs via external slices. The Virtual CPE service can also come with basic VNFs, so the ISPs can focus on value-added VNFs regarding its niche market.

4.5.2.2 Technical enablers

The following technical enablers can be identified.

- Lightweight virtualization based on containers helps the system to achieve more efficient resource allocation, allowing the Virtual CPE platform to scale.
- Dynamic deployment of new network functions in a distributed and virtualized environment, without the intervention of the network operator.
- Network data plane programmability creates a much more agile, flexible and automated network, allowing the platform to build network services chains dynamically.
- Network I/O optimization technologies can accelerate overall packet processing performance in software, enabling more efficient VNFs, with a higher throughput.
- Automatic monitoring of physical and network resources to enable workload changes in an automatic manner.

4.5.2.3 Critical success factors and KPIs

The following factors are identified as critical to success of this scenario:

- F1 - High degree of scalability (variable and dynamic number of users), with a short-time response for the deployment on the field;
- F2 - High performance standards (low jitter, low packet error rate, high availability, redundancy mechanisms, and priority schemes implementation);
- F3 - Slice management (cross-layered approach for controlling, monitoring, analysis and long-time operations); and
- F4 - Easy network elements configuration.

To accomplish this goal, several KPIs must be addressed, like:

- K1 - Average end-to-end delay (in milliseconds, measured as half average RTT); Metric – average delay.
- K2 - End-to-end slice availability (% of time); Metric – availability.
- K3 - Average slice provisioning time (in seconds); Metrics – provisioning time, decommission time.
- K4 - Performance isolation index to identify performance impact between slices in a multi-tenant environment; Metric – isolation degree.
- K5 - Average elasticity response time (in seconds) taken to provision of new resources based on real-time demand; Metric – response time.
- K6 - Average throughput (in Mbps); Metric – throughput.

4.5.2.4 Mapping to NECOS key characteristics

The following NECOS characteristics are identified as key enablers of this scenario:

- C1: Slice-as-Service model dynamically allocates and deallocates slices on-demand.
- C2: Bare-metal VIM-independent slicing makes the slice fully manageable to deploy any kind of service.
- C3: Lightweight management and virtualization systems deployable on large number of small servers and clouds at the network core and edges.
- C4: End-to-end slice provisioning enables the deployment of services across physical resources from federated cloud networking infrastructures.

4.5.2.5 Requirements

The following functional requirements were identified for the virtual CPE scenario.

- RF.vCPE.1: *On-demand slice provisioning*. Each Virtual ISP will operate using a slice, thus the slice needs to be dynamically created and removed when requested by the customer.
- RF.vCPE.2: *Manageable slice*. Virtual ISPs need to manage the slice to deploy their own VNFs and to build the network service chains.
- RF.vCPE.3: *VIM-independence*. A Virtual CPE platform has its own orchestrator and VIM to deploy the VNFs and to build the network service chains. The service platform should be able to run its own VIM and orchestrator.
- RF.vCPE.4: *Bare-metal slice*. The VNFs will be deployed within the slice. If the slice is composed of virtual resources, there will be two levels of virtualization, which brings a huge impact on performance. In this context, bare-metals should be allocated to the slice.
- RF.vCPE.5: *Lightweight virtualization*. A Virtual ISP can have hundreds of thousands of subscribers, so lightweight virtualization is important to achieve scalability.
- RF.vCPE.6: *Elasticity*. The slice should be able to adapt to workload changes by provisioning and de-provisioning resources in an autonomic manner.
- RF.vCPE.7: *Zero touch service provisioning*. The deployment of the Virtual CPE service to the slice has to be automated, so the Virtual ISP does not need to deploy it manually.
- RF.vCPE.8: *Fault detection*. NECOS should be able to detect critical failures in the slice resources and notify the services that are running within the slice.

Additionally, the following non-functional requirements were identified for this same scenario.

- RN.vCPE.1: *Isolation of slice resources*. The Virtual ISPs will be allocated to different slices, so there must be complete isolation among the slices.
- RN.vCPE.2: *SLA monitoring (QoS)*. It is important for the Virtual ISP to have the SLA's guarantee to provide the service to its subscribers.
- RN.vCPE.3: *Low latency*. All network traffic from the subscribers will travel through the Virtual ISP slice before reaching the Internet. The Virtual CPE service should minimize the impact as much as possible on connection's latency.
- RN.vCPE.4: *High throughput*. The resources allocated to the slice should provide a high throughput when processing network packets, considering that a Virtual ISP can have a large number of subscribers with high bandwidth.
- RN.vCPE.5: *High availability*. The Virtual CPE service should operate continuously for a long time, so failures in the slice level should be automatically handled without impacting the service.

4.5.3 Touristic services

High profile Metropolitan areas attract a significant number of visitors each year. The scenario involves a Metropolitan Tourist Centre (MTC) responsible for offering tourist information facilities to those visitors aiming at enriching the visiting experience of the latter, by offering state-of-the-art location-aware cultural content available in the city. These services will be offered: (i) within public transport vehicles (buses, subway, etc.), as visitors move throughout the city; and (ii) at various city locations (e.g., squares, museums) possibly taking advantage of public Wi-Fi infrastructure.

However, traditional solutions of utilizing core cloud servers for content delivery (CD), overtax network bandwidth, especially in crowded areas, and suffer from issues, ranging from non-availability to high delay response.

We envisage that the MTC will leverage the (multi-access) edge clouds deployed near crowded areas, museums, and other sights, as well as larger cloud infrastructures, potentially operated by a telco which is geographically present in the specific city, as in Figure 16. The telco's cloud infrastructure will maintain all content required by the services offered to the visitors and will further provide data processing (e.g., video transcoding) for personalized application delivery. Edge clouds will be mainly used for caching to reduce the latency while accessing data by the visitors.

Four kind of services can be envisaged:

- **Service 1: Basic sight information**, in the form of text and images, that will be transferred to the mobile device using HTTP.
- **Service 2: Advanced virtually guided sightseeing services**, in the form of downloadable location specific information, including narration (sound), video, and augmented reality facilities of tourist attractions, to the visitors' mobile device (smartphone, tablet, etc.).
- **Service 3: Touristic Social Network Services**, that involve users sharing photos, videos and comments, related to their experience on metro Social platform, including reviews of related restaurants and points of interest, cultural events, museum visits, etc.
- **Service 4: Open Traffic Management Services (IoT)**, that includes bus/metro positioning, estimated traffic load prediction based also on vehicle data, parking availability, weather conditions, emergency situations especially in underground metro transportation, etc.

Each service has particular requirements in terms of network and server resources. Service 1 mainly requires bulk-data transfers and can benefit from local copies of popular sight information (e.g., caching the information of the most popular sites across the edge clouds will reduce the processing load and network traffic towards the telco's cloud). Service 2 raises the need for certain bandwidth guarantees to sustain the large traffic volumes required for video streaming and augmented reality. The latter further requires high processing capacity that is either not available in the (visitor's) mobile device or will drain the device's battery. Hence, all CPU-intensive processing tasks should be delegated to nearby edge clouds. Service 3 entails highly time-variant data transfers (e.g., a large number of photos may be uploaded during a parade). Finally, Service 4 can be viewed as an IoT scenario with the corresponding devices /sensors mounted on the buses, with communication delay and device resource efficiency (e.g., reduced energy consumption) as main requirements.

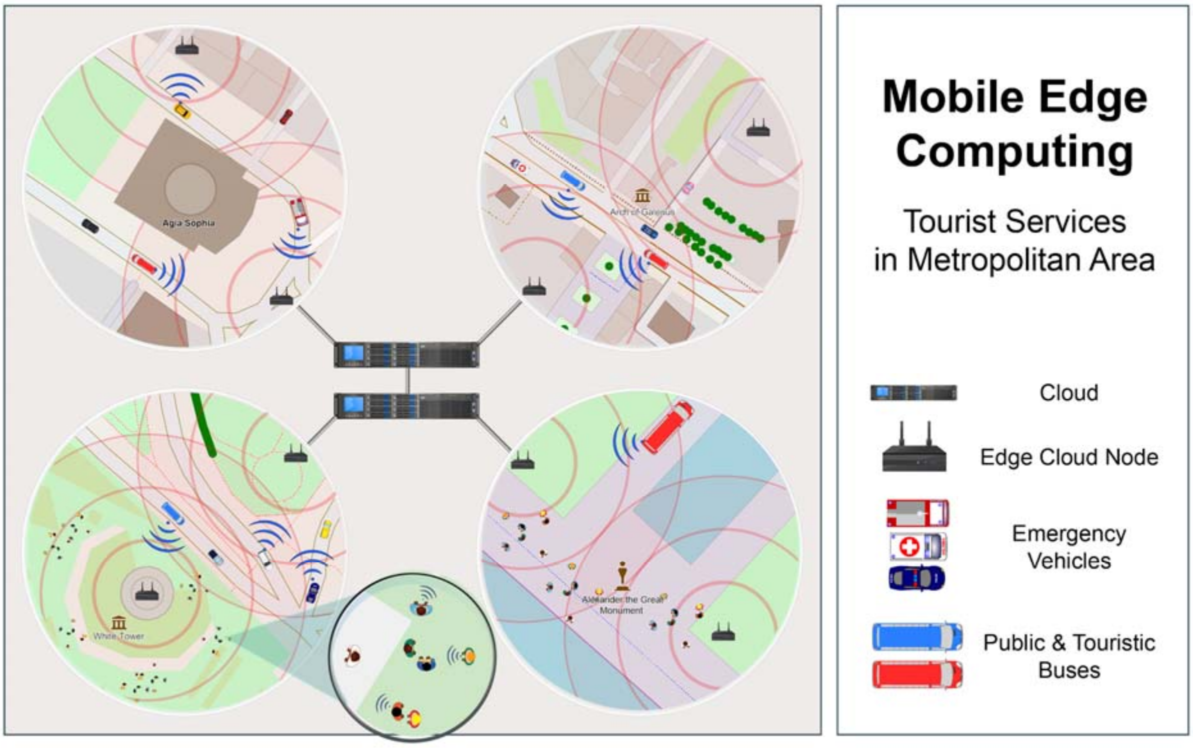

Figure 16 - Touristic Services in Metropolitan areas

NECOS can act as an enabler to the above scenario by permitting the Metropolitan Tourist Center to provision **network slices tailored to the needs of each service**. The slices will be deployed on top of the telco's cloud, the participating edge clouds, and the public Wi-Fi infrastructure. As public transport vehicles follow their itinerary in the metropolitan area, the mobile devices are connected to different geographically, distributed edge nodes, in order to achieve low latency, improved QoS and better mobility support. Updated content can be "pushed" to edge storage locations (possibly on demand) and in the case of CPU-intensive applications (like augmented reality), pre-processing tasks can be performed on edge nodes and forwarded to cloud servers for completion.

Regarding stakeholders, for services 1, 2 and 3 we can consider the following, according to the definitions provided in subsection 3.3:

- **Services End-Customer:** Visitors will be the touristic service end-users.
- **Application Service Provider (Slice Customer):** This role is assumed by the MTC Service. The MTC will deploy services for its customers-visitors and will be responsible for managing the slice.
- **Slice Service Provider:** Slice as a Service Provider (will deploy NECOS, possibly one of the infrastructure providers).

