Cloud Native Enabling Future Telco Platforms

v5.2

Cloud Native Enabling Future Telco Platforms

by NGMN Alliance

Version:	**5.2**
Date:	**17-May-2021**
Document Type:	**Final Deliverable (approved)**
Confidentiality Class:	**P-Public**

Project:	**Future Networks Cloud Native Platform**
Editor / Submitter:	**Fabrizio Moggio (Telecom Italia)**
Contributors:	**Javan Erfanian (Bell Canada), Richard MacKenzie (BT), Tim Costello (BT), Zhiqiang Yu (China Mobile), Qihui Zhao (China Mobile), Peter Weichsel (Deloitte), Nigel Young (Deloitte), Stefan Saliba (Deloitte), Shashi Bhagavathula (Deloitte), Andreas Krichel (HPE), Andreas Volk (HPE), Martin Halstead (HPE), Rodion Naurzalin (HPE), Marie-Paule Odini (HPE), Said Tatesh (Huawei), Gary Li (Intel), Sebastian Robitzsch (InterDigital), Kay Haensge (InterDigital), James Low (Keysight), Herve Oudin (Keysight), Julien Maisonneuve (Nokia), Mohamad Yassin (Orange), Fabrizio Moggio (Telecom Italia).**
Approved by / Date:	**NGMN Board, 11th March 2021**

Address:
ngmn e. V.
Großer Hasenpfad 30 • 60598 Frankfurt • Germany
Phone +49 69/9 07 49 98-0 • Fax +49 69/9 07 49 98-41

Abstract: Short introduction and purpose of document

This document analyses the transformation toward Cloud Native in the main network domains of the Telco infrastructure. The goal is to define a holistic view on the Cloud Native transformation framed in an Edge Hybrid Cloud scenario. This transformation is mainly perceived by the Telcos as an important driver for internal optimization, cost savings and to speed up vertical solution. From a broader perspective, it is also an enabler, for the Telcos, to join a wider ecosystem where Telcos, technology suppliers, developers and Hyperscale Cloud Providers work together embracing new business opportunities leveraging on their own specific and unique assets.

Contents

1 INTRODUCTION

The evolution of 5G Networks leverages on technological and modelling approaches such as cloudification, service-based architecture, network capabilities exposure, network as a service, zero touch management, just to report some of the main ones. This evolution is introducing a big technological challenge, giving hype to a paradigm change with new opportunities for a Telco. Being actively involved in a wider ecosystem of application service offering is, more than ever, an expanded role and opportunity for Telcos, to create and deliver value, along with partners. It is important, indeed, to view 5G Networks not as predesigned and static Telco infrastructure but as a flexible and Open Telco Platform.

This document analyses the ongoing transformation in the network domains of the Telco infrastructure with the goal to define a holistic view supporting the Edge Hybrid Cloud scenario. In chapter 2, an overall view of the Cloud Native adoption impacts on the Telco Platform, in term of openness and new opportunities, is summarised. In the following chapters a deepening on the most relevant areas is provided. As a basis, an analysis on the standardisation and Open Source activities is provided in chapter 3, to ensure a full support on the proposed model. Cloudification is then analysed considering the technological evolution at infrastructure level in chapter 4, and at architectural level in chapter 5. New business opportunities for a Telco, derived by this innovation, are embodied in the Hybrid Cloud exploitation as addressed in chapter 6. A distinctive Telco asset is the Radio Access Network, the cloudification in this domain is specifically deepened in chapter 7. The economical aspects and drivers for this evolution are considered in chapter 8, where new market and collaboration possibilities are deepened. Challenges and critical success factors are identified and discussed in chapter 9.

This transformation is perceived by the Telcos as an important driver for agile and efficient architecture and operation, to create value and speed up vertical solutions. From a broader perspective, it is also an enabler, for the Telcos, to join a wider ecosystem where Telcos, the technology suppliers, developers and Hyperscale Cloud Providers work together. The common goal is to embrace new business opportunities leveraging their own specific and unique strength and assets.

This document outlines how network cloudification is one of the main enablers for an Open Telco Platform. A platform that is founded on internal efficiency, cost and energy savings, fast service delivery and leverages on the cloudification process and on API exposure. It is open, supporting a new service model, where the applications are deployed at the Edge of the Telco network. Applications can be delivered to the end user over the Telco network with appropriate performance and guaranteed service level agreement over a managed Communication Services. The Network evolution toward Cloud Native is the enabling factor for internal efficiency and platform openness. It is the enabler for integration in an ecosystem where the Telco can exploit its assets e.g. providing tailored Communication Services. Because of this evolution, service differentiation and tailoring to the customer requirements is nowadays a concrete possibility that gives to the Telco a distinctive and unique role in the value chain. Also, OTT services can benefit of this new network capabilities that only a Telco can provide.

This work links together the transformation that is happening in the different Telco areas and domains, starting from the state of the art, analyzing the relevant work of the main international bodies, SDOs and Open Source communities, focusing on the activities relevant for this document scope. It is easy to understand that each international body has a specific scope and it focuses on just a part of the holistic view this document is addressing. For this reason, there are natural gaps in the international bodies work in term of covering the overall picture. Despite the specific gaps, this document shows how the whole picture is anyway supported by the overall work of the international communities. The outcome, depicted in this document, is an end-to-end vision on the Telco approach to the evolution of the network, supported by standardization and Open Source activities. The idea is to promote a new opportunity and a new model foreseeing a programmable Open Telco Platform as the basis to foster Telco distinctive assets such as tailored and assured end to end network connectivity or Edge data networks. The Telco assets are well integrated with the wide possibilities and market given by the Cloud ecosystem.

The main concept is that, embracing this opportunity, the Telco can leverage on the existing activities it is facing to update its network toward 5G and can expand its services playing a distinctive role. The Telco network surely has unique assets such as geographical distribution and unique APIs, e.g. to accelerate the network performance of a

specific service. This provides to the Telco the opportunity to play an important role in the Hybrid Cloud scenario complementing the current offering of the Hyperscale Cloud Providers (HCP).

The Edge Hybrid Cloud scenario foresees an open ecosystem where developers can distribute their applications both in the Central Cloud and in the Edge Cloud. This is done easily, choosing the solution and network distribution that best fits the application requirements. The Central Cloud and the Edge Cloud are part of an integrated environment playing a complementary role. The modern and programmable Open Telco Platform is the enabler for such a scenario.

The document focuses on specific technical areas that are considered relevant enablers for an Open Telco Platform. As a main pillar, the overall evolution toward Cloud Native is considered fundamental, with different needs and level of evolution in different network domains. To be part of a wider ecosystem, the Telco Platform must be open and programmable, so the exposure of the relevant Telco APIs is a key factor. Different approaches for the Edge Hybrid Cloud integration are identified according to different models of interaction among the stakeholders (e.g. Telcos and HCPs). Intelligence in the network plays an important role when it comes to support many Communication Services that have specific network requirements and that leverage on specific Telco assets. The stakeholders invest paying for tailored Telco services expecting a guaranteed service level that must be constantly assured in real-time. The scenario foresees different challenges in terms of technical evolution, multi-party relationships and strategic decisions, for this reason possible critical success factors are identified and discussed and the role of the Telco in the open ecosystem is explored. This transformation and the new upcoming opportunities in the selected scenario have economic impacts, this document presents a helicopter view on those aspects.

2 THE OPEN TELCO PLATFORM

The full realization of 5G to enable digital socio-economic transformation involves countless and wide ranges of use cases, many unimagined today, with specific requirements. 5G enablement has thus introduced a big technological gap, demanding a network transformation. A paradigm change is needed in the overall network design, in its operation and in the provided services. This journey is well underway.

This vast heterogeneity, with a wide range of requirements, demands agile networks. Full flexibility, scalability and efficiency are key attributes of networks capable of providing dynamic, dedicated, secure, and reliable services offered as a platform. The network shall provide shared or dedicated resources for variety of use case scenarios and innovative vertical markets. It is expected to become increasingly dynamic and fully customized, in response to the needs of each service, during its lifecycle. This requires new operational models and modern infrastructures, but it will also be the basis for new business opportunities.

To enable the new business context and the service demands, 5G architecture is expected to be Cloud Native. It is based, by design, on an intelligent and dynamic multi-access core. It is designed over a service-based control-plane architecture, separate from a flexible user plane, and with the exposure of functions through open interfaces. The cloudification, virtualization / containerization of Telco networks is evolving and requires E2E orchestration, increasing automation and enhanced lifecycle management.

Enhanced by this evolution the Telco Platform is not only evolving to simplify operation. It is becoming closer and closer, in terms of technology and automation, to the cloud platforms the developers are currently using to deploy their applications. The enhancement of the 5G network in terms of performance and deployment flexibility allows the Telcos to surely support new vertical applications. The technological evolution underneath also supports new models where those services can be more and more integrated in the Telco Platform. New services can indeed leverage on unique Telco assets from one side and on common IT/Telco technologies from the other side.

Hybrid Cloud and Edge Computing create significant opportunities within a wide range of business models. The mobile network operators, with Open Telco Platform, have expanded roles in these models, creating and delivering value with their partners. The Hybrid Cloud architecture, especially at the Edge, can be logically composed by two interworking Cloud Native environments one Telco oriented and one service oriented. Different possible technological and partnership models can be foreseen in this context.

Different aspects of this evolution are already well defined while others are currently under discussion in different bodies within the ecosystem. There is increasing synergy and joint development within these bodies; however, much is yet to be concretely realized and tested in real production environment. There is a need to leverage and further harmonize, the variety of developments defined within different bodies also to simplify the work of integration Telcos need to do. The target is a well-defined Cloud Native open and interoperable platform. It shall be increasingly cognitive, programmable and autonomous, in order to design, orchestrate and dynamically manage the delivery of value for each use case and user scenario.

A fundamental aspect of 5G is the distributed, interoperable and multi-vendor capability and service delivery, through disaggregation and open interfaces. These aspects well match a Cloud Native approach. RAN disaggregation, with interoperable interfaces, enables much flexibility, scalability, efficiency, and choice within a broad and expanded ecosystem. Distributed intelligence in a Cloud Native 5G system is a key factor leveraging Hybrid Cloud and enabling multi-access edge computing and local context analysis.

The road to open, flexible and agile networks, is expected to continue reducing **environmental** footprint and total cost of ownership, as well as time to market. It is imperative that this journey maintains a clear focus on social responsibility, sustainability and significant energy efficiency.

A Telco Platform must exploit openness both internally and toward the external ecosystems. The basic characteristics for such an openness are the same whenever you are considering internal optimization or external federation and interoperability. The traditional mantra supporting open and standard interfaces is more than ever sided by the adoption of standard de facto Open Source software especially for the management layer. The new 3GPP 5G mobile network indeed foresees a microservices based architecture that aligns the different domains and vendors' solutions. Meanwhile Open Source communities are delivering IT solutions supporting this evolution on top of the experience made on cloud architecture management.

From physical to Cloud Native Network Functions the current evolution path is fostering Cloud Native concepts deeply in the Telco world, from the central data centres to the regional and edge ones. This enriches the Telco's value chain including also distinctive assets such as the edge infrastructure. This cloudified and open Telco Edge becomes a unique selling point in the Telco proposition for the ecosystem towards developers and HCP.

A key enabler for tailored evolving networks supporting both NFs and applications is Cloud Native orchestration. It has the capability to support standardised deployment and operational procedures across various cloud data centres leveraging open multi-vendor physical infrastructure. This disaggregated model allows an independent deployment paradigm without having dependencies on hardware and applications typical of a legacy, single vendor solution.

Intelligence is a key factor for both Network Functions and application life cycle management and service assurance. Artificial intelligence capabilities are urgently needed to enable flexible network automation and network augmentation and support application deployment with guaranteed QoS especially at the Edge.

Hybrid is a key word in the cloudification process of the Telco Platform and it covers different aspects. One aspect is the coexistence of VM based and Containers based NFs deployments. Another aspect is the coexistence of Telco oriented and Service oriented Cloud Native integrated environments. Centralized, Edge and Cloud based deployments coexistence and integration is another aspect that fits in the hybrid scenario.

The interoperability of hybrid cloud solutions is of paramount importance with the need to intermix different Cloud solutions. Certainly, this playing field has less stringent standardization specifications compared to 3GPP Core Network or RAN. For this reason, it poses a very steep burden on the cloud providers to find the right balance of openness and interoperability and key differentiators. In the end, how to mix-up Hybrid Cloud is more a business decision than a technological one.

3 STATE OF THE ART AND GAP ANALYSIS

There are many activities around the Telco Ecosystem evolution, some are carried out by standardization bodies such as ETSI or 3GPP, others are forged around the Open Source communities such as ONAP or CNTT. These two

approaches to create and promote innovation, once very far from each other, are currently, more and more, leveraging on one another.

Openness and network cloudification can be exploited also to create a widely adoptable Open Telco Platform. On top of this platform new Services can be efficiently provided in a Hybrid Cloud approach. This concept is coherent with the already ongoing activities in the Telco communities whose work is the concrete enabler of the proposed model.

According to this scenario, the most relevant activities of the main international bodies are briefly analyzed and summarized in this document. The goal is to identify the level of support and validate the scenario of interest also identifying possible gaps each entity has with respect to the overall picture.

This chapter is structured as a journey through the different logical layers of the platform considering the relevant supporting activities by the international bodies.

The evolution of the network toward being an Open Telco Platform is driven by different needs. The Telco Platform should be open to new Communication Services, designed and deployed according to the **customer's requirements**. Network Slicing is a key feature to support tailored Communication Service and **GSMA NEST** (Network Slicing Task Force) issued a specification to standardise how requirements can be formalised to describe a Network Slice [2]. The purpose of the document is to provide the standardised list of attributes that can characterise a type of network slice. It proposes Generic Slice Template (SGT) as standard tool to contain slice attributes. GSMA also provides a set of examples, Network Slice Types (NESTs), that, starting from the GST, have a recommended minimum set of attributes and their suitable values. GSMA GST and inherited NEST templates do not specify attributes related to the infrastructure requirements, or Cloud Native requirements. This is not in the NEST scope neither is foreseen for a Customer to care about infrastructure topics. Anyway, for an overall view of the slice deployment, infrastructure aspects must be identified by the underlining network domains.

Those customer's requirements are subsequently evaluated by a **Service Layer**. A widely recognized reference for the layer is Telemanagement Forum (**TM Forum**). On Cloud Native, TM Forum has issued a specific white paper to explain the Communication Service Providers evolution to the Cloud. It covers aspects such as migrating applications, offering APIs, transforming the operating models and introducing Artificial Intelligence in Operation. TM Forum defines more than fifty open REST APIs ranging from infrastructure level resource management to BSS level catalog and order management. The work from TM Forum sometimes overlaps with some other standards, like 3GPP SA5, or GSMA, or some Open Source projects like ONAP or Akraino. It is up to the Telco, also according to products support and integration aspects, to choose the most appropriate standard to adopt.

For the definition of mobile network infrastructure, one of the most relevant standardization group is the 3rd Generation Partnership Project (3GPP). The 5G architecture, defined by 3GPP, both for Network Functions and for the management system is widely based on the Service Based Architecture (SBA). 5G Systems were indeed introduced in 3GPP Rel. 15 by **3GPP SA WG2 (SA2)**, the working group well known for the definition of the 3GPP network, e.g. for architectural aspects of the mobile **Core Network**. The introduction of the SBA leads to the possibility of developing Cloud Native-based systems in the Telco world. It provides functional modularity and complete separation of user plane and control plane, with core functions communicating over (preferably) RESTful APIs. The introduction of Network Slicing is another important step for the definition of an open and sharable infrastructure with native means to guarantee a measurable Service Level Agreement for the Customers. The 5G 3GPP SBA uses Cloud Native principles and modularized Network Service Function (SFs) design that allows independent scalability and evolution of the Network Service Functions that constitute the 5G System. This definition perfectly matches with the evolution of software development based on containers and microservices. These principles enable network operators and system implementers to deploy their networks using Network Function Virtualization techniques and Software Defined Networking, both well proven technologies highly successful in Information Technology Systems. One of the key factors of an Open Telco Platform is the ability to provides services to external consumers. This perfectly matches with the 3GPP defined Network Exposure Function (NEF) that provides exposure of capabilities and events. There are natural gaps in how SA2 defines the 5G System. It is not indeed in the scope of 3GPP the specification on how to implement the services. The shift from big and monolithic Network Functions towards Cloud Native services requires a further elaboration and more details over the 3GPP architecture.

Radio Access Network (RAN) is well represented in the work done by 3GPP RAN groups and it is indeed also well represented in the **O-RAN ALLIANCE**. O-RAN ALLIANCE build on 3GPP specifications to provide profiling specifications and requirements for disaggregation, virtualization, open and intelligent RAN. Virtualization decouples software and hardware RAN functionalities, enabling the RAN to be built on a general-purpose processor platform to reduce manufacturing costs. The Telco industry and related partners have defined a disaggregated RAN architecture where the baseband is split into Distributed Unit (DU) and Centralised Unit (CU). Various deployment options allowing the DU and CU to be distributed flexibly based on Telco assets and use cases are supported. The O-RAN ALLIANCE uses 3GPP architecture that splits CU according to control plane and user plane (In O-RAN these splits are called O-CU-CP and O-CU-UP). It also goes further decomposing DU into O-DU and O-RU (Remote Unit), as the very edge of the RAN. This is done to have all the hardware-based functions in the O-RU and the components that can be virtualised in the O-DU. As a gap, this split is not foreseen by 3GPP, for this reason the O-RAN ALLIANCE specified a new fronthaul interface between O-DU and O-RU.

Other important enablers for the **mobile network programmability and management** are defined by **3GPP SA WG5 (SA5)** that specifies the requirements, architecture and solutions for provisioning and management of the network (RAN, CN, IMS) and its services. To provide a modern and flexible solution, SA5 structures the 5G Management System around the concept of Service Based Architecture (SBA). This opens it to be customized according to the Telco needs. The system is based on Management Services (MnSs) providing O&M features. The MnSs are not only provided by management functions, the NFs themselves can be providers of MnSs. This approach moves the standardization of the management APIs deep down to NFs with them providing both standard functional services (defined by SA2) and standard management services (defined by SA5). To integrate the 5G management system into a wider and orchestrated system, the exposure of the management services is required. For this reason, SA5 foresees exposure governance of the MnSs with the possibility to filter, adapt and limit the exposure itself. Considering that the APIs of the Open Telco Platform should be easy to use to be adopted in the Developer's ecosystem, it is interesting to consider the work from SA5 in intent driven management. The concept studied by SA5 foresees that an API shall expose capabilities in an intent based mode, addressing requirement rather than specific configuration parameters. This approach also helps in a multivendor environment. Service assurance is a key factor for an Open Telco Platform to support tailored service. Service assurance, as Network assurance, in 5G, is based on automated control loops. The management system must guarantee the SLAs of different Communication Services leveraging of the same shared network resources. To accomplish such a complex task SA5 foresee the Assurance Service being aided by an analytic service named Management Data Analytic Service (MDAS) that collects an analyse management data such as fault, alarms and performance data. Considering a modern 5G Telco Platform, supporting different services and use cases, it is important to leverage on automation and orchestration based on Self-Organising Network (SON) principles and algorithms. SA5 defines how to manage a network to support a Communication Service, the definition of such a service in terms of requirements and the request of the service itself needs to be evaluated also considering the work done by TMF for the service layer APIs and the work done by GSMA NEST. Intent based API exposure is interesting, but it should not be limited to management. Other Telco APIs can leverage on this concept especially when exposed to a 3rd party platforms.

The **management of the RAN** domain, it is covered by both 3GPP SA5 and **O-RAN.** Focusing on O-RAN, that anyway starts from and enhances the 3GPP model, the Service Management and Orchestration (SMO) framework oversees the management of all RAN network functions. This includes near-Realtime RAN Intelligent Controller (near-RT RIC), central unit, distributed unit, radio unit as well as the cloud infrastructure. The management is done over O-RAN defined interfaces named O1, for RAN NFs management, and O2 for cloud deployment. Concerning the cloud platform, it includes networking, storage, and compute resources that are ready to host RAN network functions. It also provides the required tools to manage Virtual Network Functions (VNF) initial deployment, reconfiguration, and lifecycle management. A possible gap concerns the alignment among 3GPP and O-RAN that are very much active and relevant on the same network domain.

An overall view on Network and Service management is in scope of **ETSI ZSM** (Zero touch network and Service Management). Their scope is wide, aiming to define an E2E cross-domain and cross-technology operable framework and solutions for the management and automation networks and services. We can notice that the words Cloud Native and Container are hardly used in ZSM specs. ETSI ZSM is anyway defining lots of principles and architecture models

that apply to Cloud Native environment. It is still early stage though and it is lacking concrete interface specifications yet to enable standard implementation of automated, zero touch open multi-vendor interoperable environments.

Orchestration is one of the main pillars of a modern Telco Platform. There are standardization bodies working on this topic and available Open Source solutions. Both for the virtualization infrastructure and for the **Telco services** running above. The Open Network Automation Platform (**ONAP**) is a comprehensive Open Source solution for orchestration, management and automation of networks. It is intended for network operators, cloud providers, and enterprises. The ONAP platform allows to instantiate network elements and services in a rapid and dynamic way. It supports a closed control loop process to response to real-time events. ONAP provides tools to model the resources and the relationships that make up the service also specifying the policy rules that guide the service behaviour. For the service assurance, ONAP leverages analytics and closed control loop for an elastic management of the service. It is important to notice that the information model and framework utilities continue to evolve to harmonize with the work of many SDOs including ETSI NFV MANO, TM Forum SID, ONF Core, OASIS TOSCA, IETF, and MEF and 3GPP. **The placement and the role of ONAP** in a platform that comprises of BSS and OSS systems, E2E and domain orchestrators and the compliancy to the mobile network standardization in terms of management **is not trivial**. ONAP can operate at different level and must be integrated into an overall OSS system composed by many other entities operating the network.

When it comes to the standardization on the **management of the virtual resources** the most relevant standardization group is the European Telecommunications Standards Institute (ETSI). **ETSI NFV** has defined an architectural framework for Network Function Virtualization (NFV) management and orchestration [NFV1]. The SDO has then defined open REST APIs, descriptors and templates in TOSCA or Yang for different artifacts describing the network elements and their connections. ETSI NFV is designed to be agnostic of the type of virtualization, being hypervisor based or container based. A few published reports and specifications address more specifically the Cloud Native aspects, such as a specification of the classification of Cloud Native VNF implementations [NFV2], a report on the enhancements of the NFV architecture towards Cloud Native and PaaS [NFV3]. This report introduces a new component, the CISM (Container Infrastructure Service Management) to manage the container-based infrastructure. More recently four new specifications have been started to address OS containers. In terms of gaps toward our analysis**,** ETSI NFV is not addressing the application configuration of the NFs that run on top of the VNFs. This is not in its scope, this need is filled by the work of 3GPP SA5. Some gaps may exist to design a heterogeneous environment with physical, virtual VM based and container based VNFs. There is also some misalignment between ETSI NFV and Open Source projects in that space such as OpenStack, Kubernetes, OSM and ONAP.

Many of the novelties in the different layers of Network are enablers of new capabilities and services at the Edge. **GSMA** is working on the **application deployment at the Edge** also considering how Telcos con integrate their platforms**.** GSMA, is an industry consortium that represents the interests of mobile operators worldwide in conjunction with a broader mobile ecosystem. GSMA has two main activities working on Edge: a closed group of operators working on Telco Edge Cloud (TEC), and an open group to GSMA members working on a unified operator platform (OPG). GSMA OPG is defining a common architecture for Edge with a specific focus on interworking. OPG defines four main actors involved in the process, Application Providers, Federated Operators, Network Resources and User Equipment and defines a set of interworking APIs among the actors. The Operator Platform has the goal to federate the Edge of multiple Operators. The proposition is to give, to application providers, access to a global Edge Cloud. On top of this global platform the application providers can run innovative, distributed and low latency services over the edge of different operators in a seamless way. The GSMA OPG white paper provides not only some architecture diagrams and interfaces between operator platforms, but also a southbound interface with the edge environment. The architecture also includes the cloud management platform and a northbound interface with the application providers. Considering the overall picture, we are describing in this document, OPG is currently strongly promoting a unified Edge levering a unified, integrated Telco tailored Cloud Native platform. A similar approach for CN, RAN and OSS is not a primary goal yet.

