



Converged Heterogeneous Advanced 5G Cloud-RAN Architecture for Intelligent and Secure Media Access

Project no. 671704

Research and Innovation Action

Co-funded by the Horizon 2020 Framework Programme of the European Union



Call identifier: H2020-ICT-2014-1

Topic: ICT-14-2014 - Advanced 5G Network Infrastructure for the Future Internet

Start date of project: July 1st, 2015

Deliverable D1.2

Refined architecture definitions and specifications

Due date: 31/12/2016

Submission date: 31/12/2016

Deliverable leader: Kai Habel (HHI)

Dissemination Level

-
- | | |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | PU: Public |
| <input type="checkbox"/> | PP: Restricted to other programme participants (including the Commission Services) |
| <input type="checkbox"/> | RE: Restricted to a group specified by the consortium (including the Commission Services) |
| <input type="checkbox"/> | CO: Confidential, only for members of the consortium (including the Commission Services) |
-

List of Contributors

Participant	Short Name	Contributor
Fraunhofer HHI	HHI	Kai Habel
Ericsson	ERICSSON	Carolina Canales
Fundació i2CAT	I2CAT	Shuaib Siddiqui, Eduard Escalona
Demokritos NCSR	NCSR	Eleni Trouva
Innoroute	INNO	Andreas Foglar, Marian Ulbricht
JCP-Connect	JCP-C	Yaning Liu
University of Essex	UESSEX	Mike Parker, Geza Koczian
Intracom	ICOM	Konstantinos Katsaros
Ethernity	ETH	Eugene Zetserov
COSMOTE	COSMO	George Lyberopoulos, Konstantinos Filis

Change history

Version	Date	Partners	Description/Comments
V0.1	11/10/16	HHI	ToC created
V0.2	27/11/16	HHI, ICOM, UESSEX	Initial input for: data plane architecture, fronthaul, automotive use case, multi-tenant use case
V0.31	08/12/16	ICOM, ERICSSON, COSMOTE, HHI	Architecture overview section, control plane, data plane; first input to service workflows, Introduction to Use case refinements
V0.41	12/12/16	ICOM, InnoRoute, NCSR, JCPC	Input for architectural refinements; remarks for service workflows added input for caching service; security consideration to MT UC added
V0.42	13/12/16	HHI, ICOM, InnoRoute	polishing
V0.43	13/12/16	ERICSSON, HHI	polishing
0.5	15/12/16	ERICSSON, ICOM, HHI	Polishing of security service section, clarification in MT use case, Introduction
0.6	19/12/16	ICOM, HHI	Fixes from ICOM and HHI, Summary, Introduction, Conclusion (version for int. review)
final	21/12/16	All	Fixes after review

Table of Contents

1. Introduction.....	6
2. Architecture refinements.....	7
2.1. Overview.....	7
2.2. Control plane	9
2.3. Data plane.....	10
2.3.1. Architectural refinements for physical layer technology	13
2.3.2. Fronthaul over Ethernet	15
3. Workflows & Service Life Cycle Design refinements.....	18
3.1. CHARISMA Actors and Roles Interaction.....	18
3.2. CHARISMA Services Workflows	18
3.2.1. Network slice service.....	18
Figure 3-1 CHARISMA workflow of slice service creation	19
3.2.2. Caching service	19
3.2.3. Security service.....	25
4. Use cases refinements and clarifications.....	28
4.1. Transportation Use Case (Automotive/Buses/Trains).....	29
4.2. Multi-tenant Access and Video Broadcasting Services	34
4.2.1. Rationale of the UC and Objective	34
4.2.2. Actors.....	34
4.2.3. Network setup and operation	37
4.2.4. Requirements & KPIs	42
4.3. Updated Mandatory CHARISMA Requirements and KPIs	46
5. Conclusions.....	48
References	49
Acronyms	50

List of Figures

Figure 2-1: Refined CHARISMA architecture with integrated PHY, virtualized infrastructure, and CMO planes	7
Figure 2-2 CHARISMA CMO Technology/Solutions selection.....	10
Figure 2-3: Generalised CALn (n=0..3) node of the CHARISMA data plane architecture	11
Figure 2-4: Flexible CHARISMA data plane architecture (e.g. C-RAN instantiation)	11
Figure 2-5: CHARISMA data plane architecture instantiated for tram use case scenario.....	12
Figure 2-6: TrustNode standalone router board	13
Figure 2-7: TrustNode complete Platform.....	14
Figure 2-8: OFDM-PON layered architecture integrated into CAL nodes	15

Figure 2-9: Functional split options as discussed in 3GPP (figure: [9])	16
Figure 2-10: Upper PHY split as proposed by iCIRRUS [10]	17
Figure 3-1 CHARISMA workflow of slice service creation	19
Figure 3-2: CHARISMA Caching System Architecture	20
Figure 3-3: Work flow for initialising VNO caching services	20
Figure 3-4: Network-aware Prefetching Scenario	21
Figure 3-5: Work flow for Network-aware Prefetching Procedure Triggered by Cache Policy Manager	22
Figure 3-6: vCache peering in multi-tenancy (open-access scenarios)	23
Figure 3-7: Cache miss in vCaching (no traffic optimization)	24
Figure 3-8: Traffic optimization for vCaching	25
Figure 3-9 – Workflow for Provisioning of Infrastructure-level policies	26
Figure 3-10 – Workflow for provisioning of tenant-specific policies	27
Figure 4-1: Generic 5G transportation (rail/car) vertical sector use case scenario based on the hierarchical CAL infrastructure.....	31
Figure 4-2: Virtual Network Operators sharing the infrastructure of an access network (including an edge cloud). The inter-domain application on top is video broadcasting	36
Figure 4-3: Multi-tenancy in a video streaming application	39
Figure 4-4: Security scenarios on the multi-tenancy and video streaming use case	41

Executive Summary

This deliverable D1.2 “Refined architecture definitions and specifications” provides an update of the deliverable D1.1 “CHARISMA intelligent, distributed low-latency security C-RAN/RRH architecture” for the 5G CHARISMA architecture in its multi-layered control-, data-, and service-plane form. The CHARISMA architecture has been designed to achieve many of the 5G KPIs as defined by the 5G-PPP programme as well as other key technology drivers.

The goal of the CHARISMA architecture is to provide a 5G capability offering low latency, open access, and virtualised security (v-security). It has been designed to be hierarchical and distributed in nature by providing four converged aggregation levels (CALs). The mapping of the anticipated 5G network functions to the CAL architecture has been described in this deliverable.

The expected service and workflow lifecycle that CHARISMA will have to support is updated in this deliverable; particularly with regard to provisioning of network slices, caching and security services, which are also enablers to support the CHARISMA features of low latency, open access, and v-security.

In contrast to D1.1, where we have described nine different use cases, this deliverable focuses on two use cases namely the transportation vertical sector, and the support of VNOs in a multi-tenancy video streaming environment. The former is a typical example of a use case in a fixed access network, while the latter gives insight into a mobile scenario. Both use cases integrate the research of CHARISMA together with the features of a future converged access network as required by 5G. The selected use case scenarios will be used to verify the project results through specific and updated KPIs resulting from them. These KPIs will also be verified through the project demos, final demonstration and field trials planned for the final year of CHARISMA.

1. Introduction

In this deliverable D1.2 “Refined architecture definitions and specifications” of the CHARISMA project we present the refined results of the work package WP1, particularly focussing on providing updates to the CHARISMA architecture, to the CHARISMA service workflows, and the relevant Use Case (UC) scenarios. Since the release of D1.1 [1] the architecture has also been further detailed in WP2. A first description of the physical layer transport technologies and the architecture for the intelligent nodes can be found in the deliverable D2.1 [2]. In parallel to this deliverable the physical layer architecture for the transport technologies and the intelligent nodes has been further detailed in D2.2 [3]. A lot of architectural work has also been performed in WP3, especially with respect to the control and management plane. First, the CMO architecture has been described in D3.1 [4], the challenges related to security and multi-tenancy are given in D3.2 [5], and description of a content delivery service can be found in D3.3 [6]. All these investigations and intermediate results of WP2 and WP3 have influenced the on-going development of the general architecture of CHARISMA, which has been updated and is given in chapter 2 of this deliverable.

Two of the main actors in the CHARISMA ecosystem are the network operator and the virtual network operator (VNO). The CHARISMA control and management system enables the network operator to accommodate multiple VNOs over the same physical infrastructure. CHARISMA focuses on the provisioning of network slices, caching and security services, which are described in chapter 3.

While D1.1 described a rather large number of use cases, in this deliverable we are focusing on a subset of two main and updated use case scenarios, based on the transportation vertical sector, and the support of VNOs in a multi-tenancy video streaming environment. The selected use case scenarios will be used to enable verification of the CHARISMA project results through the specific and updated KPIs resulting from them. These KPIs will be verified through the project demos and final demonstration and field trials.

In order to be concise and to avoid the duplication of text, we refer to and cite other deliverables of CHARISMA and other projects where appropriate.

2. Architecture refinements

2.1. Overview

Since the previous WP1 deliverable D1.1 was completed there has been on-going work to refine and provide additional detail for the CHARISMA architecture, particularly as we proceed to demonstrate the features of the CHARISMA concept in the project field trials, and we need greater clarification of such design details. In this context, we report in this section the additional refinements and design detail of the CHARISMA architecture. In particular, having defined the individual PHY layer and the control, management and orchestration (CMO) layer of the CHARISMA network in earlier deliverables (e.g. D1.1 and D3.1) we now provide an integrated overview of these various layers, to provide a better understanding of the overall CHARISMA architecture.

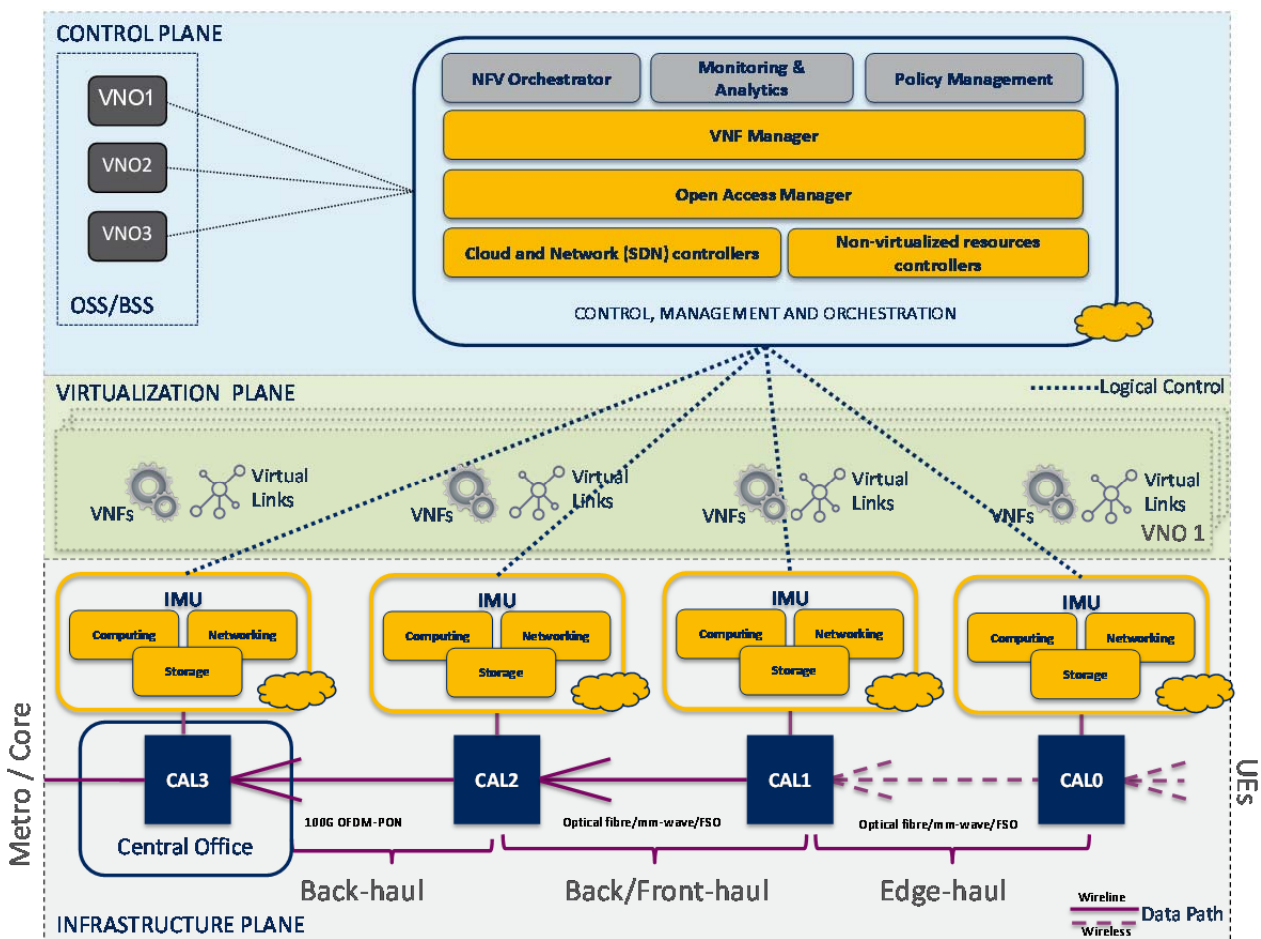


Figure 2-1: Refined CHARISMA architecture with integrated PHY, virtualized infrastructure, and CMO planes

Figure 2-1 shows the integrated schematic of the layer components of the CHARISMA architecture. Considering the various layers in turn, the bottom layer, labelled the Infrastructure Plane represents the data (PHY) layer, consisting of the hardware components, e.g. the IT and networking physical resources and devices, that make up the CHARISMA network. Here we can see the individual converged access levels (CALs) of the CHARISMA hierarchy, with multiple (tree-structure) links emerging from each CAL towards the next lower CAL level. The backhaul can be considered to lie between CAL3 and CAL2, whilst the span between CAL2 and CAL1 is a combination of the back and fronthaul. Finally, the link between the CAL1 and CAL0 is the edge-haul, at the periphery of the CHARISMA fixed network. Devices can either attach directly to the CAL0 (i.e. an access point gateway, CPE or vCPE) or to the CAL1. The purple solid/dashed (wireline/wireless as appropriate) lines between the CALs represent the user data flow (see below), whilst the blue dotted ones represented the control flow.