In Services 1, 2 and 3, we assume the MTC to be the Application Service Provider i.e., the "Slice as a Service customer", and visitors (travellers/general public), to be the final end-users of the services hosted in the slice. We consider a single tenant in a slice, and in fact MTC can own multiple slices according to the specific needs of the services deployed. MTC will need to setup slices by contacting a Slice Service Provider.

The metropolitan Tourist centre benefits from providing high tech content services to visitors, increasing the impact of its cultural assets to tourists. The MTC acting as the consumer of "Slice as a Service", will set up advanced information services in the federated cloud offered to its visitors which are going to be the end-users of these services.

Respect to service 4 stakeholders, we can summarize as follows:

- **Services End-Customer:** Emergency authorities (fire department, traffic police, emergency management centre), general public.
- **Application Service Provider (Slice Customer):** Municipal Authorities in cooperation with Public Transport operators.
- **Slice Service Provider:** We assume that there will be a Slice as a Service Provider (who will deploy NECOS, possibly a cloud infrastructure owner).

Regarding the Open Traffic Management Services, these aim at providing valuable information to local emergency authorities, reporting both non-emergency data such as traffic loads in major streets buses operate on, but also visual (video) feedback in cases where an emergency occurs (accident, fire, etc.). In this case, these authorities are exploiting the infrastructure the slice Municipal authorities and public transport providers have set up, in order to obtain an up-to-date view of the emergency site, which is especially valuable to first-responders. In these cases, the end-users are the emergency services.

Two sub-scenarios are considered as part of the touristic umbrella scenario: the distribution of touristic content, and the support of next generation touristic applications. The following sub-sections provide more insight on both.

4.5.3.1 Network Slicing for Touristic Content Distribution

4.5.3.1.1 Description

The ubiquity of mobile devices leads to a significant demand for state-of-the-art location-aware cultural content delivery (CD). The Metropolitan Tourist Centre (MTC) needs to provide high-quality network services (e.g., location-dependent content) to tourists with adaptable behaviour to the dynamic network conditions and particular resource constraints and QoS requirements, focusing also to user mobility.

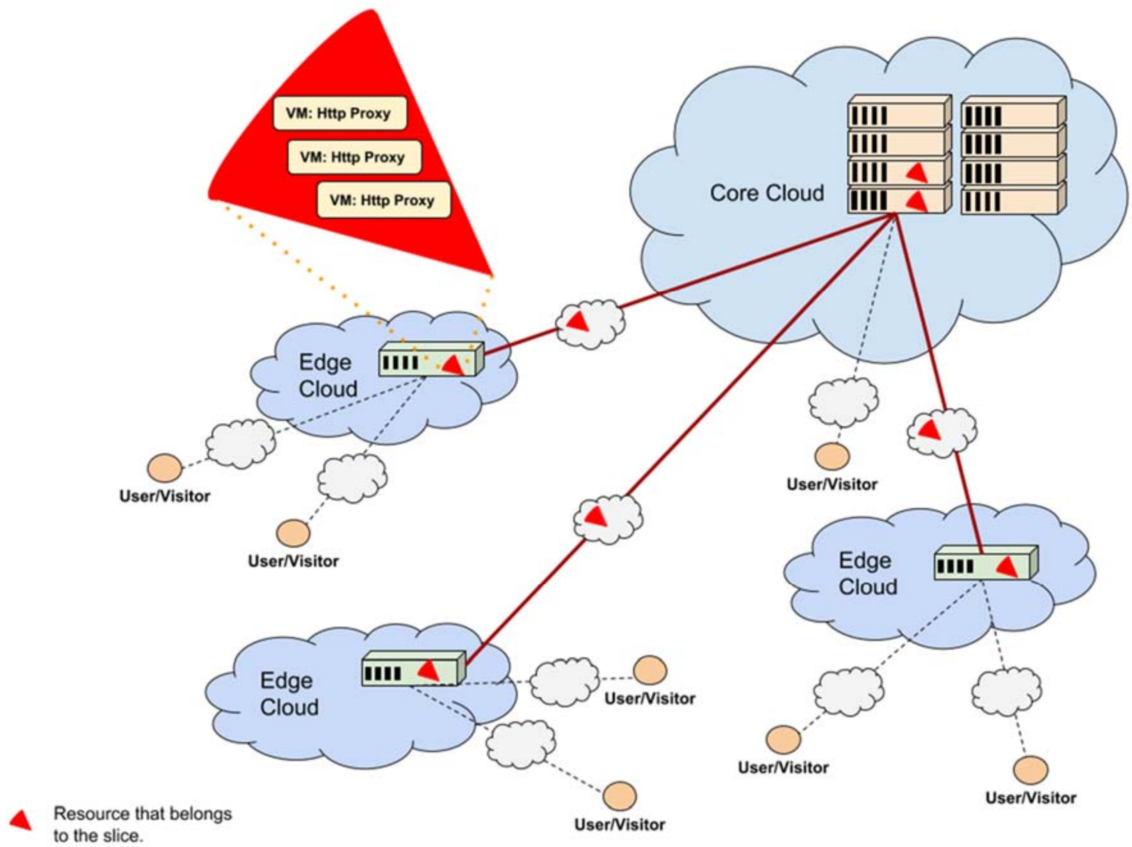

Figure 17 - Touristic Content Distribution as a NECOS Slice

In the following, we describe the different perspectives of both end-users and MTC.

End-user perspective: User Victor decides to take either a public bus or a classic hop-on-hop-off bus to have a tour of the city. Moving throughout the city, Victor downloads content related to cities monuments near his current location, e.g. text, images (Service 1) and video (Service 2). Victor, being an active social networks user, posts comments regarding his status, uploads photos, and accesses information for places of interest nearby (coffee shops, restaurants) in the Touristic Social Network (Service 3).

MTC perspective: To address the above scenario, the MTC has deployed a number of lightweight VMs hosting web servers, which offer area specific information (text, images, videos), on edge cloud servers near the Piazza. Updated content is pushed to these VMs from a central server accommodated in the same slice (Figure 17). Since the edge cloud has limited resources, content pushed is based on requests originating from visitors, thus adapting to dynamic patterns of demands. These traffic data originate from the slice's monitoring facilities and used to perform load balancing in the CD service. The main aim from the MTC point of view is to reduce the resource consumption and in turn the cost, while satisfying the end-users' requirements.

Slice Broker Perspective: The slice broker offers the slice by combining edge and core cloud resources and the network infrastructure to interconnect them. The former (i.e., the edge cloud) hosts content proxies in the form of VMs on-demand and the latter (i.e., the core cloud) hosts the content. A key aspect here is the resource-efficiency of the physical servers, while providing the resources for the deployed slices. The infrastructure providers are responsible to offer a pool of resources on demand and support the slice elasticity aspects.

For this scenario, we consider one slice/per tenant, i.e., the MTC will have some control over the slice through a set of operations/configuration actions and the slice provider will segregate the necessary infrastructure (i.e. see Figure 18).

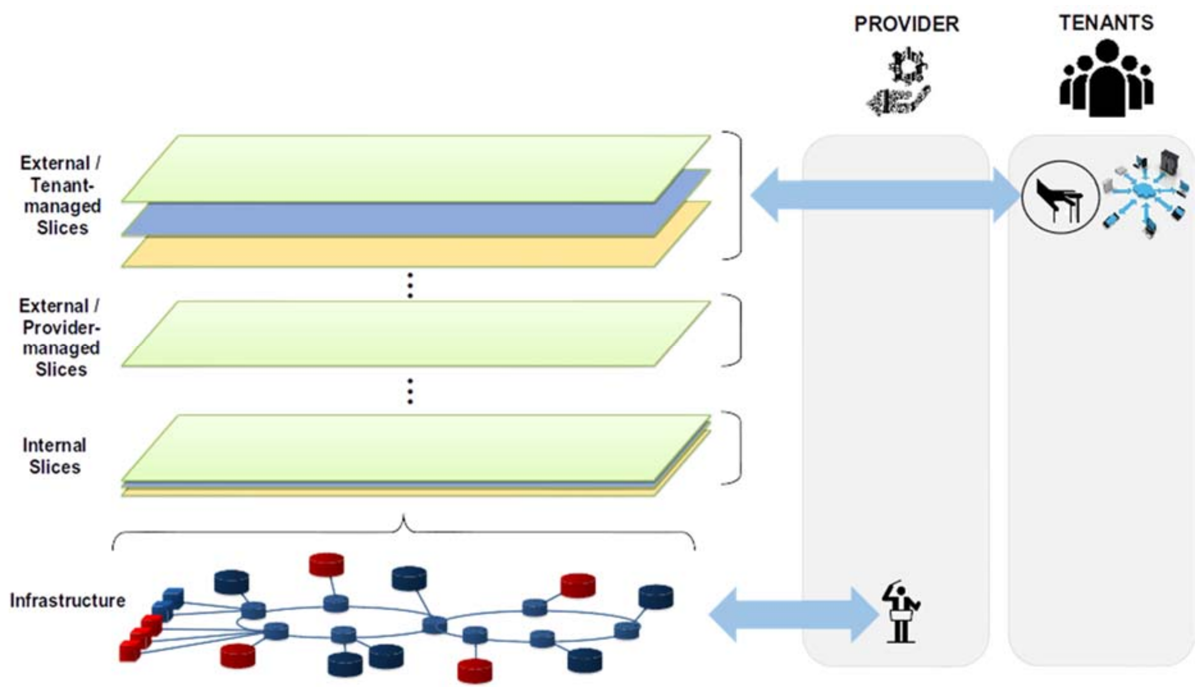

Figure 18 - Slice management in Content Distribution

User mobility is an important aspect in this scenario, as shown in Figure 19. For example, user A switches between different edge clouds, as he/she moves around the city, and utilizes sliced resources allocated on demand, i.e., the user downloads Internet content with ultra-low delays due to the locally cached content proxies that follows him/her. The content proxies are hosted in lightweight Virtual Machines (VMs) that can be deployed rapidly. As a bottom line, this scenario instantiates the Multi-access Edge Computing (MEC) use-case (i.e., see subsection 4.1), in a mobility-aware elastic content distribution context, i.e., implementing the NECOS slice as a service features for touristic content distribution.