For the application deployment at the Edge, also the work done by **3GPP WG6 (SA6)** is an important enabler. SA6 specifies a 3GPP, Telco oriented, solution to support the application integration at the Edge [6]. The proposed architecture foresees the Application Server provided by different instances located in different Edge Data Networks. The Application service is consumed by a Client in the User Equipment (UE). The architecture for Edge enablement

is a 3GPP network feature that helps the Client to locate the most appropriate Edge Application Server instance. This work from SA6 is very important in term of new network capabilities, standardized by 3GPP, to support the deployment and the discovery of the Applications at the Edge. It is important to consider that there are other initiatives in the same area e.g. from 3GPP SA2, using a simpler approach based on signaling (with less features), and from GSMA OPG with the its Operator Platform. The optimal solution, maybe integrating the different approaches, must be identified and depends on many requirements at service level, business level and platform level. The exposure of such a set of APIs and the actors that should use them is not deepened. Model of interaction among the stakeholders can be found in the work done by GSMA.

An Open Telco Platform must exploit **Telco oriented APIs**. Focusing on the Edge, this area is covered by **ETSI MEC** (Multi-Access Edge Computing). Its target is to create a standardized, open environment to allow an efficient and seamless integration of applications across a multi-vendor, Multi-access, Edge Computing platforms. ETSI MEC extends the NFV MANO at the Edge with specific enactments to enable Applications deployment at the correct location at the right time. ETSI MEC offers cloud-computing capabilities and an IT service environment at the edge of the network also exploiting Telco APIs. The ETSI MEC approach is built around technical solutions and models to foster the adoption of the defined API, anyway the adoption of such a model should be evaluated also considering the new exposure capabilities and solutions of the 5G Network.

The lower layers of the Telco infrastructure are the pillars of the 5G evolution and are also the basic enablers for the Edge exploitation and the Hybrid Cloud integration. The **virtualization** and, more recently, the **cloudification** process, start at **infrastructure level** with a look on both standardization and concrete solutions mainly from the Open Source community. To decrease costs in the Cloud Native adoption, it is important to have a Telco tailored solution otherwise the effort, time, and integration costs, risk to diminish the advantages in terms of savings. Cloud INfrastructure Telco Task Force (**CNTT**, aka Common NFVI Telecommunications Taskforce) is an Open Source community of network operators and vendors which have the goal to create a common Network Function Virtualization Infrastructure definition. A reference implementation is carried out within the (Linux Foundation) LFN framework and in GSMA. It covers both OpenStack and Kubernetes. CNTT collaborated with OPNF to test and prove the CNTT reference models. They recently merged into the Anuket Project. CNTT defines a Reference Model for the infrastructure abstraction and the exposure of a set of capabilities, resources, and interfaces to workloads. The aim of the Reference Model is to be virtualisation technology agnostic (VM-based, and, container-based) and acts as a catalogue of the exposed infrastructure capabilities, resources, and interfaces. The goal is to provide Operators with a unified consistent cloud infrastructure, vendor independent. The goal of CNTT is not easy to achieve, it has broadened its scope from generic NFV architecture to encompass Telco cloud architecture. This implies complications as Telco specific features are mapped on to a more general VNF/CNF frameworks that must also cover other environments. A Telco tailored solution is needed to reduce integration costs. The reference implementations (OpenStack and Kubernetes) each have their own characteristics that are not Telco specific. The mapping of Telco cloud functions is, for this reason, sometime complex.

Going deeper in the **Cloud Native layer,** the reference implementation comes from Google first and now it is carried on by **CNCF**, Cloud Native Computing Foundation. It is part of Linux Foundation and regroups some of the largest Open Source projects in the Cloud Native space. It includes Projects such as Kubernetes, Promotheus, Fluentd and Helm. Containers, service meshes, microservices, immutable infrastructure, and declarative APIs exemplify CNCF approach. These techniques enable loosely coupled systems that are resilient, manageable, and observable. A possible gap toward the needs of a Telco Operator, looking for interoperability and Telco standards, is in the nature of CNCF itself. The CNCF community was launched to bring together a rich Open Source community to build and deliver Cloud Native solutions. CNCF approach and working model surely guarantee faster “time to market” with respect to official standards solutions. This is quite effective with now a very rich but complex portfolio, growing rapidly. Projects have different maturity levels and their integration is not guaranteed. A challenge is the maintenance of these environments once in production. The integration with other products or tools in the Telco environment is currently not optimised in term of knowledge sharing and cost reduction. The Open Source community does not generally support telecom standard interfaces, so adaptors are often needed to convert e.g. 3GPP or ETSI interfaces with Open Source software.

A group working on **the standardization of the Telco PaaS** is XGVela looking at the capabilities needed to run, develop, manage and maintain the network function/application from a platform perspective. This group is focusing on the deployment of network functions and applications usually related to general IT services (such as firewall, load balancer, monitoring, managing). Software package including adaptation layer (telecom-level enhancement on IT software), Telco service logics (reusable Telco service among different network functions/applications), and app specific logics (unique to specific network function/application) are also considered. The identified capabilities and software functionalities could be extracted and collected as PaaS layer. XGVela focus is on the platform sharable capabilities such as common microservice functional components (e.g. database, load balancer, firewall, common NF microservices, etc.). These components could be common among different vendors/applications/network functions. Observability, PaaS level management capabilities, infrastructure level capabilities provided as a service are also in focus. xGVela work started recently and it still must address many aspects considering that being Cloud Native is not only containerization, but also micro-service and DevOps, which would bring challenges into telecom industry on network function design model, delivery model, operational model, procurement model and etc. XGVela, as a Cloud Native PaaS platform, is trying to add Cloud Native value to Telco cloud platform and aim to make it easier to run/create/manage network functions/application. It faces the challenge of bringing Cloud Native into telecom industry.

For the management of the infrastructure, DMTF (previously known as the Distributed Management Task Force) creates open manageability standards for emerging and traditional IT infrastructures including cloud, virtualization, network, servers and storage. DMTF goal is to foster more integrated and cost-effective approach to management through interoperable solutions. The most relevant outcome is Redfish, a suite of specifications providing a RESTful interface for the management of servers, storage, networking, and converged infrastructure. Redfish is designed to deliver simple and secure management for converged, hybrid IT and the Software Defined Data Centre (SDDC). Both human readable and machine capable, Redfish leverages common Internet and web services standards to expose information directly to the modern tool chain. The standard defines a protocol that uses RESTful interfaces to provide access to data and operations associated with the management of systems and networks. DMTF Redfish® has strong industry support for compute infrastructure management across all of the major IT vendors, however there is less adoption for SNIA Swordfish (an extension of Redfish) and currently no support for Ethernet based networking. Both deficiencies are being actively worked on within the DMTF, but will be subject to industry adoption, once ratified.

3.1 Analysis summary

Going through the work done by the SDOs and Fora, it is clear how the cloudification process, as a key enabler of the Network evolution toward the Open Telco Platform, is well embraced. Many aspects are well supported as represented in the following table. The table is not intended to be complete, it just identifies the relevant players for each domain according to the scope of this document:

Table 1: SDO Fora Summary

Domain	SDO – Fora (considered in this analysis)	Relevant aspects
Service	GSMA NEST	It provides a standardized list of attributes that can characterize a type of network slice. The GSMA GST and inherited NEST templates, focus on the service level requirement and do not specify requirements related to the infrastructure.
	TM Forum	TM Forum has issued a white paper on Cloud Native to explain Communication Service Providers evolution to the Cloud, migrating

		applications, offering APIs, transforming the operating models and introducing AI in operation.
Core Network	3GPP SA2	The introduction of the SBA leads to the possibility of developing Cloud Native-based systems in the telecom world, providing functional modularity and complete separation of user plane and control plane, with core functions communicating over (preferably) RESTful APIs. The introduction of Network Slicing is another important step for the definition of an Open and Sharable infrastructure with native means to guarantee a measurable Service Level Agreement for the Customers. Network Exposure Function (NEF) is another key component for the network programmability and openness providing exposure of network capabilities.
RAN	O-RAN ALLIANCE	O-RAN ALLIANCE uses 3GPP architecture that splits CU according to control plane and user plane (In O-RAN ALLIANCE these splits are called O-CU-CP and O-CU-UP). It also goes further decomposing DU into O-DU and O-RU (Remote Unit), as the very edge of the RAN. This is done to have all the hardware-based functions in the O-RU and the components that can be virtualised in the O-DU.
Management and orchestration	3GPP SA5	SA5 structures the 5G Management System around the concept of Service Based Architecture (SBA) opening it to be customized according to the Telco needs. The SBA concept is based on the idea of having specific Management Services (MnSs) that offer capabilities for management and orchestration of network and service. This approach is well aligned with a Cloud Native environment. Considering that the APIs of the Open Telco Platform should be easy to use, to be adopted in the Developer's ecosystem, it is interesting to consider the work from SA5 in on Intent Driven management. A Hybrid Cloud platform supports fast Service deployment and it is intended to give specific network performance, e.g. at the Edge. For this reason, Service assurance is a key factor for an Open Telco Platform. Service assurance, as Network assurance, in 5G are based on automated control loops.
	ETSI ZSM	An overall view on Network and Service management is in scope of ETSI ZSM (Zero touch network and Service Management) which is working on the automation of operational processes & tasks for emerging and future network and services.
	ONAP	ONAP is a comprehensive Open Source solution for orchestration, management, and automation of network and edge computing services. The solution instantiates network elements and services in a rapid and dynamic way, together with supporting a closed control loop process that supports real-time response to actionable events. This is a relevant tool in the network automation landscape enabling openness and fast deployment of network and services to support Telcos integrations into a dynamic ecosystem.
	ETSI NFV	ETSI NFV has defined an architectural framework for Network Function Virtualization. More recently new specifications have been started to address Cloud Native technologies such as containers.
	O-RAN ALLIANCE	O-RAN ALLIANCE, the Service Management and Orchestration (SMO) framework oversees the management of all RAN network functions. This includes near-Realtime RAN Intelligent Controller (near-RT RIC), central unit, distributed unit, radio unit as well as the cloud infrastructure. The management

		is done over O-RAN ALLIANCE defined interfaces named O1, for RAN NFs management, and O2 for cloud deployment.
Edge	GSMA OPG	It is defining a common architecture for Edge enablement with a specific focus on interworking and federation. The Operator Platform has the goal to federate the Edge of multiple Operators. The proposition is to give, to application providers, access to a global Edge Cloud. The federation approach and the unified platform are essential concepts for the Telco offering to be widely adopted in the ecosystem. The Cloud Native infrastructure and principles are key enablers for this vision.
	3GPP SA6	The architecture for Edge enablement proposed by SA6 is a 3GPP compliant network feature that helps the Application Client to locate the most appropriate Edge Application Server instance. This work from SA6 is very important in term of new network capabilities, standardized by 3GPP, to support the deployment and the discovery of the Applications at the Edge. Telco openness and Edge exploitation with a standard support at network level are supported by SA6 work.
	ETSI MEC	Its target is to create a standardized, open environment to allow an efficient and seamless integration of applications across a multi-vendor, Multi-access, Edge Computing platforms.
Platform and Infrastructure	CNTT	CNTT aims to define an agnostic Cloud Native infrastructure, removing the dependencies between workloads and the deployed cloud infrastructure. CNTT defines a Reference Model for the infrastructure abstraction and the exposure of a set of capabilities, resources, and interfaces to workloads. One important task of the group is translating the reference architecture into a reference implementation covering both OpenStack and Kubernetes.
	CNCF	It is part of Linux Foundation and regroups some of the largest Open Source projects in the Cloud Native space with a set of graduated projects such as Kubernetes, Promotheus, Fluentd and Helm.
	xGVela	Its goal and to define an infrastructure tailored to the Telco needs tu support the deployment of network functions or applications usually related to general IT services (such as firewall, load balancer, monitoring, managing). This work supports the concept of a Cloud Native infrastructure defined one and for all and Telco tailored to minimize integration costs.
	DMTF	It creates open manageability standards spanning diverse emerging and traditional IT infrastructures including cloud, virtualization, network, servers and storage.

4 CLOUDIFIED OPEN INFRASTRUCTURE

The evolution of the Telco infrastructure is a process that always happens periodically driven by different aspects such as the introduction of new advanced network features, new communication services for the customers, internal efficiency or better network operation. The current evolution is driven by the same needs, but it is also exploiting a deeper and deeper integration of IT and Telco worlds. Cloud based approaches proved to be fundamental for the existence of complex and heterogeneous systems. The automation and orchestration features developed for such a complex Cloud system is becoming one of the most relevant factors of today's Telco infrastructure evolution. This Chapter is intended to describe a cloud ready Open Telco Infrastructure that will enable adoption of Cloud Services

leveraging a cost-efficient infrastructure as a whole; delivering the description of a multi-vendor open infrastructure and its capabilities up to the VIM Layer:

- Adopting a Cloud Native approach for the network implies having integrated in the infrastructure a service oriented, platform that can also support 3rd party service-oriented applications (not in the same domain of the Telco network functions)
- It is important to choose the key technologies that are the real enablers, the ones that give to the Telco advantages for its core business (the Telco network) and that are also the enablers for the opening of the network to the developers.

4.1 From Physical to Cloud Native Network Functions

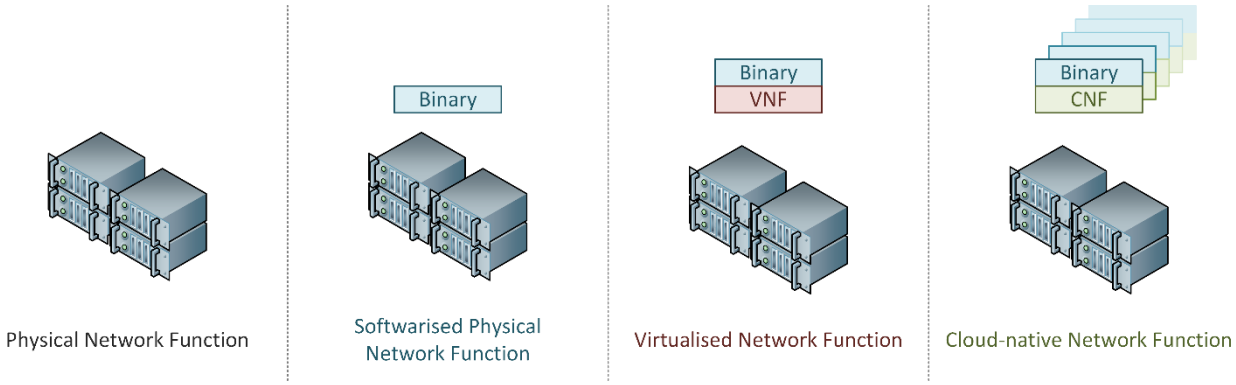


Figure 1: From physical functions to Cloud Native applications

4.1.1 Physical Network Functions

In legacy wireless telecommunication systems such as 2G and 3G the functionality of an architectural component or network entity, was realised as a physical function, connected to other physical functions over point to point functional interfaces, released by a vendor as a single physical closed system. The vendor had all the freedom to customise hardware as well as software (kernel and function itself) to offer an all-in-one solution. This is understood to be a Physical Network Function (PNF). Extensibility on customer side is therefore only possible with strong alignments with the physical function provider and/or extensions to standardised interfaces.

4.1.2 Softwarised Physical Network Functions

Softwarising a physical function decouples the functionality it offers from the compute hardware it operates on, but these functions retain their monolithic characteristics that makes it hard or impossible to easily decouple services. With the success of Linux and support for a wide range of computer architectures (x86, arm) softwarised functions only have dependencies to other software libraries or drivers. Software and operating system components can be updated, upgraded, and replaced through software pipelining concepts only. Even though there is no dedicated term defined in the industry, one can argue to classify this as a softwarised PNF.

4.1.3 Virtual Network Functions

The ability to virtualise a softwarised network function allows to offer the compute hardware resource to more than one softwarised function and abstracts its underlying operating system dependencies in a constraint manner depending on the required level of isolation (mainly virtual machines, containers). Furthermore, Virtualize Network

Functions, particularly those used in 4th Generation wireless telecommunication, start to show separation of concerns, e.g., separation of Data and Control planes, albeit still maintaining monolithic SW blocks. As a result, multiple Virtualised Network Functions (VNFs) can be orchestrated in a "as-a-service" fashion onto a wide range of Commercial off-the-shelf (COTS) compute hardware and frameworks to automate this procedure through open APIs and programmable infrastructures.

4.1.4 Cloud Native Network Functions

The term Cloud Native originates from the ability to realise an economy at scale – hyperscale – through agile code development and code integration design patterns. At the core is the idea to decompose a function into microservices that can exist as multiple instances to allow to scale with demand. Cloud-native is commonly agreed to define applications that follow the 12-factor [29] methodology, as outlined by various market leaders [30] [31] and summarised in Table 2. Thus, if VNFs follow the aforementioned 12-factor code development and integration methodology, they can operate as Cloud Native Network Functions (CNFs).

Table 2: 12-factor app properties

Number	Property	Description
1	Codebase	One codebase tracked in revision control and being able to deploy it into different production stages (development, staging, production).
2	Dependencies	Explicitly declare and isolate software dependencies through packaging.
3	Configuration	Software configuration stored in environment and not "hard coded" inside binary allowing different deployment scenarios.
4	Backing Services	Any service an individual function relies on must be treated as an attached (remote) service that can be reached over a network. Examples are databases or external service such as Twitter or Google Maps.
5	Build, release, run	Separation of software development into separate stages disallowing changes to code after build phase to enforce proper code integration workflows.
6	Processes	The application is decomposed into individual stateless processes that can be packaged as individual microservices.
7	Port binding	Mapping function from internal port to public port, e.g. public HTTP Port 80 is mapped inside instance to port 8080 where the function is listening.
8	Concurrency	Microservices of same type can be scaled out to meet demand.
9	Disposability	Maximise robustness of microservice with fast start-up and graceful shutdown.
10	Dev/prod parity	Keep development, staging, and production as similar as possible.
11	Logs	Treat logs generated by a microservice as event streams that can be analysed outside of the application.
12	Admin Processes	Run admin/management tasks as one-off processes such as database migration.

In addition to the 12 factors, three more have risen in the cloud community which are listed in Table 3.

Table 3: Additional three properties to the 12 factor app properties

Number	Property	Description
13	API First	Make everything a service. Assume your code will be consumed by a front-end client, gateway, or another service.
14	Telemetry	Ensuring that the microservice is designed to include the collection of monitoring, domain-specific, and health/system data as part of the logs.
15	Authentication/ Authorization	Implementation of identity across all microservices that form the application.

4.1.5 Cloud Native vs Cloudified Network Functions

It becomes apparent that VNFs implementing network functions such as firewalling, IP address assignment or switching and routing might not be able to comply entirely with the 12-factor paradigm. For instance, aiming at implementing a 3GPP SA2 Service Communication Proxy (SCP) as a CNF, a component performing proxy-like routing tasks can be certainly decomposed into micro services based on their workload type (e.g. long-running tasks versus short logical operation to determine an outcome); however, by decomposing a network function into microservices the newly created CNFs need to be addressable among each other based on stateless protocols like HTTP. The result is a typical “chicken and the egg” problem, as the CNFs were supposed to implement service routing but relies on a service routing among them. Other factors such as port binding and dev/prod parity simply do not apply to functions that sit below the transport layer where ports are exposed. Furthermore, for networking related tasks (routing, firewalling, etc.) packets from senders such as the UE that are supposed to be handled must be encapsulated in a stateless protocol to reach the next microservice that forms the networking application. Thus, not all VNFs can be ported to CNFs to enable an economy at scale.

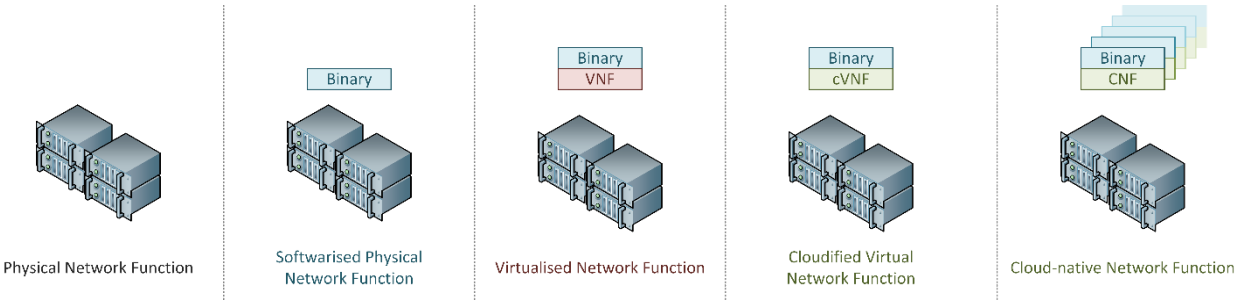


Figure 2: Revised evolution of PNFs to CNFs and cVNFs

However, even though not all 12 factors can be fulfilled for some VNF types, VNFs can be cloudified aiming at a high adoption of the Cloud Native factors without the notion of decomposing a VNF into microservices (CNFs) that form the application. Thus, this paper argues for the introduction of the term cloudified VNF (cVNF) indicating the adoption of the Cloud Native factors 1-5, 10 and 11.

4.2 Cloud-native Orchestration in the Telco Edge

This section aims at putting a stake in the ground when talking about Cloud Native service orchestration within the Telco world and the impact this is going to have on the current Telco technologies enabling orchestration and lifecycle management of VNFs and CNFs. The discussion is structured with the question of what Cloud Native concepts in

the Telco world entails and a mapping to a value chain that includes the operator enabling the communication at the edge and disregarding today's Over the Top (OTT) business proposition.

From a value chain perspective, it can be stated that (public cloud) Platform-as-a-Service (PaaS) offerings like Google and Azure as well as (public cloud) Infrastructure as a Service (IaaS) offerings like Amazon Web Services (AWS) have become the norm with service providers like SAP and Salesforce being able to scale their offerings based on the cloud value chain [32]. However, this paradigm does not constitute for the connection from the cloud (data centre) to the terminal itself where telecommunication operators enable the data pipe between the UE and the internet but are left out in the IP value chain. Instead, by acquiring access to the internet, the user essentially financially compensates the Internet Service Provider (ISP) (including mobile operator) with a monthly flat fee for the delivery of all services that are available on the internet. It becomes apparent that such a flat fee must compensate for the entire infrastructure and platform which is – in comparison with a cloud *-as-a-Service offering –physically even more distributed and in case of mobile operators also requires more facilities for base stations and cabinets. Understandably, adding vertical services into Telco networks requires economic and technological advances to integrate both value chains, the ISP/operator's and the vertical's one. The current approach to on-board Verticals into the Telco world is a bottom-up approach (technologically speaking) where it is expected that services integrate into an existing Telco system following frameworks like ETSI Multi-access Edge Computing (MEC) or 3GPP's Common API Framework (CAPIF); not to mention that all services end up becoming a VNF/CNF orchestrated via the infrastructure orchestrator, i.e. not location-aware and does add additional complexity for Telco-Verticals to bring their service to the edge or just closer to the required consumption point. However, when turning the approach upside down and putting the vertical service provider at the starting point, it becomes apparent that challenges around building scalable services and following state of the art programming and library usage paradigms are key for Verticals, as outlined by the Cloud Native Computing Foundation (CNCF) [33].

A key characteristic of Cloud Native concepts for orchestration is the ability to follow standardised deployment and operational procedures across various cloud data centres. The orchestration procedures are fully decoupled from the service that implements how to respond to requests. In more detail, for both operations the key is the separation of deploying and managing service instances, and the operation of the service itself inside an instance. In a Cloud Native orchestration world, services are pre-packaged (offline or at run-time) images that have no notion of the deployment and operational procedures required to orchestrate a service, to scale it based on demand, failover procedures, or economic incentives. The ability to build services in such a way is what the paradigm shift from monolithic functions to microservices entails. When mapping this to the Telco, it means that a service is simply waiting for a service request to arrive to process it without any additional logic to talk to the underlying system for location, billing, orchestration, identity management, or any other purposes.

Scaling a service is key for a cloud service, and cloud providers such as AWS, Google or Azure allow service providers to define Service Level Agreements (SLAs) under which services must scale in (less instances of the same service) or out (more instances of the same service). The underlying scheduler then takes over the task of monitoring instances and takes respective decisions on what to scale when and where within the data centre. The major challenge that the Platform presented in this paper addresses is to enable location-aware orchestration including scaling in all three dimensions, i.e. up/down, out/in vertically, and out/in horizontally with the clear separation of the service management operations from the operations of the service itself.