Located at each CAL is an Intelligent Management Unit (IMU), which can either contain physical networking functions (PNFs) of computing, networking and storage resources, or alternatively, these resources can be virtualised via VNFs. As envisaged in the CHARISMA concept, functions such as security (virtualised security functions, VSFs etc.) and caching (also vCaches etc.) are components of the IMU, which can be realized using the IT and networking resources located there. For example, the combination of IT and networking resources indicated in the IMUs of Figure 2-1 support the security and caching functions that are important features of the CHARISMA architecture. Although indicated as present at each CAL, the functions (virtual or physical) can also be chosen not to be present or instantiated at each CAL node of the CHARISMA hierarchy (see section 2.3 below for more detail).

Thus the Figure 2-1 indicates the Virtualization Plane above the PHY layer, where all the component elements of the IMU present at each CAL are realised. It should be noted that each of the IMUs above the CAL3, CAL2, CAL1, and CAL0 can be considered as cloudlets of the overall CHARISMA CMO, where networking intelligence has been devolved towards the edge of the network (indeed, intelligence is distributed between *all* the various CALs located from the central office (CO) to the edge.). We note that CAL3 may be associated with the main cloud infrastructure (e.g. a data centre) where the overall “centralised” M&O system is located; i.e. the CMO can be located at a DC, which may (or may not) be co-located with the CAL3 central office.

The Control Plane (in SDN terminology) is above the virtualised infrastructure layer, as indicated in Figure 2-1, and represents the CMO comprising the appropriate cloud and network (SDN) controllers. With the CMO only moderately centralised in a DC close to the CAL3 (i.e. the CO), this means that certain functions or “CMO agents” are therefore also distributed in multiple micro DCs (μ DCs) closer towards the edge of the network, each μ DC assumed to be co-located with a CAL and offering the associated IMU capability. This also means that not all CMO elements are therefore necessarily present in the DC, but rather that there can also be certain functions or “CMO agents” that are physically distributed in multiple μ DCs, so as to create a quasi-distributed control system.

The CMO also comprises Policy Management (which can also include the Service Management and Security Management) as well as Monitoring & Analytics, and the NFV Orchestrator. In addition, the CMO manages the NFVs via the NFV Manager (at both central and the distributed locations), and also manages the virtualised infrastructure (VI) via the VI Manager (VIM). Again, the VIM can be centralised or also have a distributed character, e.g. at the IMUs of the CALs further down the hierarchy.

Finally, on the left, we have the instances of the virtual network operators (VNOs) who can access the CHARISMA networking functions via the CMO, i.e. for their VNO slices and service instantiations (inc. cache

and security components). In this case, the CMO enables the VNOs to provide the required operations and business support systems (OSS/BSS) as well as have appropriate control over the network inventory, configuration and fault management, and service provisioning.

Data flow

Considering the PHY (infrastructure) layer of Figure 2-1 from the lowest levels (at the right-hand side), data flow upstream (i.e. towards the core) from devices, and enter the CHARISMA network from an access element such as a home gateway or access point at CAL0, or alternatively from a radio base station at CAL1. These take the user data and transfer them upwards to another aggregation level (CAL2, CAL3 or even staying in CAL1) where the intelligent element(s) of the IMU are able to route and decide the final destination point of the user data. Intelligent elements controlling the data flow are the combination of: 1) the Management and Orchestration elements; 2) the VNFM; 3) the VIM elements; and 4) any service-specific VNF (providing “services” such as telephony, video streaming, security etc.).

2.2. Control plane

Another key part of CHARISMA’s architecture being developed within the project is the control plane, an overview of which we present here for completeness. CHARISMA’s approach on the design and implementation of its Control, Management and Orchestration (CMO) plane takes the most cutting-edge technologies and open source tools, and extends them to provide the required 5G functionalities envisioned by the project, namely low latency, open access (multi-tenancy support), and virtualised security management. As indicated in Figure 2-1, CHARISMA’s CMO sits on top of the access network infrastructure and cloud resources implementing slicing and management functions end-to-end in an integrated and homogeneous way. CHARISMA’s initial CMO architecture was first reported in D3.1 in M6, where the deliverable D3.1 presented a state-of-the-art survey of the different technologies related to the management of virtualised resources, the key drivers for the CHARISMA CMO, and the initial architecture of the CMO. A more detailed description of the CHARISMA CMO was presented in D3.2 at the end of the first year (M12), with the deliverable D3.2 reporting CHARISMA’s security and multi-tenancy scope from the perspective of the CMO plane. D3.2 also presented a detailed description, and the requirements and workflows of the different components of the CHARISMA CMO including the Service Policy Manager, Monitoring and Analytics, Open Access Manager and the Virtual Security Functions (VSFs). The D3.2 deliverable also detailed the different technologies and solutions that were selected to provide the basis for achieving the goals of the CMO, specifically for the NFVO and VIM (see Figure 2-2 for an updated view of the CMO technologies and solutions selection). In the first year, intermediate demonstrations were conducted to present the partial achievements related to the Open Access Manager and VSFs.

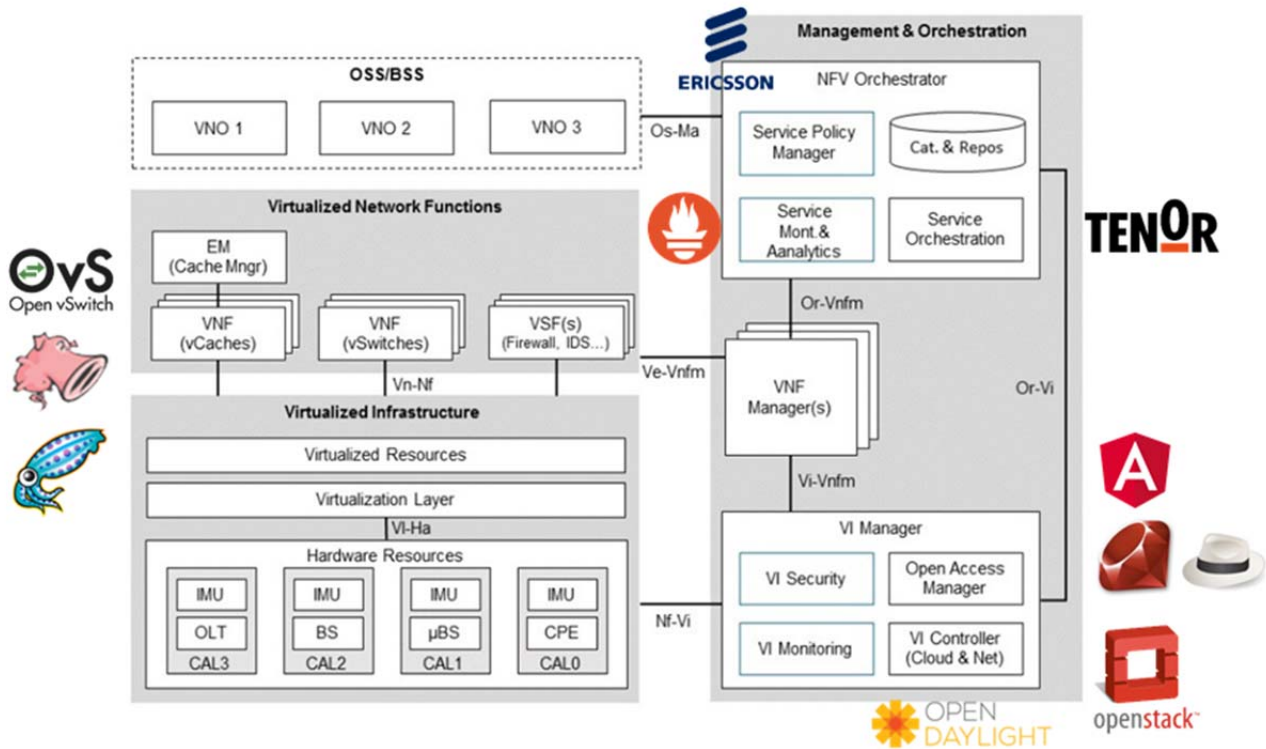


Figure 2-2 CHARISMA CMO Technology/Solutions selection

As already described in D3.1 and D3.2, the CMO presents three main architectural blocks that follow the ETSI standard recommendations: Management and Orchestration (MANO), Virtualized Infrastructure (VI) and Virtual Network Functions (VNFs). OSS/BSS functionalities are out of the scope of CHARISMA, but the CMO provides integration with OSS/BSS systems through its Service Policy Manager. Currently, CHARISMA’s CMO continues to be in active development, and this progress will be particularly presented in the future deliverable D3.4 in M24 of the project.

2.3. Data plane

The CHARISMA data plane architecture for 5G consists of the intelligent CAL nodes that provide: firstly, connectivity at the lower layers (L1-L3) of the data plane protocol stack; and secondly provide the intelligence for virtualising or accelerating network functions. Ethernet is the dominant L2 technology in the access domain today, which supports a large variety of L1 technologies (e.g. fibre, copper, RF wireless at different frequencies, etc.). Passive optical networks are the exception this, since they use a different L1/L2 technology, but can transport IP packets (L3) natively.

The added value of a CAL node in the CHARISMA architecture is the integration of an IMU, which provides computational resources, storage capabilities and switching/routing functions. The computational resources are not limited to CPUs or micro-controllers, but also include hardware acceleration by means of FPGAs or dedicated digital circuits. In CHARISMA it is foreseen, that the IMU provides the protocol layer of L3 and above by means of a virtualised infrastructure and hardware acceleration capabilities, as and when required (see Figure 2-3 below). Together with a CMO, this enables a distributed cloud with its high flexibility, where virtualised resources can be added or removed on demand. In addition these functions can also be shifted to other aggregation levels.

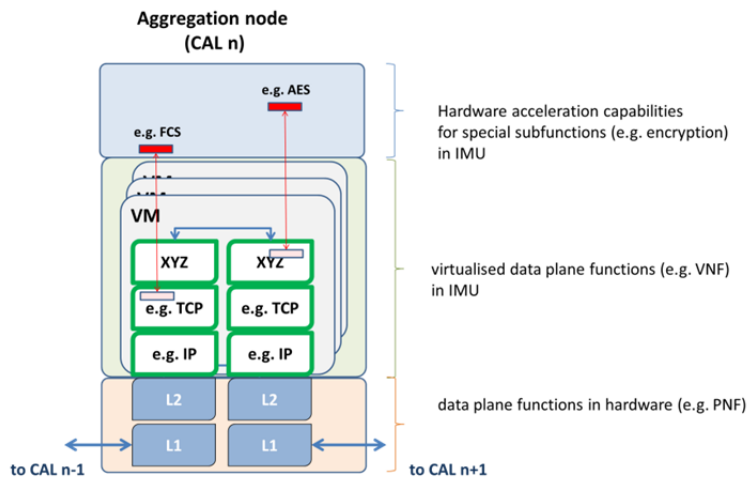


Figure 2-3: Generalised CALn (n=0..3) node of the CHARISMA data plane architecture

The CAL resources are controlled and managed by the sophisticated CHARISMA CMO system, an overview of which is offered in section 2.2. In a 5G environment the CHARISMA architecture allows the flexible mapping of 5G functions, including the new functional split of the fronthaul for a C-RAN implementation [10]. The explicit mapping of these (e.g. C-RAN functional split) functions to the CAL nodes is depicted below in Figure 2-4.