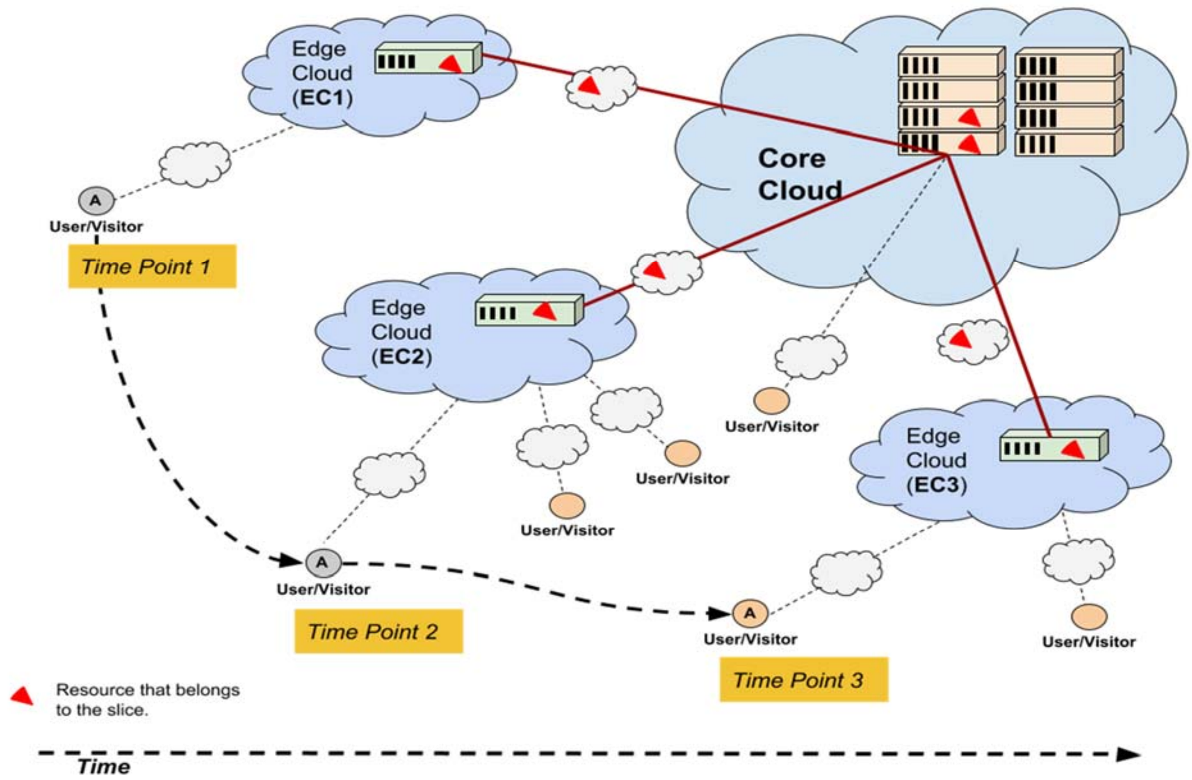

Figure 19 - User Mobility and Association with Multiple Edge Clouds

4.5.3.1.2 Technical enablers

In order to realize this scenario, the platform should provide:

- Flexible network technologies that allow traffic load balancing, for associating the end-users (visitors) with the appropriate http proxies, deployed at the edge clouds.
- Network and physical server monitoring facilities that enable intelligent decisions for http proxy deployment / management and efficient slice operation.
- Lightweight virtualization technologies for hosting Content Distribution (CD) services on the edge cloud that can be easily deployed and managed.
- Efficient and simple slice manipulation facilities allowing the implementation of alternative elastic CD approaches.
- Mobility handling features to allocate content proxies to end-users as they move around a wider geographical area, while avoiding the service disruptions due to the involved handovers between different edge clouds, as depicted in Figure 19.

4.5.3.1.3 Critical success factors and KPIs

The Network Slicing for the Touristic Content Distribution scenario is assessed from the point of view of the involved stakeholders. Relevant critical success factors are:

- F1 - Quality of content provisioning, i.e., the delay the visitor experiences when retrieving a webpage, or uploading photos, should be under application-specific limits. This allows an enhanced viewing experience to the user (Slice Efficiency);
- F2 - Heterogeneous resource management over multiple VIMs deployed at both edge and core clouds. Especially in the case of edge clouds, the resource utilization should be as efficient as possible, allowing the slice provider and the MTC to offer quality services with a minimum cost (Slice Efficiency / Cost Reduction);

- F3 - Elastic slice operation targeting a large number of supported end-users without compromising the level of provided service (Elasticity); and
- F4 - User mobility handling aspects that enable new touristic applications with locality considerations (Flexibility).

The above success factors can be quantified using a number of KPIs, such as the following:

- K1 - Content delivery performance to highlight the end-user satisfaction, e.g., for video streaming content, such as video start-up time, buffering percentage, switching between bit rates, etc.; Metrics – end-user QoE.
- K2 - Overall consistency to SLA index, expressing the Quality of Service to demonstrate the efficient network performance over heterogeneous resources with respect to the SLA; Metric – SLA fulfilment index.
- K3 - Physical server utilization to quantify the resource-efficiency of the NECOS Slice as a Service capability in the touristic content distribution scenario, while considering the case of multiple coexisting VIMs (i.e., CPU utilization, memory allocation, link utilization); Metrics – physical server utilization.
- K4 - Service disruption index to evaluate the efficient mobility handling, i.e., expressing end-to-end availability (% of time in which a ping gets a response over total time); Metrics – service disruption index.

4.5.3.1.4 Mapping to NECOS key characteristics

The aim of the scenario is to utilize the NECOS platform (NECOS objective 1) to demonstrate a novel Service Provisioning (NECOS Objective 2), i.e., elastic content distribution in a Multi-access Edge Computing context, through Uniform and Efficient management of infrastructure resources (NECOS Objective 3), with emphasis on automatic reallocation of resources and services across geographically distributed and computing/storage/networking infrastructures. The scenario validates the full impact of the NECOS approach to the network slicing (NECOS Objective 4).

We consider the following key NECOS characteristics for the scenario:

- C1: The MTCs can create, based on the **slice-as-service model** (Characteristic 1), a novel elastic content delivery platform for geo-specific content that is adaptable to the number of expected visitors. For example, the deployed slices can scale up when large scheduled events occur (e.g., New Year's Eve) or in an "on-demand" fashion.
- C2: Since the MTC services require both edge and core cloud resources, thus the ability to **configure slices across heterogeneous physical resources and VIM technologies** (Characteristic 2) is critical.
- C3: The MTC **manages everything via software** (Characteristic 3), including network resource allocation, monitoring slice/service performance and the slices' (re-)configuration. For instance, the MTC can dynamically adapt to an unexpected large number of requests for popular content or to support user mobility.
- C4: Content delivery http proxies are based in **lightweight virtualization technologies**, deployed on small edge cloud servers, closely operating with core cloud content delivery services. Such a deployment, requires a **unified view and management of core cloud, edge cloud and network components** (Characteristic 4), a key NECOS platform characteristic.

4.5.3.1.5 Functional and Non-functional Requirements

The following are considered to be functional requirements for the specific scenario:

- *RF.Touristic(CD).1: Slice and slice-resource management*. Definition and lifecycle management of a slice as a service over a certain geographical area that includes both core and edge-cloud resources. The MTC sets up the desired slices around areas of high interest in its city.

- *RF.Touristic(CD).2: Automated Virtual Machine deployment.* Deployment of lightweight and regular VMs on particular slices, based on real-time and historical monitored data. This involves the creation/deletion of VMs on strategically selected physical nodes.
- *RF.Touristic(CD).3: Traffic load-balancing for content delivery*. Automated assignment of lightweight VMs providing cached content to Users, i.e., load-balancing of new bursts of user requests to newly created content-hosting VMs or to follow end-users as they move.
- *RF.Touristic(CD).4: Slice resource and service monitoring*. Monitoring of resource specific traffic patterns/service usage to detect changes in service requests demands.
- *RF.Touristic(CD).5: Service planning.* The MTC should be able to request additional resources to be included in its slice, to cope with occasional circumstances in which high traffic is anticipated. For instance, scheduled festivities that attract an increased number of visitors (e.g., the New Year's Eve) require more infrastructure resources for a relatively short period of time.

Respect to non-functional requirements, these are the ones identified so far:

- *RN.Touristic(CD).1: Transparent end-user performance*. End-user performance transparency, with respect to service usage, i.e. visitors should be unaware of changes due to load-balancing actions (introduction of new VMs, changes in routing, etc.), and experience minimal service delays and interruptions.
- *RN.Touristic(CD).2: Heterogeneity handling*. The service provisioning should be transparent with respect to details of the physical infrastructure, VIM used, load balancing and slice management, i.e. the MTC must "see" and control slice resources in a uniform manner, regardless of the underlying (different) technologies used.
- *RN.Touristic(CD).3: Elasticity*. The slice as a service should adapt to changes in the network (e.g., delays due to congestion) or user context (e.g., new end-users appear, content becomes viral etc.).
- *RN.Touristic(CD).4: Resource-efficiency*. The slice resources should be efficiently utilized, i.e., no over-provisioning of resources.
- *RN.Touristic(CD).5: Scalability*. Touristic content distribution should be scalable, to offer quality services to a large-number of users.

4.5.3.2 Multi-Domain Network Slicing for Next Generation Touristic Applications

4.5.3.2.1 Description

Next generation touristic services, involve augmented reality (AR) applications that offer a technologically advanced, enhanced travellers' experience. These applications aim to increase the value of the touristic product of high profile metropolitan areas, by adding a layer of content to locations captured by the traveller's mobile device (e.g., smartphone or PDA). It is expected that municipal authorities are going to adopt the provisioning of such services in the near future. However, augmented reality applications demand significant CPU power for image processing and content to annotate camera video, that is either not available in the (visitor's) mobile device or will overtax the device's battery.

In the NECOS project, all CPU-intensive processing tasks should be delegated to nearby edge or core clouds, offering AR applications as a service.

As in the previous case, the scenario assumes a Metropolitan Tourist Centre (MTC) acting as a main provider of high-quality network services (e.g., location-dependent content) to tourists with particular resource constraints and QoS requirements, but now focusing on next-generation touristic services, such as those utilizing augmented reality.

Victor is the *end-user* of the services residing in the slice, while the MTC is the *Public Service Customer* (i.e., slice as a service consumer), that has setup all the above Services in appropriate slices through the *Slice Provider*. In the following, we describe the different perspectives of both end-users (visitors) and MTC.