4.3 Service Routing

The routing of service requests and responses among Cloud Native service instances must be enabled by a service routing approach that operates transparently and independently of the orchestration – similar to the cloud. Packet routing among services is driven by SLAs defined for the service and an interplay between the adherence of aforementioned SLAs (Quality of Service (QoS) thresholds for various layers). Key for any service routing realisation is to offer throughput at high data rates for bandwidth intense services (e.g. XR media applications), low latency and extremely low jitter (e.g. time sensitive applications), and – most importantly – an instantaneous policy-driven switching of the path to a new service instance in case of a lifecycle management change. With the work in

3GPP/SA2, described in Deliverable D0, concepts around service routing are emerging now for the control plane of a 5G core network, but only recently including the user plane in pre-standard activities, as worked on in Horizon2020 Innovation Action projects [34]. Clearly, improvements to triangular routing approaches, dependency on Domain Name System (DNS) and routing based on flows or labels with a stateful switching/routing fabric, are desperately needed to offer a service routing for both control and user planes that meet the requirements for 5G and beyond.

4.4 Physical Infrastructure

Defining Factors of an Open Multi-vendor Physical Infrastructure

There are several factors that define open infrastructure. The following physical attributes apply to the Telco Industry as a minimum: Industry standard form factors that adhere to common Enterprise IT environments for rack width and depth. There are however exceptions to this rule for Telco environments that are more space constrained. In this case, shorter depth servers (around 500 mm) may be appropriate. Support for DC in addition to AC power supplies is critical, especially for infrastructure that is to be deployed outside more 'enterprise-like' data centres. For those remote data centres, environmental factors may need to be considered and support for NEBS/ETSI/ASHRAE A3 for the hardware chassis as well as all infrastructure components contained within it may be required. The component sets should be industry form factor compliant, adhering to specifications from organizations like OCP. This is particularly appropriate for server NICs and allows for a greater choice of NICs to be used than those that require adaption to vendor specific form factors.

Dependent on workload, the server NICs themselves should be capable of supporting open hardware assisted acceleration technologies such as Data Plane Development Kit (DPDK), Remote Direct Memory Access (RDMA) and Single-Route Input/Output Virtualization (SR-IOV). In addition, offloading key networking functions and protocols such as Open vSwitch (OVS), Virtual Extensible LAN (VXLAN) and IPSec can be beneficial for Telco specific use cases. Management of infrastructure should tightly adhere to open standards such as Distributed Management Task Force (DMTF) Redfish with minimal vendor proprietary properties carrier as part of Redfish OEM.

Hardware Disaggregation & Common Industry Components

Fully decoupled software running on abstracted general-purpose hardware is likely inevitable for the Telecoms industry due to a number of factors covered in the Total Cost of Ownership (TCO) section below. Hardware disaggregation consists of hardware that allows for the independent operation of multiple communications or other applications from one or more software vendors. Those applications are typically virtualized and should be capable of being dynamically deployed, updated and replaced, based on business need.

In a disaggregated model, support for the applications themselves should allow for that independent deployment paradigm and not have single vendor dependencies on integrated support across both hardware and application level components. The component sets themselves should be general purpose in that they are not tied at the application layer to a single network function or vendor. Where possible, applications should function on de facto standard commodity chipsets, rather than vendor specific custom ASICs. The major chipset vendors in the Telco space are increasingly developing Telco specific SDKs along with vendor eco-systems to enable this change.

Optimizing Infrastructure Total Cost of Ownership?

Telcos are spending heavily on operational expenses (OPEX) due to an ageing infrastructure, with irregular capital expenses (CAPEX) to pay when equipment requires replacing. As an example, mobile network operators spend up to 80% of their capital expenditure and up to 60% of their operating expenditure on the Radio Access Network (RAN) alone. Constant traffic growth and flattening revenue per subscriber has put enormous pressure on mobile network operators to evaluate alternative ways to reduce capital and operating expenses. Typically, services are siloed with each application or network type requiring a unique infrastructure. Virtualizing and sharing general purpose infrastructure can help reduce OPEX via streamlined operations and management and reduce power usage through improved asset utilization.

An example of this is virtualised RAN (vRAN) which brings significant TCO improvement into RAN economics with up to 44% lower TCO than conventional distributed RAN deployments Source: ACG Research 'Economic Advantages of Virtualizing the RAN in Mobile Operators' Infrastructures' Given that RAN TCO constitutes around 50% of overall operator spending on a cellular network, vRAN significantly improves overall CAPEX and OPEX of a Telco operator. With fully decoupled software, vRAN infrastructure scales horizontally and does not require vertical hardware upgrades to address evolution of network functionalities, typical for appliance RAN. That optimizes investment into RAN infrastructure and decreases truck rolls and tuning efforts associated with hardware upgrades. Superior TCO across both the RAN and core networks is achievable by repurposing general-purpose infrastructure for additional use cases as well as leveraging open, industry aligned tooling for maintenance and lifecycle management, versus a closed, vendor specific ecosystem where specialized knowledge is required.

By virtualizing multiple network functions on general purpose infrastructure, Telcos can test and launch new services without the acquisition and test cycles typical of hardware-based vertically integrated services, reducing service deployment times. CAPEX reduction can in turn occur via procuring general purpose infrastructure capable of hosting multiple network functions while enabling consistent sparing. This cost, coupled with Telco requirements for long life (typically 5 year) COTS infrastructure Stock Keeping Units (SKUs), significantly offsets the long lifespans of legacy vertically integrated appliances. Infrastructure virtualization inherits an additional cost in terms of frequency of patch updates along with complexity of managing a virtualization layer in combination with the actual network functions in use.

Once the network is virtualized, Telcos can however optimize service configuration in real-time to meet customer demand instead of incurring the expense of deploying single function infrastructure that is scaled to support peak periods but operates at minimal utilization for most of the time. Optimizing multivendor procurement policies across technological suppliers provided Telcos with considerable commercial advantages vs being locked into single vendor per service procurement cycles. The potential economic impact of new procurement policies is marked by many Telcos as the number one driver for vRAN introduction as a minimum.

4.4.1 Networking

In a (B)5G Telco cloud with many tenants, the programmability of the networking fabric is of paramount importance enabled by a software-driven realization of switches and routers. Commonly known as Software-defined Networking (SDN), the ability to program the data plane behavior of network components based on the header information of packets has been around since the move to packet-switched networks but only manifested in a standardized and open approach in the early 2000s with OpenFlow, [35], and the softwarisation of switching and routing functionalities. OpenFlow standardizes the communication between the network controller including switch bootstrapping, pro-active and reactive switching/routing rule management and monitoring. The set of rules implement Layer 2, 3 and 4 protocol header offsets standardized within IETF and 3GPP, e.g. Internet Protocol Version 6 (IPv6), Transmission Control Protocol (TCP) or GPRS Tunnelling Protocol User Plane (GTP-U). However, it became apparent that the realization of Software-defined Networking (SDN) controllers became an almost proprietary race where OpenFlow could not guarantee that SDN controller of Vendor A could work with Vendor B. Not to mention that northbound SDN controller APIs were not standardized beyond the usage of a Yang model [36] and any northbound application would either require to become a module of a controller or implement controller specific Application Programming Interfaces (APIs). In addition to the non-standardized northbound API of SDN controllers, the existence of parent and child SDN controllers is also left to the developer community realizing a controller allowing controllers to serve a particular tenant with a selected topology view [4].

Ethernet Virtual Private Network (EVPN) is seen as the savior for SDN and realizes a controller-less SDN solution that aims at making more scalable and interoperable in comparison to OpenFlow. EVPN is control-plane solution for VXLAN tunnels that uses the BGP routing protocol and does not require any changes to the switches and routers deployed in an infrastructure.

Either way, neither OpenFlow nor EVPN allow to program switches with an arbitrary header offset and matching rule definition allowing deviations of existing header offsets or entirely new protocols to be used. Especially in the growing

area of private network deployments, innovative switching and routing approaches, e.g. 3GPP Service-based Architecture implementations, could greatly benefit from that. Unsurprisingly, a step towards this ambition is underway in the likes of P4, [37], which also follows a separation of control and data plane (similar to OpenFlow) and allows for programming how a switch processes packets.

In combination with the shift towards virtualized networking components, the move towards fully programmable network switches, routers and ports have some challenges to overcome though related to functionality pushed into hardware for performance reasons, i.e. checksum offloading, header fragmentation and segmentation as well as VLAN acceleration. When operating in a (nested) virtual environment hardware supported features such as checksum offloading have severe implementation flaws and must be disabled which significantly impacts the performance negatively though.

4.4.2 Compute

Key for a cloudified open Telco infrastructure is the ability to offer compute that can be universally utilized either through bare metal deployments or the virtualization of a Central Processing Unit (CPU). Especially the latter allows the abstraction of hardware for tenants and the programmability of the usage of virtual CPUs and is a key enabler for slicing the compute resources of a unit. The two CPU architectures that have emerged in the market are x86 and ARM-based architectures. While ARM offers better power consumption and higher speed without the necessity for excessive active cooling, x86 comes with a wider range of capabilities through its extended set of instructions.

Over recent years, the usage of Graphical Processing Units (GPUs) became more popular for CPU intensive tasks such as AI/ML execution or video/image processing. While the inclusion of GPUs as part of an IaaS offering comes with its own current limitations related to virtualizing them, there has been a paradigm shift by the key GPU vendors to offer Docker-based solutions.

4.4.3 Storage

The cost per gigabyte storage on a disk has literally fallen since the introduction of computes in the early 1980s and is at an all-time low. Additionally, the increase in R/W speed with the help of solid-state drives allowed a wide range of storage solutions ranging from smartphone-based to public cloud ones. However, the core concept to deal with the tremendous increase of (streamed) video traffic is frontloading content to the edge where it is required allowing for spatial content distribution to ease the load on the core of today's communication networks. While the isolation and virtualization of disk operations and its space are widely supported in modern operation systems, a disk (similar to read access memory) cannot be overallocated on a best effort basis, as it can be done with virtual CPUs (vCPUs). This poses a challenge to IaaS providers to equip the edge with the right amount of storage. One could also argue that the edge should become even more powerful than a centralised data centre, as all CPU and Disk intense processing/content will be required to be offloaded to the edge in order to meet the desired Key Performance Indicators (KPIs) on latency, jitter and throughput per user.

4.4.4 Infrastructure Monitoring

The challenge for moving PNFs to VNFs or CNFs is to provide the same level of reliability and availability on separately developed hardware and software. A PNF provides a tight integration that includes all necessary software and services, purpose built to its specific function. The risk is very low that an integration issue will impact the performance of a PNF. Since a VNF is provided independent of hardware, it requires that both the host infrastructure and the VNF are qualified independently, then correlated to assess performance and identify faults.

Unlike PNFs and their accompanying defined specifications, VNFs do not have identified performance caps as applied to infrastructure. This presents a challenge to defining performance thresholds for fault management related to infrastructure that is not yet being heavily utilized. Benchmarking tools in place of VNFs are applicable to assessing

performance of physical, virtualized and Cloud Native infrastructures. Benchmarking provides the ability to characterize VNF work-load performance metrics to assist with optimizing the assigned infrastructure resources for VNF deployment as well as define infrastructure performance monitoring and alert thresholds. Furthermore, benchmarking can deliver better capacity planning for VNFs that auto-scale and for VNFs that are sensitive to latencies from remote peers, network attached services, and storage [1].

4.5 Virtual Infrastructure

4.5.1 High Level Overview of Trends in Telco Virtualization (DeFacto-Standard and Market Drivers)

Telco virtualization workloads are in the main currently focused on deploying network functions on Virtual Machines (VMs). OpenStack has become the de facto Open Source based standard in that space alongside VMware's offerings. Decisions by Telcos on which virtualization stack(s) to deploy usually fall into balancing the need for mature, full infrastructure lifecycle management solutions from proprietary commercial vendors vs selecting OpenStack distributions from industry leading vendors and supplementing those with Open Source based assurance tools.

As proprietary commercial offerings typically own the full solution stack, they can offer fully integrated lifecycle management and associated SLAs from a single vendor whereas this may not fit with some Telco's desire to make independent decisions on technology component choice at a more disaggregated level. As the market moves towards 5G and associated containerized workloads, the current status quo will shift towards Kubernetes based deployments with stated directions from the prime vendors to run both VM and container-based Network Functions (NFs) on the same control plane using technologies such as KubeVirt. This move will accelerate the migration of workloads from VMs to containers and may help increase the Telco market share of additional virtualization vendors including the major Hyperscalers deeper into Telcos networks.

4.5.2 Bridging Management of physical and virtual infrastructure in an open and standardized way.

Reducing the number of integration points across data centres for infrastructure management is a key concern for Telcos. The evolution of a Cloud Solution Providers (CSPs) network to 5G and Edge Computing requires IT compute, storage and networking infrastructure from multiple vendors, hosting component sets from multiple manufacturers to be deployed across potentially thousands of geographically distributed and diverse points of presence including central offices, cell tower huts, wiring closets as well as traditional data centres.

This sharply contrasts with the comparatively few homogeneous Hyperscale data centres of the cloud service providers. Most data centre infrastructure management solutions originate from the Enterprise, are monolithic, vendor specific, sometimes embedded and do not scale to support the massively distributed, heterogeneous data centres as are prevalent in telecoms and other networks today. Furthermore, the infrastructure they manage is typically tightly integrated with few, well defined interfaces that allow for consistent, repeatable management of identically configured racks of infrastructure. As such, those management solutions are ill equipped to deal with heterogeneous multi-vendor infrastructure environments in an open, standard-based manner.

CSPs are demanding open, scalable infrastructure management for this distributed multi-vendor infrastructure based on widely adopted industry standards such as the DMTF's Redfish API and data model. There is however minimal alignment between DMTF Redfish which has been targeted towards supporting compute infrastructure and networking as well as storage management. This is being addressed within the DMTF with new standards for Ethernet switch fabric support as well as Non-Volatile Memory Express over Fabrics (NVMeOF) support for shared storage. In addition, new Open Source communities such as Open Distributed Infrastructure Management (ODIM), hosted within the Linux Foundation are moving to build Open Source tools that exploit DMTF Redfish standards. This space is a work in progress for Telcos but will become more and more critical to address as IT based workloads are deployed outside the core 'Enterprise Like' data centres prevalent in Telco networks today.

4.5.3 Prominent VIM, NFVO and Service Orchestrator Technology Enablers

With 5G and the evolution to NFV and Cloud Native for every parts of the network except maybe the antennas, the 5G virtual infrastructure becomes a set of 'data centres', Network Function Virtualisation Infrastructure (NFVI), distributed across the edge and the core as described in Figure 3. These different instances of NFVI are managed by a Virtual Infrastructure Manager (VIM) as defined by ETSI NFV that provides an abstraction of these physical data centres into a set of virtual data centres with a mix of physical and virtual resources.

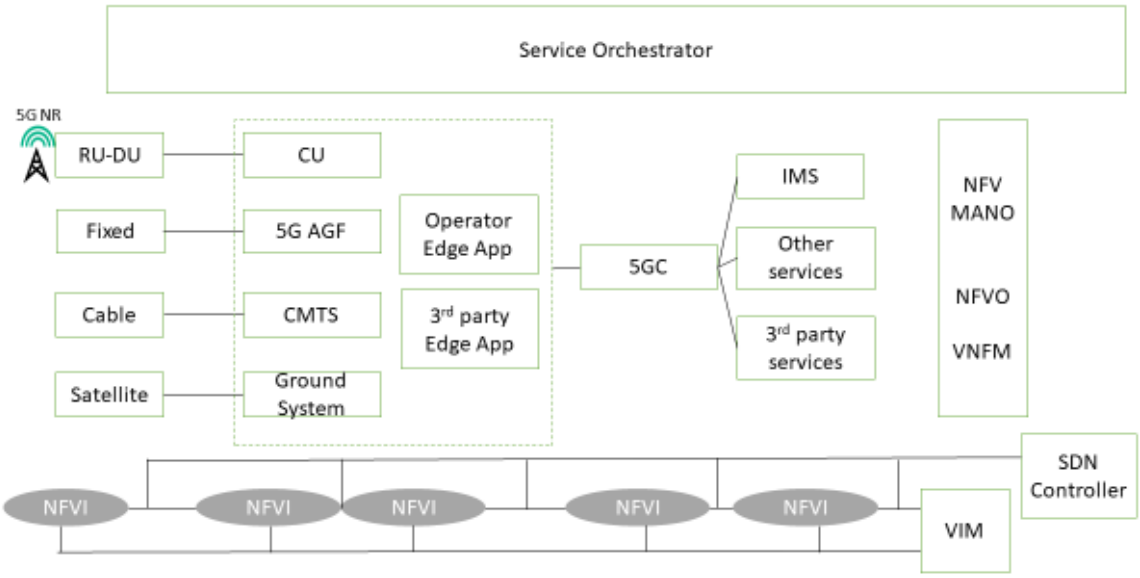


Figure 3:Distributed 5G virtual infrastructure

The virtual network functions that are deployed on these virtual data centres and the network services that are built out of the combination of those VNFs and PNFs are managed by the VNF Manager (VNFM) and NFV Orchestrator (NFVO) of the ETSI MANO. The connectivity between the different data centres or NFVI is managed by a set of network controller or SDN controllers. The overall orchestration of the network is performed by a Service Orchestrator.

New paradigms are introduced with 5G. Typically NFVIs are more and more hybrid with a mix of bare metal Model #1, hypervisor based Model #2, container based infrastructure, either independent Model #3, or containers inside VMs, Model #4 as depicted in Figure 4.

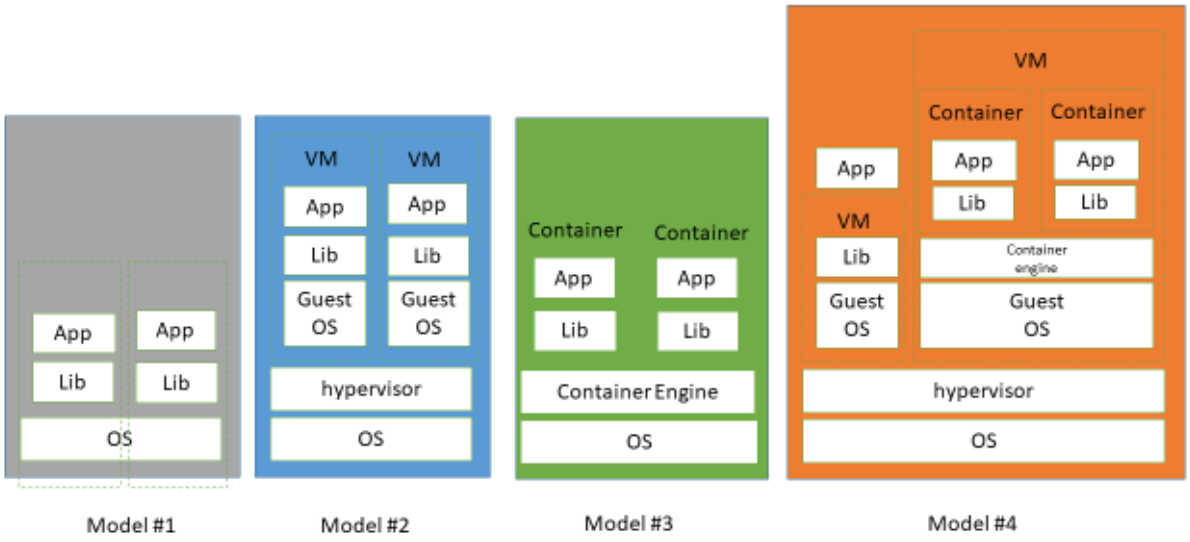


Figure 4: NFVI deployment models

To manage the co-existence of these different infrastructure models in an NFVI or across multiple NFVI inside an operator network, operators need to either deploy independent VIMs or VIMs that can manage hybrid environments or a hierarchy of VIMs with a layer of abstraction towards the NFV MANO. Typically access real time is Model #1, Edge operator environment with CUs for instance is most often Model #2 today for SR-IOV and DPDK, while Edge 3rd party applications is moving to Cloud Native either Model #3 or Model #4, and 5G Core is definitely moving to Model #4 or Model #3. One approach is to deploy a VIM for each of these environments and then either connect directly those RANs to the NFVO, or to have an umbrella VIM that integrates all those VIMs and provides a layer of abstraction to the NFVO.

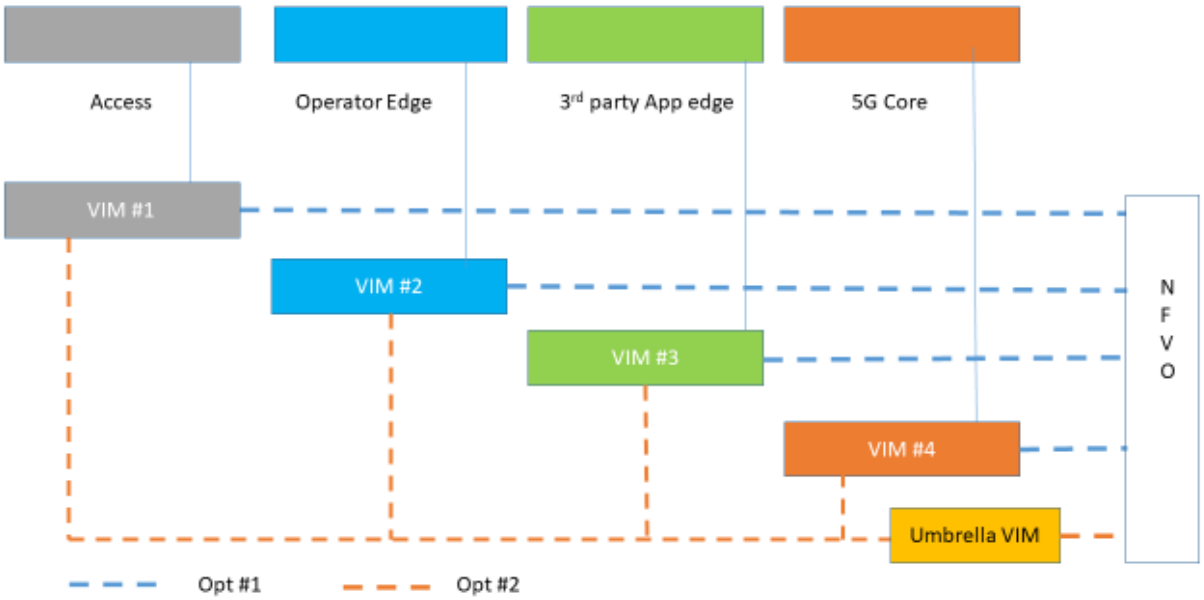


Figure 5: NFVO and VIM Deployment

With the evolution towards lighter containers, more automation, and reducing time to deploy or update instances, NFV Instractuctures are being enhanced with more value added services, ie load balancer as a service, firewall as a

service, logging, tracing, with open APIs: Infrastructure as a Service (IaaS) when hypervisor based or Container as a Service (CaaS) when container based. This implies an evolution of the VIM towards a IaaS manager, i.e. OpenStack in combination with a CaaS manager, i.e. Kubernetes or OpenShift.

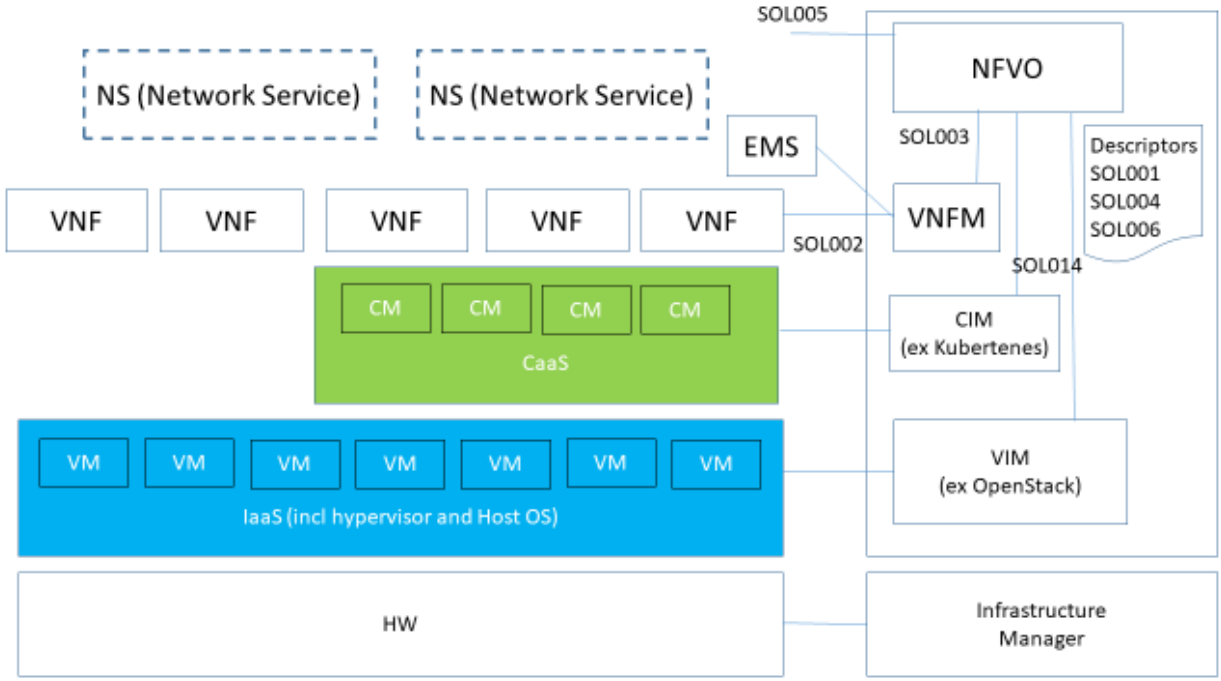


Figure 6: IaaS and CaaS

The principles of the ETSI NFV MANO stack remain, with the layers Virtual Infrastructure Management (VIM & Context Information Management (CIM)), VNF Manager (VNFM) for the lifecycle of the VNF being VM or Container (CM) based and the NFV Orchestrator for the lifecycle of the Network Services and the interactions with infrastructure via VIM or CIM and with the Operations Support System (OSS).