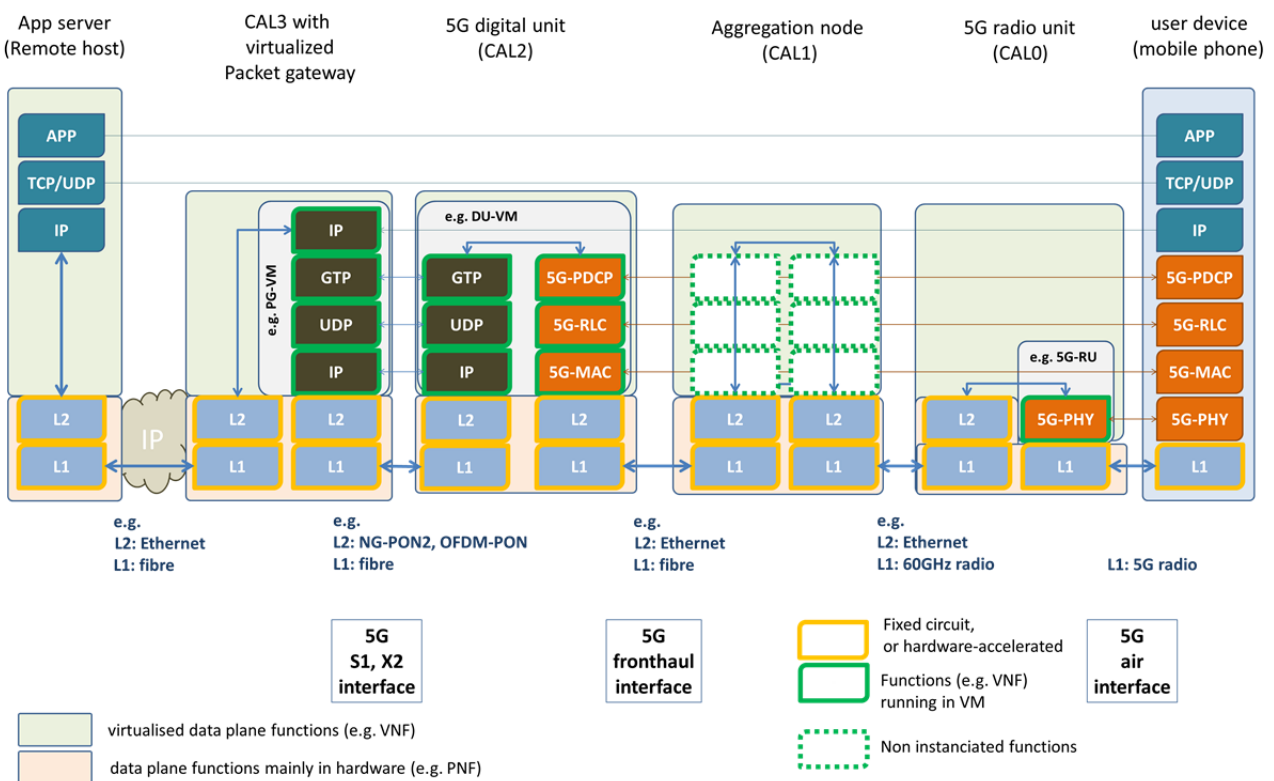


Figure 2-4: Flexible CHARISMA data plane architecture (e.g. C-RAN instantiation)

A 5G mobile user can be seen (on the right), that is connecting via an access network to an application server (left) as an example for a particular instantiation of the CHARISMA data plane architecture. We assume that 5G will use a similar protocol stack as compared to LTE in this example. The respective protocol layers are depicted in orange and prefixed with '5G'. The S1/X2 interface is depicted in brown, it is also assumed to be similar to LTE. The end user device contains the complete 5G protocol stack and the

TCP/IP stack for this application. The physical layer function (5G radio) is terminated in the 5G radio unit (RU), which performs the calculation of the radio waveforms; this function can also be virtualized. In the upstream direction, the 5G-MAC information is transported over Ethernet via the CAL1 node to the 5G digital unit (DU). The CAL1 node aggregates the traffic from a few radio units. At the digital unit (CAL2) the upper 5G protocol layers are terminated for all radio units connected in the downstream direction. The digital unit functions can reside in a virtual machine, which allows for high flexibility, e.g. if the traffic pattern changes, the entire VM can be moved to another node. From CAL2 going upstream, the IP packets from the end user are encapsulated into GTP packets and passed to the CAL3 node, which provides in this example a virtualised packet gateway (denoted by PG-VM in Figure 2-4). There, the original user data is transported via an IP network to its final destination – the application server. It must be noted that a possible hardware acceleration as depicted in Figure 2-3 is not explicitly shown here.

In Figure 2-5 the data plane architecture is mapped to another (alternative concrete) example; this time optimised for the transport use case scenario, which is also described in section 4.1. Here, it is assumed that an end user is using a device and connects via Wifi to an access point (CAL0), which is located in a bus or a train. The vehicle is connected in the upstream direction via 5G to the radio unit (CAL1) and the digital unit (CAL2). The packet gateway function resides again in CAL3, which manages the connection to the application server via an IP network.

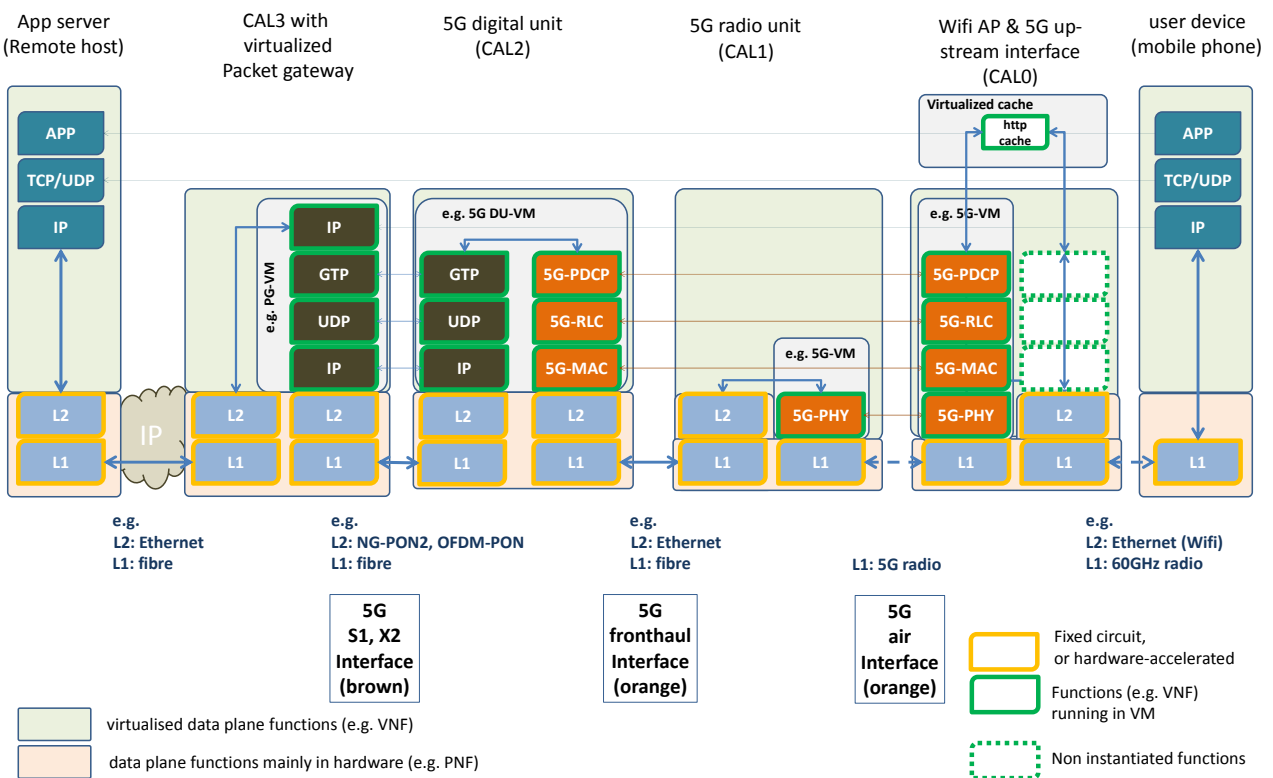


Figure 2-5: CHARISMA data plane architecture instantiated for tram use case scenario

In this example the CAL0 node hosts at least two virtual machines: firstly, the VM that provides the 5G connectivity to the mobile network; and secondly a VM to cache the content for the end user. CAL1, CAL2, and CAL3 provide the required functions for the 5G network, namely DU, RU and packet gateway. It must be noted that the functional split between 5G-MAC and 5G-PHY is just exemplary; also the mapping between virtual machines and protocol layer functions is shown as an example here.

If required, an additional aggregation layer between CAL2 and CAL1 can be inserted in order to connect a larger number of RUs to a digital unit.

2.3.1. Architectural refinements for physical layer technology

In this section we present the architectural refinements for the physical layer technologies developed in CHARISMA that have occurred since the earlier deliverable D1.1 was issued.

2.3.1.1. TrustNode architecture

Whereas D1.1 described the router architecture that was verified using a combination of development boards, the TrustNode router exists now as stand-alone device, as shown in Figure 2-6 that shows the first version of the circuit board. First tests of the FPGA configuration running on the board have showed promising results for the low latency IPv6 packet forwarding. The measurements were verified using an IXIA-Network analyser lent from HHI. The cut through latency for routing 6Tree IPv6 packets was measured to 2.5 μ s. The schematic of the router board was improved to fix minor bugs for the CPU junction. The second version is currently in fabrication. Figure 2-6 shows the complete router platform, including case, front panel and mainboard. The front panel which is used to configure the 6Tree prefix of the device is also ready to use. To make this feature accessible through the SDN orchestration of CHARISMA this value will be exposed using a REST-interface on the TrustNode-Atom-CPU.

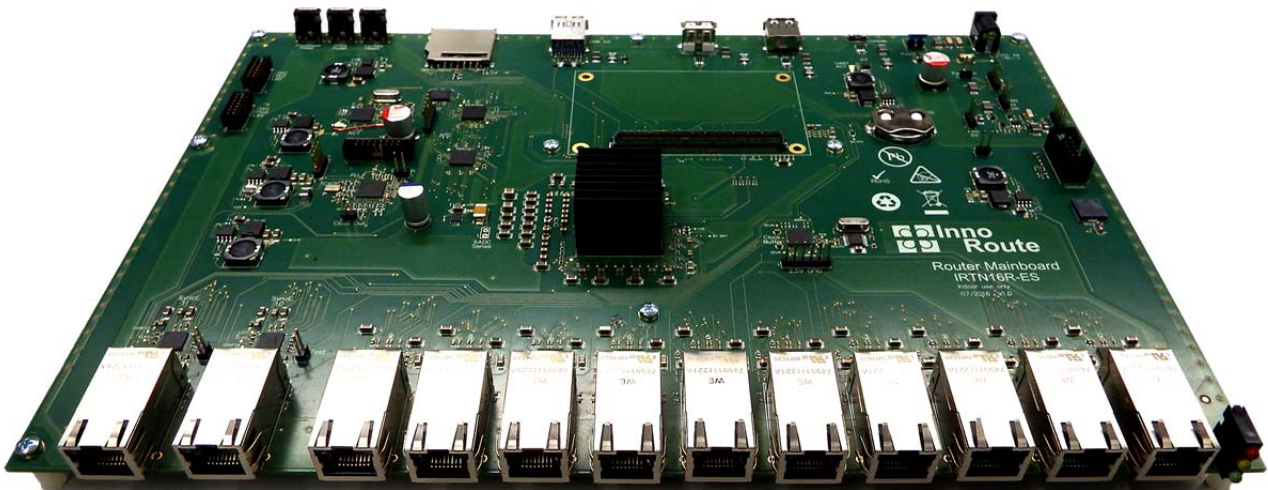


Figure 2-6: TrustNode standalone router board



Figure 2-7: TrustNode complete Platform

The logical position of the TrustNode router in the updated CHARISMA architecture shown in Figure 2-4 is at any of the CALs; however for the proposed CHARISMA demonstrators, the TrustNode will be located at the physical layer between CAL2 and CAL3 and on top of CAL3. To meet the requirements of the CHARISMA open-access slicing concept, the hardware-accelerated 6Tree routing will be extended to work in combination with VLAN-tagging without an impact on the fast routing.

2.3.1.2. 100G OFDM-PON

The general 100G OFDM-PON architecture has not changed much since the publication of D1.1. However, we can now apply the more detailed data layer architecture of Figure 2-4 to the OFDM-PON integrated into a virtualized 5G radio access network, as shown below in Figure 2-8. The CAL3 node consists of a packet gateway which runs in a VM using the IMU resources. The OFDM-PON connects a packet gateway located at CAL3 and a radio unit located in CAL2 in this application example. It can be seen, that the interface between the packet gateway and the OLT is exchanging IP packets, which are transported over the PON to the ONU. The CAL2 node integrates both a virtualized 5G digital unit together with the OFDM-PON ONU.

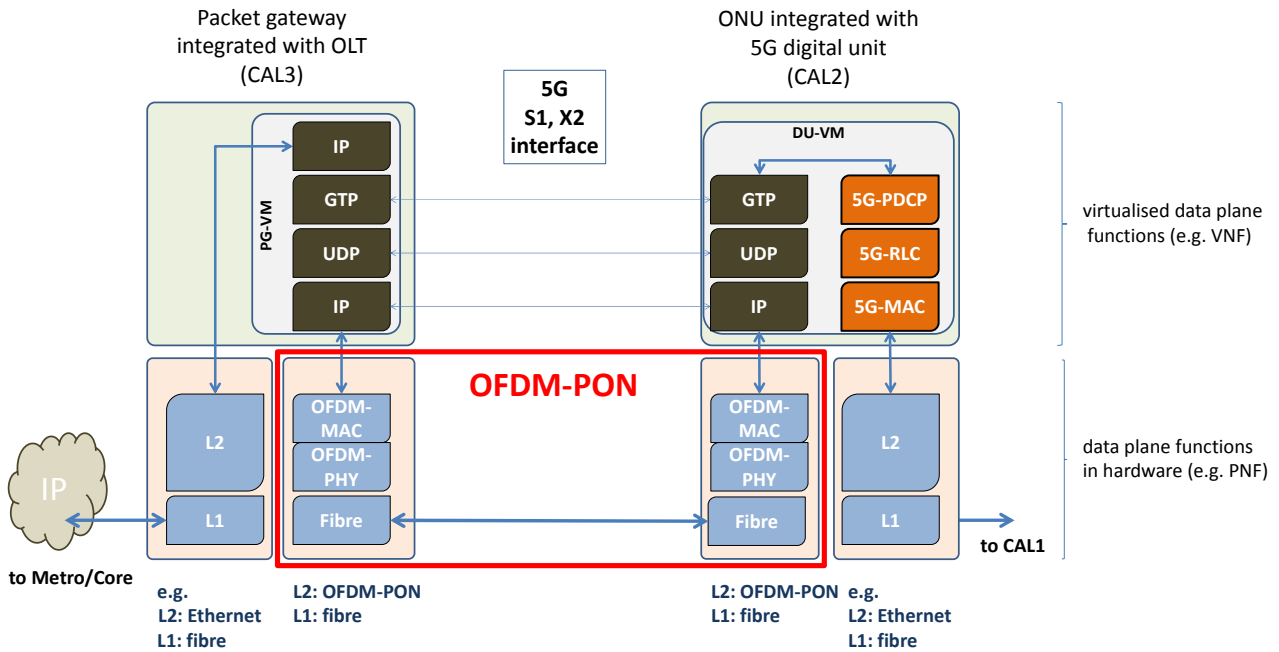


Figure 2-8: OFDM-PON layered architecture integrated into CAL nodes

The *data flow* is indicated by the blue arrows. Data coming from the metro/core is passed through the CAL3 and CAL2 nodes and targeted towards the end user who is connected beyond the CAL1. The description of the physical layer parameters of the OFDM-PON was already included in deliverable D2.1.