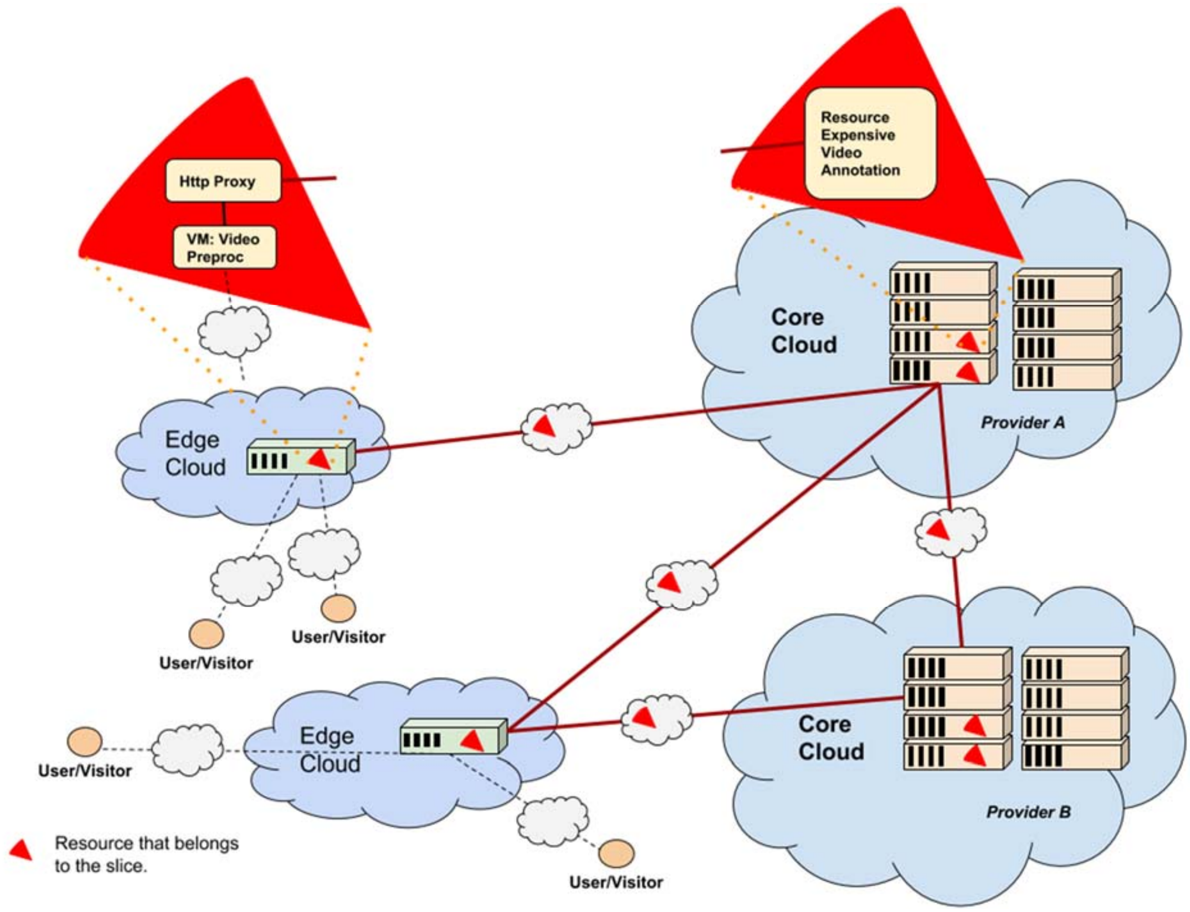

Figure 20 - Augmented Reality Touristic Services in the Multi-Domain Cloud

End-User Perspective: User/visitor Victor arrives at touristic location Piazza, where monument(s) Cathedral is located. Upon arrival, Victor downloads from the MTC web server a map to navigate in the Piazza, information (e.g., historical and architectural data) regarding the Cathedral (Service 1), related videos and sound recordings (Service 2). Victor points his cell phone to the facade of the Cathedral and the Virtual Guide app displays on his screen (augmented reality) information regarding particular architectural details (Service 2). Being fascinated by the facade's beauty posts his selfie, status and comments on the metro Social platform (Service 3). Once his visit is completed, he checks traffic conditions (Service 4) to decide whether to continue his tour to another place of interest or rest at a Piazza cafe (Service 3). If he decides to follow the first option, then upon his arrival to the new location, he will be offered a similar set of services, connecting to its new location edge cloud (possibly to a different one than that of the Piazza).

MTC Perspective: The MTC has deployed a slice hosting a number of VMs that realize a touristic augmented reality application. Augmented reality content delivery is a multistep process that can be implemented with service function chaining, i.e., pre-processing of content input data is performed on a dedicated VM (located on the edge cloud), then the data is transmitted to a cloud server for computational intensive processing and finally the enriched content is sent back through the edge cloud VM to the end-user. This communication involves a number of VMs deployed to both edge and core clouds, owned possibly by different cloud providers, so multi-domain orchestration is a main requirement.

Slice Broker Perspective: The slice broker sets up and manages the slice involving resources from multiple providers. Since the connectivity between different edge clouds and core cloud servers requires high link capacity and ultra-low delay, a single provider may not be present in all city areas. The involved data centre infrastructure and network connectivity providers are motivated to participate in the resource federation through the reduced cost due to the efficient resource allocation and new novel touristic services that bring profit. A market-place type of resource offering and reservation balances the requirements and directions of different cloud providers.

In this case, we still consider one slice/per tenant, i.e., the MTC will be the single tenant in the slice, but due to the federation the slice control is on the slice (broker) provider side.

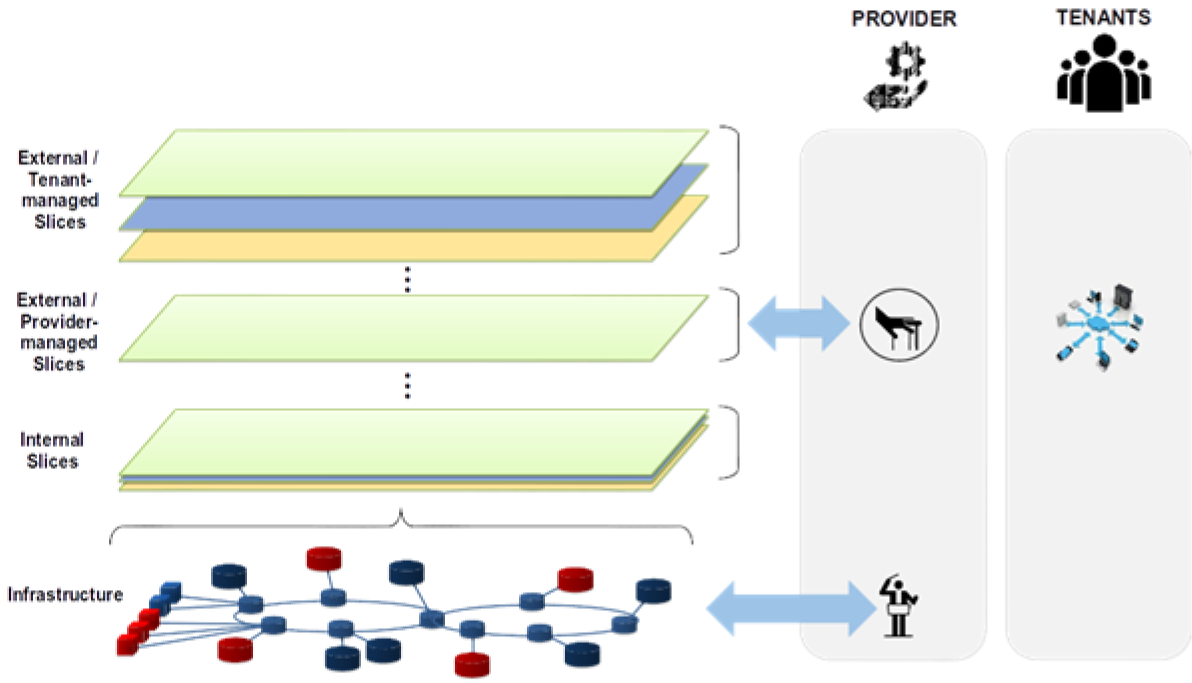

Figure 21 - Slice Management in AR Service

This is another instantiation of the Multi-access Edge Computing use-case that pushes the boundaries of the "Network Slicing for Touristic Content Delivery" scenario towards multi-domain orchestration and heterogeneous resources' federation at edge and core clouds (i.e., for both processing and communication). Its main focus is to provide novel next-generation touristic services.

4.5.3.2.2 Technical enablers

We see the following technical enablers in this scenario:

- Discovery and allocation of resources to form a slice over multiple domains, respecting a number of constraints, such as geographical location, bandwidth and delay connectivity demands, CPU requirements, etc.
- Flexible network softwarization technologies that allow efficient traffic load balancing that improves resource utilization of both network and cloud servers.
- Network-Function Virtualization technologies to support the service chaining necessary for the operation of the augmented reality service (i.e., VNFs in both edge and core clouds).
- Multi-domain orchestration facilities for the efficient slice manipulation over federated resources that allow a unified view of the underlying resources in the created slice.
- Intelligent orchestration techniques and prediction mechanisms on expected usage regarding both infrastructure resources and end-user demands. The latter aim to support advanced techniques in multi-domain orchestration, by detecting patterns of service usage, that indicate changes in demands in order to intelligently infer resource orchestration decisions.
- Network and physical server monitoring facilities that enable prediction mechanisms and decision making regarding elasticity and efficient multi-domain slice operation.

4.5.3.2.3 Critical success factors and KPIs

The critical success factors of the "multi-domain network slicing for next generation touristic applications" scenario follow:

- F1 - Quality of service in next-generation touristic applications, i.e., the slice should sustain the *large traffic volumes* and offer the *processing* required for video streaming and augmented reality, in order for the end-users to enjoy real-time, low-latency and low jitter service delivery (Slice Efficiency);
- F2 - Resource efficiency and balanced resource allocation between the different infrastructure providers, motivating different providers to realize multi-domain slices (Slice Efficiency);
- F3 - Scalability and reduced cost in slice setup and management that will allow infrastructure providers to extend their client-base and the MTCs to offer quality services to more citizens (Cost Reduction/Scalability);
- F4 - Efficient resource offloading between the different clouds to support a large number of users, i.e., the service should be able to scale to satisfy a significant number of visitors within a limited geographical area (Lifecycle Efficiency/Flexibility); and
- F5 - Intelligent multi-domain orchestration that supports prediction of service demands and resource utilization, to cope with time-varying service demands (Elasticity).

The above success factors can be quantified using KPIs, such as the following:

- K1 - AR application performance in terms of response time of receiving annotated video, to highlight the end-users' satisfaction, e.g., for video streaming AR applications such as: video start-up time, buffering percentage and switching number between bit rates, etc.; Metric – end-user QoE.
- K2 - Number of application users serviced within an area to demonstrate the scalability aspect; Metric – number of application users.
- K3 - Time to adapt network service provisioning through multi-domain orchestration, in cases of a sudden rapid increase in the service requests number; Metric – service adaptation time.
- K4 - Accuracy of prediction of service demands and resource utilization for the intelligence aspects of multi-domain orchestration. Metric - service demands prediction accuracy.
- K5 - Physical server utilization per cloud to quantify the resource-efficiency and the balanced resource offloading between the different clouds providers (i.e., CPU utilization, memory allocation, link utilization); Metric – physical server utilization.

4.5.3.2.4 Mapping to NECOS key characteristics

The scenario addresses all NECOS objectives and key characteristics. For example, the NECOS platform (Objective 1) enables novel next-generation touristic applications (Objective 2) through realizing multi-domain orchestration over multiple providers and heterogeneous clouds (Objective 3). Furthermore, it demonstrates all the main NECOS characteristics (Objective 4), such as:

- **C1: Slice as a Service provisioning** (Characteristic 1). Since the MTC has to specify and deploy its services in the cloud, through integrating Cloud and Edge cloud resources from multiple providers.
- **C2: Configuration of slices across the physical resources in the cloud** networking infrastructure to better accommodate various service demands (Characteristic 2). The slice is dynamically configured to support augmented reality applications that pose stringent requirements in terms of network and cloud server resources.
- **C3: Everything is managed via software** (Characteristic 3), from the network configuration and adaptability to the service function chain orchestration that enables novel next-generation touristic content delivery.
- C4: It supports **uniform management over lightweight virtualized resources** (Characteristic 4), through realizing service function chains over edge clouds with limited physical server resource availability. In case of more demanding processing, a nearby core cloud, residing in the same or different infrastructure provider, is involved.

4.5.3.2.5 Functional and Non-functional Requirements

The functional requirements of this particular scenario follow:

- *RF.Touristic(APP).1: Service function chain orchestration,* i.e., to deploy and configure the service components realizing the augmented reality applications. For example, NECOS can support automated deployment of lightweight VM on slice resources based on real-time and historical monitored data.
- *RF.Touristic(APP).2: Resource and user-demand prediction capabilities,* i.e., driving the NECOS slice as a service capabilities and offering resource-efficient next-generation touristic services.
- *RF.Touristic(APP).3: Resource offloading between edge, core clouds and cloud providers.* The synergy between cloud providers and hierarchical clouds can make a range of new touristic services possible, e.g., ultra-low delay augmented reality applications.