However this NFV MANO needs to evolve to support some new parameters on the SOL interfaces, and SOL TOSCA and Yang specifications for the Network Service and VNF descriptors, on going work in ETSI NFV. The NFVO and typically vendor NFVO or Opensource (ONAP or OSM) need also to evolve to not only interface with hypervisor based VIMs but also CIMs and CaaS management environments, and support the onboarding of container based applications, as well as the evolution of their internal data models to support the different topologies described earlier.

The Service Orchestrator that sits on top of the NFVO interacts also with the network controllers and SDN controllers that control the connectivity across the network. Typically to deploy a new 5G RAN instance, assuming the antenna and RU have been deployed, the Service Orchestrator needs to request the NFVO to deploy different Distributed Unit (DU) instances in specific edge locations for instance with Central Unit (CU) user plane (CU-UP) for different slices and a CU control plane (CU-CP) more centrally. It also needs to ask the network controller to provide connectivity between those different entities and configure the connection points accordingly, but also the management parameters. Because of the heterogeneity and dynamicity of the environment as described earlier, interfaces with the Service Orchestrator should be intent based and not go into the details on how to do things. The information model and instance modeling inside the Service Orchestrator should also be intent based, for the same reason, and very flexible to support the dynamicity of the environment and the automation of service management, as defined in ETSI ZSM005 ‘means of automation’.

The Service Orchestrator needs to evolve also to support not only the orchestration of network services, but also the orchestration of network slicing with NSMF (Network Slice Management Function) and subnetwork slices with NSSMF (Network Subnet Slice Management Function) and interactions with Network Function Management Function (NFMF), as defined by 3GPP and in alignment with ETSI NFV.

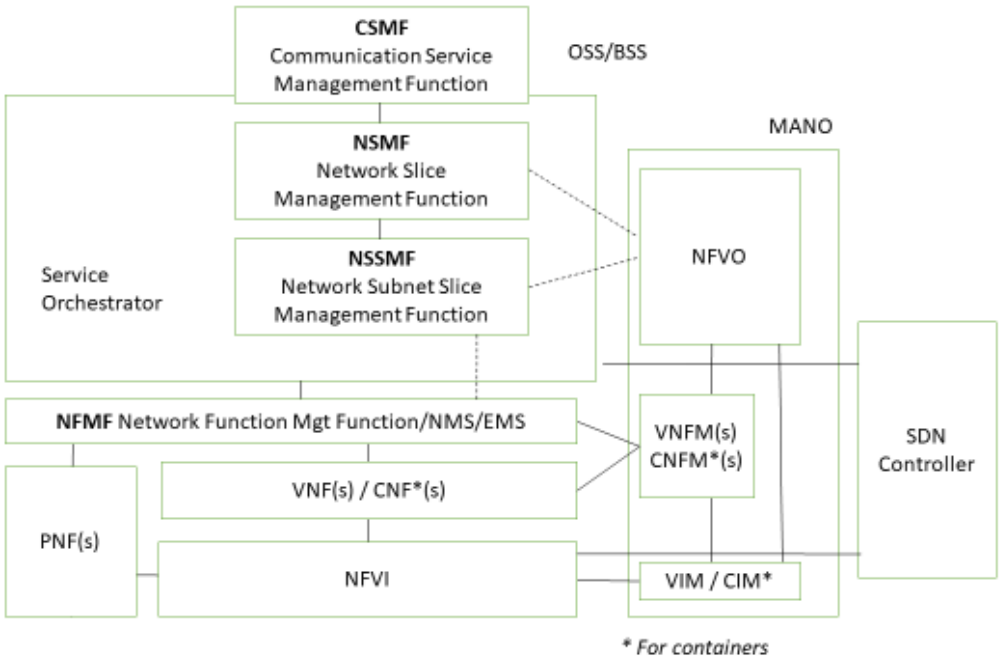


Figure 7: NFVO & NFV MANO

5 CLOUDIFIED OPEN ARCHITECTURE

Different aspects, at different layers of the Telco Platform, concur to the Cloudification process. The innovation at infrastructure level, described in the previous chapter, is one of the main pillars of this evolution. The Cloudification process also includes architectural aspects relevant for the openness of the Telco Platform such as the adoption of a Service Based Architecture (SBA). This leads to modularity, interoperability, adopting standardized interfaces, and an open and multi-vendor ecosystem. Operating an open platform comprising of different modules and functions is a challenge that requires the adoption of modern and Cloud based management solutions. Orchestration, for this reason, is a fundamental factor for a platform to manage this complexity. This is a key enabler for an open integration of new components provided not only by different network vendors but also by developers and Verticals.

5.1 Service-based Architecture

Web technologies and cloud concepts, i.e. *-as-a-Service, have a relevant impact on the Telco world and have seen a significant adoption. Architectural changes for the 3GPP-based mobile system are underway since Release 15. 3GPP SA2 is one of the main contributors for the paradigm shift towards a service-based architecture as explained in chapter 3. This section describes the core changes to a 3GPP system with a focus on the control plane only (user plane is not SBA). Anyway a brief outlook on user plane possible evolution is provided too.

5.1.1 Control Plane

5.1.1.1 Service Mesh

In a Cloud approach, applications are composed of multiple microservices. These microservices need to interact with each other, e.g. to transfer user data or context from one point to another. These communication could be implemented independently in the different microservices or can leverage on the architecture itself. As the system becomes more complex, with many microservices interacting, the second option is the most efficient. To address this issue, service mesh architecture and solutions are being used. Service mesh provides a proxy (also referred as SideCars) that is deployed with each microservice and helps to route communication requests to the destination microservice proxy using the optimum route. Nowadays PaaS generally include a service mesh functionality.

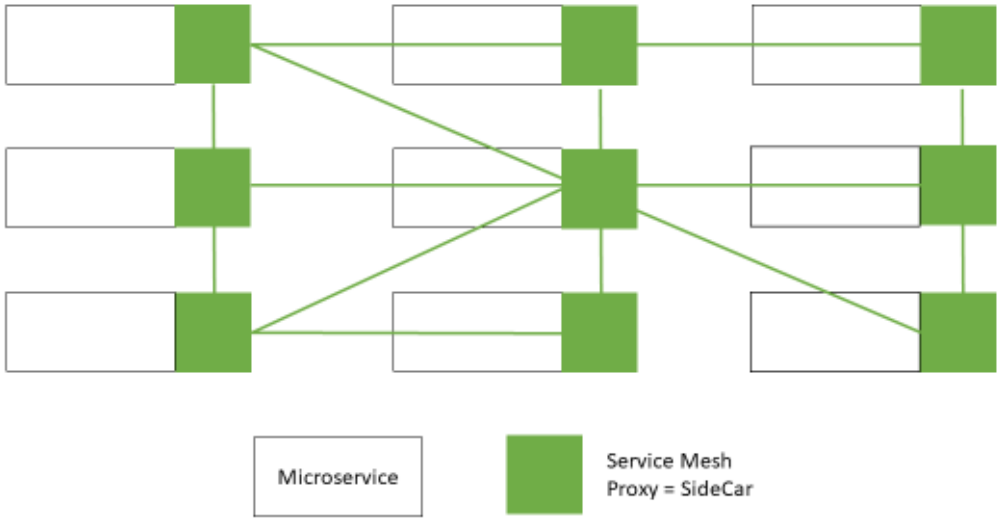


Figure 8: Service Mesh

5.1.1.2 Independent Deployment Units

The 3GPP specification TS23.501 Rel. 16 [5] provides has an alternative approach for microservices communication. 3GPP defines the independent deployment unit scenario where a Service Communication Proxy (SCP) deployment unit can make use of microservices. This allows the microservice to be independent of the message forwarding platform. The SCP agents implement the HTTP intermediaries between service consumers and service producers. It is acting as HTTP proxy which registers services on behalf of the producers in Network Repository Function (NRF). NRF is a 3GPP component very much in line with the Cloud Native approach. It is indeed used to discover the services offered by other network functions. The SCP agents are controlled by the SCP controller. As depicted in Figure 9, communication between SCP controller and SCP agents is via SCP internal interface (4) and up to vendors implementation. The SCP interfaces (1), (2) and (3) are service based interfaces. SCP itself is not a service producer. Interface (2) represents same services as (1) however using SCP proxy addresses. Interface (3) is interfacing NRF e.g. for service registration on behalf of the 5GC functions or service discovery. The SCP determines the routing and forwarding procedures based on the delegated discovery request by interacting with the NRF.

Since the SCP is acting in proxy mode, there is a need to explicitly address the SCP within the 5GC functionality for leveraging the SCP's functionality. This creates further configuration and complexity within the 5GC consumers and producers. However, this option envisages direct communication which can coexist in the same deployment based on 3GPP specified mechanisms – a SCP is not in use in this special case.

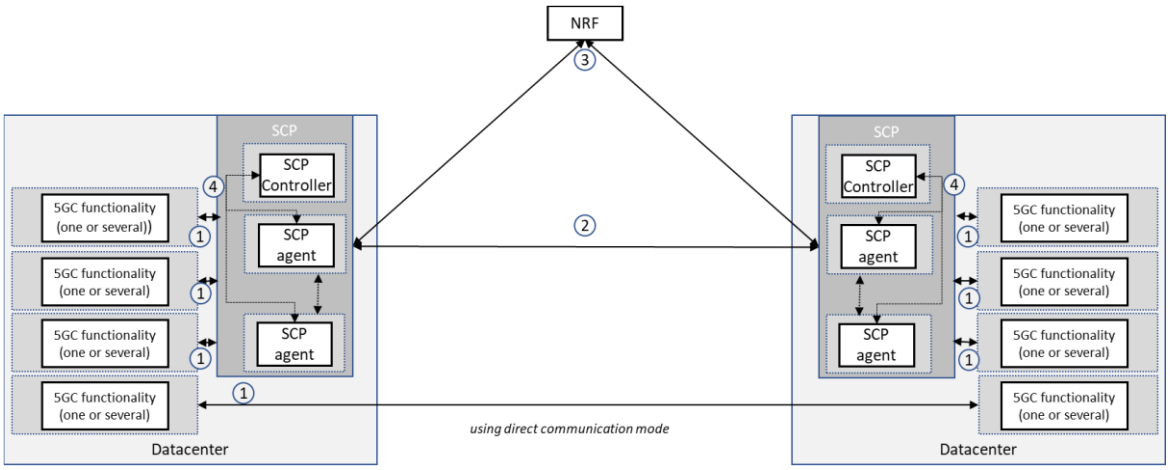


Figure 9: Overview of SCP deployment [5]

5.1.1.3 Name-based Routing

A third option within [5], again based on Service Communication Proxy (SCP), is Name-based Routing (NbR). It utilises a particular information-centric networking (ICN) flavour and operates straight on top of Layer 2 using path-based forwarding identifiers (Bloomfilters) [38]. Fully compatible with OpenFlow 1.3, this approach uses different information from upper layers to enable the internet (i.e. all TCP/IP-based communication), as illustrated in Figure 10. The figure illustrates a traditional IP stack on the left and the Name-based Routing stack on the right. NbR offers service routing as well as lightweight E2E (edge termination to edge terminal) flow and error control if desired. As can be seen on the right the stack has been "flattened" and should be read as follows: If non-TCP traffic arrives NbR uses the destination IP address as the information identifier to find the correct destination. If TCP traffic arrives (but not TLS and not HTTP) the service router (SR) that implements the traffic termination of NbR terminates the TCP session and uses the destination IP address to find the appropriate destination. If TLS or HTTP are present, the SR terminates the underlying transport session and uses the Full Qualified Domain Name (FQDN) to find the most appropriate service endpoint that could serve the request. The logic translating the IP world to ICN and vice versa at the edge is implemented in the SR component. It is well understood that the flattened stack looks controversial but the intention here is to highlight which protocols are used for the translation into ICN. Once translated the task of finding potential subscriber (logic implemented in SRs) to the information offered is realised in the Path Computation Element (PCE) which is a rendezvous (matching publishers and subscribers) and a topology manager (calculating the most suitable path between SRs).

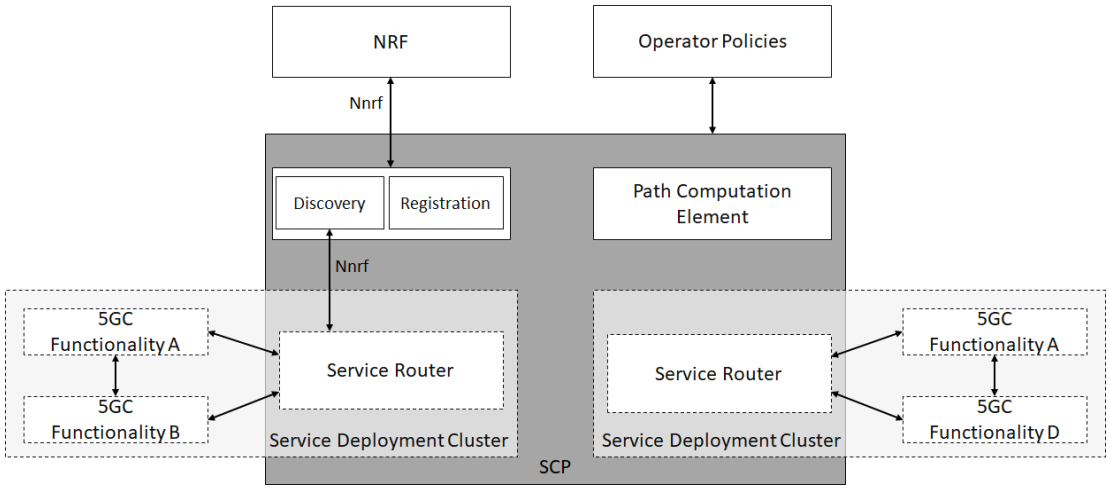


Figure 10: Service Communication Proxy implementing Name-based Routing

An eSBA deployment option for an SCP is depicted in Figure 10 and illustrates the deployment option NbR, as described in [5]. The other SCP deployment options are Service Mesh-based SCP and an Independent Deployment Unit-based SCP. All these deployments follow the principles of the communication model of indirect communication with delegated discovery, also described in [5]. The requesting client adds any necessary discovery and selection parameters required to find a suitable producer to the service request. It is based on the described eSBA platform principles to interconnect 5G control-plane services (or a subset of the respective services). The control-plane services are running as microservices on cloud/deployment units (service hosts for microservices). An SR is the communication node (access node/gateway) between the SCP and the services, e.g. 5G core, and resides as a single unit within a service deployment cluster. The SR acts as communication proxy and serves all services within the service host.

5.1.2 User Plane Outlook

Considering the 3GPP work on 5G user plane, e.g. for 5GLAN, the UE communicates with a private network. The destination network in 5G is referred as Data Network (DN). It hosts the application the UE wants to access. The user plane data flow, from the RAN (gNB), reaches the User Plane Function (UPF) in the CN, which eventually routes it to the DN.- Providing E2E Quality of Service (QoS) trough different NFs and protocols could be a challenge. Also considering the transport network of an ISP toward the Cloud, it is evident how user plane E2E QoS requires a special attention. The inherit complexity of a 3GPP user plane, which – in contrast to the internet – focuses on QoS enforcement to meet very strict KPIs, poses a significant challenge to be included in an open platform. An example of the complexity is some 3GPP specific protocols are not supported by SDN, e.g. OpenFlow does not support GTP-U.

Another impact to consider is a potential reduction of the performance, e.g. in term of latency. Since 3GPP Rel. 15 more than one UPFs can be chained via N9 tunnels for covering the needs of a PDU session. However, when such a chaining is introduced, the trade-off between the processing latency and low latency requirements must be considered. There is a trade-off between flexibility and latency. Many “low latency” deployments options prefer to integrate in the UPF some basic service chains functions (e.g. DPI, NAT/firewall) to address latency requirements. The so-called "crack-the-packet" once approach. Coming to performance, "cracking" each packet more than once and transferring all packets east-west via forwarders certainly will decrease throughput KPIs. Thus, distributed UPF deployments should receive further considerations within the standard body to address the challenges outlined above.

5.2 Platform Orchestration

Chapter 4 describes the cloudification of an infrastructure enabling an Infrastructure-as-a-Service (IaaS) offering utilizing SDN and NFV technologies. Not only has IaaS the ability to offer open APIs for the support of the ETSI MANO reference model, it also allows the realization of multi tenancy solutions. This demands resource isolation and resource configuration with exposed APIs. The API exposure itself must be coherent with the tenancy approach providing API isolation. In an Open Telco Platform specific attention is required for the orchestration of its basic components, the NFs, considering their different deployment options as VNFs, cVNFs or CNFs.

5.2.1 CNF Orchestration

With microservices, applications are evolving from monolithic applications, running e.g. on VM, to simpler independent components, running on containers. Some elements of the application are specific to the vendor for an application or for a set of applications, such as user interface. Others may become open microservices for other applications to use. An example is the state management in 5G Core, that can be implemented as a shared data layer open to applications that need to store states or context data. Other common service elements are relying on the infrastructure, such as security, logging, tracing, etc. These elements are typically features offered by a Platform as a Service (PaaS) or CaaS (Container as a Service) platform. The PaaS or CaaS support different Cloud environment, private or public, including Hyperscalers.

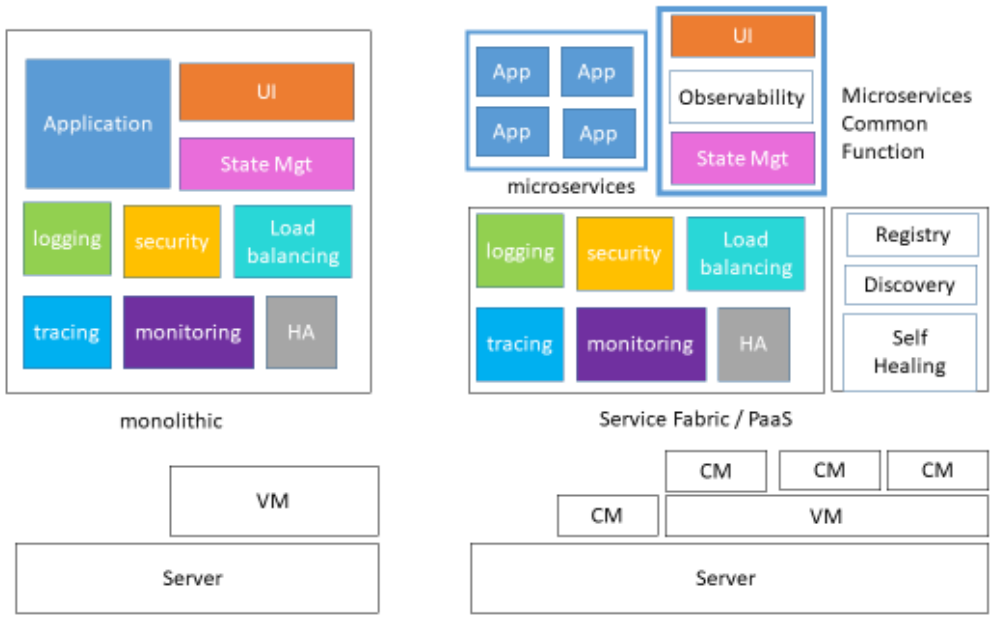


Figure 11: Microservice Orchestration via Services Fabric vs monolithic Virtualization

The orchestration of microservices is achieved with via the orchestration of these different components. Orchestration of the infrastructure common services, which is typically performed by the PaaS/CaaS. Orchestration of the VNFs/CNFs microservices, as described in Section 4.5.1, are orchestrated by the NFV MANO. Considering E2E, Service Orchestrator and the SDN Controller stiching network services, are other elements of the puzzle.

The management and orchestration interface with the NFVI is now with the PaaS/CaaS management embracing the NFVI classical resource management (virtual compute, network, storage) but also the NFVI PaaS/CaaS services such as load balancing, firewall, self healing, logging, etc.

When defining new microservices, metadata are being provided. NFV descriptors and packages provide list of resource requirements, configuration file artifacts, virtual links within a VNF or affinity rules. With microservices and PaaS, PaaS metadata it can also be provided more specific instance dependencies e.g. requirements for a load balancer or scaling triggering conditions. With this approach lots of the lifecycle operations on the microservices can be automated and provided by the PaaS/CaaS management platform.

With microservices architecture, new services are composed by connecting microservices endpoints. Microservices are onboarded on the NFV MANO and registered on the PaaS/CaaS, with the appropriate metadata. Network slices and underlying network services will be designed to deliver some Communication Service to the customer. Intent based interfaces between the Service Orchestrator, the NFV Orchestrator and the PaaS will allow the selection of most appropriate resource instances to meet the requirements.

Typically the PaaS/CaaS offers discovery services to identify available microservice instances to be used. If microservice underperforms, the PaaS/CaaS is able to scale that microservice or activate a new instance and reroute the traffic to it.

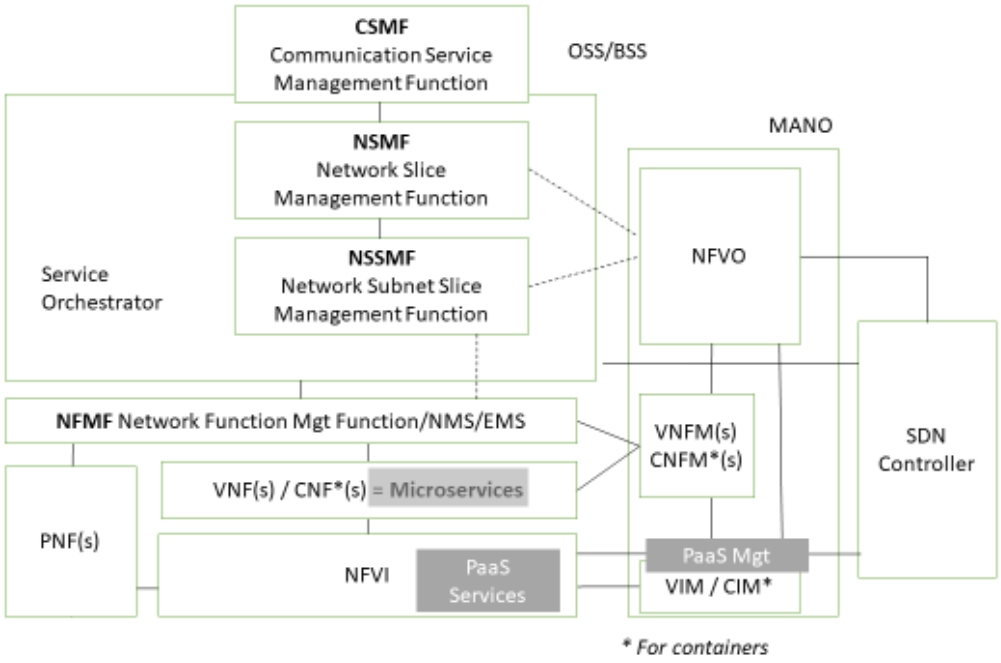


Figure 12: NFV MANO Microservice onboarding via PaaS

5.2.2 VNF Orchestration

As described in chapter 4, not all softwarised functions can see the full transformation to a CNF by applying the 12-factor app methodology. Functions that implement routing and switching functionalities are facing software engineering challenges to be truly converted into Cloud Native. Thus, they are referred to as cloudified VNFs (cVNFs).

Considering a VNF offering routing and switching services, the requirement on where the VNF is deployed is of paramount importance. We are indeed considering scenarios with computing resource are available across various locations in the Telco cloud. Policies should enable location-aware rules that respects security and failover goals.