The *control flow* is not shown here, but can be separated into two parts. The external control flow exchanges information between the CMO and the PON control instance using an external interface. The most recent description of the respective physical layer and the predefined messages is described in the parallel deliverable D2.2. It can retrieve information from the OLT and also set some configuration parameters.

The internal control flow passes information between the control entities located at the OLT and ONU, and is transported in-band over the OFDM-PON physical layer. Its main function is to adapt to changing external parameters and to changes of the physical channel.

2.3.2. Fronthaul over Ethernet

Fronthaul design of 5G is currently a topic of great research interest and is being intensively investigated. The background context for the issues associated with fronthaul-over-Ethernet has already been discussed in D1.1 and can be summarized as follows: Typically, the data between RRH and antenna is uncompressed, leading to very high fronthaul data rate requirements. Extrapolating from current requirements to target requirements for 5G networks suggest that using the same techniques may require data rates in excess of 100 Gbit/s in future fronthaul networks; which is considered both unnecessary and impractical. Therefore CPRI over Ethernet is should only be considered for legacy applications, like LTE.

We can employ the 3GPP use of the LTE protocol stack as the basis for this discussion. In Figure 2-9 we can see the potential splitting points within the stack, as indicated by numbering the options from 1 to 8. Here a splitting point means that the lower (right) part is located in the radio unit (sometimes called the distributed unit) and the higher (left) part is located in the data unit (sometimes also called central unit). The concept of a flexible functional split has also been discussed, such that the splitting point can change

over time in order to adapting to different load situations, etc. However, this high flexibility comes with increased implementation and validation costs; and it is therefore beneficial to specify only one or two splitting points for the 5G fronthaul architecture.

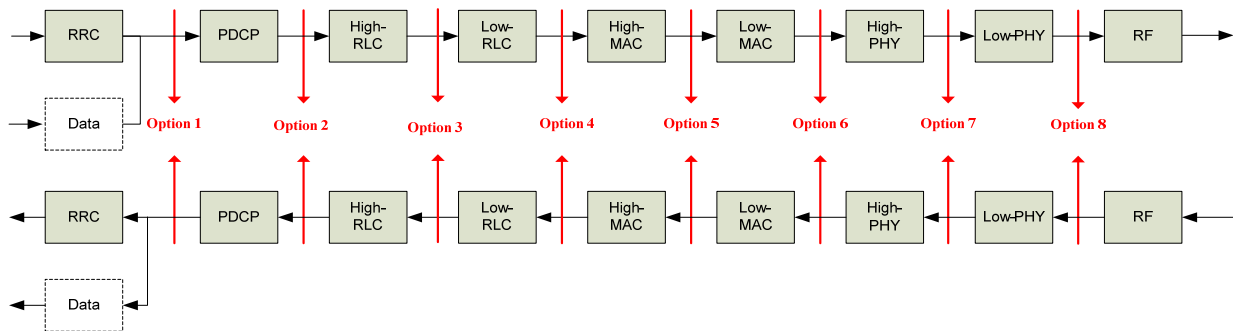


Figure 2-9: Functional split options as discussed in 3GPP (figure: [9])

The project H2020 iCIRRUS (grant agreement No 644526) has already performed an excellent overview of Ethernet-based fronthaul design. Therefore CHARISMA adapts the approach followed by iCIRRUS. However, we make the point that the innovative CHARISMA architecture can support any functional split of radio and data unit in principle, since the virtualisation layer is available at all aggregation nodes. This design allows the instantiation of a VM holding the data or radio unit functions. The different split options have also been discussed in the reference [10]: “... with the main finding being that a split between MAC and PHY might be most appropriate in terms of data rate reduction, and that other split interfaces at higher layers such as between MAC and RLC or between PDCP and RLC layers might be more interesting in terms of latency.” Therefore the experimental investigations as planned in the CHARISMA WP4 will also include an Ethernet-based fronthaul design with a functional split at option 7 (upper PHY split), as shown in Figure 2-10. Parts of physical layer function and RF are in the radio unit, while the upper layers are in the central unit.

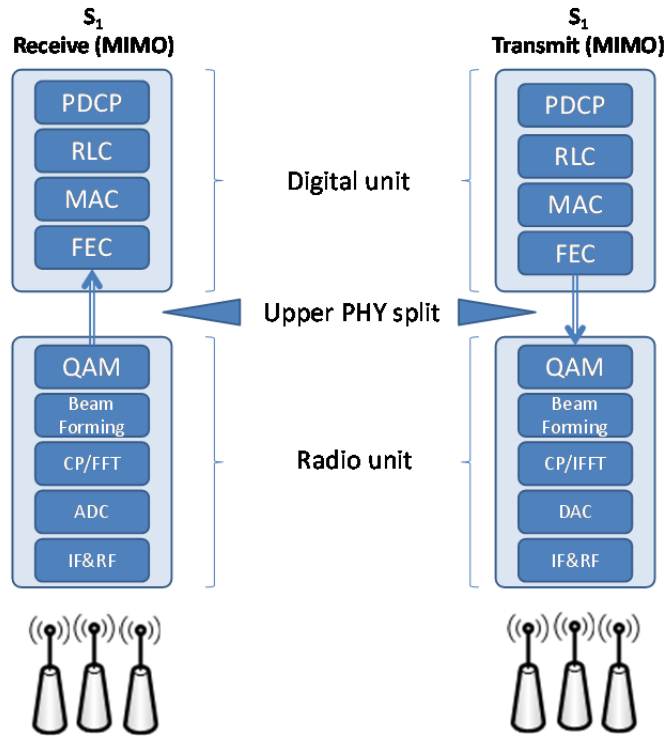


Figure 2-10: Upper PHY split as proposed by iCIRRUS [10]

3. Workflows & Service Life Cycle Design refinements

3.1. CHARISMA Actors and Roles Interaction

There is no update on the CHARISMA actors and roles interaction apart from what already was reported in the D1.1. However, as CHARISMA focuses more on the Network Operator (NO) and virtual network operator (VNO) actors and the interactions between them, the next section presents the workflows related to the main interactions.

3.2. CHARISMA Services Workflows

The two main actors in the CHARISMA ecosystem are NO and VNO. The CHARISMA CMO enables an NO to accommodate multiple VNOs over the same physical infrastructure, with CHARISMA allowing the provisioning of network slices, caching and security services. The subsequent subsections detail the workflow of each of the mentioned interactions.

3.2.1. Network slice service

Slicing of resources (physical or virtual) is one of the core requirements of 5G, and which therefore paves the way for multi-tenancy operation. As mentioned above, CHARISMA approaches slicing as a combination of the IT and network resources, i.e. a slice consists of a network part and an IT part. The CMO component at the heart of the slicing of the infrastructure is the Open Access Manager (OAM). A registered VNO can request a slice according to its needs via the CHARISMA portal. At the backend, the OAM first checks for any resource conflicts and proceeds with the creation of both a Network and IT slice in the case of there being no conflicts. The workflow of the slice creation is shown in Figure 3-1. As CHARISMA considers an infrastructure consisting of heterogeneous devices (SDN and non-SDN), the OAM takes into account the creation of a slice via the network controller for SDN-based devices as well as via the customized API for devices that do not support SDN. In this way, the OAM is able to stitch up a slice (Network and IT) across the CHARISMA CALs

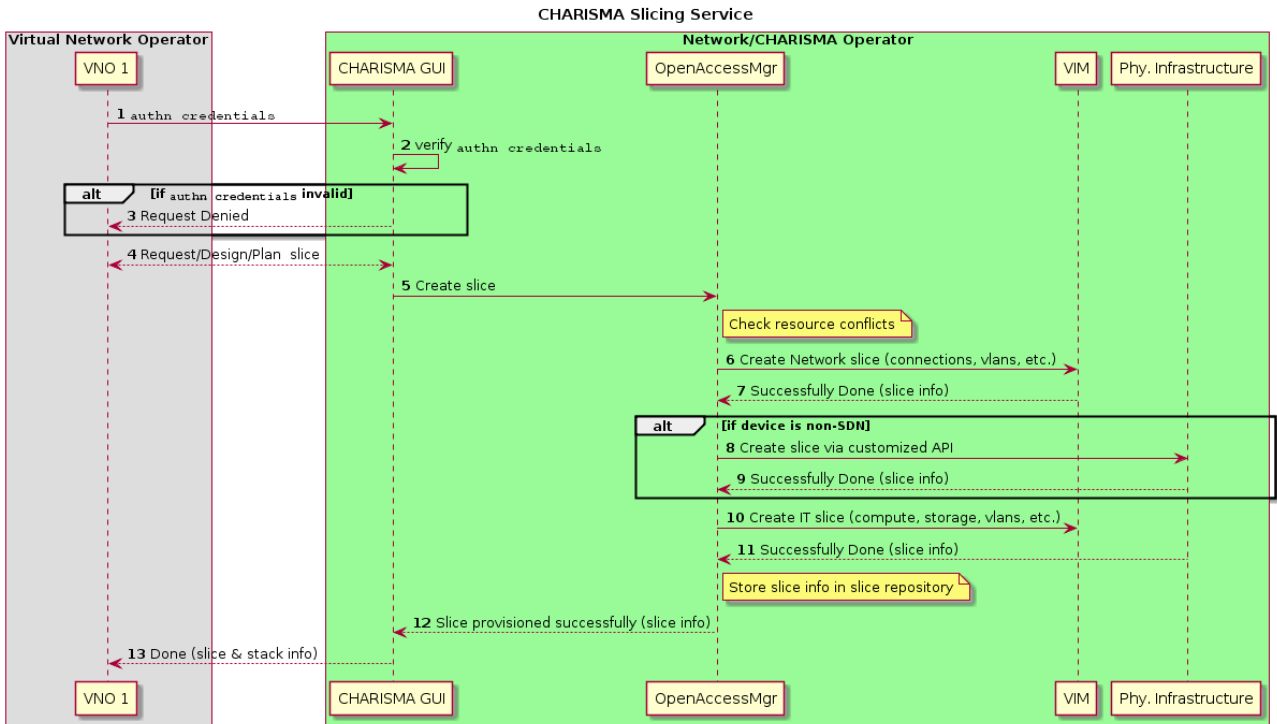


Figure 3-1 CHARISMA workflow of slice service creation

3.2.2. Caching service

The CHARISMA caching system targets the provision of an open cache access solution by virtualization of caches and the Cache Controller (CC). The cache nodes with instances of virtual services including caching and prefetching are provided in the different levels of the CHARISMA CAL architecture. This allows the dynamic allocation of virtualized caches and CC to different service providers (SPs) or Virtualized Network Operators (VNOs) over the same common infrastructure. Figure 3-2 shows the necessary components (in yellow) of the CHARISMA architecture required to realize the CHARISMA virtualised caching system.

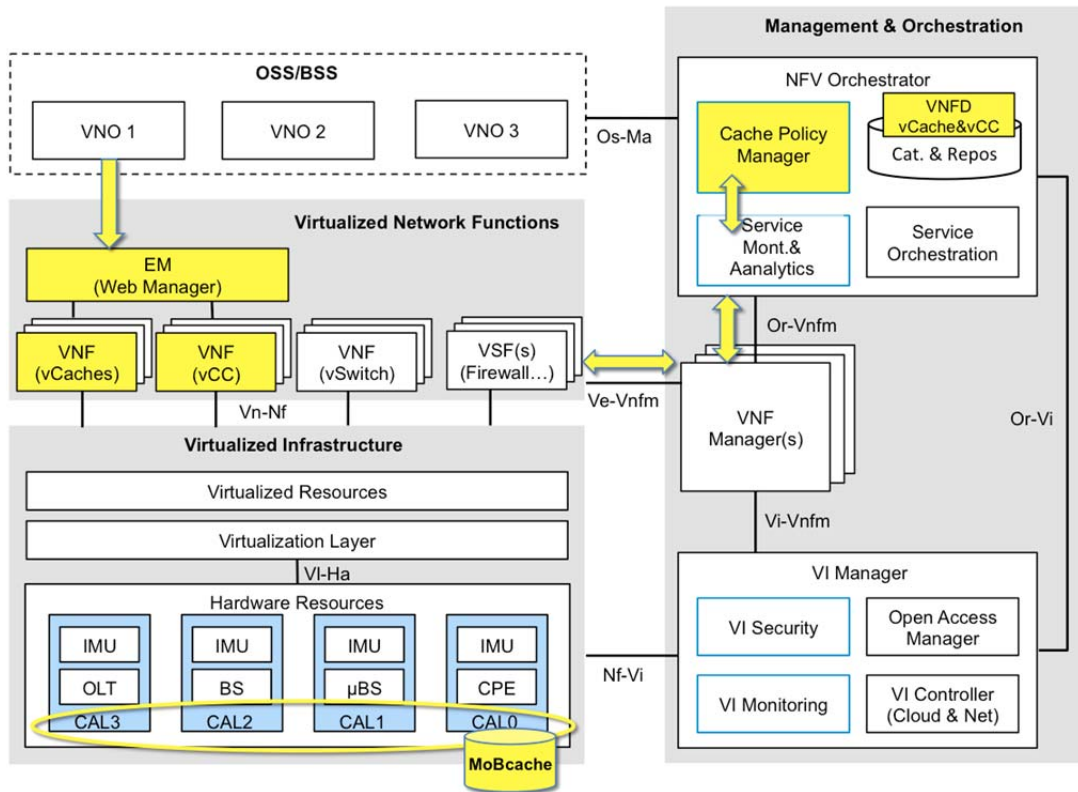


Figure 3-2: CHARISMA Caching System Architecture

The CHARISMA management system for caching services provides two levels of operations: i) at the global level managed by the Infrastructure Provider applying to all the tenants/VNOs; or ii) at the tenant-specific level managed by the specific VNO.