We consider for the scenario the following non-functional requirements:

- *RN.Touristic(APP).1*: *Resource federation and intelligent multi-domain orchestration*, to enable novel touristic services that were not possible before.
- *RN.Touristic(APP).2: Scalability*. Touristic augmented reality services should be offered at a scalable manner to support a large-number of users.
- *RN.Touristic(APP).3: Efficient next-generation touristic application performance*. Next-generation touristic applications, such as augmented reality, are very resource-demanding.
- *RN.Touristic(APP).4: Elasticity*. Such novel resource-demanding applications can operate only over slices that adapt to changes in the network or user contexts.

4.5.4 Emergency Scenario

This scenario addresses a Command, Control and Communications (C3) Public Safety Centre (PSC) [11], responsible for enhancing incident management and resolution to a significant number of first responders (e.g., police department, fire & rescue services, emergency medical services, and public works), as well as their interaction with civilian population in a metropolitan area. The C3 PSC aims at leveraging the information flow level between citizens, responders, and agencies, by quickly offering the ability to receive, correlate, and share information.

These services will be offered *(i)* within public transport vehicles (buses, subway, trains, etc.), as they move throughout the city, *(ii)* to citizens and visitors at various public buildings (e.g., city hall, hospitals, public schools, and museums) possibly taking advantage of public Wi-Fi infrastructure, or *(iii)* by storage, computing and infrastructure resources deployed by C3 PSC agencies (for instance, public agency mobile and/or stationary data centres, C3 applications, and tactical radio terminals).

Thus, as depicted in Figure 22, C3 PSC scenario may utilize either (mobile) edge clouds deployed near crowded areas, museums, and other public buildings, on an opportunistic manner or based on agency-owned infrastructure. Additionally, it may use larger cloud telco-operated infrastructure, maintaining all content (e.g., weather channel, traffic updates, and emergency alerts) required by the services offered to the civilian population, as well as data processing (e.g., intelligence reports, suspects' identification, video transcoding) for on-demand secure application delivery to authorized personnel. On the other hand, edge clouds caching will reduce latency to access data for stakeholders (civilians and first responders) when deployed on the field.

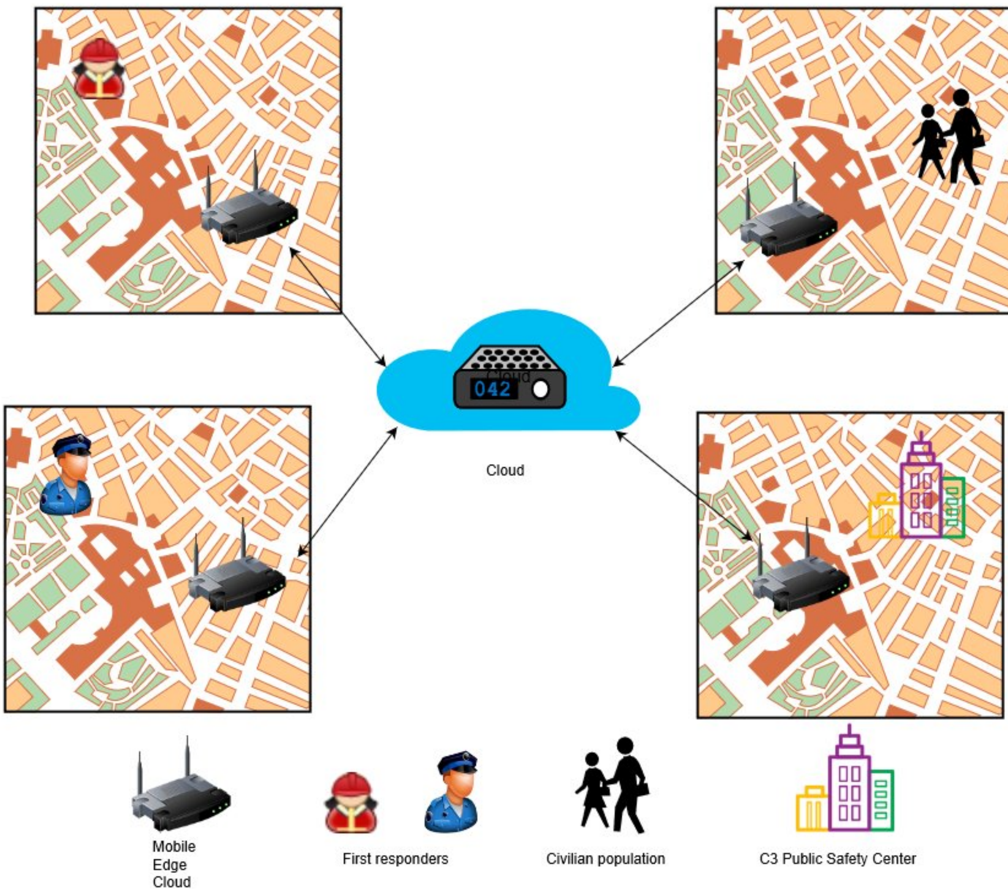

Figure 22 – Command, Control and Communications scenario

In brief, the C3 PSC may offer the following services:

1) Service 1 - Basic community information, in the form of text and images sent to mobile devices using HTTP. Targeted audience: civilian population.
2) Service 2 - Advanced GIS mapping and location services, in the form of downloadable location specific information, including narration (sound) and video of safety incidents, targeted to first responder's mobile equipment, as well as to civilian mobile device (smartphone, tablet, etc.). Targeted audience: civilian population and first response officers.
3) Service 3 - C3 Network Services, that involves users sharing photos, videos and comments, related to their experience on different city spots, including reviews of criminal incidents, comments regarding suspect personnel, traffic updates, weather updates, etc. Targeted audience: civilian population and first response officers.
4) Service 4 - Advanced Analytics & Incident Management Services (i.e., IoT), that includes information sharing to/from the C3 PSC out to/from responders at the scene, gathering audio, text, image, video, and reports produced by responding agencies, citizens and officers into the command centre, as well as by sensors/devices deployed on the field with the public safety staff. Targeted audience: first response officers and C3 PSC command staff.

Thus, several enablers (a.k.a. stakeholders) shall be listed to provide the aforementioned features, namely:

- C3 PSC command staff representatives (police department, fire & rescue services, emergency medical services, and public works);
- First response officers, as well as sensors/devices deployed on the field (transport vehicles, public buildings, public Wi-Fi infrastructure)

EUB-01-2017

- Civilians in general; and
- Service providers, responsible for content provision and data processing infrastructure, which can be from 3 different types:
    - T1 – Non-stationary C3 PSC systems: edge clouds and radio resources from C3 PSC agencies deployed close to emergency incident spots;
    - T2 – Mobile public-owned systems: edge clouds and radio resources from museums, public buildings, transportation facilities, or near crowded areas, opportunistically used under emergency needs; and/or
    - T3 – Private-owned systems: namely, cloud telco-operated infrastructure.

Based on the preconditions stated above, there are several advantages on the adoption of NECOS network slicing mechanisms, since each user/group of users may behave as a real and independent network.

Also in this case, two sub-scenarios are considered as part of the Command, Control and Communications - Public Safety scenario: the smart city data content distribution, and the Metropolitan Integrated Monitoring (MIM). The following sub-sections provide more insight on both of them.

4.5.4.1 Network Slicing for Smart Cities Data Content Distribution

4.5.4.1.1 Description

According to United Nations 2014 World Urbanization Prospects [12], 54 per cent of the world's population lives in urban areas, a proportion that is expected to increase to 66 per cent by 2050. Public safety is one of the most relevant issues for this large population, whether in more developed centres or in developing countries. Thus, the widespread usage of mobile devices by the population in urban areas leads to an increase in demand for novel techniques related to location-aware public safety content delivery. However, traditional solutions, ranging from SMS alerts to over-the-top smartphone applications, experience side-effects like useless data sharing, high delay response, and low availability rate. Similarly, topics like traffic, major sport and artistic events, floods, fires, manifestations and strikes present great relevance and interest.

Given the scenario stated above, we introduce the Network Slicing for Smart Cities Data (SCD) Content Distribution scenario, which addresses a Command, Control and Communications (C3) Public Safety Centre (PSC) responsible for enhancing metropolitan authorities' interaction with civilian population in a metropolitan area. The C3 PSC aims at leveraging the information flow level between citizens and agencies, by quickly offering the ability to receive, correlate, and share timely information in an ecosystem of lightweight edge clouds to be in support of the mobile user applications, paving the way of the Lightweight Slice Defined Cloud paradigm to demonstrate NECOS suitability to the aforementioned MEC use case.

To accomplish its goal, the SCD scenario may utilize either (mobile) edge clouds deployed near crowded areas, museums, and other public buildings, on an opportunistic manner or based on agency-owned infrastructure. Additionally, it may use larger cloud telco-operated infrastructure, maintaining all content (e.g., weather channel, traffic updates, and emergency alerts) required by the services offered to the civilian population. On the other hand, edge clouds caching will reduce latency to access data for the urban population. NECOS will provide the necessary solutions for the C3 PSC to deploy SCD services over slices with adaptable behaviour to the dynamic network conditions and application requirements, focusing on user mobility and quality of experience.

In order to cope with the services described above, one can envision the creation of different slices like:

- Internal slices: targeted to C3 PSC command staff and first response officers in charge of network management tasks. Those slices should care of operational control tasks – namely, network functions monitoring/configuration/performance evaluation, as well fault & security management; and

- External slices: targeted to broader audience (civilian population and first responders small task groups), which can be of two classes:
  - Managed: where consumers only use the network resources provided by C3 PSC, without any management or control capabilities. Suited for civilian population, which only consumes the resources provided by C3 PSC central body.
  - Unmanaged: here, consumers have partial or even complete control of the network functions allocated by C3 PSC central body. Suited for the public-owned metropolitan small task groups, due to mission dynamics, which requires features like real-time VNF provision in order to provide C3 capacities to the public agents deployed on the field.

Thus, several service and infrastructure providers shall be listed to provide the aforementioned features, responsible for content provision and data processing infrastructure, which can be from 3 different types:

- T1 – Non-stationary C3 PSC systems: edge clouds and radio resources from C3 PSC agencies deployed close to emergency incident spots;
- T2 – Mobile public-owned systems: edge clouds and radio resources from museums, public buildings, transportation facilities, or near crowded areas, opportunistically used under emergency needs; and/or
- T3 – Private-owned systems: namely, cloud telco-operated infrastructure.

Several and distinct perspectives can be foreseen at SCD scenario for different stakeholders, according to NECOS ecosystem categories:

Slice consumer perspective: User Alice, living in a large urban centre, wants to get updated traffic and incident reports to guide her daily home-to-work commuting. Moving throughout the city, Alice downloads content related to main points of interest near her current location, e.g., text, images (C3 PSC Service S1) and video (C3 PSC Service 2). Besides that, Alice posts comments regarding her status with family, friends, and co-workers, uploads photos, and accesses information for places of interest nearby (weather reports, car accidents) in the C3 Social Network (Service 3). The main focus is to improve the Quality of Experience (QoE) of users like Alice.

Infrastructure Resource Provider Perspective: The infrastructure resource provider offers edge and core clouds, data centre (e.g., compute, storage, network) and/or wired and wireless connectivity resources to build slices. Several key aspects may be listed, like physical and virtual servers resource-efficiency, on-demand and timely resource offers to support NECOS elasticity aspects, stringent requirements for data isolation and integrity, and adoption of low-cost data centre solutions to provide the resources for the deployed NECOS slices.