5.2.3 Tenant Models in Telco Oriented Virtual Infrastructure Managers

In Open architecture, with open APIs enabling programmability and multi tenancy, compute, networking and storage resources will be sliced. Slicing is also about resource isolation and QoS enforcement (where applicable mainly networking and compute) allowing infrastructure providers to assure service-level agreements. Tenants can work freely within the constraints attached to the acquired slice. However, when aiming for the operation of VNFs in a cloudified infrastructure, a requirement for the tenant is to freely program the resources within the given slice. This requirement poses a critical challenge as many resources are currently managed by admins in a rather manual fashion and are usually not programmable by the tenant. For example (c)VNF flavors or locations where it should be initiated. A possible requirement for the tenant is to have some management control over the slice and the VNF of a slice: location placement of the VNF, configuration of the VNF, etc.

5.3 Vertical Application Orchestration

The orchestration of vertical applications deployed in a Telco Edge, poses the challenge of integrating it with a 3GPP system. The establishment of a session from a UE to a vertical application is indeed always handled by a 3GPP Core. The CN takes cares of the configuration of the user plane between the UE and the vertical application located in a Data Network (DN). This is enabled by the framework specified in 3GPP allowing an Application Function (AF) to communicate with a 5G Core registering the application including its services. ETSI MEC [39] and 3GPP [40] are defining specifications and APIs for Telco edge application management and orchestration and federation mechanisms for these APIs. The scenario foresees application providers to deploy and to orchestrate applications seamlessly across multiple operators and offering to end users roaming on Telco edge applications. This scenario is also covered by GSMA [3].

Vertical service providers (e.g. OTT service providers) expect APIs and on-boarding procedures that are coherent with how their services are designed, implemented and orchestrated in the Cloud service providers. It is of paramount importance to offer orchestration and lifecycle management APIs to the Verticals allowing them to programmatically utilize the Telco Edge in a similar fashion as public Clouds do. Current approaches have followed a rather strict client – server principle, with the UE always being the client. APKs built into the UE's application usually interact with server components that manage and control the application instances in the DNs. Current OTT realization of most applications available to UEs (smartphones, tablets, laptop/desktops, TVs, wearables) have a very minimal (not to say zero) interaction with a 3GPP system to locate a specific server-side instance in the Cloud (as promoted by 3GPP SA6 for instance). Ideally, the orchestration of a vertical application into the Telco Cloud (into a local DN) should follow similar principles, as the orchestration into one of the big cloud providers. Ultimately a convergence of Telco Edge API and public cloud API could simplify management and orchestration of vertical applications across those hybrid clouds.

5.4 Intelligence

With the proliferation of resource-constrained mobile devices with diverse emerging computation-intensive broadband applications, including virtual reality, augmented reality, and online gaming, MEC is promising solution to revolutionize existing computing infrastructure. MEC is typically bridging the capacity of cloud and requirement of device by pushing the computation/storage resources to the proximity of the user equipment, thus deploying application functions on MEC locations shared with RAN virtual functions. How to orchestrate wireless radio resources between MEC and traditional mobile services requires a careful design and adds another dimension of complexity to the network management. In a software-defined RAN, the SDN-orchestrator, providing infrastructure flexibility as well as service-oriented customization, handles all control plane decisions. Software-defined RAN is also beneficial for network sharing, which has been studied by 3rd Generation Partnership Project (3GPP) Technical Specification Group (TSG) Service and System Aspects (SA) 1 - Services in [16]. Based on such a study, the sharing paradigm introduced by 3GPP TSG SA 5 - Telecom Management considers that an infrastructure provider (InP), which is referred to as a master operator, is responsible for the configuration of a shared physical network [17]. As

such, the same network infrastructure is able to host, on a RAN-as-a-service basis, multiple service providers (SPs), which is also known as multiple tenants.

The Telco Platform is open as an infrastructure to deploy third party Application. For example, an over-the-top (OTT) application provider (e.g., Netflix and Google) can become a SP so as to lease wireless radio resources from the InP to improve the Quality-of-Service (QoS) and the Quality-of-Experience (QoE) for its subscribed users. If the Telco is providing an Edge DC for a Developer to deploy an App on and it also provides the Telco components to have the right network performance according to an SLA (e.g. Edge routing toward the application via UPF), it is important for the OSS to guarantee that SLA.

With an open and scalable infrastructure management, intelligent edge-cloud resource orchestration platforms with data collection, distributed processing, and artificial intelligence capabilities are urgently needed to enable flexible network automation and network augmentation, and support Application deployment with guaranteed QoS at the Edge. To do so, the intelligent mechanisms that efficiently exploit the decoupling of control plane and data plane under a software-defined architecture must be developed to achieve optimized radio resource utilization across logically independent RAN slices.

To adapt to the QoS/QoE requirements and fluctuated wireless channels of mobile users, the AI-enabled edge-cloud resource orchestration will need to intelligently learn the network dynamics originating from the mobilities as well as the random computation task and data packet arrivals of users. AI can be important enablers to open the Telco Platform to the developers. The most important AI algorithms in networking include (Deep) Reinforcement Learning for diverse network configurations, CNNs for input dimension reduction, RNNs (inc LSTMs) for temporal correlated data prediction, and Artificial Intelligence Programming (e.g. algorithmic synthesis, DeepCoder). In research, they are used in a supervised, unsupervised & reinforcement fashion to predict/control various parts of the RAN/CN/TN. In industry, they are used to populate standardised architectures, such as NWDAF in 3GPP SA2. To allow for assurance, this AI-enabled edge-cloud resource orchestration is required to be designed with quick training efficiency and response time for service delivery with guaranteed QoS for Edge applications.

6 HYBRID CLOUD

6.1 Definition of Hybrid

Hybrid is a key word in the cloudification process of the Telco Platform and it covers different aspects. One aspect is the coexistence of VM based and Containers based NFs deployments. Another aspect is the coexistence of Telco oriented and Service oriented Cloud Native integrated environments. The coexistence and integration of Centralized, Edge and Cloud based deployments forms another aspect that fits in the Hybrid scenario.

Considering the internal operation of the Telco Platform, the cloudification process is proceeding at different speeds according to technology maturity, needed performance or operational aspects, among others. This requires the coexistence of different technologies for NFs virtualization. The deployment of an E2E slice, e.g. in a 5G NSA flavor, may require the management of a hybrid platform (leveraging on VMs and containers) that must be integrated in the whole orchestration process. Hybrid resource managers are involved in the process supporting the different virtualization technologies.

Viewing the Telco Platform as an open environment to offer a complete communication service leveraging on both advanced network features and broad service integration, includes hybrid environment as a key component, especially at the Edge. This hybrid scenario foresees autonomous and integrated edge nodes for Telco capabilities and applications exploitation. Leveraging on a common technological approach toward cloudification, Telco and service nodes can be deeply integrated and both Network Functions (NFs) and Application Functions (AFs) can coexist on the platform taking advantage of the Edge characteristics.

Hybrid also means coexistence of different deployment options bringing together the peculiarity and opportunities of Edge, Centralised and Cloud environments all integrated in a single workflow and operated seamlessly because of the Telco cloudification. The 5G evolution in terms of NFs and Management Services, based on virtualisation and

SBA-based APIs leads to the opportunity to shape the network and its management, given technology as well as operational and business-oriented drivers. A Hybrid approach can be adopted because of the coherence of the underlying technologies so the choice can be based on a strategy that goes beyond technological boundaries.

6.2 Challenges for a Unified 5G Hybrid Cloud

When designing the architecture for a communication system, it is of paramount importance to avert any technology-specific details in its description allowing a wide applicability without constraints beyond component and interface definitions. The work in 3GPP on Service-based Architecture is one of these examples which defines the components and interfaces for the 5G core's control plane. Furthermore, it describes components such as the Service Communication Proxy that enables service routing capabilities among 5GC network functions. The same applies to the ETSI MANO reference model which provides rich set of components and interfaces. On the other hand, there is a rich and often crowded landscape of technologies defined by various architecture groups. And as a result, the realization of an architectural model via a multi-vendor solution is very often prone to interface incompatibilities and reveals the complexity of feature-rich interfaces.

The interoperability of hybrid cloud solutions is no exception to this statement, with an even more isolated with lack of openness and the ability to mix controller and compute nodes e.g. from different ETSI MANO-based realizations. Certainly, this playing field has less stringent standardization specifications compared to 3GPP, but it perfectly demonstrates how technology is being realized and how interoperability and openness becomes a problem of the tenant/user aiming to orchestrate a Telco platform across multiple cloud solutions. There are apparent differences such as virtual machine vs containers with their string of dependencies for the underlying technology enabling an automated and programmable environment. This poses a steep burden on the cloud providers to find the right balance of openness and interoperability and key differentiators. At the end it is a business decision whether to open up a technology platform.

Indeed, the telecommunication sector is rather agnostic to the actual cloud solution that enables a virtualized infrastructure to orchestrate cloudified Telco components. Furthermore, most environments are private cloud deployments with a stronger focus on virtualizing networking functionality due to the nature of an operator compared to the typical vertical service provider accessing a cloud. However, given the significant number of services and their instances, the usage of hybrid clouds even for private deployments becomes important for Telco players. The services and applications which should be hosted in the (private) cloud ranges from system analytics, firewalling, user management/policy control, to networking (routing/switching) and resource scheduling. The required SLA for each category varies as their actual requirements and/or support for certain drivers or toolkits become important criteria to decide on a specific cloud solution for a Telco player.

Most of the interactions between a Telco entity and the cloud will utilize the NFVO (when using the ETSI MANO reference model as the baseline) which already demonstrates the fragmentation in the market when it comes to a homogenous approach of this interface. In other words, when switching from cloud solution A to solution B, the interfaces on how to orchestrate and manage a set of virtual instances are incompatible in description language and feature set. Additionally, when adding a new compute node to an existing cloud in a new location, the set of abstractions from a cloud solution A for enabling a controller to manage the compute node, is specific to a cloud solution and does not allow any hybrid scenario.

In summary, enabling hybrid cloud scenarios can currently be achieved through an aggregator that offers the ability to access specific clouds via their NFVO, including an abstraction and translation of the cloud solution specific semantics. Mixing cloud solutions by means of swapping NFVOs, VIMs or VNFMs is rather impossible and only allows hybrid cloud scenarios by the aforementioned aggregator approach. This applies to cloud solutions for VNF and CNFs.

6.3 Hybrid Cloud at the Edge

Edge data centres are key assets for Telcos and an important component of the Open Telco Platform. Hybrid Cloud at the edge represents a concrete opportunity for the Telco to be part of an ecosystem where developers and HCP collaborate to enrich Communication Services leveraging on 5G distinctive features. They include such features as low-latency radio optimization and Core Network local breakout toward Edge Data Networks.

The Hybrid Cloud architecture at the Edge can be logically composed by two interworking Cloud Native environments, a Telco Edge Node (TEN) and a Service Edge Node (SEN). The TEN hosting Network Functions and the SEN hosting cloud-based applications (a possible implementation of the 3GPP AFs) are potentially provided by different players.

This picture can be realized in different ways with different partnership models among Telcos and HCPs and must adopt technological openness and business level collaboration and agreements. A tight collaboration is indeed the basis to guarantee a wide adoption of the solution.

There are many SDOs working on this topic and supporting this vision. GSMA OPG, [3], is defining a common architecture for edge enablement with a specific focus on interworking. OPG defines four main actors involved in the process, Application Providers, Federated Operators, Network Resources and User Equipment. It defines a set of interworking APIs among the actors.

3GPP SA6 is also working on an edge enablement architecture with a specific focus on application deployment and discovery [6].

CNTT is considered by GSMA OPG as the base architecture for the virtualization layer on top of which the network resources and the applications are deployed.

ETSI MEC was the first group working on Edge computing and it is currently enhancing its architecture to support MEC interworking in a Hybrid environment.

The federation among the Telcos, promoted by GMA, is a key enabler for a proficient Edge Hybrid Cloud ecosystem. developers and Verticals need a “public Cloud” like approach to deploy their application seamlessly in different geographical areas leveraging on service and Telco Edge capabilities of different providers. The goal is to place solutions and applications near the end users independently on the underlying serving MNO. The MNOs differentiate themselves providing the most efficient and performing TEN, deeply integrated with the SEN providing the Data Network and the applications. GSMA underlines, with the Telco federation concept, the importance of Telcos level interworking to enable multi regional end-to-end service delivery. This implies defining a common way of enabling actors to interact with each other.

TEN and SEN integration and deployment can be approached in different ways with different partnership models and technological solution but in any scenario the underling technological enabler is the cloudification of the infrastructures.

Telco and Service Edge Nodes are deployed at the edge of the Operator’s network jointly or separately in the data centres at the edge. Interaction between Service Edge Node and Telco Edge Node is needed at networking level to ensure the routing of the LBO traffic. Telco API exposure is required to ensure LBO and mobility coordination.

The integration between SEN and TEN foresees different coordinated actions that shall be exposed, to external systems, via an Intent Based API to facilitate the usage of the system. The request for application deployment at the Edge, by a developer, shall be expressed using high level requirements, e.g. the latency requested in a geographical area.

Considering the Open Telco Platform evolution and the Intent Based API approach, as suggested by 3GPP SA5 [7], a Service Layer is foreseen. It acquires the Communication Service requirements and supports applications deployment end enablement in the Edge Hybrid Cloud environment.

TEN and SEN can be both provided by the Telco or they can be provided in partnership with HCPs, this leads to different collaboration models that are further discussed in the following chapter.

7 CLOUDIFIED RAN

A vRAN already describes a concept of decoupling the infrastructure running the RAN and the SW-components that provide the RAN functions (DU, CU) and protocol stacks for the transport interfaces. This enables the services to run in VNFs on a commodity HW that forms the base station.

A RAN where software functions are disaggregated from the underlying infrastructure (as described in vRAN) is a prerequisite for virtualisation. But in addition to virtualization it is required that the network functions are delivered as micro services which can scale independently and be orchestrated by cloud management technologies.

A cloudified vRAN means that cloud technologies and hyper-scale models known from web-services are applied to the RAN. This scale is seen as a requirement to enable the full potential of 5G.

A Cloud Native Radio Access Network enables RAN network functions delivered as microservices in containers over bare metal servers, supported by dynamic orchestration such as e.g. Kubernetes. Software application life cycle management relies on DevOps principles and Continuous Integration and Continuous Delivery (CI/CD).

For the Functional Disaggregation of virtualized RAN please refer to [8] and to [9] on network architecture, transport options and dimensioning for more detailed RAN background.

Openness describes a design of the RAN functions based on interoperable interfaces that are based on open and industry standards without any proprietary code.

In this chapter we would like to have a more detailed look at RAN specifically under afore mentioned aspects of Cloud Native.

7.1 NG-RAN overview

7.1.1 3GPP NG-RAN

3GPP has defined the architecture of the 5G next generation RAN, NG-RAN [15] with a reference architecture as described in Fig 13 below with 2 keys components:

- a gNB, providing 5G NR user plane and control plane protocol terminations towards the UE
- an ng-eNB, providing enhanced 4G E-UTRA user plane and control plane protocol terminations towards the UE

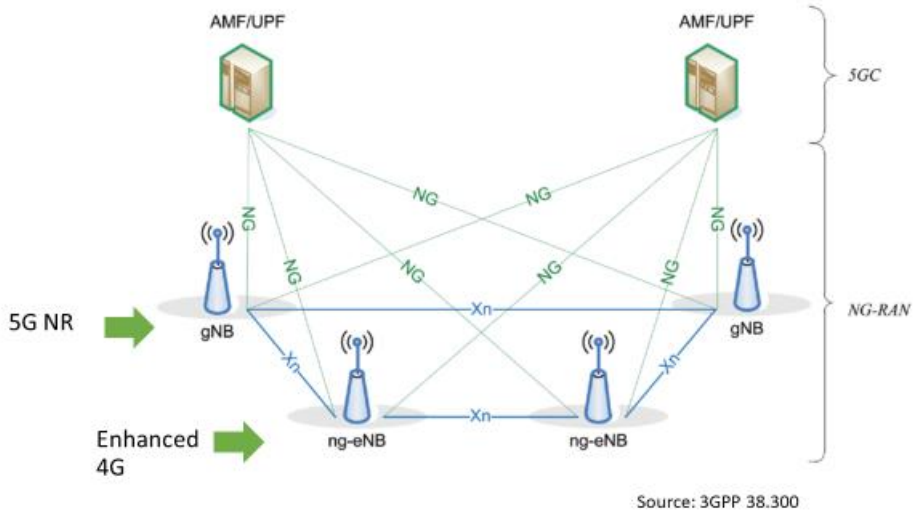


Figure 13: 3GPP NG-RAN architecture

This NG-RAN connects to the 5G Core (5GC) via an NG interface for control plane to the AMF and user plane to the UPF. The principle being that the device can either connect using 5G NR and connect directly to a gNB or use 4G radio and connect to an ng-eNB. Then different deployment model can be supported, with communication between the different radio nodes via Xn to support secondary nodes and handover mechanism.

DU and CU functional split:

3GPP has also defined a functional split [13] inside the gNB with 2 components: the DU (Distributed Unit) and the CU (Centralized Unit), communicating via a standard interface F1, as described in Fig x2 below:

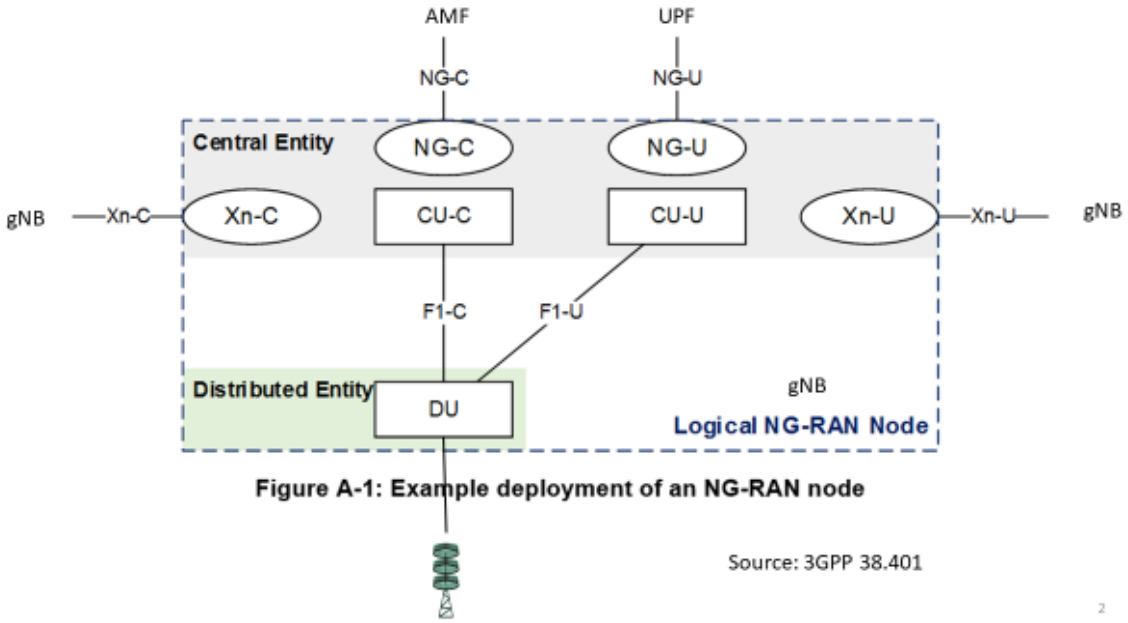


Figure A-1: Example deployment of an NG-RAN node

Figure 14: 3GPP DU CU functional split

The CU can also be split in 2 entities: a CU-C for control plane, and a CU-U for user plane.

This architecture allows for the RAN to be more and more virtualized and a number of functions to run in the cloud, either close to the antenna on edge location if low latency is being required, or further down in more centralized data centre as explained in fig x3 [14] with different split options between central unit (CU) and distributed unit (DU).

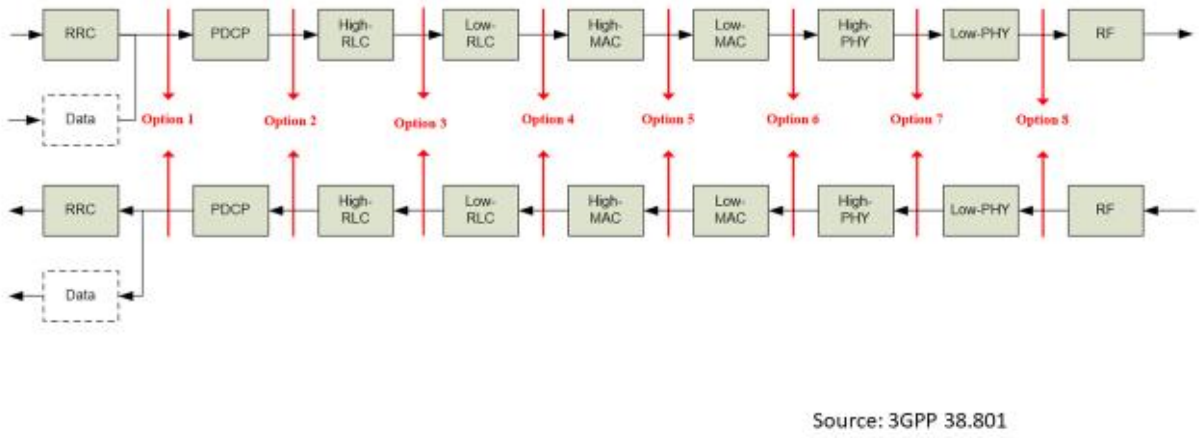


Figure 15: RAN Split Option

Different split options have merits and drawbacks depending on the different operators deployments. 3GPP could not converge on a single split option. However, O-RAN selected option 7 while Small Cell Forum selected option 6.

RAN slicing:

3GPP has defined network slicing with end to end network slicing spanning across devices, access, transport and core networks. A UE is associated with an NSSAI and connects to a slice, up to 8 slices per UE. A given network can support many slices. The NG-RAN shall support slicing [14], multiple slices and isolation, leveraging RRM policies but the way slicing is performed in the NG-RAN, which functions are selecting slice resources in the NG-RAN or which resource are restricted to a given slice, are implementation specific, which may be a challenge for designing multi-vendor Cloud Native environment if no further standard is defined.

Depending on the split option, the number of functions to be performed in a given entity, the traffic load, the latency requirements or the slicing and isolation requirements, the design of the virtualized or Cloud Native infrastructure for these disagregated vRAN network functions will be different.

Management of the NG-RAN:

Following the decomposition of the NG-RAN as described earlier, 3GPP SA5 has defined an information model for 5G resources [12] including the NG-RAN resources such as DU, CU-C, CU-U and respective interfaces as well as different combinations and relationships depending on the split option. These resources are objects, network functions, also called vNF, virtual network functions, being managed following the architecture and interface specifications of ETSI NFV and NFV MANO (Management and Orchestration) for lifecycle management (LCM), fault management (FM) and performance management (PM).

3GPP has defined a hierarchy with E2E network slice management (NSMF), subnet slice management (NSSMF), and network function management (NFMF), as detailed in Fig x4. NG-RAN Network functions have 2 management interfaces: the NFMF interface for NF application provisioning (LCM, CM, FM, PM) and the virtualization management interface, [11] for Ve-Vnfm-em and Ve-Vnfm-vnf reference points with the NFV MANO.

Depending on the deployment model of this NFV MANO and NFMF, centralized or distributed for instance, but also the number of entities to manage, the isolation of management functions, or special requirements in terms of location and security aspects, or latency, the design of the Cloud Native communication infrastructure will be different.

7.1.2 O-RAN ALLIANCE

In this section we focus on 5G RAN architecture as specified by O-RAN ALLANCE [20]. We first provide a high-level description of O-RAN network architecture; then, we describe why interfaces standardization is essential to achieve interoperability, and to promote dynamic RAN ecosystem. We also summarize Cloud deployment scenarios, and considerations for improved virtualized network functions openness.

The O-RAN ALLIANCE (O-RAN) was founded in 2018 to provide specifications and requirements for disaggregation, virtualization, open and intelligent RAN. Virtualization decouples software and hardware RAN functionalities, enabling the RAN to be built on a general-purpose processor platform to reduce manufacturing costs. RAN component disaggregation enables Telco to select network components individually. RAN openness requires standardized interfaces to enable multi-vendor deployments and increase network agility. Improving network intelligence is a must for Telcos to be able to handle increasingly complex RAN deployments.

The overall O-RAN architecture is illustrated in Fig.16. The Service Management and Orchestration (SMO) framework oversees the management of all RAN network functions. This includes near-Realtime RAN Intelligent Controller (near-RT RIC), central unit, distributed unit, radio unit as well as the Cloud infrastructure. The management is done over O-RAN defined interfaces named O1 and O2. The O-RAN Alliance also provides profile specifications for 3GPP-defined interfaces, such as E1, F1, X2, and Xn, with the aim to ensure interoperability in multi-vendor environments. Concerning the Cloud platform, it includes networking, storage, and compute resources that are ready to host RAN network functions. It also provides the required tools to manage Virtual Network Functions (VNF) initial deployment, reconfiguration, and lifecycle management.