In the global management, the CMO is responsible for the creation/removal/configuration of slicing for each VNO, including a virtual Cache Controller (vCC) along with one or more virtual Caches (vCaches) located at different CALs. For example, Figure 3-3 shows the procedure of initialization of caching services including vCaches and vCC required by a VNO with the assumption that the network slice has been created and provisioned successfully.

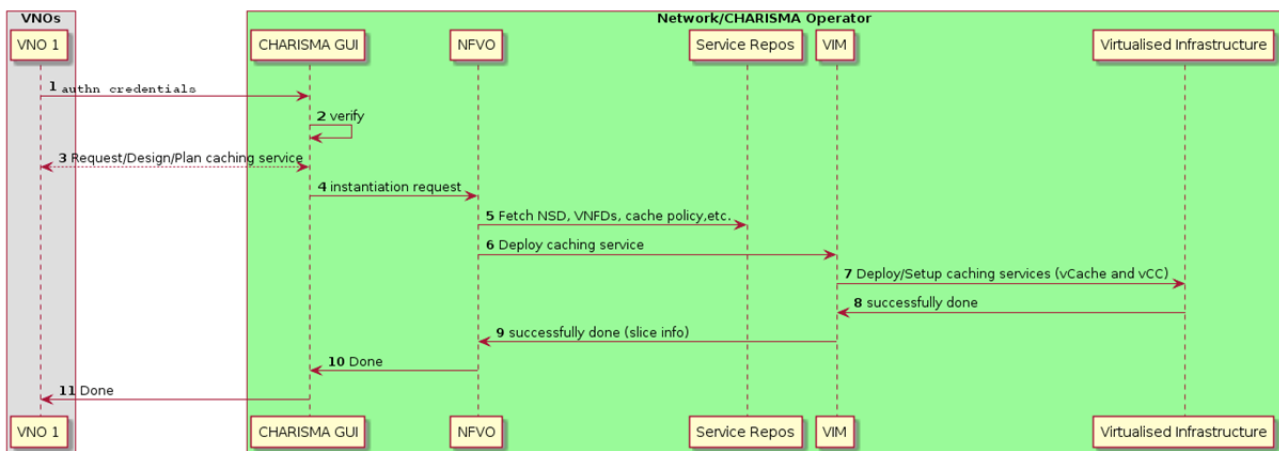


Figure 3-3: Work flow for initialising VNO caching services

Furthermore, the cache policy manager is designed to set caching policies that will apply to some or all of the virtualised and physical caching resources under its administration. Through interfacing to the Service M&A module, the cache policy manager is able to help improve user QoS by reconfiguring caching/prefetching settings. For example, a network-aware prefetching policy can improve QoS by downloading user requested content closer to the user, even as the user’s network performance is deteriorating. This is further presented in the following descriptions of Figure 3-4 and Figure 3-5.

The tenant-specific level management is operated by the vCC assigned specifically to a VNO. The vCC configures and manages the vCaches (caching and prefetching) allocated to the VNO. The vCC is able to configure the caching and prefetching programs running in the vCaches like service port, caching/prefetching algorithm. Figure 3-4 describes a network-aware prefetching scenario where MoBcache is located at CAL1 (a mobile router with multiple wireless interfaces and caching functionalities), and is considered as a mobile CPE switching from one AP to another, or from a LTE network to a WiFi network. The prefetching is performed even as the cache policy manager detects a network handover on MoBcache. The network-aware prefetching mechanism is applied in the bus use case described later in this deliverable in in section 4.1.

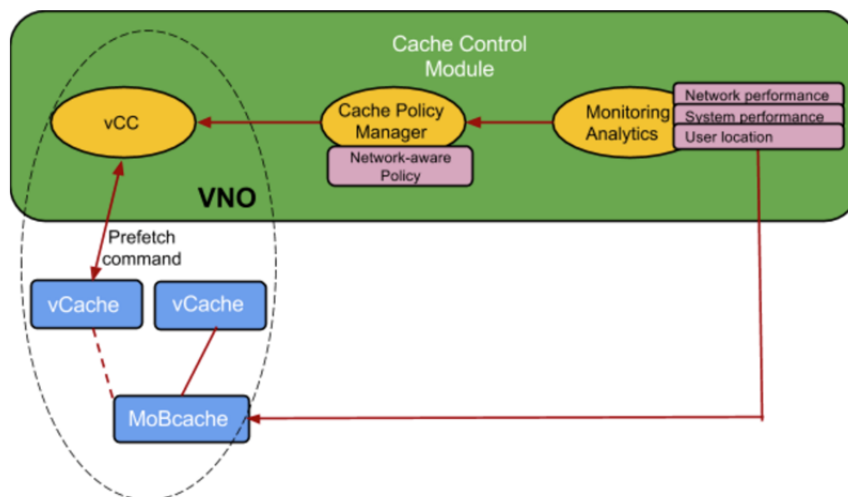


Figure 3-4: Network-aware Prefetching Scenario

Figure 3-5 further presents the procedure of triggering a prefetching action operated by the vCC while the network status meets the predefined rules in the cache policy manager. The MoBcache, vCaches and vCC have been dedicatedly assigned to a VNO that could specify caching policies that apply only to its network slice. The Service M&A module collects the information of the network performance (such as wireless signal strength, throughput and loss rate, etc.) at the MoBcache. The cache policy manager periodically communicates with the Service M&A module. The MoBcache at CAL0 connecting to an AP (or eNodeB) on CAL1 is able to switch to another AP on CAL1 where the vCache has been deployed. The cache policy manager is able to detect this network switch by the information of the networks status at the MoBcache as provided by the Service M&A module. Subsequently, the cache policy manager requests the vCC to trigger a prefetch that is performed on the vCache deployed at the AP that the MoBcache will switch to. As soon as the MoBcache connects to the new AP, the request of the end user connecting to the MoBcache can be directly served by the vCaches in this AP.

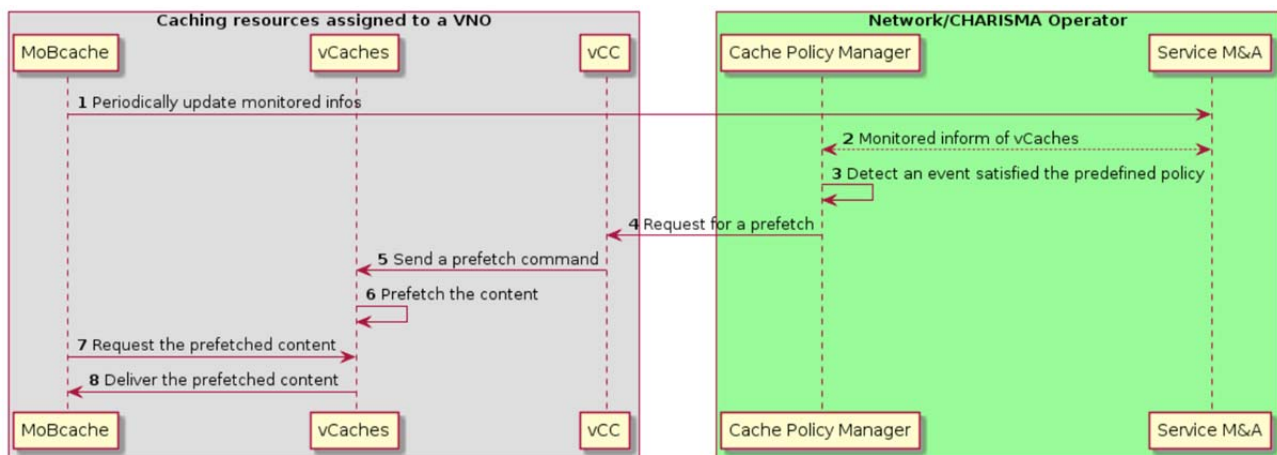


Figure 3-5: Work flow for Network-aware Prefetching Procedure Triggered by Cache Policy Manager

3.2.2.1. Cache peering in multi-tenancy environments

As has already been described in the CHARISMA deliverable D3.3, the virtualised character of resources and the potential collocation of vCaches of different VNOs in the same IMU, incentivise and facilitate the establishment of peering relationships between collocated vCaches. In the envisioned scenario, VNOs sharing the same infrastructure establish peering relationships between their vCaches. As a result, a cache miss (i.e., a requested item is not found in the repository of a vCache) at a vCache results in the vCache requesting the content item from its peering vCache. Figure 3-6 shows the envisioned environment. The figure shows a typical setup of the virtualised IT infrastructure (IMU) in OpenStack: a Network Node is responsible to route traffic towards the various VNFs in the available servers of the infrastructure, i.e. the Compute Hosts. At each Compute Host, a set of internal switches is responsible for delivering traffic to the appropriate VNF, ensuring the isolation of traffic between different tenants of the IMU. In this environment, vCaches of VNOs #1 and #2 are instantiated on the same Compute Host. Initially, and based on the default multi-tenancy configuration (supported by currently available tools e.g., OpenStack), the two vCaches are isolated and cannot communicate. The following steps are followed to support cache peering:

1. A shared network is created by the OpenStack administrator i.e., the infrastructure operator.
2. vCaches are each configured with an additional virtual network interface. This step is performed again by the OpenStack administrator.
3. The vCaches' new virtual network interfaces are configured to interact with the newly created shared network. This step is performed again by the OpenStack administrator.
4. The application-level components of the vCache instances, i.e. the Squid cache instances, are configured so as to established the cache peering relationship. This is performed through either the EM or the vCC, which applies this configuration using the IP addresses of the newly added virtual network interfaces (step 2). The IP addresses are provided to the EMs/vCC by the infrastructure operator.
5. The application-level components of the vCache instances are configured to apply certain rate limits so as to avoid the overconsumption of local resources due to peering requests. This is again accomplished through the EM/vCC.
6. (Optional) The application-level components of the vCache instances are configured with authentication credentials so as to enable peering interactions only with authenticated vCaches.

This is again accomplished through the EM/vCC. It is noted that this step is optional and targets an extra level of security assurance, i.e. the network configuration already assures that the vCaches receive requests from a legitimate counterpart (i.e. vCaches are explicitly given access to the shared network by the infrastructure operator); in this case, the mechanism here offers protection against malicious users getting unauthorised access to the shared network.

7. Upon successful configuration of the cache peering relationship, a cache miss results in the content request getting conditionally redirected to the peering vCache through the shared network. The decision on the redirection is made based on either re-actively or proactively received information on the content availability at the peering vCache. In the former case, the Internet Cache Protocol (ICP)¹ is employed by vCaches which actively query their peers about the availability of a requested content item. In the former case, Bloom-filter based descriptions of cached contents (i.e., the *Cache digests*) are pro-actively and periodically exchanged, denoting the availability of content at each side of the peering link.