Service provider perspective: To address the above scenario, the C3 PSC has deployed a number of lightweight VMs hosting web servers, which offer widespread area specific information (text, images, videos), on edge cloud servers scattered throughout the city. Updated content is pushed to these VMs from C3 PSC central server accommodated in the same slice. The main goal from the C3 PSC point of view is to provide timely and consistent information to population in general. Moreover, it shall accomplish resource consumption reduction with a low-cost approach.

4.5.4.1.2 Technical enablers:

We see some boundary conditions in this scenario:

- Widespread usage of Software-Defined Networking technologies by operators and providers to support data security policy enforcement (e.g., between agency-owned core data centres and edge clouds) and traffic load balancing (e.g., between the end-users and the http proxies deployed at the edge clouds).
- Automatic network and physical server monitoring by operators and providers to enable elasticity and efficient slice operation.

- Lightweight virtualization technologies (e.g., UCL's VLSP, unikernels or containers) for the edge clouds' virtual resources
- Elastic content distribution technologies for end-users that consider NECOS' Slice as a Service.
- Mobility handling aspects by operators and providers to allocate content proxies to end-users as they move around the area, while avoiding the service disruptions due to the involved handovers between different edge clouds.

Given the aforementioned support, we envision the following technical enablers:

- Dynamic slice creation and tear-down in any given SCD network slice.
- Services and capabilities definition and on-demand update supporting interest data retrieval.
- Support to slice merging, isolation, and modification, w/ on-demand capacity redefinition and function insertion/removal supporting slice consumer mobility and disruption.
- Slice monitoring and maintenance to accomplish service provision and resource management scalability supporting typical fluctuations on slice consumer population in a timely manner.
- Operation onto a multi-domain environment, to allow missions gathering several service providers (namely, public safety agencies), with previously agreed interfaces;
- Provision of an orchestration-enabled environment to deal with resource requests from different slice consumers.
- Infrastructure services provider support to typical device density in metropolitan areas to attend low speed vehicles and pedestrians
- High-level service and slice reliability, survivability, and availability, for different types of traffic to accomplish service customer requirements.
- Provision of a secure data exchange environment, enhanced by extreme levels of confidentiality, message authentication, operation registry and auditing, and detection of violation of authorized policies to deal with service provider security policy.
- Exposition of features like resource discovery, selection and allocation to support slice provision on a timely basis.

4.5.4.1.3 Critical success factors and KPIs

The Network Slicing for Smart Cities Data Content Distribution (SCD) scenario is assessed from the point of view of the involved stakeholders. In brief, the most critical success factor is situational awareness - "all knowledge that is accessible and can be integrated into a coherent picture, when required, to assess and cope with a situation" [13], which imply in relevant critical success factors as:

- F1 - Elastic slice operation providing a high degree of scalability (variable and dynamic number of users)
- F2 - Elastic slice operation providing short-time response for the deployment on the field;
- F3 - High QoE (low jitter, low packet error rate, high availability, redundancy mechanisms, and priority schemes implementation) for end-users;
- F4 - Network management (cross-layered approach for monitoring, analysis, and update) to endure long-time operations under different mobility patterns; and
- F5 - Easy and low-cost network elements configuration.
- F6 - Security management based on proper data segregation and authentication patterns.

The above success factors may be derived in a set of KPIs, such as the following:

- K1 - Physical server utilization to quantify the resource-efficiency of the NECOS Slice as a Service capability in SCD scenario. Metric - physical server utilization;
- K2 - Content delivery performance to highlight the end-user satisfaction. Metric – end-user QoE;
- K3 - Service disruption to evaluate the efficient mobility handling. Metric – service disruption index;
- K4 – Timely and cost-effective slice deployment, conjugated with end-user QoE to demonstrate the efficient network performance. Metric – slice time deployment; and
- K5 – Slice isolation to highlight data segregation and authenticity over a multi-domain environment. Metric – slice isolation index.

4.5.4.1.4 Mapping to NECOS key characteristics

The particular service needs for the scenario are the following:

- Secure content availability
- Support for alleviating bulk data transfers of public safety data to end-users
- Low latency content-delivery to end-users (civilian population) by low-cost dynamic resource allocation according to end-user demands.

Thus, NECOS service provisioning approach (Objective 2) is suitable for SCD scenario, since it foresees the integration of resources within the collection of independent slices. A goal of the LSDC approach is to reduce the complexity and timescale for service provisioning and deployment in federated DCs, thus reducing the OPEX for the infrastructure owner.

Moreover, NECOS also envisions service and resource orchestration and management methods for the LSDC infrastructure resources that are located within and at the edge of the SCD network. This service orchestration and management approach, which comprises NECOS Objective 3, includes the automatic re-allocation of resources and services across distributed and geographically separated computing, data storage, and network infrastructures within separate slices to support population requests for SCD application data.

Finally, NECOS objectives 1 (develop and build a Lightweight Slice Defined Cloud (LSDC) platform enabling computing, network, and storage elements in the cloud through the Slice as a Service across federated clouds) and 4 (a pilot-based impact demonstration) will be achieved as well to demonstrate the provision of location-aware public safety content delivery to population.

In order to accomplish the aforementioned objectives, we consider the following key NECOS characteristics for the scenario:

- C1: Slice as a Service, mapping enhancing component (re) configuration to support E2E SCD applications;
- C2: Lightweight Edge Cloud, with small footprint components, deployable on both core and edge small servers and cloud systems, to support a user environment characterized by mobility and disruption
- C3: Software Management under a per-slice approach, to support features like user demand elasticity (i.e., Load Balancing through multiple VMs) and monitoring.

4.5.4.1.5 Functional and Non-Functional Requirements

Based on C3 PSC scenario, this section presents the main requirements for the smart cities data domain. These requirements are listed under functional and non-functional scopes:

The functional requirements considered are the following ones:

- RF.emergency.1 – *Dynamic slice management*: support to dynamic slice creation and tear-down in any given PSC network slice.

- RF.emergency.2 – *Dynamic service definition*: support to services and capabilities definition and on-demand update.
- RF.emergency.3 – *Timely slice management*: support to slice monitoring and maintenance to accomplish service provision and resource management scalability.
- RF.emergency.4 – *Orchestration*: provision of orchestration capabilities to deal with resource requests from different users, under a multi-domain environment, to allow missions gathering several public safety agencies, with previously agreed interfaces.

Regarding non-functional requirements, these are the ones identified:

- RN.emergency.1 – *High Reliability:* provision of high-level service reliability (> xx %)
- RN.emergency.2 – *High Availability:* provision of high-level service availability (> xx %)
- RN.emergency.3 – *High Survivability:* provision of high-level service survivability (> xx %) for the scenario described in requirements RN.emergency.1 and RN.emergency.2

4.5.4.2 Network Slicing for Metropolitan Integrated Monitoring

4.5.4.2.1 Description

In addition to SCD scenario, we introduce the Network Slicing for Metropolitan Integrated Monitoring (MIM) scenario, which addresses a Command, Control and Communications (C3) Public Safety Centre (PSC) responsible for enhancing metropolitan authorities command and control capabilities in a metropolitan area. MIM aims at leveraging the information flow level between distinct public agency teams – namely, first responders deployed on the field and major staff authorities headquartered at C3 PSC buildings, by quickly offering the ability to receive, correlate, and share timely information in an ecosystem of lightweight edge clouds to be in support of the mobile user applications, paving the way of the Lightweight Slice Defined Cloud paradigm to demonstrate NECOS suitability to the aforementioned MEC use case.

To accomplish its goal, MIM scenario may utilize either (mobile) edge clouds deployed near crowded areas, police stations, fire department stations, town hall, and other public buildings, on an opportunistic manner or based on agency-owned infrastructure. Additionally, it may use larger cloud telco-operated infrastructure, maintaining all content (e.g., C3 state-owned applications, which gather data from sensors deployed on the field, as well as inputs from weather channels, traffic updates, and emergency alerts) required by the services offered to public authorities.

On the other hand, edge clouds caching will reduce latency to access data for the first responder’s teams. NECOS will provide the necessary solutions for the C3 PSC to deploy MIM services over slices with adaptable behaviour to the dynamic network conditions and application requirements, focusing on user mobility and quality of experience.

In order to cope with the services described above, one can envision the creation of internal and external (managed and/or unmanaged) slices like proposed for the SCD scenario.

Moreover, several service and infrastructure providers shall be listed to provide the aforementioned features, responsible for content provision and data processing infrastructure, which can be from 4 different types - the 3 types described in SCD scenario and S4 - Advanced Analytics & Incident Management Services (i.e., based on the IoT), that includes information sharing to/from the C3 PSC out to/from responders at the scene, gathering audio, text, image, video, and reports produced by responding agencies, citizens and officers into the command centre, as well as by sensors/devices deployed on the field with the public safety staff. Targeted audience: first response officers and C3 PSC command staff.

Based on the preconditions stated above, there are several advantages on the adoption of network slicing mechanisms, since each user/group of users may behave as a real and independent network.

Several and distinct perspectives can be foreseen at MIM scenario for different stakeholders, according to NECOS ecosystem categories:

Slice consumer perspective: User Julian, a policeman serving at an urban garrison, wants to get and share updated crime reports downtown with his hierarchical superiors based at C3 PSC. Moving throughout the city, Julian shares data to/from the C3 PSC out to/from responders at the scene, gathering audio, text, image, video, and reports produced by responding agencies, citizens and officers into the command centre, as well as by sensors/devices deployed on the field with the public safety staff (C3 PSC Service 4). The main focus is to improve the Quality of Experience (QoE) of users like Julian.

Infrastructure Resource Provider Perspective: Similarly, to SCD scenario, the infrastructure resource provider offers edge and core clouds, data centre (e.g., compute, storage, network) and/or wired and wireless connectivity resources to build slices.

In addition to SCD key aspects, we may list specific C3 requirements, ranging from advanced analytics with Augmented Reality support provided by C3 PSC headquarters to a plethora of field data provided by IoT sensors (e.g., IR camera, video, speed limit sensors) deployed on the field to provide the resources for the deployed NECOS slices.

Service provider perspective: Several public agencies shall be motivated to participate in the resource federation due to the novel analytics services upspring, enhancing their C3 capabilities through service/resource cooperation and data sharing, providing cost reduction due to service/resource/infrastructure sharing. The main goal from the C3 PSC point of view is to provide timely and consistent information to public agencies in general. Moreover, it shall accomplish resource consumption reduction with a low-cost approach.

4.5.4.2.2 Technical enablers:

We see some boundary conditions in this scenario:

- Widespread usage of Software-Defined Networking technologies by operators and providers to support data security policy enforcement (e.g., between agency-owned core data centres and edge clouds) and traffic load balancing (e.g., between the end-users and the http proxies deployed at the edge clouds).
- Network-Function Virtualization technologies to support the service chaining providing C3 PSC analytics and incident management services (i.e., VNFs in both edge and core clouds).
- Automatic network and physical server monitoring by operators and providers to enable elasticity and efficient slice operation.
- Lightweight virtualization technologies (e.g., VLSP, unikernels or containers) for the edge clouds' virtual resources.
- Prediction mechanisms for the resource allocation and end-user requirements, to support the intelligence of the multi-domain orchestration.
- Intelligent multi-domain orchestration for the efficient slice manipulation over federated resources.
- Mobility handling aspects by operators and providers to allocate content proxies to end-users as they move around the area, while avoiding the service disruptions due to the involved handovers between different edge clouds.