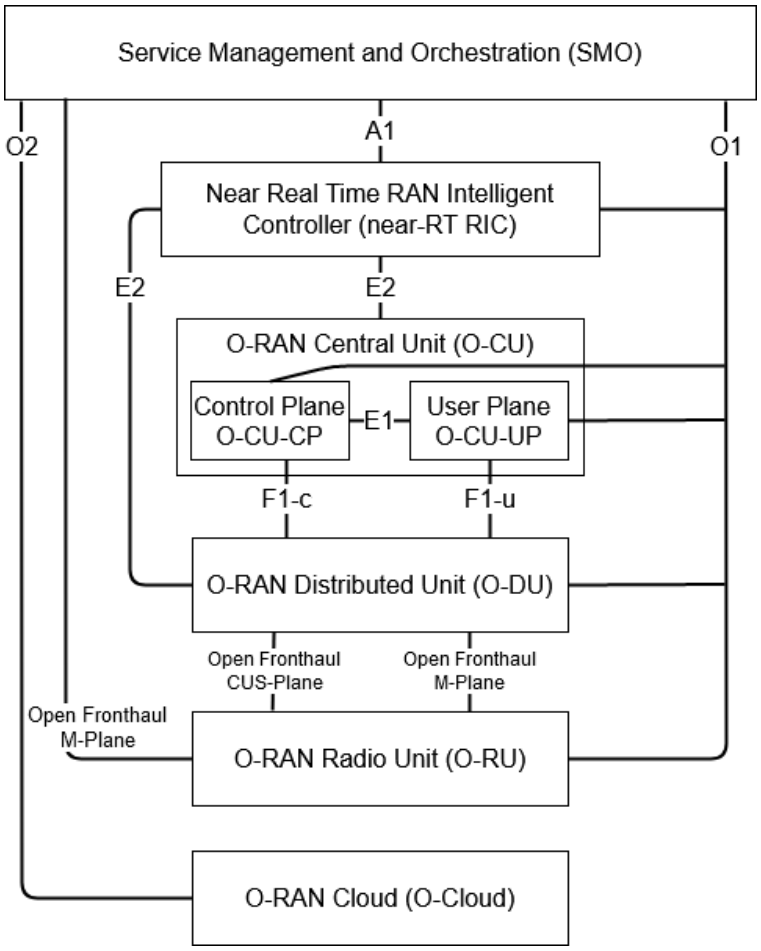


Figure 16: Logical architecture of O-RAN

One major objective for O-RAN is to leverage dynamic multi-vendor ecosystem for the RAN. Network agility and interoperability require having standardized interfaces, which is the reason why O-RAN Alliance focuses on RAN architecture disaggregation and provides interfaces profile specifications that are consistent with 3GPP specifications [21, 22].

In addition to standardized open interfaces, Cloud-related constraints need to be satisfied to ensure open and interoperable RAN. For instance, data centre servers hosting RAN components shall have their chassis open for blades from multiple vendors. Moreover, the virtualized network functions shall be able to run correctly on servers provided by different vendors.

Decoupling Software components from the underlying Hardware resources is an enabler for flexible RAN deployment and cost reduction. When deploying Virtual Network Functions (VNFs) on commodity hardware, three layers need to be considered: Hardware, virtualization, and VNF layers. Given that VNFs should be able to run on commodity hardware, specifying hardware requirements is a must to simplify VNF deployment and maintenance, and to allow for automating VNF deployment and management. To learn more about Cloud deployment scenarios specified by O-RAN, refer to [23].

Relationship between 3GPP and O-RAN

O-RAN specifications are built based on the 3GPP specifications by defining interface profiles, additional new open interfaces, and new nodes, in three RAN areas: disaggregation, automation, and virtualization.

One of the key new interfaces standardized by O-RAN is open interface of fronthaul, connection between RU and DU. In addition, new open interfaces and functions introduced by O-RAN are not covered by 3GPP SA3 security standards.

To learn more about commonalities and difference of functions, interfaces, and features between O-RAN and 3GPP, please refer to [24].

7.1.3 Open Source

Open Source initiatives has proven to be a key pillar of accelerating the evolution of telecommunications infrastructure into the world of open, flexible, disaggregated technology. From Core NFV programs to open OSS/BSS and IT transformations many initiatives are active.

Open Source communities and other industries are providing foundation for radio access network disaggregation and openness. For this reason, it is expected that Open Source initiatives will be key enablers in accelerating actual implementation of Open RAN solutions.

Many functional domains of cloudified Open RAN benefit from embracing Open Source approach. Currently there already valuable examples of Open Source initiatives addressing them.

Open Source projects around cloud layer has proven tremendous value and demonstrate high degree of maturity in NFV world. Examples of such Open Source projects are hosted by Open Infrastructure Foundation [25] Linux Foundation [26], infrastructure projects of Open Networking Foundation [27] and more. Open RAN standards and deployment architectures bring new requirements to the well-established world of network functions virtualization. These requirements are coming from the specific nature of Open RAN workload (e.g. real time baseband processing of O-DU function) and from new Open RAN deployment models (e.g. deployment of limited compute footprint at remote far edge site). While many existing deliverables of Open Source projects from NFV world are applicable for Open RAN, some of the Open Source initiatives are targeting to address specific new requirements to O-Cloud for Open RAN:

- StarlingX under Open Infrastructure Foundation (https://www.starlingx.io/) creates cloud infrastructure software stack for the edge, specifically addressing needs of Open RAN workload (incorporating real-time low latency OS stack which enables O-DU deployment) as well as Open RAN deployment model (optimizations for remote far edge deployment on limited compute footprint)
- BBdev extension of DPDK specifically addresses a need for open standard API for Open RAN O-DU software acceleration, enabling full decoupling between O-DU software layer and underlying general purpose compute hardware including accelerators

Considering DU and CU split, while various Open Source base station implementations for 2G, 3G and 4G have been available on the market for some time, they typically lack wide adoption and established developer community as well as industry standards to back up the code. O-RAN Software Community [28] under Linux Foundation targets to address the need for a standard-based O-DU and O-CU implementation while driving wider adoption of Open Source O-DU and O-CU project among Open Source community.

Producing an Open Source implementation of near real time RIC has been targeted by SD-RAN project under Open Networking Foundation [27]. Another Open Source initiative which targets Non-RT RIC as well as nRT RIC is O-RAN Software Community [28].

Management and Orchestration is another area where opens source communities are very active and relevant. One of the existing Open Source orchestration projects specifically targeting Open RAN service management and Orchestration is ONAP under Linux Foundation [18]. Alignment of ONAP functionality with O-RAN architecture is being developed around compliance with O-RAN SMOF (SMOF - Service Management and Orchestration Function) the O-RAN group addressing the management and orchestration of Open RAN solutions via standard interfaces (e.g. A1) as well as incorporating O-RAN SMOF functionalities such as non-RT RIC.

Accelerating deployments of cloudified Open RAN requires growing maturity and adoption of above Open Source projects, as well as Open Source implementation for Open RAN domains not addressed by the initiatives above.

7.2 Design requirements

7.2.1 Cloudification & Openness requirements

Given that RAN is the most transaction-intensive and time-sensitive part of the network, there is little tolerance for any performance slack to impact user experience. To this end, an open, cloudification RAN needs to meet the following requirements:

- Virtualization - Decoupling software and hardware RAN functionality enables the RAN to be built on a general-purpose processor platform to reduce manufacturing costs. Virtualized RAN functions deliver deployment flexibility and efficiencies. It also allows E2E network slice, tailored to the specific service and QoS requirements
- Component Disaggregation – RAN components need to be decoupled to allow operator to select network components individually. It also enables great flexibility to place network element whenever is appropriate for particular network deployment scenarios
- Open RAN architecture - Interfaces between network components should be open and standards-based, and there should be as well as open interoperability on the GPP-based baseband processing platform, radio hardware, and software
- Decomposing a RAN service into microservices - with microservice structure, an application is composed of microservices, each of these microservices is built individually and deployed separately, leading to isolation and resilience. Failure of one microservice is isolated and will not impact to other services. There could be an option to use another service and the application will continue to run independently. Therefore, it is easy to resolve performance issue and scale up/down
- Containerization – Network functions are virtualized, one or more isolated containers are used, orchestration dynamically supported by Kubernetes
- DevOps and CI/CD: Application and service development and delivery attend to DevOps practices, integration, upgrade of software and introduction of new features/functionality following CI/CD framework
- Reliability and availability – Should be at 99.999%, on a par with traditional RAN system
- Security Measures – Security measures should be developed from design and by vendors to protect the trust network components and interfaces introduced by decoupling of functions and share common understanding and implementation of security requirements
- Vendor-neutral - All software should be interchangeable and able to operate on any vendor hardware
- Support all 3GPP-compliant and O-RAN RAN split options – There should be ability to use a different protocol stack distribution flexibly between CU and DUs, depending on use case, fronthaul availability, and KPI requirements
- Open interoperability validation - Standardized set of tests should be in place for the certification of software, regardless whether it is hosted on cloud and/or bare-metal infrastructure
- Responsibility - Since RAN components are from different vendors, the responsibility among different vendors needs to be defined clearly for deployment. For instance, vendors will not only test their own products alone, but also shall test their products under an overall CI/CD umbrella
- Organization – Network deployment team needs to develop knowledge and skill set to integrate specialized processors like FPGAs, ASICs, and application acceleration techniques with GPP to achieve optimal RAN performance

7.2.2 Management and Orchestration

In this chapter an overview of the 3GPP management system is considered to understand how RAN is managed as a part of a Network Slice and how it is modeled. The 3GPP service-based architecture is briefly described to understand how APIs are defined. A possible integration between the 3GPP management system and the MANO NFV system is provided. This are the basic notions to understand a 5G management system that O-RAN complements.

7.2.2.1 Introduction

The 5G E2E Network Management and Orchestration system for mobile networks is well defined with a set of open APIs, resource models and services. Network Slice Instance lifecycle managemnt is defined condidering different phases that range from Preparation (including design), Commissioning, Operation (including run-time Activation, Supervision and De-activation), and Decommissioning as defined in 3GPP [11].

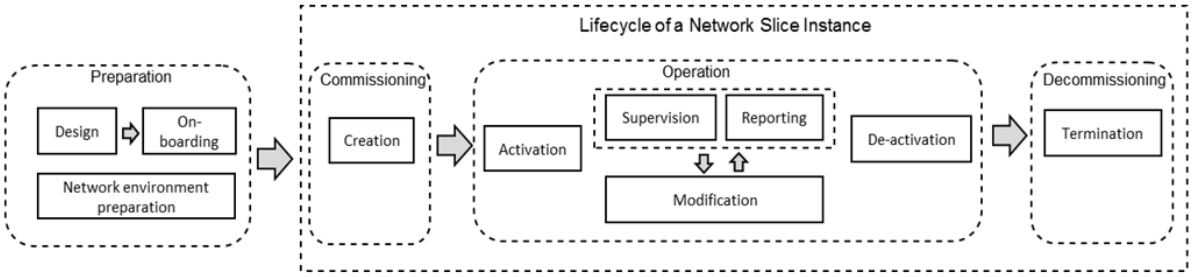


Figure 17: management aspects of network slicing

Because of the service based approach of the 3GPP management system, different deployment scenarios are possibile. An example by 3GPP [10] is represented in Fig xx. showing a consumer which may be any trusted party in the OSS domain, interaction with the E2E network management system.

And as part of the system is the E2E umbrella , ie Network slice management, interfacing southbound with the different subnetworks including the one in the RAN domain.

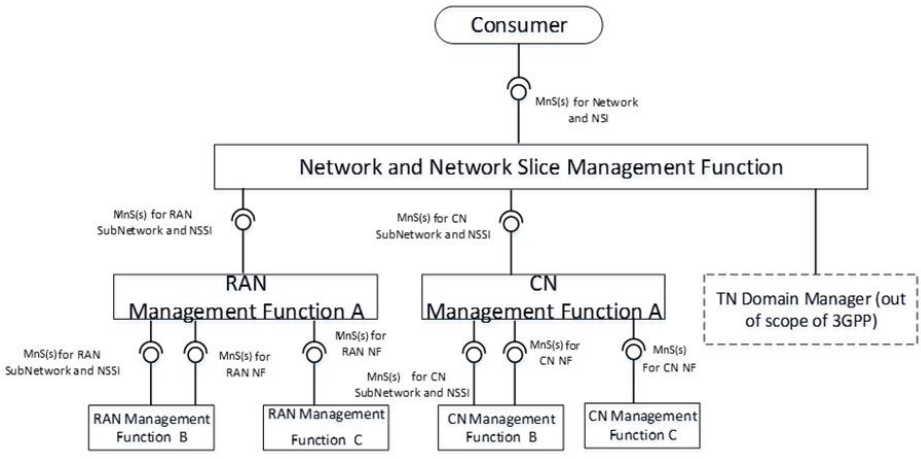


Figure 18: Example of deployment scenario for 3GPP management system

This implies that each of the different processes mentioned earlier cascade down and apply across the different network domains including RAN subnetwork.

To enable these processes to apply consistently, a common information model needs to be implemented across the different layers, different network domains and also across Telcos. RAN subnetwork indeed could be in a different management domain inside a Telco network or it could even be provided by another Telco according to the Slice as a service model.

3GPP has defined a Network Resource Model [12] for NR and NG-RAN in XML, Json and YANG, to model each RAN NF according to its configuration parameters. The model for gNB and en-gNB foresees for 3 scenarios:

- Non-split NG-RAN deployment scenario, represents the gNB defined in [13]. In this scenario, a gNB is represented by a combination of a GNBCUCPFunction, one or more GNBCUUPFunctions and one or more GNBDUFunctions.
- 2-split NG-RAN deployment scenario, represents the gNB consist of gNB-CU and gNB-DU defined in [13]. In this scenario, a gNB-CU is represented by a combination of a GNBCUCPFunction and one or more GNBCUUPFunctions, whereas a gNB-DU is represented by a GNBDUFunction.
- 3-split NG-RAN deployment scenario, represents the gNB consist of gNB-CU-CP, gNB-CU-UP and gNB-DU defined in [13]. In this scenario, a gNB-CU-CP is represented by a GNBCUCPFunction, a gNB-CU-UP is represented by a GNBCUUPFunction, and a gNB-DU is represented by a GNBDUFunction.

7.2.2.2 Service-Based Management Architecture

Service-based management architecture is specified by 3GPP in order to enable management and orchestration, of 3GPP networks. In fact, components management is achieved over a standardized service interface composed of individually specified Management Service (MnS) components. A MnS is provided by a MnS producer and can be consumed by one or multiple MnS consumer(s) [10].

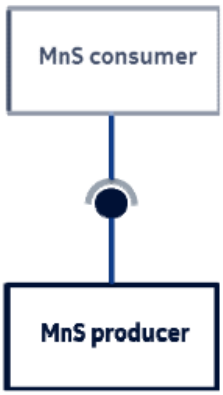


Figure 19: MnS producer and MnS consumer [10]

A concrete MnS component is composed of two or more independent components. Three different component types are defined; they are described in the following table.

Table 4: Management Service component types

Component type	Description	Example
MnS component type A	Group of management operations and/or notifications that is agnostic with regards to the entities managed. These operations and notifications are not involving any information related to the managed network (network agnostic)	Creating, reading, updating, and deleting managed object instances
MnS component type B	Management information represented by information models representing the managed entities	Network resource models [16]
MnS component type C	Performance information and fault information of the manged entity	Alarm information, performance data [17]

Therefore, a MnS is composed by a MnS component type A and a MnS component type B, or a MnS component type B and a MnS component type C. Figure 20 illustrates an example of MnS instances with various MnS component types.

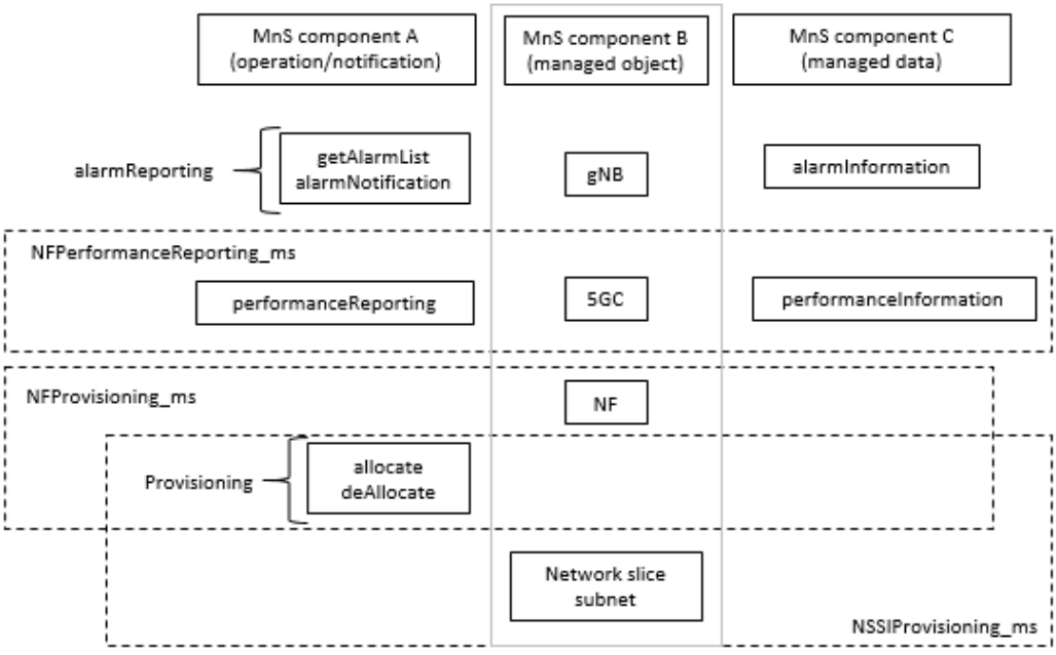


Figure 20: MnS and MnS component types [10]

In order to enable MnS instances to be discovered by MnS consumer, the MnS needs to be discoverable to the operator's management system when the MnS instance is operative. This is achieved using MnS discovery service that enables MnS consumer to discover management capabilities of MnS instances provided by MnS provider(s).

7.2.2.3 Interaction with NFV MANO and ONAP

Figure 21, [10] provides a possibile example for the creation of a Network Slice subnetwork in terms of intergration between the 3GPP Managemet System and the virtualisation infrastructure managed according to NFV MANO. The figure shows the 3GPP management services and their interfaction with the management interfaces provided by the NFV MANO.

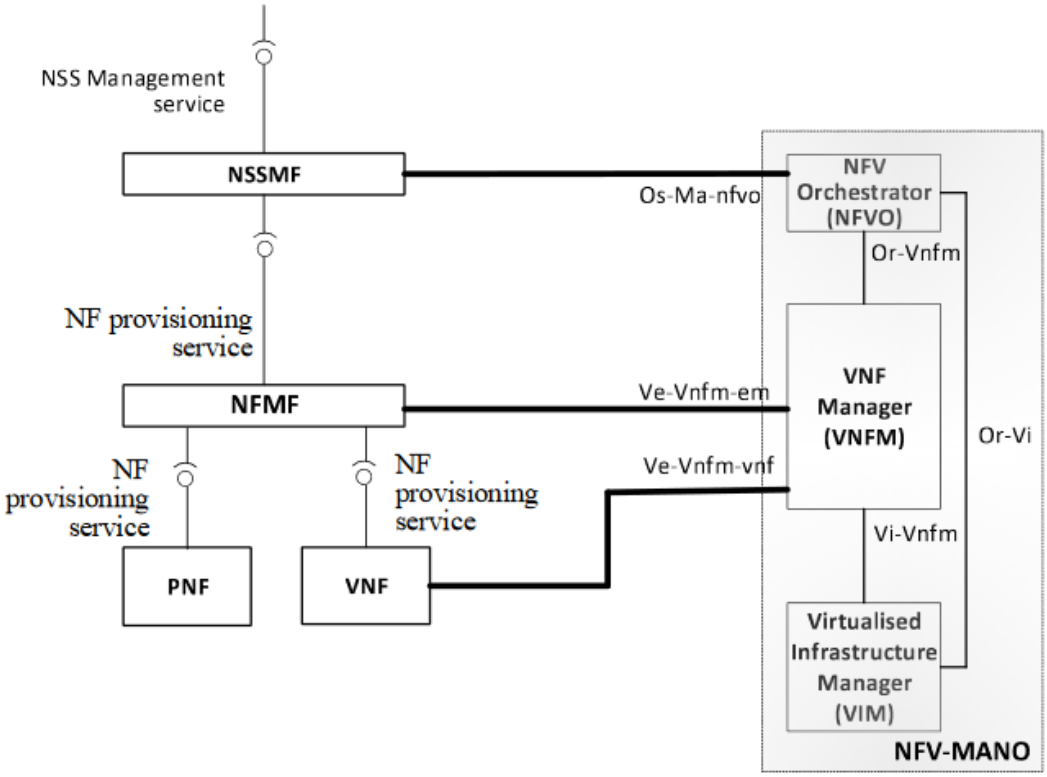


Figure 21: NSSI creation and management with interface to NFV MANO

In this example, Network Slice Subnet Management Function (NSSMF) acts as:

- Consumer of Life Cycle Management (LCM) related services provided by the NFV MANO NFVO
- and provider of Network Slice Subnet management services

Similarly, Network Function Management Function (NFMF) is a consumer of Network Function provisioning service produced by VNFs and PNFs, and a producer of the NF provisioning service including Configuration Management (CM), Fault Management (FM), and Performance Management (PM).

Considering the relevance of ONAP as Open Source network orchestrator, it important to notice the possibile integration with 3GPP. The management services provided by 3GPP Data Report MnS producer can be consumed by ONAP [18] Data Collection Analytics and Events (DCAE) acting as MnS consumer. Examples of these management services include: performance data file report MnS, performance data streaming MnS, and Fault Supervision data report MnS. Moreover, ONAP controller, such as APPC, can be integrated with 3GPP provisioning management service producer.

7.2.2.4 O-RAN Service Management and Orchestration

Leveraging on the conecept intrudced by 3GPP for the management system, O-RAN Alliance has a specification dedicated to the RAN management and orchestration called Operations and Maintenance Architecture [19]. It applies to 4G and 5G RAN. O-RAN has defined a management function called the Non-RealTime RAN Intelligent Controller (Non-RT RIC). This component is part of the Service Management and Orchestration (SMO) layer. The Non-RT RIC with the SMO collect metrics from the RAN to optimize operations and network performance for a better user experience.

O-RAN [19] specification is aligned with 3GPP SA5 and ETSI NFV for LCM (lifecycle management).

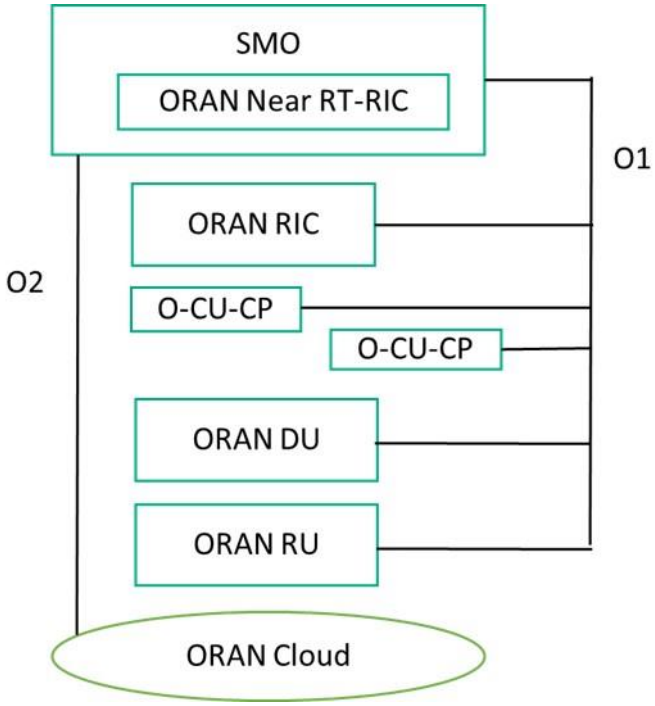


Figure 22: O-RAN Operation and Management Interfaces O1 and O2

O-RAN has defined 2 operation management interfaces produced by the SMO (Service management & orchestration): O1 for O-RAN RU, DU, CUs and RIC and O2 for O-RAN Cloud. Typically O1 either directly interfaces with the O-RAN components or with the aggregated function (ex DU+CU combined), depending on the deployment model. Similarly O-RAN RU may be managed by O-RAN DU via the Fronthaul Management I/F and O1 would only interface with the O-RAN DU. O1 interface performs FCAPS functions : Fault, Configuration, Accounting, Performance, Security and File and software management functions, for both physical and virtual functions of O-RAN. The O1 interface is part of the SMO functionality with the SMO domain scope. Typically if the RAN deployment is across different management domains, their main be one SMO for all domains or multiple SMO. The SMO can be a vendor management & orchestration platform.