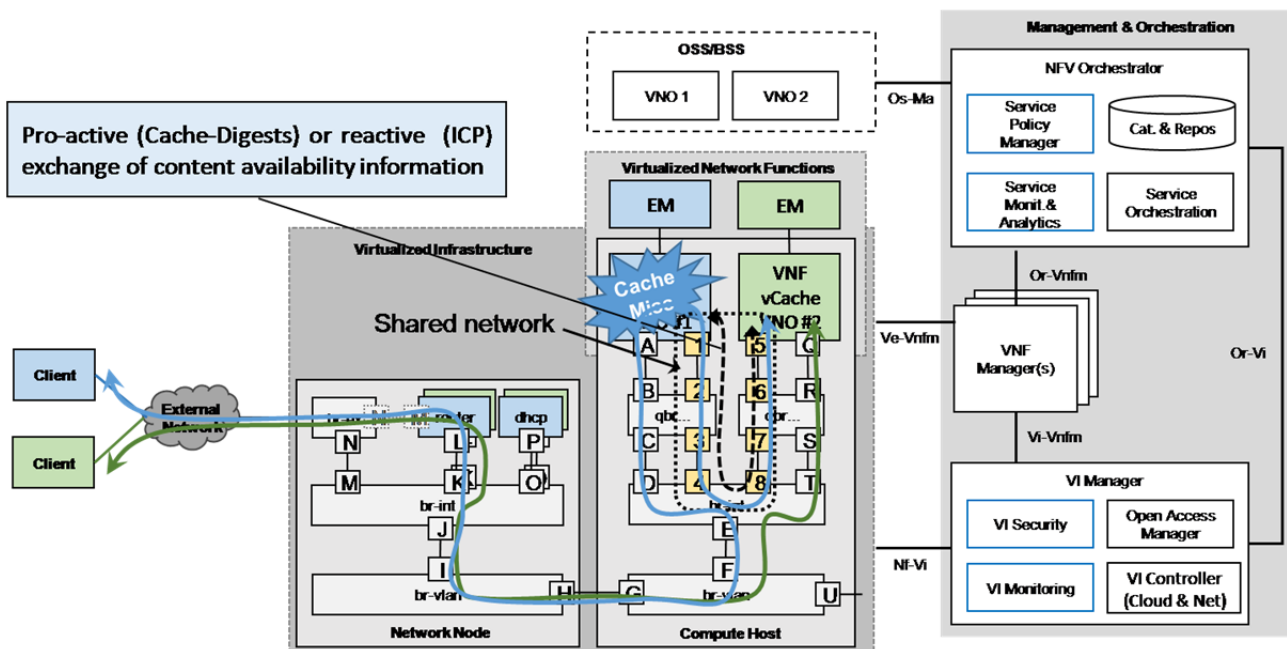


Figure 3-6: vCache peering in multi-tenancy (open-access scenarios)

3.2.2.2. Traffic optimisation for vCaching

As already described in D3.3, CHARISMA enables the optimised handling of traffic passing through virtualised caches. The objective is to take advantage of SDN capabilities and the availability of virtualised caches, so as to decide which traffic flows get through to the vCaches. The rationale of the designed solution builds on the observation that not all traffic leads to a cache hit and in this case, traffic suffers a delay overhead owing to the traversal of the vCache. This is shown in Figure 3-7 below: a cache miss results in the request packet traversing the protocol stack of the network node up to the vCache, only to return back to the network node on its way to the original content server. The content traverses the network node and compute host's protocol stacks once more in order to reach the vCache, before it follows the reverse route towards the requesting client. This requires the traversal of the virtualised infrastructure by 4 times per cache miss. If we consider that the majority of content items are typically of a low popularity (i.e. Zipf-

¹ RFC2186

law), it follows that this delay overhead is suffered for a substantial proportion of requests, i.e. typically most requests are for content that is not cached and will not be requested again.

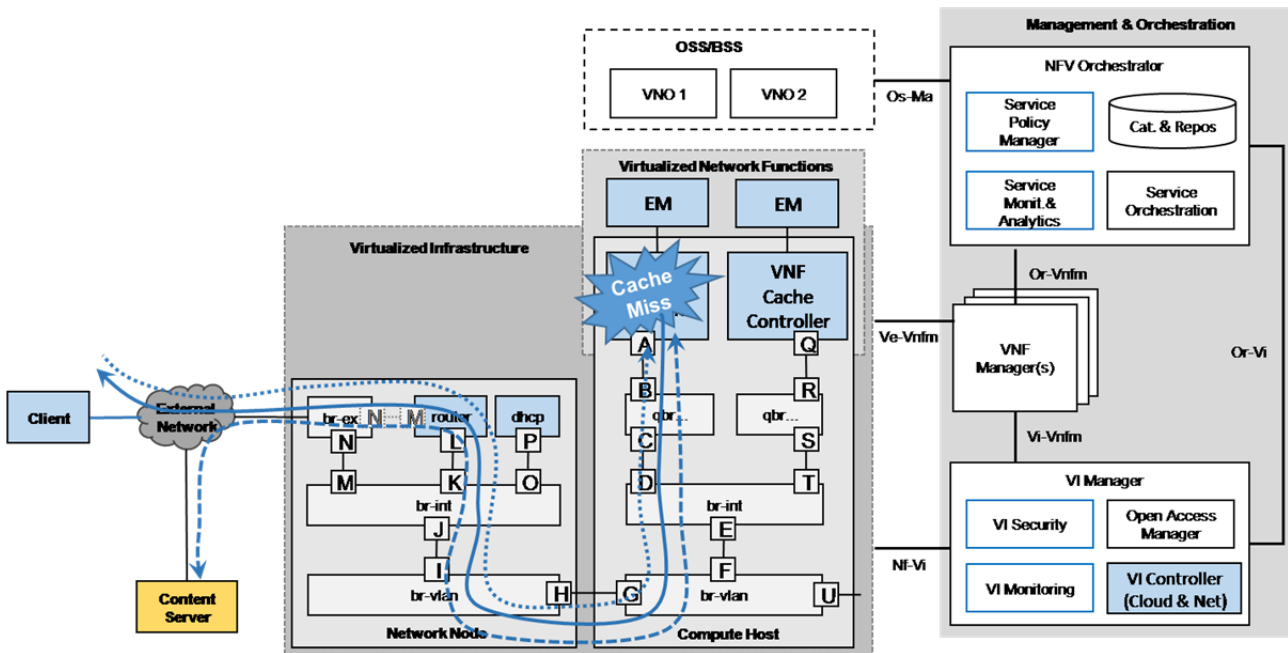


Figure 3-7: Cache miss in vCaching (no traffic optimization)

Based on this observation, the CHARISMA solution aims to identify those traffic flows that are likely to lead to a cache hit or miss, and therefore to subsequently decide on bypassing the vCache (or not, as the case may be.). The entire procedure for the support of this functionality is shown in Figure 3-8:

1. The vCC inspects the cache access logs of the vCaches (e.g., Squid instances) to retrieve information about cached content. Each entry is created per request and includes the following information:
 - URL column → IP of origin server
 - result code = HIT/MISS

A processing script at the vCC identifies all URL-HIT and URL-MISS entries which result in potential rules for flows traversing or not (correspondingly) the cache. A URL-HIT entry shows that the URL is related to cached content; subsequently the IP address leading to the URL belongs to the IP destinations that are likely to lead to a cache hit. A URL-MISS entry shows that the requested item was not cached, which subsequently triggers the content to be fetched from the content origin server and be cached for future reference. If this URL does not appear in subsequent URL-HIT entries, then the corresponding content item is not asked again, which contributes to the likelihood of an unpopular item. Subsequently the IP address leading to this URL is considered as unlikely to lead to a cache hit in the future. Future traffic flows to this IP destination will bypass the vCache, so as to save the delay overhead.

2. The identified IP destinations are delivered by the vCC to the VNFM (over the Ve-VNFM interface), annotated appropriately, so as to indicate if the flows with the corresponding IP destinations should bypass the vCache or not.

3. The VNFM delivers the annotated IP destinations to the SDN controller of the VIM (over the Vi-VNFM interface). In the designed solution the OpenDayLight controller is employed, receiving this information over its north-bound REST interface.
4. The SDN controller transforms the received information into flow rules and corresponding OpenFlow control messages, applying the rules to the `br-int` switch of the network node (over the Vf-Vi interface).

Note: As the delivery of the flow rules in steps 3 and 4 results in duplication of control messages and increasing complexity, we consider the direct delivery of this information by the vCC to the SDN controller.

Based on this configuration, flows destined to IP addresses associated with unpopular content will never reach the compute host where the vCache is instantiated.

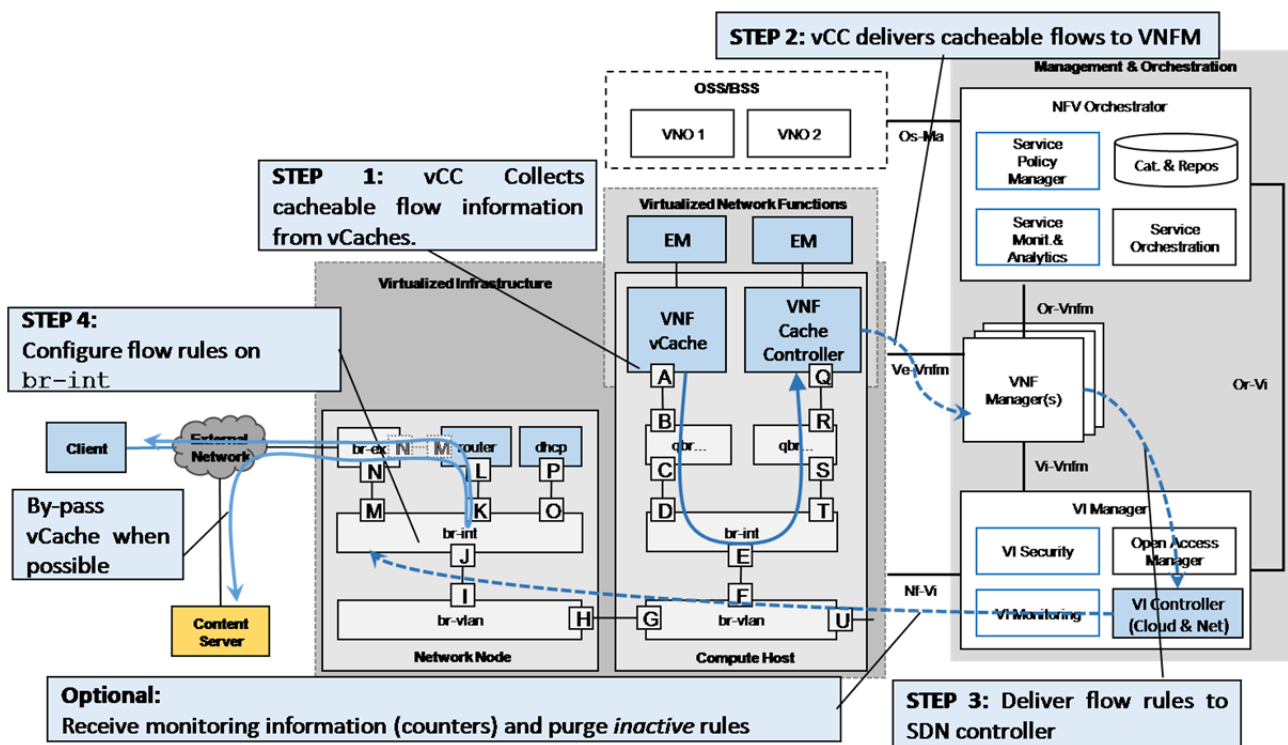


Figure 3-8: Traffic optimization for vCaching

3.2.3. Security service

The management and operation of the security-specific services provided by the 5G network can happen either at the general level (Infrastructure Provider level, and possibly applying to all the tenants/VNOs making use of the infrastructure) or at the tenant-specific level.

3.2.3.1. Security Workflow: Infrastructure Provider View

The infrastructure provider may set some basic security policies that will apply to all virtualised and physical resources under its administration, such as Tenant Isolation (Network slice) policies or Resource Hardening policies. Examples of the first case are policies governing the Authentication and Authorization of VNO users when access part of their Slice (IT and Network), policies applying to the Cloud Controller (every VNO will have its own project with their respective security groups in OpenStack to have at least minimum level of resource isolation among them), policies governing Workload isolation (all workloads

associated to a Tenant or Network Slice must remain isolated inside such Network Slice, at all levels: processing, networking, and storage), policies regarding Data encryption and confidentiality (at Network Slice/Tenant level) etc.

Examples of the second case (Resource Hardening) can be that all the physical machines part of the infrastructure must be securely hardened (including physical security), or that all VNFs to be deployed by the VNOs that are available in the Catalogue must be signed and verified. If a Signature of a VNF cannot be verified, the Catalogue should not allow it to be stored, or similarly, an orchestrator shall not proceed with its deployment.

Figure 3-9 reflects the workflow associated with the provisioning of Infrastructure-level policies. Once the policy is provisioned and instantiated, it should be triggered and enforced when the predefined conditions are met.

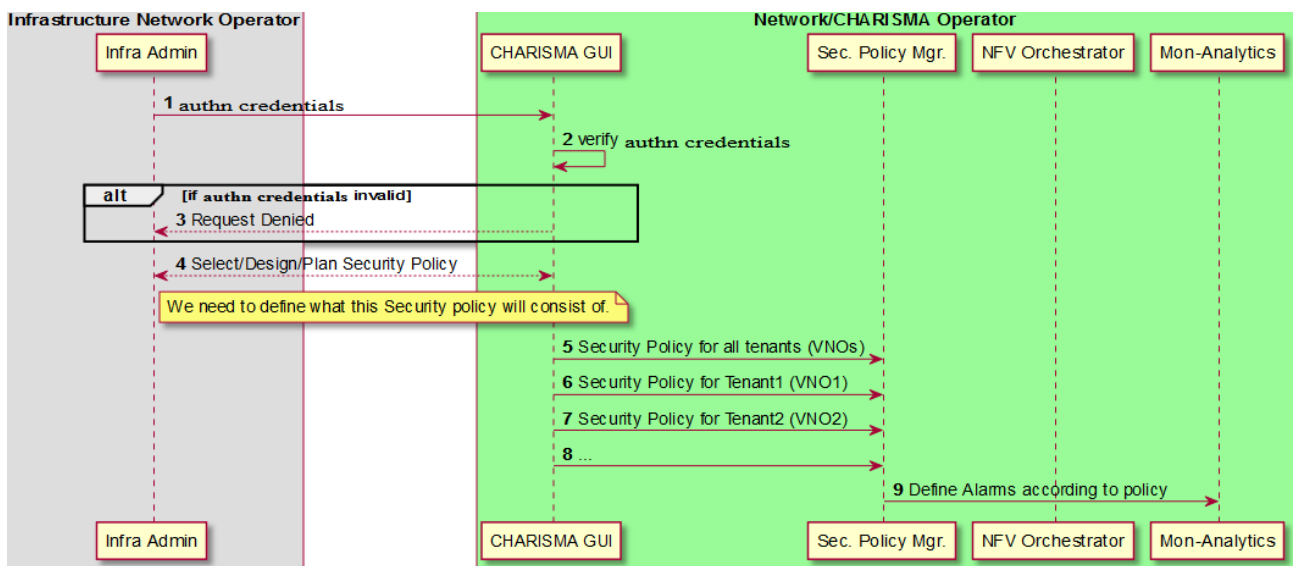


Figure 3-9 – Workflow for Provisioning of Infrastructure-level policies

3.2.3.2. Tenant-level Security Policy Examples

Additionally, each tenant could specify the policies that apply only to their network slice. Some examples of pre-configured security policies available at the tenant-level are:

- Deploy a virtual Firewall type in a specific Point of Presence in the case of where a DDoS attack is detected (If "alarm_of_attack = DDoS_attack_type1" then "deploy" "vFW1" in "Point of Deployment1");
- Deploy a virtual Firewall type in a specific Point of Presence in the case where two different DDoS attacks are detected, and also terminate VNF1 (If "alarm_of_attack = DDoS_attack_type1" and "alarm_of_attack = DDoS_attack_type2" then "deploy" "vFW2" in "Point of Deployment2" and "terminate" "VNF1")

Figure 3-10 reflects the workflow associated with the provisioning of tenant-specific policies.