Given the aforementioned support, we envision the following technical enablers:

- Dynamic slice creation and tear-down in any given MIM network slice.
- Services and capabilities definition and on-demand update supporting interest data retrieval.
- Support to slice merging, isolation, and modification, w/ on-demand capacity redefinition and function insertion/removal supporting slice consumer mobility and disruption.
- Slice monitoring and maintenance to accomplish service provision and resource management scalability supporting typical fluctuations on slice consumer population in a timely manner.

- Operation onto a multi-domain environment, to allow missions gathering several service providers (namely, public safety agencies), with previously agreed interfaces;
- Provision of an orchestration-enabled environment to deal with resource requests from different slice consumers.
- Infrastructure services provider support to typical device density in metropolitan areas to attend low speed vehicles and pedestrians
- High-level service and slice reliability, survivability, and availability, for different types of traffic to accomplish service customer requirements.
- Provision of a secure data exchange environment, enhanced by extreme levels of confidentiality, message authentication, operation registry and auditing, and detection of violation of authorized policies to deal with service provider security policy.
- Exposition of features like resource discovery, selection and allocation to support slice provision on a timely basis.

4.5.4.2.3 Critical success factors and KPIs

The Network Slicing for Metropolitan Intelligent Monitoring (MIM) scenario is assessed from the point of view of the involved stakeholders. In brief, the most critical success factor is situational awareness as well, which implies in relevant critical success factors as:

- F1 - Elastic slice operation providing a high degree of scalability (variable and dynamic number of users);
- F2 - Elastic slice operation providing short-time response for the deployment on the field;
- F3 - High QoE (low jitter, low packet error rate, high availability, redundancy mechanisms, and priority schemes implementation) for end-users;
- F4 - Network management (cross-layered approach for monitoring, analysis, and update) to endure long-time operations under different mobility patterns;
- F5 - Easy and low-cost network elements configuration;
- F6 - Security management based on proper data segregation and authentication patterns; and
- F7 - Intelligent multi-domain orchestration that supports prediction of service demands and resource utilization.

The above success factors may be derived in a set of KPIs, such as the following:

- K1 - Physical server utilization to quantify the resource-efficiency of the NECOS Slice as a Service capability in MIM scenario; Metric - physical server utilization.
- K2 - Content delivery performance to highlight the end-user satisfaction; Metric – end-user QoE.
- K3 - Service disruption to evaluate the efficient mobility handling; Metric – service disruption index;
- K4 – Timely and cost-effective slice deployment, conjugated with end-user QoE to demonstrate the efficient network performance; Metric – slice time deployment.
- K5 – Slice isolation to highlight data segregation and authenticity over a multi-domain environment; Metric – slice isolation index.
- K6 - Number of application users to demonstrate the scalability aspect; Metric – number of application users.
- K7 - Accuracy of prediction of service demands and resource utilization for the multi-domain orchestration; Metric – service demand prediction accuracy.

4.5.4.2.4 Mapping to NECOS key characteristics

The particular service needs for the scenario are the following:

- Secure content availability;
- Support for alleviating bulk data transfers of public safety data to first responders;
- Low latency content-delivery to first responder's teams by low-cost dynamic resource allocation according to end-user demands;
- E2E service chaining in the multi-domain cloud environment; and
- Bandwidth guarantees to support different connectivity levels required by teams deployed on the field.

Thus, NECOS service provisioning approach (Objective 2) is suitable for MIM scenario, since it foresees the integration of resources within the collection of independent slices. A goal of the LSDC approach is to reduce the complexity and timescale for service provisioning and deployment in federated DCs, thus reducing the OPEX for the infrastructure owner.

Moreover, NECOS also envisions service and resource orchestration and management methods for the LSDC infrastructure resources that are located within and at the edge of the MIM network. This service orchestration and management approach, which comprises NECOS Objective 3, includes the automatic re-allocation of resources and services across distributed and geographically separated computing, data storage, and network infrastructures within separate slices to support requests for MIM application data.

Finally, NECOS objectives 1 (develop and build a Lightweight Slice Defined Cloud (LSDC) platform enabling computing, network, and storage elements in the cloud through the Slice as a Service across federated clouds) and 4 (a pilot-based impact demonstration) will be achieved as well to demonstrate the provision of location-aware public safety content delivery to distinct public agencies teams.

In order to accomplish the aforementioned objectives, we consider the following key NECOS characteristics for the scenario:

- C1: Slice as a Service, mapping enhancing component (re) configuration to support E2E MIM applications;
- C2: Lightweight Edge Cloud, with small footprint components, deployable on both core and edge small servers and cloud systems, to support a user environment characterized by mobility and disruption
- C3: Software Management under a per-slice approach, to support features like user demand elasticity (i.e., Load Balancing through multiple VMs) and monitoring.

4.5.4.2.5 Functional and Non-Functional Requirements

Based on C3 PSC scenario, this section presents the main requirements for both smart cities data and metropolitan intelligent monitoring domains. These requirements are listed under functional and non-functional scopes:

The functional requirements are:

- RF.emergency.1 – *Dynamic slice management*: support to dynamic slice creation and tear-down in any given PSC network slice.
- RF.emergency.2 – *Dynamic service definition*: support to services and capabilities definition and on-demand update.
- RF.emergency.3 – *Timely slice management*: support to slice monitoring and maintenance to accomplish service provision and resource management scalability.
- RF.emergency.4 – *Orchestration*: provision of orchestration capabilities to deal with resource requests from different users, under a multi-domain environment, to allow missions gathering several public safety agencies, with previously agreed interfaces.

On the other hand, the non-functional requirements considered are:

- RN.emergency.1 – *High Reliability:* provision of high-level service reliability (> xx %)

- RN.emergency.2 – *High Availability:* provision of high-level service availability (> xx %)
- RN.emergency.3 – *High Survivability:* provision of high-level service survivability (> xx %) for the scenario described in requirements RN.emergency.1 and RN.emergency.2

5 Digest: Requirements, Critical Success Factors and Key Performance Indicators

This section provides a digest of all requirements, critical success factors and key performance indicators relevant to the NECOS project. All of them will serve as input for forthcoming architectural work as well as validation in future steps of the project.

5.1 Requirements

The following notation has been followed in order to collect all the requirements listed previously during the scenarios description.

Notation Summary:

R<1>.<word>.<#>

where:

1: F for functional and N for non-functional

Word: a unique word associated to that particular scenario

#: a sequence number

5G Infrastructure (vRAN) Scenario, 3 Functional Requirements and 3 Non-Functional Requirements	
RF.vRAN.1	*Service Level Agreement*
RF.vRAN.2	*Accountability*
RF.vRAN.3	*On-demand slice provisioning*
RN.vRAN.1	*Isolation of slice resources*
RN.vRAN.2	*Fairness*
RN.vRAN.3	*Fault detection*
5G Services scenario, 4 Functional Requirements and 3 Non-Functional Requirements	
RF.5G.1	*Service Level Agreement*
RF.5G.2	*Accountability*
RF.5G.3	*On-demand slice provisioning*
RF.5G.4	*External control and management of the offered slices*
RN.5G.1	*Isolation of slice resources*
RN.5G.2	*Fairness*
RN.5G.3	*Fault detection*
vCPE Scenario, 8 Functional Requirements and 5 Non-Functional Requirements	
RF.vCPE.1	*On-demand slice provisioning*

EUB-01-2017

RF.vCPE.2	*Manageable slice*
RF.vCPE.3	*VIM-independence*
RF.vCPE.4	*Bare-metal slice*
RF.vCPE.5	*Lightweight virtualization*
RF.vCPE.6	*Elasticity*
RF.vCPE.7	*Zero touch service provisioning*
RF.vCPE.8	*Fault detection*
RN.vCPE.1	*Isolation of slice resources*
RN.vCPE.2	*SLA monitoring (QoS)*
RN.vCPE.3	*Low latency*
RN.vCPE.4	*High throughput*
RN.vCPE.5	*High availability*
Content Delivery Touristic Services Scenario, 5 Functional Requirements and 5 Non-Functional Requirements	
RF.Touristic(CD).1	*Slice and slice-resource management*
RF.Touristic(CD).2	*Automated Virtual Machine deployment*
RF.Touristic(CD).3	*Traffic load-balancing for content delivery*
RF.Touristic(CD).4	*Slice resource and service monitoring*
RF.Touristic(CD).5	*Service planning*
RN.Touristic(CD).1	*Transparent end-user performance*
RN.Touristic(CD).2	*Heterogeneity handling*
RN.Touristic(CD).3	*Elasticity*
RN.Touristic(CD).4	*Resource-efficiency*
RN.Touristic(CD).5	*Scalability*
Applications Touristic Services Scenario, 3 Functional Requirements and 4 Non-Functional Requirements	
RF.Touristic(APP).1	*Service function chain orchestration*
RF.Touristic(APP).2	*Resource and user-demand prediction capabilities*
RF.Touristic(APP).3	*Resource offloading between edge, core clouds and cloud providers*

RN.Touristic(APP).1	*Resource federation and intelligent multi-domain orchestration*
RN.Touristic(APP).2	*Scalability*
RN.Touristic(APP).3	*Efficient next-generation touristic application performance*
RN.Touristic(APP).4	*Elasticity*
Emergency Scenario, 4 Functional Requirements and 3 Non-Functional Requirements.	
RF.emergency.1	*Dynamic slice management*
RF.emergency.2	*Dynamic service definition*
RF.emergency.3	*Timely slice management*
RF.emergency.4	*Orchestration*
RN.emergency.1	*High Reliability*
RN.emergency.2	*High Availability*
RN.emergency.3	*High Survivability*

5.2 Critical success factors and KPIs

CSFs and KPIs are closely related. What is a feature that determines the success (understood as a success among its users) of a solution is what the users would what to assess in order to keep using a service, and what the provider want to monitor to stay being a "success".

5.2.1 NECOS' CSFs

The following CSFs have been identified so far:

Isolation

Isolation is the factor that distinguish slicing from other cloud-based solutions. Since the slices are isolated from each other in all network, computing, and storage planes, the user experience of the slice will be the same as if it was a physically separate infrastructure.

Cost Reducing

Any virtualization solution, being sliced or not, must benefit from its economy of scale, a multi-domain solution such as NECOS has to take advantage of the benefits of the market economy reaching scale through multiple providers, potentially specialized in their own virtualization domain (i.e., virtual networks, virtual computing, etc.) and therefore reducing its own costs and at the end, the costs transferred to the users.

Reliability

The redundancy, geographic distribution, and technology diversity provided by a service that orchestrates different connectivity and computing services from different providers is a key factor for its reliability. As long as the orchestrator has the ability to quickly re-provision a slice after a failure is detected the reliability of the overall slice service may be higher than the reliability offered by each provider.

Flexibility

Flexibility of cloud-based services should be one of its main distinguishing features. Much more on a multi-domain, diverse platform, as far as it works efficiently and transparently for the user. It is also closely related with reducing the costs for the user, as the resources used by the slice, and therefore its cost, can be reduced or augmented following the demands of the service deployed on top of it, avoiding overprovisioning. Additionally, the tenant must consider the flexibility of a lightweight solution such as NECOS, which intends to offer a basic service inside of which the tenant deploys its own services following its own policies and requirements.

Scalability

A particular case of such flexibility is the scalability of a provisioned slice. There are several definitions of scalability and elasticity, the CSF described below. For the scope of this project we define scalability as the ability to increase workload size within existing infrastructure (hardware, software, etc.) without impacting performance. We can think in scalability in two different dimensions: scalability of a particular provisioned slice and scalability of the number and size of the slices provided.