The O2 interface provides management and orchestration interface with the O-RAN Cloud. It will be defined in more details in future O-RAN specifications.

O-RAN O1 and O2 OAM interfaces provide a number of capabilities including :

- Provisioning and Instantiation, incl file management, software management, configuration/startup/termination of Physical and Virtual function
- Fault and Performance management including monitoring of the communication links

- Life cycle management including scaling
- tracing

in line with 3GPPSA5 and ETSI NFV.

7.2.3 CI/CD aspects to vRAN

API models for management and orchestration of NFs present the opportunity for operators to leverage software CI/CD best practices to reduce service costs and improve time for service delivery. Implementation of CI/CD in RAN requires that execution and delivery of development and test results that is initialized by operations real time using automated process pipelines. Service delivery is then accelerated through the automation of common manual tasks that are typically functionally diverse. Examples of these tasks include:

- Identification and certification of software build candidates for upgrades
- Service change validation
- Security policy change validation
- Support and maintenance change validation

The challenge for realization of CI/CD in RAN is that not all functional aspects can be delivered as software. This requires additional consideration into create demarcations in the operational network to reduce the risk of including development and test into operational processes. Demarcations can be physical and virtual depending on the NFs and the level of risk. The traditional demarcation is that all development and test is completed in a lab by a separate team. The following figures demonstrate how development and test can be integrated into an operational process with minimal operational impact:

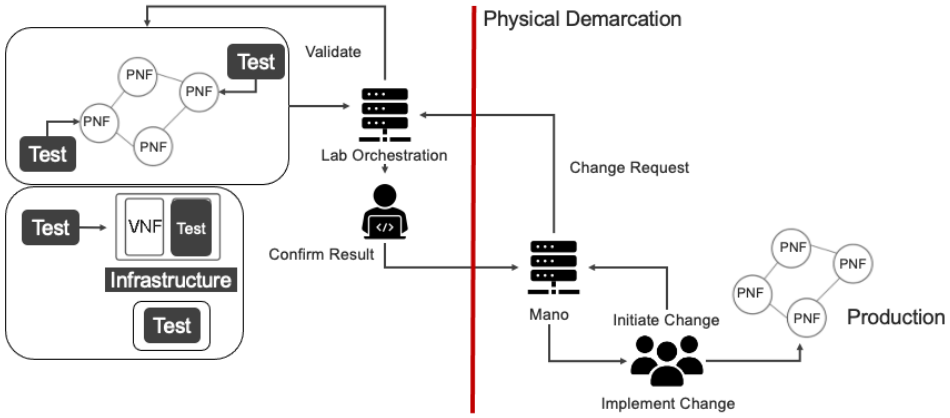


Figure 23: Physical Lab Demarcation Physical and Virtual NFs

A physical demarcation does not inhibit the ability to include development and test into a CI/CD pipeline. It requires that operational teams agree to automation and implement an API for MANO to request validations and pull results. The benefit to this model is for operators that already have the lab in place and just need to implement the automated procedure between the MANO and the lab orchestration.

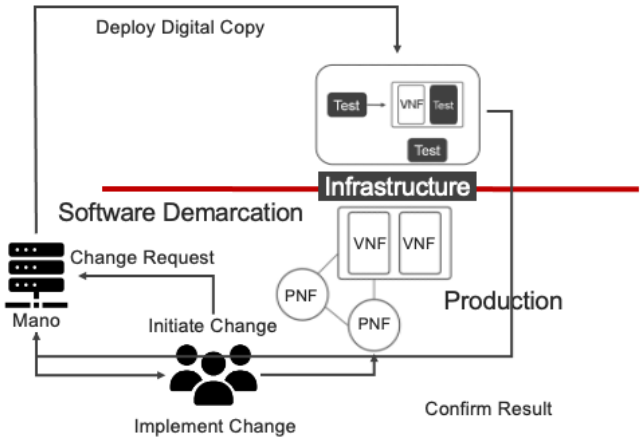


Figure 24: Software Demarcation

Software demarcation leverages native tenant demarcations to allow development and test to use the same platform resources. The change request deploys either a pre-defined development architecture or simple clones the production network function(s).

Software demarcation development and testing is also at a disadvantage in that there is no lab orchestration to pull and deploy images, executes tests and validate results. It is not ideal to place that workload on the MANO. A solution for this is it implement a development test manager to accept requests from MANO, execute and deliver a result.

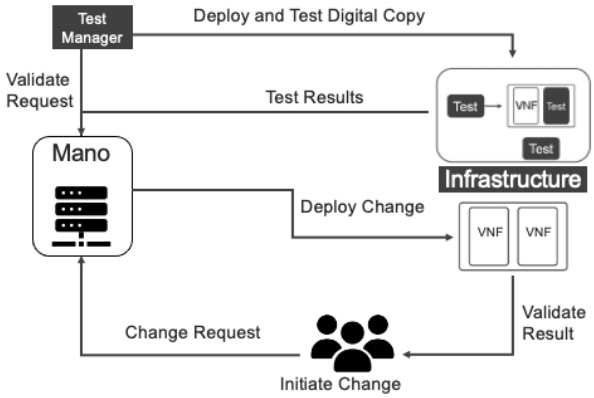


Figure 25: Test Management Cloud Native

The Test Manager fulfills two tasks:

1. It assembles a “release” to be tested, composing VNF packages and descriptors, the specific parameterization for instantiation, and infrastructure-, operator- or other specialization for operational tests.
2. It automates the test execution, performs regression tests and provides test and regression results.

The test manager will advise the MANO to execute the service tests on a separate tenant (for example own sandboxes) on the infrastructure. While adapting the tested releases (e.g. descriptor versions, parameters etc.), it may run several cycles until desired test results are achieved. The final release can go into the production tenant.

Application examples:

- Separate integration and validation infrastructure – push/pull API.
- In production upgrade - staging process instantiate integrate and validate in test availability zone, cutover.
    - Mirrored availability zones replicate cNFs in service chain, build and test change, flag success and apply in production.
    - Utilize network slice for in production isolation for service change management.

CI/CD is critical to address security, ensure thorough trust certification and integration of open software

7.2.4 Security in an Open RAN environment

Traditional RAN deployment scenario assumes deployment of highly specialized appliance hardware deployed at remote RAN sites. The proprietary nature of the appliances as well as high degree of hardware specialization and tight coupling with software allow for a high degree of "security by obscurity". Therefore historically majority of attack vectors in traditional Radio Access Network are focused on tampering the radio air interface (e.g. jamming or high-jacking signaling channels by means of rogue base station transmitting malicious messages on air interface, attacks on L2 layer for traffic re-direction such as e.g. aLTEr - an active cryptographic attack that allows an attacker to redirect network connections by performing DNS spoofing due to a specification flaw in the LTE standard., etc.).

Cloudification of RAN assumes deployment of standard open IT COTS infrastructure at the remote RAN locations, running open cloud software. Typical cloud infrastructure is deployed in centralized data centres with high degree of physical security as well as specialized infrastructure to wall off a security perimeter (e.g. firewalls around secure zones and DMZ, anchors of trust within those, etc.). Therefore, traditional security practices for open cloud mainly address scenarios of remote attacks (DoS, DDoS, phishing and other social engineering-based attacks, SQL injection, cross-site scripting are some of typical examples).

However, with deployment of Cloud RAN at physically non-secure locations (unmanned remote RAN site shelters, outdoor cabinets, etc.), an attacker has unprecedented opportunity to have direct physical access to open cloud infrastructure stack, enabling a fundamentally new attack vectors via direct physical access to the infrastructure (some high-level examples are illustrated on Figure below).

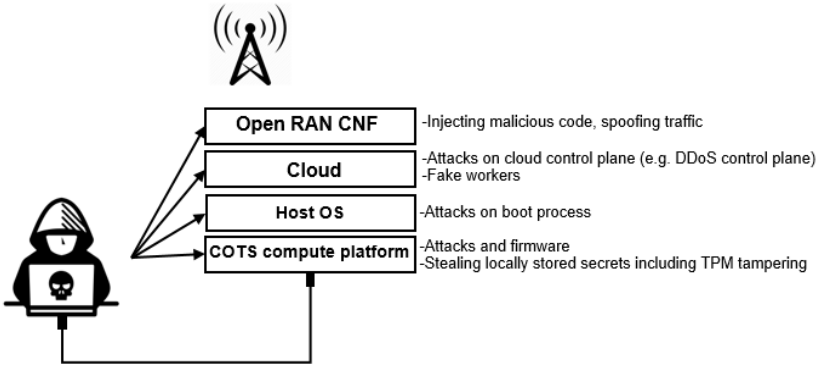


Figure 26: RAN Attack Vectors

Addressing new challenges of Open RAN security requires careful evaluation and adaptation of existing security frameworks against the changes in RAN threat model brought by Open RAN. Some of the fundamental approaches which might prove useful to address this changing threat model are following:

- Zero trust policy for every element deployed at RAN site. As physical security of RAN sites is not guaranteed, one cannot assume any element of Open RAN infrastructure deployed at RAN site can be secure and trusted by default

- Local chain of trust for Open RAN technological stack deployed at RAN site should be anchored to an immutable source of trust in underlying infrastructure itself. No configurable, removable or otherwise mutable source should anchor chain of trust of local technological stack
- As physical infrastructure deployed at RAN site becomes an anchor of trust of Open RAN software layers above it, whole supply chain between infrastructure manufacturing facilities and end RAN site (including local warehouses, staging areas, etc.) should be secured
- Security practices should cover entire lifecycle of the Open RAN technical stack, including secure decommissioning at End of Life. Decommissioned equipment should be fully wiped to a default state and not have any leftover shared secrets (e.g. authentication tokens, passwords,) stored after decommissioning
- Open RAN deployment architecture should limit exposure of overall network accessible from local RAN site – a compromised Open RAN site should not provide means to mount an attack on larger segment of network or the entire network. Therefore, architectural principles which promote atomicity of Open RAN infrastructure should be preferred
- Traditional elements and approaches to network and infrastructure security do not lose their fundamental relevance and should be upheld for every element of Open RAN environment (e.g. authentication policies, securing networking fabrics, etc.)

While giving above initial examples of what a comprehensive Open RAN security framework should study and address in further details, it is understood that these guiding principles for securing Open RAN should be continuously evaluated and adapted to rapidly evolving threat models in RAN.

The O-RAN architecture adds many new interfaces as well as the incorporation of the 5G-XHaul transport to connect small cells to the core network, all of which increase the security attack surface of the RAN. In addition, O-RAN's intelligent non-real-time/near-real-time RAN intelligent controller (RIC) adds new complexities such as potential xApps conflicts and root of trust concerns. O-RAN's deployment on a virtualized or Cloud Native environment adds additional layers of security issues, such as infrastructure, container, and cluster security, among others.

Work in progress through multiple agencies:

- 3rd Generation Partnership Project (3GPP) SA3 security assurance specification
- O-RAN Alliance Workgroup 1 focused on security test specifications
- The Open Test and Integration Centre (OTIC) jointly operated by the O-RAN Alliance and the Telecom Infra Project (TIP)

7.3 Cloud RAN Success Factors

Global alignment of SDOs and ecosystem to design, test, and enable Cloud Native realization of RAN architecture, operation and value delivery roadmap, with

- Virtualization
- Disaggregation
- Openness and
- Intelligence

Which is agile and scalable, leveraging Cloud Native technologies, with orchestration of containerized microservices, and lifecycle management, within DevOps and CI/CD value creation and delivery environment, unified across products

Maturity and standardized realization of end to end dynamic network slicing, and the associated value provided by MNOs along with partners, particularly to Vertical markets

Maturity and realization of hybrid cloud and distributed intelligence across the disaggregated and open network, particularly at the edge

Emergence of carrier-grade Open Source software solutions enabling AI-based RAN management

Security models and practices across layers, from design to products, deployments and operation, confidentiality, integrity, replay protection, authentication, zero-trust and rule-based access, physical and virtual, network-based as well as endpoints, malware and social, zoning (e.g. local site confinement), and probing monitoring, analytics and anomaly detection / mitigation

8 ECONOMICAL DRIVERS

This chapter evaluates, from an economical perspective, the Cloud Native trend in connection with 5G, aims at quantifying key effects and outlines options for Telcos. A view of existing communication services and the opportunity for new revenue sources is included.

Impact on Existing Market

Adoption of Cloud Native practices in Telcos aligns directly with addressing CAPEX and OPEX pressures from customer data growth and evolution of their networks for 5G. Open, software defined systems, with vendor neutral hardware and standard solutions can accelerate innovation by displacing vendor lock-in and by encouraging competition, and potentially disrupting market norms, providing benefits to Telcos, their partners, and customers.

Table 5: Redefining Market Norms

PROVIDES VENDOR DIVERSITY	Greater freedom to partner with multiple vendors and paving the way to a 'plug and play' network where 'best in breed' solutions can be implemented across the value chain to provide a differential proposition.
INCREASING COMPETIVE INTENSITY	New vendors are increasingly entering the market as barriers to entry soften. Downward pricing pressure on hardware, software and services is widely anticipated with a greater focus on providing differential and innovative solutions / services.
FOSTERS INNOVATION	Cloud native approach will encourage shorter, more agile R&D cycles which will accelerate the emergence of a new and broad ecosystem of truly innovative products that can be deployed into market more rapidly and at a reduced cost point.

5G Revenue Opportunity and the Emergence of Edge

Enhanced video, real-time automation, connected vehicle, monitoring & tracking, hazard sensing, autonomous robotics, remote operations, smart surveillance and augmented reality are cited as typical use cases for 5G. Latency, bandwidth, security and compute efficiency requirements for many of these services push the optimal location for processing close to the source of data. So many new services, and new revenue sources, are only enabled by 5G Edge capability

8.1 Economic Case for Cloud Native Telco

Enabling the new market opportunity in "5G Edge services" is the new business territory for many Telcos. This is further discussed in this section.

8.1.1 Market Dynamics

Driven by 5G, edge services are emerging as a new segment within the value chain. Figure 27 shows the emergence of this 5G edge opportunity as a new enabler for products and where Telcos and Hyperscalers may contend. Telcos and Hyperscalers are both potentially vying for the market to provide communication services (and connectivity) and for the edge platforms that will support the smart homes, lifestyle, work, transportation, manufacturing and environment of the near future. Whether the customer products are powered by Hyperscalers or Telcos, or potentially a hybrid model, customers will expect a fully digital experience to set-up, pay-for and manage their services, and expect the service level that can be provided by the Cloud Native characteristics of reliability, elasticity, openness and manageability. Historically these are the strengths of Hyperscalers and Telcos should continue to move forward on those, to become part of a larger eco-system.

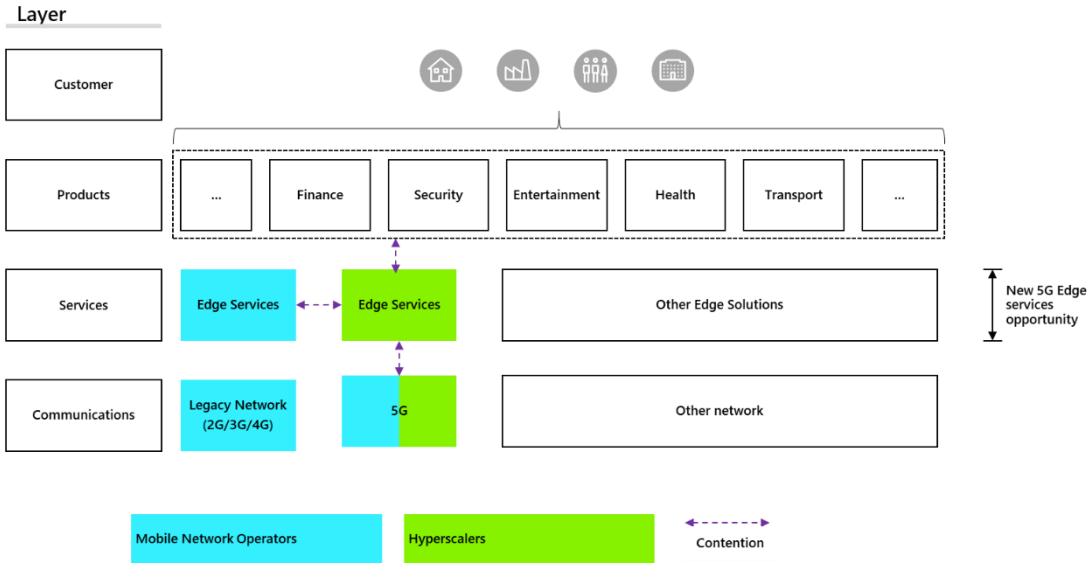
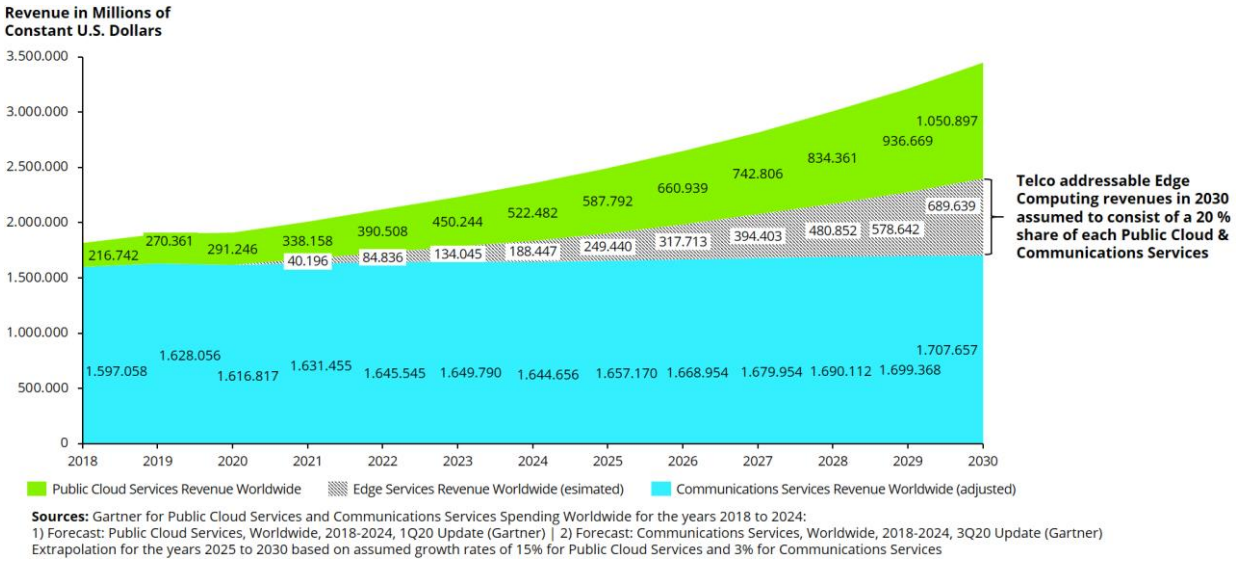


Figure 27: The Edge Services Market Segment

8.1.2 Market Opportunity

A projection for the edge market opportunity is shown in Figure 28 and presents revenue growth in relation to public cloud and communications services. Based on predicted public cloud and communications services revenue [41] the addressable edge opportunity for Telcos is in the order of 700 billion USD in 2030. Telcos and Hyperscalers will potentially be competing or partnering for this market, as determined by their business model. (The figure of 700 Billion USD is consistent with Ericsson's "5G for business: a 2030 market compass" October 2019 [42] prediction of Telco addressable edge service revenue).



Sources: Gartner for Public Cloud Services and Communications Services Spending Worldwide for the years 2018 to 2024:
1) Forecast: Public Cloud Services, Worldwide, 2018-2024, 1Q20 Update (Gartner) | 2) Forecast: Communications Services, Worldwide, 2018-2024, 3Q20 Update (Gartner)
Extrapolation for the years 2025 to 2030 based on assumed growth rates of 15% for Public Cloud Services and 3% for Communications Services

Figure 28: Assumed Revenues for Edge Computing in 2030

8.1.3 Future Business Models to Support Edge Services

Three business models for Telcos and Hyperscalers in connection with edge services are possible, reflecting different value propositions, end customer relationship and revenue flows. Figure 29 below shows these potential models.

As the figure shows, for Model 1, the Telco manages the relationship with the customer and is able to pull together all components of the value chain. The Telco is providing communication services and either directly providing applications and cloud platform or managing other parties such as a Hyperscaler to provide parts of the product, effectively as sub-contractors. This model has the highest potential for revenue and the highest demand on Cloud Native competence.

Model 2 is a partnership approach with joint engagement with the customer. Specific division of responsibilities for providing the customer product for application, platform and communication services are agreed according to partnership agreements. This model brings together the different strengths and capabilities of Telcos and Hyperscalers and still places emphasis on Cloud Native competence for the Telco to peer with Hyperscaler services.

Model 3 is a mirror of Model 1 but with the Hyperscaler or other party managing the relationship with the customer and determining the extent of involvement of other parties. In this model the Telco is potentially providing only some connectivity and communication services and as a sub-contracted entity to the Hyperscaler.

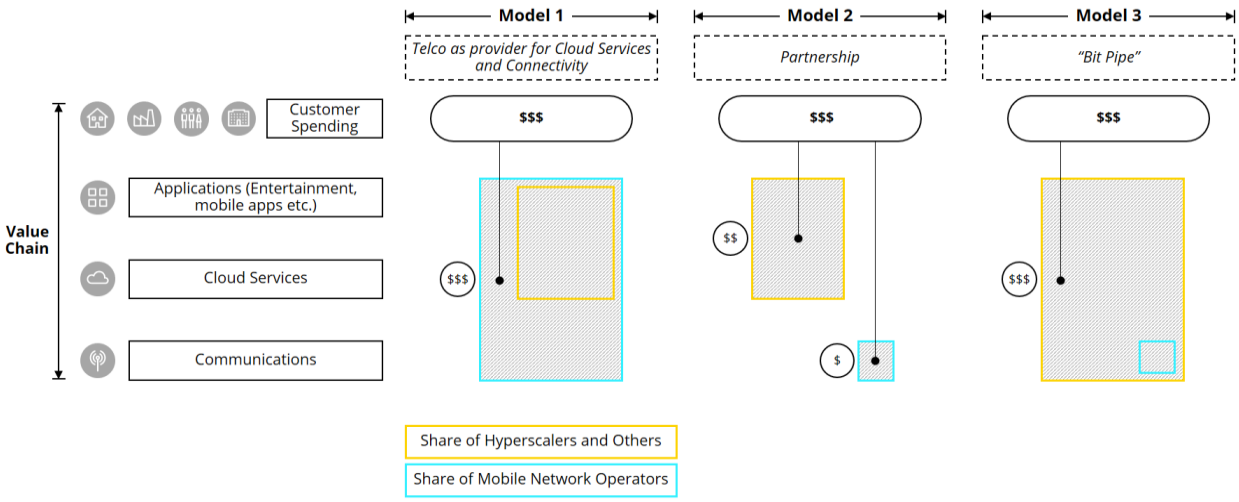


Figure 29: Future Business Models for Edge Services

Table 6: Telco – Hyperscaler Business Models

	Model 1: Telco as Provider for Cloud and Connectivity	**Model 2: Partnership with Hyperscaler**	**Model 3: Telco as Bit-Pipe**
Attributes	End customer relationship owner and integrator: Telco The Telco provides Cloud Native connectivity services and brokers Telco, Hyperscaler and other clouds based on customer requirements. The customer pays the Telco directly for the cloud services and communication services elements.	End customer relationship: Telco and Hyperscaler Integration lead: Telco or Hyperscaler Telco provides connectivity services and the Hyperscaler provides the cloud services platform and Over-The-Top services. The customer pays the Telco directly for the communication services elements, and pays the Hyperscaler for the cloud and application elements.	End customer relationship owner and integrator: Hyperscaler The Hyperscaler provides cloud services and brokers Telco and Hyperscaler communication services based on customer requirements. The customer pays the Hyperscaler directly for the cloud services and communication services elements, and the Telco receives payment from Hyperscaler.

All models require Cloud Native competence, with Model 1 pushing that competence and application developer engagement to be on a level with Hyperscalers. Further, Model 1 allows Telcos to differentiate from Hyperscalers by bundling integration with other clouds for edge services and simplifying relationships and integration for customers. Model 2 is a component also of Model 1, but on its own is potentially less demanding for Telcos and discussed in the

next section. Model 3 implies a focus on cost-reduction, and potentially leads to Telco revenue shrinkage as Hyperscalers evolve to provide more connectivity and communication services themselves.