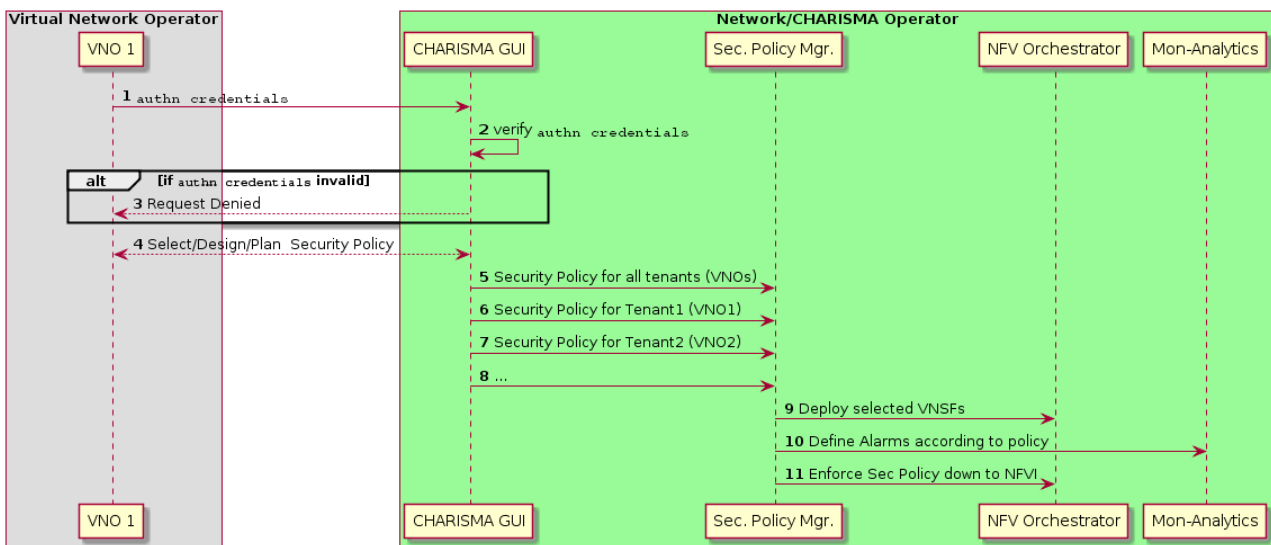


Figure 3-10 – Workflow for provisioning of tenant-specific policies

4. Use cases refinements and clarifications

In deliverable D1.1 nine use cases were defined representing the set of 5G use cases that the CHARISMA project has considered for the definition of the CHARISMA architecture. The purpose of these 9 use cases were twofold:

- First, to highlight how the key innovations of the CHARISMA solution would benefit the various actors and stakeholders (e.g., end-users, network/service providers) involved;
- Second, to be used to define the widest possible range of performance and functional requirements that the CHARISMA architecture should meet.

The nine use cases considered in D1.1 were the following ones:

1. Automotive – Trains
2. Automotive – Platooning, Vehicle Collision Avoidance
3. Automotive – Buses
4. Big Event
5. Emergency - Fire Fighters
6. Factory of the Future (IoT)
7. Video Streaming
8. Remote Surgery
9. Smart Grid

In this deliverable D1.2 we are now focusing on a subset of two main and updated use case scenarios, based on the transportation vertical sector, and the support of VNOs in a multi-tenancy video streaming environment. The reasons for selecting and updating the aforementioned use cases are the following:

- Since deliverable D1.1 was written a number of 5G KPI white papers have been published, with revised KPIs and a more detailed description set of requirements for the various vertical sectors. For example, the Expert Advisory Group of the European Technology platform Networld 2020 have issued their Strategic Research and Innovation Agenda (SRIA) document on “Pervasive Mobile Virtual Services” in July 2016 [1], and “Service Level Awareness and Open Multi-Service Internetworking” in September 2016 [2]. These are WPs emerging from the ongoing global effort into 5G research and exploitation, and reflect the latest thinking in this rapidly developing technology area.
- The selected use case scenarios can be used to enable verification of project results through specific and updated KPIs resulting from them. These KPIs can be verified through the project demos and final demonstration and field trials.

The rest of this chapter is organised as follows:

- Section 4.1 describes the updated Automotive Use Case scenario.
- Section 4.2 describes the updated Video Broadcasting Use Case scenario through the emergence of Virtual Network Operators in multi-tenancy environments.

Section 4.3 summarises the specific updated requirements and KPIs that will be supported by the CHARISMA architecture and will enable the verification of project results.

4.1. Transportation Use Case (Automotive/Buses/Trains)

4.1.1 Rationale of the UC and Objectives

The key objective of the transportation use case is to ensure that CHARISMA can support provide the technology developments so as to provide 5G digital connectivity to transport infrastructures and passengers; primarily rail and road, e.g. as outlined in the “5G Automotive Vision” white paper (October 2015)[13]. In addition, from a societal perspective, another objective of the use case is to make the time people spend while traveling (e.g. on trains, trams, buses etc.) productive and/or rewarding, e.g. as specified by the S4 5G-PPP KPI for the stimulation of new economically-viable services of high societal value, such as UHD TV. Particularly with respect to the transportation vertical sector, 5G advances for massive connectivity can also aid more continuous, real-time monitoring and supervision through sensors on trains and track, or on the roadside, for intelligent transportation systems (ITS). In this regard, ultra-reliable and low-latency vehicle-to-vehicle and vehicle-to-infrastructure communication is needed for future ITS systems and automated driving, which are closely related to CHARISMA’s key targets of low-latency, alongside security and open access. Achieving these requires utilizing the full benefits of network functional virtualization (NFV) and orchestration, with increased softwarisation and slicing of network resources, alongside the developments in mobile edge computing (MEC). Together, these create the rationale for the CHARISMA use case scenario for the transportation vertical.

Key objectives for the transportation use case therefore include:

- Ultra-high-speed multi-Gb/s data connectivity for train/tram/bus passengers, and on-board data systems, along with low-latency e2e transport and application (e.g. video content) access times.
- Provision of secure, ultra-reliable and low-latency communications for transportation management and control systems, assuring safe operation, over a multi-tenancy (both for VNOs, as well as different (potentially competing) transportation operators) environment.
- Use of network orchestration to deploy and optimize the CHARISMA 5G network according to varying demands, e.g. allocate network capacity to follow transport traffic patterns, and provide network slices with appropriate QoS to passengers (end users), and network management and infrastructure control systems.

In the following sections, we discuss quantification of these transportation use case objectives via appropriate key performance indicators (KPIs), as well as the CHARISMA technology approaches being adopted to achieve these objectives. We note that these technology approaches are considered more generic and abstract, as compared with the specific technology solutions that will be adopted for the CHARISMA field-trials and test-bed demonstrators (which are the subject of WP4 investigation and are described in detail in the deliverables associated with WP4.) We also discuss how these transportation use case KPIs can be appropriately verified, both in the generic case, as well as for WP4 field-trials and test-bed cases.

4.1.2 Actors

Infrastructure Operators (Network and Transportation):

The transportation use case is interesting since the overall (vertical) infrastructure required for a 5G solution will tend, inevitably, to be owned by two current operators, who are normally independent and operate in very different spheres to each other: the transportation infrastructure, e.g. train operator(s), e.g. rail track infrastructure and rolling-stock owners. Indeed, for the automotive scenario, the transportation infrastructure would tend to be a state or municipal authority that owns and maintains the road infrastructure, with the vehicles (e.g. private cars, or a fleet of buses and trams owned by a bus/tram operator) representing another infrastructure owner. The telecoms infrastructure, e.g. mobile network operator, is the other main infrastructure actor of the transportation use case. Together, the transportation and the telecoms network owners can either co-operate, or indeed coalesce to create a single overall infrastructure operator. The operators need to work together to be able to map all the infrastructure equipment into the pool of available resources that can then be leased to virtual network operators in the form of virtual infrastructures. In addition, it could be that the Network Operator also has to lease space (e.g. gantries, bridges, and other street furniture (lamp and signage posts), or other spaces of land) from the transportation operator, in order to install their telecommunications equipment (e.g. RRHs, base stations, and even active remote nodes (ARNs) and larger intelligent aggregations nodes higher in the CHARISMA CAL hierarchy) alongside the train trackside or roadside.

Virtual Network Operators:

Multiple VNOs can lease the available infrastructure (either from a single unified infrastructure owner, or from the individual owners of the various infrastructure component parts) to operate an overall virtual infrastructure and provide network services to the end users (travelling public, and vehicles). The VNO actor can be considered as the integrating actor, to seamlessly operate the composite infrastructure components as a single virtual infrastructure, to provide network services to end users. In order to achieve this, it leases the available resources from the various infrastructure (Network and Transportation) Operators.

Service and Content Providers, including Application provision:

For the transportation use case, the service/content/application provider is an entity that produces and gives to end users specialized services, e.g. intelligent transportation systems (ITS), software for collision avoidance, internet access, and video streaming services. Such a provider can either offer a service direct to a human end-user (e.g. internet, video streaming, etc.) via human-type-communications (HTC) data formats, or to a machine/sensing-device, (e.g. ITS, and signalling/control services) for a machine-type communications (MTC) system. The different HTC and MTC data flows will each have their own performance requirements and quality of service/experience requirements, according to the specific services being provided.

End Users:

The end users of the CHARISMA services for the transportation use case are the travelling public, either journeying on trains, buses and trams (i.e. public transport) or using their own private vehicles. They subscribe to the VNOs for access to the broadband services and to the service and content providers. From an intelligent transport system (ITS) and IoT perspective, the actual vehicles (cars, buses, trams train carriages) can also be considered as the end-user devices, since all the multiple sensing networks, monitoring and control/signalling systems for a secure, reliable, high-availability, and ubiquitous ITS system will also need direct access to the CHARISMA infrastructure.

4.1.3 Setup and Operation

Figure 4-1 shows the CHARISMA concept for the 5G transport infrastructure use case scenario. Here, we can see the multiple challenges of providing a 5G architecture with technologies supporting communications for seamless end-to-end high QoS and QoE services across a shared, and heterogeneous 5G network for the transportation (trains and cars) vertical sector. Particularly for the high-speed transportation sector (both trains and cars), the main challenges of ITS systems are speed, latency, and the volume of data, and also security & integrity of the links. When a communications link between any of the trains is disrupted then all or part of the system might have to enter a standalone state until the problem is remedied. Depending on the severity of the communication loss, this state can range from vehicles temporarily reducing speed, coming to a halt or operating in a degraded mode until communications are re-established. If the communications outage is permanent, special operations must be implemented which may consist of manual operation using absolute block or evacuation. As a result, high availability is crucial for proper operation. There is also a need for extensive upgrading for information analytics at high speeds; i.e. milliseconds can be the difference between life and death. Thus, very robust systems management has to be adopted, including the SDN approach of MANO, with redundancy on communication links, backup of states and support of standalone mode for at least a short time period with situation prediction.