Elasticity

As with scalability, there are more than one definition for elasticity. In the context of this project, elasticity is the ability to grow or shrink slice resources dynamically as needed to adapt to workload changes in an autonomic manner, maximizing the use of resources.

Security

Being a standard critical feature for any system, the security must be of particular attention for a solution based on sharing resources, isolation, and avoiding side channels must be a priority.

Slice Efficiency

The critical factors mentioned above are no relevant if the communications and computing services provided by a slicing solution do not reach the performance expected and paid by the tenants/users.

Life-cycle Efficiency

Additionally, a slicing service must be efficient regarding the own slice life-cycle: provisioning, monitoring, recovering from failures, etc. This efficiency is closely related with most of the factors mentioned above, from it depend factors such as cost, flexibility, and reliability.

Simplicity

Simplicity is obviously a factor of importance from the point of view of the client but also an aspect with impact on other CSFs such as *cost efficiency*, but also on *reliability* and *security*. Understanding clearly what is being configured and provisioned is the first step to implement a service inexpensive, reliable and secure.

5.2.2 NECOS' KPIs

The Network, Network Cloud, and Cloud slicing model designed by NECOS is in line with the vision of an open ecosystem. The objective is to be able to "mix & match" everything from bare metal to an architected Cloud consisting of computing and storage virtualization resources and virtualized networks from different vendors. Integrating multiple VIM and WIM software stacks and hardware from different vendors need to be validated by a set of Key Performance Indicators (KPI) in order to have tools to assess the SLAs made between the slice tenant and the slice provider.

Specifically, KPIs are required to address these challenges in order to achieve operational capabilities are close to those coming from the multi-domain NFV solutions [14]:

- Interoperability and portability between different network appliance vendors, hardware vendors, and with different hypervisors.
- Achieving the target performances.
- Elasticity based on automation.
- Ensuring the appropriate level of resilience to hardware and software failures.

EUB-01-2017

- Ensuring security against attack and misconfiguration.
- Operation & Manageability of the end-to-end solution.

We are going to refer here to overall, end to end KPIs. In a multi-domain environment such as the one envisioned by NECOS, to monitor domain per domain KPIs is mandatory for the NECOS solution, but from the point of view of the user may not be useful nor a good indicator of the quality of the service.

The other interesting aspect of the metrics that are relevant for a service such as the NECOS solution is that they are not related with the services being deployed inside a slice but about the slice itself. For example, NECOS KPIs are not about the response time of a web-service but about the time needed to create a VM with the delegated VIM. However, relevant KPIs include also performance some indicators inside the slice that are requirements of the services for which the slice is provisioned, such as end-to-end delay or throughput.

The end-to-end aspects to assess by a multi-domain solution such as NECOS could be:

Functional Assessment

- Functional Acceptance
  - Basic capability (CPU, NIC, COTS)
  - Computing capability, network service capability, storage capability, and security management capability.
  - Manage virtual machines.
  - VIM basic capability
  - WIM basic capability
  - Interface interoperability capability for components
- Solution Test
  - Management, operation and maintenance.
  - Network Readiness for service integration.
- Performance Assessment
- Network Performances
  - User plane Bandwidth
  - User plane Throughput
  - User plane RTT
  - User plane Packet Loss
  - User plane Jitter
- WIM Stack Performance
  - VNF / CloudOS deployment management lifecycle duration.
  - COTS layer: CPU, memory, network card performance indicators.
- VIM Stack Performance

Reliability Assessment

- Network Reliability
  - VN-NF interface reliability, Ve-vnfm interface reliability, Vivnfm interface reliability, Vl-Ha interface reliability and other components Interface reliability
- Component Reliability
  - VM reliability, CloudOS reliability, and server reliability
- Virtualization Reliability
  - Virtual machine isolation,
  - virtual machine anti-affinity function,
  - virtual machine watchdog function
- Data reliability
  - Orchestrator data reliability,
  - compute node data reliability,
  - VNFM data reliability.

Security Assessment

Some security KPIs are related to general management:

- Number of Implemented Preventive Measures
- Number of Major Security Incidents
- Number of Security-Related Service Downtimes
- Number of Security Tests
- Number of Identified Shortcomings during Security Tests

Others are layer specific:

- Network Security
    - Physical security
    - Network slice isolation
    - Network Segmentation
    - Network Access Control
- Component Security
    - DoS vulnerabilities,
    - Client-end vulnerabilities
    - Server-end vulnerabilities
    - Cloud-end vulnerabilities
    - Component Access Control
- Virtualization Security
    - Virtual machine isolation,
    - virtual machine anti-affinity function,
    - Privileged insiders.
- Data Security
    - Orchestrator data security.

Those above are an overview of most of the possible KPIs for a system such as NECOS, however, for the scope of the project, we have selected a subset that was presented previously for each scenario. Bellow there is a summary of those KPIs.

Code	KPI Name
KPI1	Average elasticity response time (in seconds).
KPI2	Average end-to-end delay (in milliseconds, measured as half average RTT);
KPI3	Average service provisioning time (in seconds);
KPI4	Average slice provisioning time (in seconds);
KPI5	Average throughput (in Mbps).
KPI6	End-to-end slice availability (% of time);
KPI7	Monitoring-data availability (%)
KPI8	Number of application users (TBD).
KPI9	Physical Server utilization (TBD)
KPI10	Service demand prediction accuracy (TBD)
KPI11	Service disruption index (TBD)
KPI12	Service QoE (TBD)
KPI13	SLA fulfilment index (TBD)
KPI14	Average Slice Decommission time (in seconds).
KPI15	Slice isolation index (TBD)
KPI16	Average Slice Provisioning time (in seconds).

5.3 KPIs and Scenarios

In this section, we instantiate the general KPIs described above for each scenario and its relation with the requirements enumerated in Section 5.1. The NECOS project will work to create the means needed to monitor these KPIs, leaving out of its scope all other metrics.

Req. Name	Related KPIs
RF.5G.1	KPI13
RN.5G.1	KPI15
RF.5G.2	KPI7
RN.5G.2	KPI15
RF.5G.3	No KPI
RN.5G.3	KPI11
RF.5G.4	No KPI
RF.emergency.1	No KPI
RN.emergency.1	KPI11
RF.emergency.2	No KPI
RN.emergency.2	KPI6
RF.emergency.3	KPI7
RN.emergency.3	KPI11, KPI6
RF.emergency.4	No KPI
RF.Touristic(APP).1	No KPI
RN.Touristic(APP).1	No KPI
RF.Touristic(APP).2	No KPI
RN.Touristic(APP).2	KPI1
RF.Touristic(APP).3	No KPI
RN.Touristic(APP).3	No KPI
RN.Touristic(APP).4	KPI1
RF.Touristic(CD).1	No KPI
RN.Touristic(CD).1	KPI6
RF.Touristic(CD).2	No KPI
RN.Touristic(CD).2	No KPI
RF.Touristic(CD).3	No KPI
RN.Touristic(CD).3	KPI1
RF.Touristic(CD).4	KPI7
RN.Touristic(CD).4	KPI9
RF.Touristic(CD).5	No KPI
RN.Touristic(CD).5	KPI1
RF.vCPE.1	No KPI
RN.vCPE.1	KPI15
RF.vCPE.2	No KPI
RN.vCPE.2	KPI13
RF.vCPE.3	No KPI
RN.vCPE.3	KPI2
RF.vCPE.4	No KPI
RN.vCPE.4	KPI5
RF.vCPE.5	KPI1
RN.vCPE.5	KPI6
RF.vCPE.6	KPI1
RF.vCPE.7	No KPI
RF.vCPE.8	No KPI
RF.vRAN.1	KPI13

RN.vRAN.1	KPI15
RF.vRAN.2	No KPI
RN.vRAN.2	KPI15
RF.vRAN.3	No KPI
RN.vRAN.3	KPI11, KPI6

6 Summary and outlook

This deliverable presents the description of the two use cases considered as a starting point of the NECOS project, namely the Telco Cloud and the Multi-Access Edge Computing (MEC). These use cases act as platforms on top of where a number of relevant scenarios are presumed to be supported, controlled and managed through the NECOS platform, leveraging on the concept of slicing as a form of segregating multiple services in the same (federated) cloud networking substrate. It presents also the descriptions of relevant service scenarios as target to be supported by NECOS platform. These scenarios assisted the consortium to identify the functional and non-functional requirements that NECOS platform should satisfy. The priority and relevance of those requirements will be further elaborated by WP3. Last but not least, this deliverable has presented an analysis of the ecosystem enabled by NECOS as a first step to study the viability of business models associated to it.

The next WP2 deliverable D2.2 will refine the information contained here and provide a consolidated version of the service scenarios and related requirements, including their technical and business analysis.

7 References

[1] 3GPP TR 28.801, "Study on management and orchestration of network slicing for next generation network", v15.1.0, January 2018.

[2] NGMN, "5G white paper", February 2015, available at: https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf

[3] TMForum report IG1152, "Dynamic Network Slices Management and Business Models", r17.0.1, September 2017.

[4] ETSI white paper, "Mobile-Edge Computing", September 2014, available at: https://portal.etsi.org/portals/0/tbpages/mec/docs/mobile-edge_computing_-_introductory_technical_white_paper_v1%2018-09-14.pdf

[5] L. Peterson, "CORD: Central Office Re-Architected as a Datacenter", IEEE Softwarization, November 2015, https://sdn.ieee.org/newsletter/november-2015/cord-central-office-re-architected-as-a-datacenter

[6] Broadband Forum, "Cloud Central Office Reference Architectural Framework", TR-384, January 2018.

[7] ITU-R Recommendation M.2083-0, "IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond", September 2015, available at: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-I!!PDF-E.pdf

[8] 3GPP TS 22.261, "Service requirements for next generation new services and markets", Release 15, March 2017.

[9] 5G-PPP white papers on vertical industries, available at https://5g-ppp.eu/white-papers/

[10] L.M. Contreras, D. López, "A Network Service Provider Perspective on Network Slicing", IEEE Softwarization, January 2018, available at: https://sdn.ieee.org/newsletter/january-2018/a-network-service-provider-perspective-on-network-slicing

[11] L. Suchman, "Centers of Coordination: A Case and Some Themes", in L.B. Resnick, R. Säljö, C. Pontecorvo, B. Burge (eds), *Discourse, Tools and Reasoning*, NATO ASI Series (Series F: Computer and Systems Sciences), vol. 160, Springer, 1997.

[12] https://esa.un.org/Unpd/Wup/

[13] N.B. Sarter, D.D. Woods, "Situation Awareness: A Critical but Ill-Defined Phenomenon", International Journal of Aviation Psychology, Vol. 1, Num. 1, pp. 45-57, 1991.

[14] Huawei white paper, "3 Dimensional KPI Assessments for High Efficient NFV Multi-Vendor Integration", 2017, available at: http://carrier.huawei.com/~/media/CNBG/Downloads/Services/nfv/3-Dimensional-KPI-Assessments-for-High-Efficient-NFV-Multi-Vendor-Integration.pdf

Version History

Version	Date	Author	Change record
0.1	04/04/2018	L.M. Contreras	Creation
0.2	20/04/2018	L.M.Contreras, R.A. López	First integrated draft
1.5.2	24/05/2018	L.M.Contreras, R.A. López	Final version release