To achieve new revenue and deliver MNO's value creation, Telcos will need to pursue Model 1 or Model 2.

8.1.4 Emerging Predominance of Partnerships

Previously Telcos had different levels of success in offering cloud services. Some exited offering their own cloud services years ago whilst some are making increasing investment in their own cloud services. However, cloud services on a par with Hyperscalers such as Google, AWS or Microsoft Azure require a broad set of capabilities and substantial global scale. The Hyperscalers on the other hand are already partially developed towards being network operators with global transport networks and CDNs being an integral part of the Hyperscaler's value chain and they generally lack edge deployments.

Gaps in Telco capability with respect to offering public cloud services, and the gaps for Hyperscalers with respect to providing communication services mean that partnerships between Telcos and Hyperscalers are the most likely way forward to provide edge services. Hyperscalers are keen to partner with Telcos to capture network and communication services, as the Telcos move forward with 5G. The gaps in the Hyperscaler's capabilities with regard to access network assets, field support teams and enterprise customer reach, further point towards the partnership being beneficial to both parties. Figure 30 outlines the potential contribution of Hyperscaler and Telco in a win-win partnership.

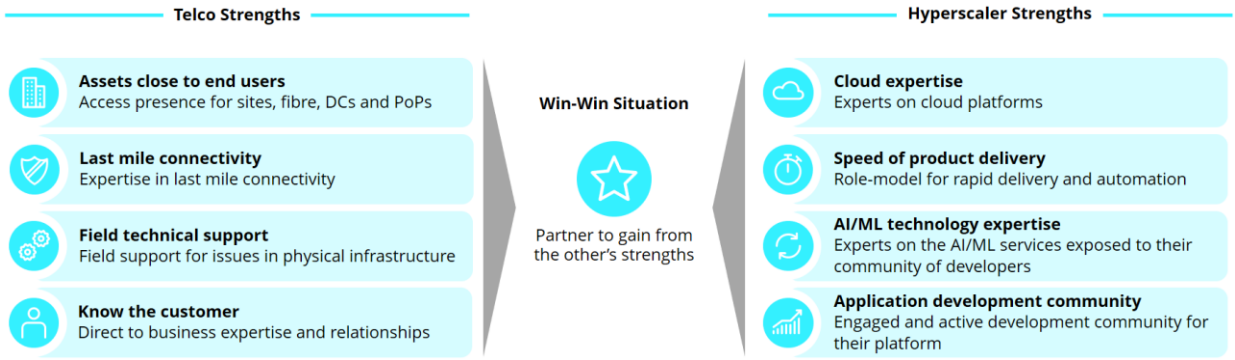


Figure 30: Win-Win Potential of Partnership

The Telco's path towards development of a hybrid cloud strategy is consistent with this model, to create and deliver end-to-end value in partnership.

Figure 31 shows the positioning of computing according to requirements for latency and analytics and illustrates the role that edge computing has for enabling new communication services. The figure suggests there are various deployment models for edge services, balancing requirements of latency, device processing, network bandwidth and cost. Telcos with a high presence of sites and fiber in the access network are best placed to support a range of options, driven by customer requirements and business opportunities.

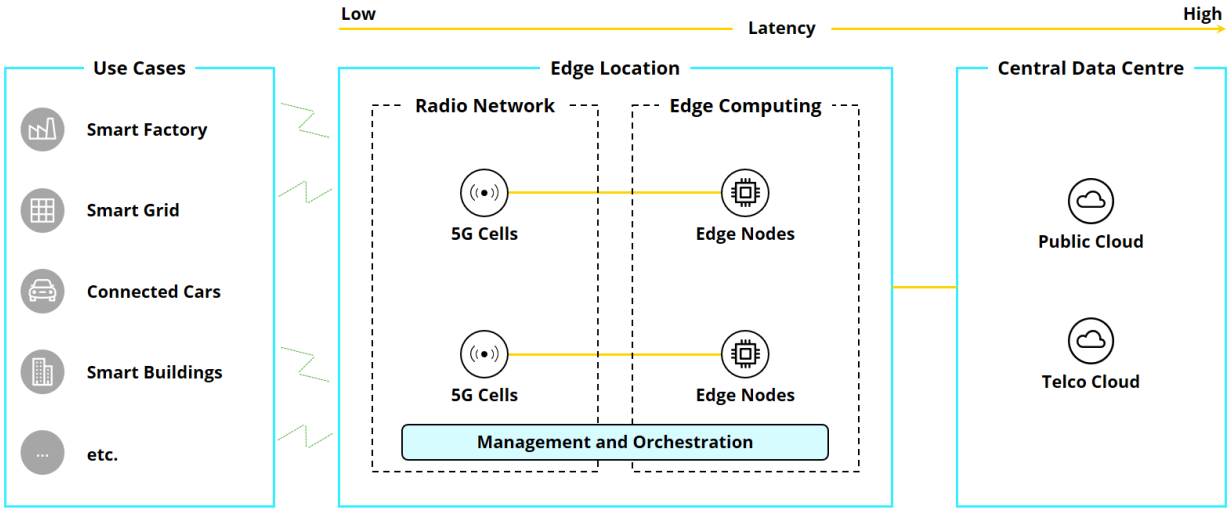


Figure 31: Edge Compute Positioning Supporting Latency Requirements

Providing flexible resources is however not enough. Assuming most future services will be provided by third parties on top of the Telco value chain, integration of very flexible agile development processes is necessary to foster the required innovation which will lead to future services. Edge should therefore be implemented with a strong API based approach to support the developer communities and should also be supported by seamless agile DevOps tool chains to support an iterative way towards innovative future services.

There is potential to use the same virtualization platform for edge services and open RAN, further simplifying automation and orchestration across edge applications and network functions. Telcos could therefore target a shared virtualization platform to provide communication efficiency between edge and Radio and core network functions.

8.1.5 Telco Organisation Transformation Dependency

For Telcos, the Cloud Native path is the natural path for 5G and beyond and promises ability to create and deliver value for countless and diverse use cases, with cost and energy management, and opportunities including the potential to capture new revenue enabled by 5G edge services. The challenges and requirements concerning technology maturity, deployment and operations, and organization, however, will need to be managed and addressed. Of these challenges the Telco's organizational transformation to Cloud Native, as regards Hyperscalers, is seen as the biggest challenge ahead. Telcos continue to evolve and transform to align with and drive the new 5G enabled cloud services, and to deliver a service according to Cloud Native principles and Hyperscalers levels of cloud competence. A change in talent profile will be needed, and Telcos are focusing on:

Cloud expertise and DevOps: Engineers will require new technical capabilities to operate in the cloud. In parallel Telcos will increasingly be more agile and adopt a DevOps mindset and ways of working.

Analytics and artificial intelligence: Building an Analytics and AI engine with Machine Learning and automation solutions to optimize management of the network.

Upskilling existing fieldforce: Retraining existing employees to ensure they have the knowledge, skills and expertise required for 5G and Cloud-Native.

9 CHALLENGES, CRITICAL SUCCESS FACTORS & DEPENDENCIES

This chapter addresses the main challenges, critical success factors and dependencies that are relevant for the adoption of a Cloud Native approach for Telcos.

9.1 Challenges

5G and beyond require unprecedented levels of flexibility, agility, scalability, and automation. Cloud Native architecture is a paradigm shift empowering new services from network architecture perspective, requiring the adoption of new operational and organisational models. It also requires new cooperation strategies among stakeholders to fully embrace the business possibilities enabled by such a technological shift.

The evolution toward a Cloud Native network is full of challenges and requirements from architecture, implementation, and security perspectives.

In terms of architecture, considering 3GPP Release 16, some network functions still have some interfaces (not service-based) using protocols which are not compliant with the Cloud Native architecture.

While many vendors claim open interfaces in their network components, they are still limited by their own customized network management tools. Even though a network may have been deployed with multi-vendors elements, it is still not entirely vendor-neutral. This is because orchestration and network management software for different components may still be vendor specific.

Aiming a full Cloud Native ecosystem, it is important to notice that it may not be possible to convert every network function into Cloud Native in near term. This can be due to stringent latency and throughput requirements. Heterogenous network architectures will exist for a long time in the network evolution path toward Cloud Native.

Computation-intensive RAN nodes performing Layer1 and Layer2 functions may still need custom hardware accelerators to deliver services for some use cases. These special purpose resources/nodes (e.g. FPGA, SmartNIC) may not be fully optimized according to the Cloud Native principles. Heterogenous network architectures also incur high overhead in network management.

In contrast to core network which has a general accepted nodal structure, RAN architecture has many options to split the protocol stack, each with different demand on the underlying interface and transportation link. This provides flexibility in terms of deployments balancing central and edge distribution of the RAN network elements. This also provides different possibilities on the RAN component that can be Cloud Native.

It is conceivable that Cloud Native transition will be gradual, i.e. VNF and CNF will co-exist in a network for some time. This means that orchestration tools must be able to manage both VM based and container-based NFs. Otherwise different orchestrators must be used. This coexistence also requires different teams and operational models to coordinate and to manage the network.

Nodes with multiple sockets have different latency, Cloud Native system allocates resources without knowing these variances in latency. This may cause synchronization issues.

Considering the implementation of Cloud Native, migration aspects must be considered. It is indeed important to ensure a smooth migration through analysis of workloads and the inter-dependencies between legacy and hybrid-cloud systems. It is critical to ensure existing services are not interrupted by the development and production of new services.

Considering the opportunity and flexibility a Cloud Native system provides, Network operations and service teams will need to adopt the DevOps mindset and CI/CD processes.

Security for Cloud Native functions must be considered over their whole lifecycle. Development, deployment, and operational cycles must be consistent with a unified security framework (central policy management and visibility) across the system over which the application exists and runs.

9.2 Critical success factors

To make a smooth and successful evolution to Cloud Native, many factors are critical to embrace this evolution capturing all the possibilities and maintaining the overall cost efficiency of the transition. In the following some of these factors are identified.

- The harmonization of orchestration and other tool sets, such as assurance and analytics, is critical for multiple reasons. From a technical perspective it is required to optimise the service delivery and operation process. From a business point of view, it is important to guarantee the overall SLA agreed with the customer. In terms of costs it is important to have a coherent and consistent well-defined solution to avoid integration costs that, in such a complex scenario, can be very high considering the many components involved.
- End-to-end global standards including open interfaces need to be in place. This enables the Cloud Native architecture to deliver agile, resilient, flexible, and scalable services. This also enables the integration process in a multi-vendor scenario. Open interfaces, easily accessible and exploitable by other Telcos and HCPs, are mandatory for a Telco to be part of a wider ecosystem.
- Global roaming agreements, edge federation and strong partnerships among the stakeholders need to be established. It is indeed important to ensure the availability of the services at global level. A customer shall experience the same level of services regardless which Cloud Native architecture the services are leveraging on. developers must be allowed to deploy their application seamlessly over the different Telco Platforms.
- Consistent security policies and capabilities are required. There should be agreement on security policies and capabilities among different infrastructure vendors, network operators, and service providers involved in the delivery of a Cloud Native service.
- Training is a key success factor. Cloud-native service development, network operation and maintenance, and security teams must be aligned with the same mindset and they must augment their skills in Cloud Native architecture and technologies.
- The Cloud Native platform is the natural technological glue allowing different players to integrate their resources. To fully embrace this opportunity, it is important to set up a win-win and shared model among Telcos and HCPs to engage developers and customers to build an ecosystem.

10 CONCLUSIONS

The evolution of the Telco Platform toward Cloud Native is a key innovation process for Telcos. This evolution is well supported by the work done by the international SDOs and For a. It is already part of the ongoing activities a Telco is facing nowadays to modernize its infrastructure to fully embrace 5G.

This innovation does not start from scratch, it inherits the standard IT components and platforms from IT cloudification. This implies the adoption of open and industry standard interfaces allowing management and operation of NFs and applications from one or more vendors. A Cloud Native standard platform provides Telcos the means to integrate multiple vendors solutions into a software defined pool that can be provisioned, managed and monitored “as codes”.

An open infrastructure leverages a flexible virtualization layer that entirely abstracts the physical layer. On top of it, Telco services can be deployed fully decoupled, according to disaggregation and control-plane/user-plan separation methodologies. Depending on the level of maturity and technological requirements, virtual machines or containers can be adopted. This will allow Telcos to participate in the innovation cycles of enterprise IT infrastructure. It also is the foundation to create an environment that allows to open the ecosystem for developers. If deployed properly, OPEX of operating the infrastructure and TCO can be reduced by creating a zero-touch orchestration on top of a fully flexible and highly scalable pool of resources.

Leveraging this technological evolution, the network is becoming an Open Telco Platform, both for internal efficiency and to be active actor of an external ecosystem.

The openness of the Telco Platform is defined by the adoption of interfaces for internal usage or to be exposed to the external entities. The current evolution embraces both standardized interfaces as well as those based on well-documented de facto standards. The ability to program the service offered by the platform must be exploited while preserving tenancy relationships without any implication or increasing in security risks.

The cloudification of the Telco Platform is producing a harmonization, driven by the technology, between the centralized Telco infrastructure and the Edge deployments. The adoption of a Cloud Native infrastructure is also bringing the Telco data centres more and more close to the HCP Cloud technology. A Hybrid Cloud model is naturally emerging from the current network evolution and it is going forward the expectation of new business opportunities. Hybrid has different flavors; one aspect is the coexistence of VM and Containers. Another aspect is the coexistence of Telco and Service oriented environments. Edge and Cloud coexistence and integration is another aspect that fits in the Hybrid scenario. The main aspects of this Hybrid Cloud transformation are a stronger focus of enabling a fluid –and potentially seamless – interoperability among the coexisting entities. Cloud solutions allow a Telco to form a single continuous cloud stratum.

When the Cloud Native evolution will reach maturity, the focus must be shifted back to the actual service being virtualized. For cost-effective and agile value creation and delivery, it should not be necessary to spend resources on assessing which cloud solution is the most appropriate, sustainable and future-save. Currently, the mix of cloud solutions on the management plane can only be achieved by realizing a unified translator for interfacing with various orchestrators. However, it is of paramount importance for the Telco industry to have both the ability to “speak the same language”, and the same set of features available.

Whether the customer services are powered by Hyperscalers or Telcos, or potentially through a hybrid model, customers will expect a fully digital experience. This experience embraces the service set-up, payment and the delivery on a managed platform. The overall management of their products is a key factor. Customers expect Cloud Native characteristics of reliability, elasticity, openness and manageability. Historically, these are Hyperscaler platform strengths. Nowadays Telcos that manage to achieve Cloud Native competence and levels of service, are well placed to secure a leading role in providing 5G services and application support especially leveraging on their distinctive Edge resources.

The future of mobile networks is being re-shaped by the rise of Cloud architectures that extends levels of efficiency and scale from the datacentre to the mobile network and Edge. With Software-Defined Networking (SDN) and Network Function Virtualization (NFV), General Purpose Processors (GPP)-powered cloud servers have the flexibility to change workloads based on demand. This allows Telcos to exploit the flexible infrastructure, with different kinds of NFs or applications. The infrastructure guarantees long term stability for NFs and dynamicity for applications.

To make a smooth and successful evolution to Cloud Native, technical, organizational and operational aspects must be considered. Global standards, including open interfaces, need to be in place for a smooth adoption on technical side. Service development, network operation and maintenance, and security teams must all be aligned with the same mindset for a smooth transition at organizational and operational level. Those teams must augment their skills in Cloud Native architecture, technologies and processes. Telcos must collaborate with one another creating synergy with HCPs. All the stakeholders are enabled by the Cloud Native technology to embrace a collaborative approach to business towards developers and customers. Cloud network providers and application developers shall take a collaborative approach across the value chain to accelerate the pace of innovation and to establish a robust ecosystem. The current evolution process is indeed not just technical, it is also driven by economical drivers that are related to internal Telco factors (e.g. savings) and to new business opportunities. A favorable future for a Telco will demand transformation to Cloud Native and a change in skills, knowledge, governance, funding, leadership and culture. In return, a Cloud Native skillset and organization will drive cost efficiency, innovation, agility, automation, customer engagement and data-driven decision making. For a traditional Telco, the transition may be challenging, but Telcos that fail to make the transition to Cloud Native will likely see competition erode existing revenue and lose much of the 5G opportunity.

In summary, the Telco being part of wider ecosystem together with the developers and HCPs is a scenario well supported by the current Telco network cloudification process, by the work of the standardization bodies and Fora, and by the software developed by the Open Source communities. This is a challenge for the Telcos to embrace, to enhance the value of the network cloudification investments, to catch a unique opportunity that, driven by new service opportunities, adopts a common technology platform among the stakeholders with a solid base to succeed.

To make a smooth and successful evolution to Cloud Native, many success factors are critical to embrace this evolution as outlined in this document and summarized in Section 9.

ABBREVIATIONS

5GC	5G-Core
AF	Application Function
AMF	Access and Mobility Management Function
API	Application Programmable Interface
BBU	BaseBand Unit
BSS	Business Support Systems
CAPEX	Capital Expenditure
CI/CD	continuous integration continuous deployment
CM	Configuration Management
CNF	Cloud-native Network Function OR Containerised Network Function
COTS	Common of the shelf
CU	Central Unit
CU-C	CU Control Plane (or CU-CP)
CU-U	CU User Plane (or CU-UP)
cVNF	Cloudified Virtualized Network Function
DDoS	Distributed Denial of Service
DMZ	Demilitarized Zone
DN	Data Network
DoS	Denial of Service
DPDK	Data Plane Development Kit
DU	Distribution Unit
eMBB	enhanced Mobile BroadBand
E-UTRA	Evolved Universal Terrestrial Radio Access
EVPN	Ethernet Virtual Private Network
FM	Fault Management
gNB	Next Generation NodeB
GNBCUCPF	Next Generation NodeB Central Unit Control Plane Function

GNBCUUPF	Next Generation NodeB Central Unit User Plane Function
GNBDUF	Next Generation NodeB Central Distribution Unit Function
GSMA	GSM Association
HCP	Hyperscale Cloud Providers
ICT	Information and Communications Technology
ISP	Internet Service Provider
LBO	Local Breakout
LCM	Lifecycle Management
MAC	Medium Access Control
MANO	Management and Orchestration
MEC	Multi-Access edge Compute
mMCT	massive Machine Type Communication
MNO	Mobile Network Operator
NbR	Name-based Routing
NF	Network Function
NFMF	Network Function Management Function
NFV	Network Function Virtualisation
NRF	Network Repository Function
ng-eNB	Next Generation evolved NodeB
NG-RAN	Next Generation RAN
Non-RT RIC	None Realtime RIC
n-RT RIC	Near-Realtime RIC
NSA	Non-Stand-Alone
NSMF	Network Slice Management Function
NSSMF	Network Slice Subnet Management Function
O-Cloud	Open Cloud SW
O-CU	Open Central Unit
O-DU	Open Distribution Unit
OPEX	Operational Expenditure
OSS	Operational Support Systems
OTT	Over-the-Top
PaaS	Platform-as-a-Service
PDCP	Packet Data Convergence Protocol
PHY-H	Physical Layer - Higher
PHY-L	Physical Layer - Lower

PM	Performance Management
PNF	Physical Network Function
QoS	Quality of Service
RAN	Radio Access Network
RDMA	Remote Direct Memory Access
RF	Radio Frequency
RIC	RAN Intelligent Controller
RLC	Radio Link Control
RRH	Remote Radio Head
RU	Radio Unit
SBA	Service Based Architecture
SCP	Service Communication Proxy
SDN	Software-defined Networking
SEN	Service Edge Node
SMO	Service Management and Orchestration
SMOF	Service Management and Orchestration Function
SR-IOV	Single-Route Input/Output Virtualization
TCO	Total Cost of Ownership
TEN	Telco Edge Node
UE	User Equipment
UPF	User Plane Function
URLLC	Ultra Reliable Low Latency Communication
VNF	Virtualized Network Function
vRAN	Virtualized RAN

DEFINITIONS

Service Edge Node	Edge location in the network where Application Server are deployed. SEN hosts the application level services, maybe by 3rd party. These Application Server offers services generally to end users.
Telco	Telecommunications service provider
Telco Edge Node	Edge location in the network where Network Functions are located (e.g. component of the Core Network such as UPF)

REFERENCES

[1] ETSI GS NFV-MAN 001 – 4.5.1 Fault and Performance Management

[2] PRD –G116v4 - Generic Network Slice Template

[3] OPG.01 - Operator Platform Telco Edge Proposal

[4] Mamoun, M. B., & Benaini, R. (2016). An Overview on SDN Architectures with Multiple Controllers. *Computer Networks and Communications*.

[5] 3GPP TS 23.501 Rel. 16. System architecture for the 5G System (5GS)

[6] 3GPP TS 23.558. Architecture for enabling Edge Applications (EA)

[7] 3GPP TR28.812. Study on scenarios for Intent driven management services for mobile networks

[8–] NGMN - Overview on 5G RAN Functional Decomposition

[9]– "NGMN - 5G RAN CU – DU network architecture, transport options and dimensioning"

[10] 3GPP, "5G; Management and Orchestration; Architecture Framework," 3GPP TS 28.533, V16.5.1, Rel. 16,Nov. 2020.

[11] ETSI, "Restful protocol specification for Ve-Vnfm reference point"

[12] 3GPP TS 28.541: "Management and Orchestration; 5G network resource model (NRM); stage 2 and 3"; rel 16

[13] 3GPP TS "8.401: "NG-RAN; Architecture description"

[14] 3GPP TR 38.801: "Study on new radio access technology; radio access architecture and interfaces"

[15] 3GPP TS 38.300: ""NR; NR and NG-RAN overall description; stage 2"

[16] 3GPP, "UMTS, LTE, Generic Network Resource Model (NRM), Integration Reference Point (IRP), Information Service (IS)," 3GPP TS 28.622, V15.4.0, Rel. 15, Jan. 2020.

[17] 3GPP, "5G; Management and Orchestration; 5G Performance Measurements," 3GPP TS 28.552, V16.7.0, Rel. 16, Nov. 2020.

[18] Open Network Automation Platform (ONAP), https://www.onap.org/

[19] O-RAN Alliance, "O-RAN Operations and Maintenance Architecture," Technical Specification, v03.00, Apr. 2020.

[20] O-RAN Alliance, https://www.o-ran.org/

[21] O-RAN Alliance, "O-RAN Operations and Maintenance Interface Specification v03.00," Technical Specification, Apr. 2020.

[22] O-RAN Alliance, "WG4 Management Plane Specification v03.00," Technical Specification, Apr. 2020.

[23] O-RAN Alliance, "Cloud Architecture and Deployment Scenarios for O-RAN Virtualized RAN," Technical Specification, Apr. 2020.

[24] https://www.ericsson.com/en/blog/2020/9/ran-what-policy-makers-need-to-know

[25] Open Infrastructure Foundation, https://openinfra.dev/projects

[26] Linux Foundation, https://www.linuxfoundation.org/projects

[27] Open Networking Foundation, https://opennetworking.org/onf-sdn-projects/

[28] O-RAN Software Community, https://o-ran-sc.org/

[29] https://12factor.net/

[30] https://docs.microsoft.com/en-us/dotnet/architecture/cloud-native/definition

[31] https://tanzu.vmware.com/cloud-native

[32] https://www.cisco.com/c/dam/en_us/about/ac79/docs/sp/Cloud-Value-Chain-Exposed_030512FINAL.pdf

[33] https://www.cncf.io/announcement/2015/06/21/new-Cloud-Native-computing-foundation-to-drive-alignment-among-container-technologies/

[34] www.fudge-5g.eu

[35] https://www.opennetworking.org/software-defined-standards/specifications/

[36] https://tools.ietf.org/html/rfc6020

[37] https://p4.org

[38] Trossen, D. and Robitzsch, S. and Hergenhan, S. and Riihijarvi, J. and Reed, M. and Al-Naday, M., "Service-based Routing at the Edge", arXiv preprint arXiv:1907.01293, 2019, http://arxiv.org/abs/1907.01293

[39] https://www.etsi.org/technologies/multi-access-edge-computing

[40] https://www.3gpp.org/common-api-framework-capif

[41] Gartner for Public cloud Services and Communications Services Spending Worldwide for the years 2018 to 2024: 1) Forecast: Public cloud Services, Worldwide, 2018-2024, 1Q20 Update (Gartner) | 2) Forecast: Communications Services, Worldwide, 2018-2024, 3Q20 Update (Gartner)

[42] Ericsson's "5G for business: a 2030 market compass" October 2019