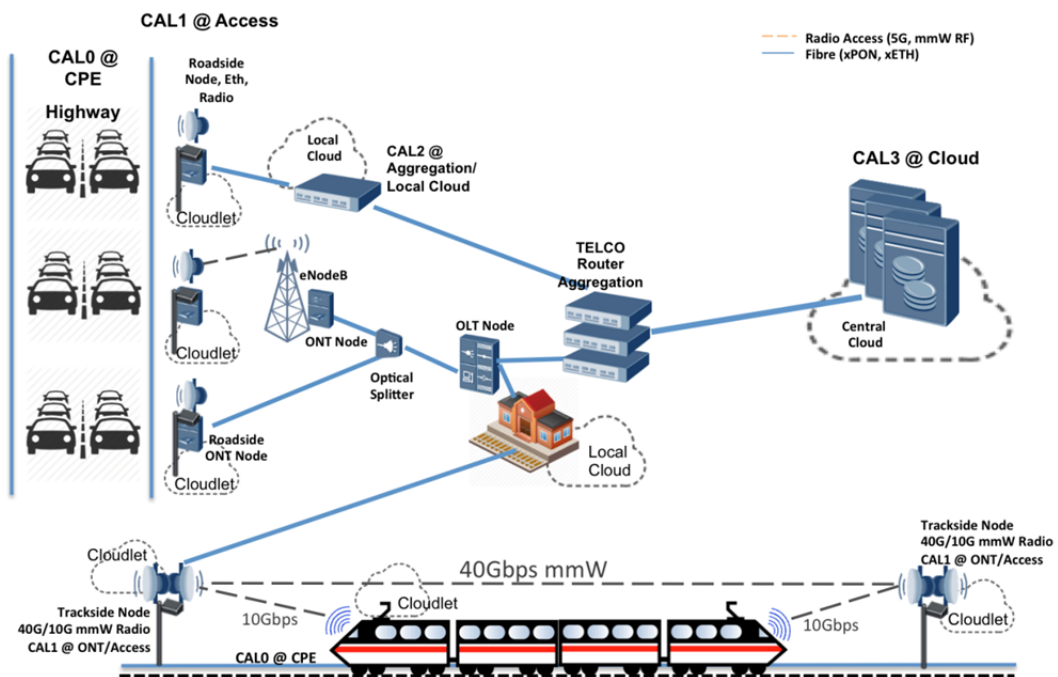


Figure 4-1: Generic 5G transportation (rail/car) vertical sector use case scenario based on the hierarchical CAL infrastructure.

A key architectural objective for CHARISMA, particularly for the transportation use case is to include the flexible distribution of network functions, from both core and radio access network, aided by functional virtualization, such that the use of virtualized network functions (VNFs) allows the fast development and deployment of network services (inc. security). Distributed security (inc. resilience) across the network, requires the CMO to feature virtualized security functions (VSFs) in its security management/policy system for precise control at the edge, as well as monitoring & analytics, and isolation of network resource slices.

CHARISMA’s scaleable hierarchical approach, characterized by its CALs (each having its own scaled intelligent management unit (IMU) performing data storage/caching, processing and routing functionalities) therefore is key to enabling such transportation use case performance. Achieving low latency (KPI <5ms) requires data be handled (i.e. routed and/or processed) as close to where it is required or at its lowest common aggregation point; i.e. cloudlet and fog computing (mobile edge computing, MEC) concepts are therefore well supported. Each CAL node is designated with a number, to signify its level in the hierarchy. For CHARISMA, the CALx hierarchical approach enables technical solutions to challenges such as service continuity and high QoS to moving vehicles; varying network conditions and performance, for coverage, throughput, and low latency; network resource usage optimization; and intelligent network services, such as hierarchical and virtualized caching (vCache); and multi-tenancy with isolation.

In order to achieve efficient content delivery services with high throughput and low service latency, caching functionality can be virtualized and enabled on network devices located in different levels of the CHARISMA architecture. The SDN/NFV based cache controller enables dynamicity and flexibility for a caching solution, as the SDN allows flexible provisioning for the caching functions itself whereas the NFV provides the required rapid automatic network provisioning. Furthermore, context-aware cache proxy management enables dynamic caching resource management, cache service optimization and improved service scalability.

Overall, CHARISMA’s hierarchical cloud infrastructure can permit high data-rates while enabling low latency, secure and scalable services, and provide MEC at various levels of abstraction, as a means to satisfy the transportation vertical use case requirements.

4.1.4 Requirements & KPIs

The automotive vertical use case has its own set of requirements, based upon the dynamic and high-speed environment that the 5G network architecture is expected to operate in, e.g. [13]. The associated KPIs and verification metrics are still a matter of definition, but we provide some initial parameters that will enable verification of the performance of the CHARISMA architecture in the transportation use case context. We have identified those KPIs from reference [13] that are of particular relevance (and overlap) with CHARISMA’s objectives, and which we can assess and measure within the current scope of the CHARISMA project. Naturally, within the transportation vertical, there are many more associated KPIs (e.g. as identified in [13]) that lie outside of the scope of the CHARISMA project, and which we therefore do not explicitly mention here; rather, these can be referenced in the citation [13].

Index	Description	Importance	KPIs and Verification
1	High bandwidths	Mandatory	1 Gb/s peak per user; 10 Gb/s peak to vehicle/train video streaming 100 Mb/s or 50 Mb/s [8]
2	Low latencies (particularly for driverless cars)	Mandatory	<5ms (e2e) <10ms (application access time)
3	Multi-tenancy (both VNOs and transportation operators)	Mandatory	Ability to host 2 or more VNOs.
4	Security	Mandatory	- Isolation between slices, e.g. see the note associated with KPI #3 of the Video Streaming

			use case, section 4.2, below.
5	Large no.s of connections (both MTC, and HTC) (Industry 4.0, FOTF) D2D, V2V, V2X	Mandatory	- 10-100x more connected devices - 1000x higher mobile data volume per geographical area
6	Up to 500 km/hr speeds Location precision <0.3m [<0.1m for vulnerable road users, e.g. pedestrians/bikes]	Optional	KPIs as per Description [Note, that CHARISMA use case is for buses/trams, and demos will be <100 km/hr.]
7	Predictable usage patterns and temporal timings	Optional	QoS/E, Mean Opinion Score (MOS) PV, PS, BR+BS [8]
8	Patchy coverage due to physical barriers (cuttings, tunnels etc.), & Reception within vehicle i.e. High availability & reliability	Optional	99.999% packet loss rate of 10^{-5} or less.

In this context, of describing how the transportation vertical use case can be mapped onto the CHARISMA architecture, it is also worth noting the table below, copied from the reference [11] that indicates a more generic approach to formulating KPIs in a rapidly evolving 5G networking infrastructure scenario.

Table 4-1: KPI general definition

Layer	Technology/Standard	KPI
Physical	PHY of wireless and wired IEEE standards 802, 4GPP, 5GPP, radio propagation models, fiber propagation, Modulation, link budget	SNR, Availability, Coverage, propagation delay, attenuation, spectral efficiency, Collisions, power.
MAC	MAC wireless and wired IEEE standards 802, 4GPP, 5GPP, BoD, FEC, QoS, Synchronization, SDN, Security.	BER, Access Delay, Contention, queuing time, Overhead, framing efficiency, switching time, frame loss, frame priority.
Network	IP, IPv6, NAT, Flow Classification, IPSec, mobility, handover, Proxies, Network Coding, Firewall, SDN,	PER, Inter-networking adaptations needed, integration effort, goodput, APIs functions for operator interactions, QoS mapping.
Transport	TCP, UDP, other protocols, Proxies, Security (TLS), firewall, CDN	RTT, protocols efficiency, throughput, delivery time, channel utilization, fairness, friendliness
Application	SIP, HTTP, FTP (and all other main internet protocols), Streaming, Proxyng, SCADA, Shaping, DPI, firewalling,	Timeouts, QoE, overhead, protocols malfunctions, application blocks, quality stepping and degradation, service availability
N.A.	Other aspects of network integration, deployment, costs assessments,	System Capabilities, costs, complexity

4.2. Multi-tenant Access and Video Broadcasting Services

4.2.1. Rationale of the UC and Objective

The high level objective of this use case is to ensure that CHARISMA can support the emergence of Virtual Network Operators (VNOs) in multi-tenancy environments. The virtualized physical resources of a 5G network infrastructure operator are shared by the VNOs enabling the rapid deployment of services, the flexible and efficient utilization of the required resources, and the differentiation of the offered services against competitors. In this broad context, CHARISMA builds on the observation that the coexistence of VNOs over the same infrastructure results in inter-domain traffic scenarios, where traffic flows between VNOs so as to accommodate the communication needs of the corresponding VNO end users. In the particular context of multi-tenancy, this brings up a series of challenges related to both the latency and security aspects of the envisioned 5G networks. This use case subsequently aims to highlight these issues in the particular context of video broadcasting services. Such services are operated by Over-The-Top (OTT) (third party) service providers that utilise resources from the underlying VNOs, including both network and compute/storage resources in the form of caching services (cache VNFs). The emerging challenges then relate to:

- Traffic and resource isolation between the multiple tenants (i.e., VNOs), so as to ensure only legitimate consumption of network resources;
- Routing optimizations that enable the direct communication between network elements of different VNOs as close to the edge of the infrastructure as possible, so as to reduce perceived latencies at the end user side as well as traffic overheads in the network;
- Inter-domain peering of caches collocated at the micro-datacentre (μ DC) level, so as to enable VNO cooperation for the further improvement of cache performance and the corresponding latency reduction, however also raising issues regarding the legitimate use of the corresponding compute/storage resources of peering VNOs;
- Security threats such as denial of service attacks, as well as side-channel attacks.

In the following we provide a detailed description of this use case, including the different actors in the envisioned 5G landscape along with their business and network level interactions. Subsequently we further elaborate on the requirements stemming from this setup, associating key CHARISMA features with measurable KPIs.

4.2.2. Actors

As shown in Figure 4-2, the use case involves the following actors:

- An access network infrastructure operator, namely SliceNet. SliceNet is a company that owns and operates infrastructure for country-wide wireless access. SliceNet has augmented its infrastructure with compute and storage resources in the form of micro-data centres (μ DCs) deployed at multiple locations in the network (see Figure 4-2). SliceNet leases slices of that entire infrastructure to virtual network operators, like FixTel and MobiCom (see next). These slices include all types of compute, storage and network resources. Compute and storage resources are leased in the form of Virtual Network Functions (VNFs) instantiated at the μ DCs. Network resources are leased in the form of access network services enabling connectivity through isolated slices of the network infrastructure. This can

include both the fixed and wireless part of the network². All VNFs and virtual slices of the network resources comprise the overall, end-to-end Network Service (NS) provided to each VNO. A NS supports end-to-end connectivity to each VNO's users, allowing their traffic to pass through the instantiated VNFs. The exact type of VNFs (e.g., caches, firewalls, intrusion detection systems, deep packet inspection, etc.) is selected by the VNOs, subject to their intended services. The overall NS instantiation is performed by the Control, Management and Orchestration (CMO) framework of CHARISMA.

- Two virtual network operators (VNOs), namely FixTel and MobiCom. FixTel provides Internet access to residential users and small businesses. MobiCom on the other hand focuses on mobile communications at a country scale. Both VNOs are interested in improving the performance of content delivery applications, such as video broadcasting ones, e.g. in terms of the experienced latency, while also reducing traffic in their (virtual) networks. To this end, the instantiated VNFs support caching functionality (vCaches). In our baseline business model, VNOs utilize vCaches for opportunistic and transparent caching i.e., vCaches opportunistically cache content delivered to their users with the purpose of reducing download times and reducing traffic in their network; in this case, caching is deployed transparently to OTTs³. In a future, more elaborate business model, this capability is offered as a service to Over-The-Top (OTT) content providers and/or applications that desire enhanced performance for their subscribers. In this case, the provisioned service follows a Software-as-a-Service (SaaS) model, where a certain type of software service (i.e., caching) and an agreed volume of resources (i.e., number of VNFs/VMs, CPU/Storage/RAM per VNF/VM) is provisioned to OTTs under a service level agreement (SLA) (e.g., describing the agreed availability of the resources). As CHARISMA focuses on the technical implications related to multi-tenancy, low latency (in this particular use case, by utilizing caching) and security, the details of the employed business model followed are considered out of scope. In this sense, the following definition of requirements will follow the simplest model of transparent, VNO-operated caching (see next).

The decision mechanism for the placement of vCaches depends on the selected business model. In the simple case of transparent caching, vCache locations are defined subject to the traffic observed by VNOs across their network, i.e. increased traffic at certain areas of the network suggest the possibility of caching benefits. In the case of service provisioning to OTTs, such as show.me (see next), the exact vCache locations are selected during service setup, subject to mobility profiles of end users, along with information on their application level behaviour and profile (e.g., frequency/duration of video broadcasts, set of recipients and their network locations). The traffic corresponding to video streams is considered to be continuous, i.e. constant bit rate for the entire duration of the transmission.

- An application, namely *show.me*. *show.me* is a social network application where people create their personal live video channel and watch the channels of their friends⁴. Additionally, past video broadcasts are made available for later viewing. All video viewing requests head towards the application back-end that is responsible for the setup of the communication with the broadcasting users. For live streaming, the back-end is responsible for providing the video stream recipients with the IP address of the streaming user, thus enabling the establishment of direct end-to-end connections.

²As solutions for the virtualization of the wireless network in the context of 5G are still immature, and out of CHARISMA's scope, we hereby consider only solutions based on integrated WiFi technology e.g., through virtual WiFi SSIDs.

³ E.g., <http://qwilt.com/solutions/transparent-caching/>

⁴Similar to applications like Meerkat and Periscope.

Video streams are stored in the application back-end via application-level mechanisms⁵. For asynchronous view of (past) videos, the back-end itself is responsible for streaming the video to the requesting users. show.me does not use a content delivery network operator to support this operation, but its performance can be improved with the utilization of VNO resources that are closer to the end users. In our case show.me’s service is enhanced by caching services provided by both MobiCom and FixTel, i.e. one or more vCaches at selected locations of the two VNO’s virtual infrastructure support the scalable streaming of the video sent by an end user’s device (in our case, Bob’s smart phone, see next).

- End users (Bob, Alice and their colleagues at the office). Bob and Alice are civil engineers frequently visiting construction sites to inspect construction progress. Both Bob and Alice are subscribers of MobiCom. Bob and Alice’s office is one of the subscribers of FixTel. When on site, Bob and Alice use show.me on their smartphone to show each other and their team back at the office how construction is progressing. CHARISMA considers limited or no mobility for end users, i.e. no handover events or fluctuations of the wireless signal due to mobility are considered.

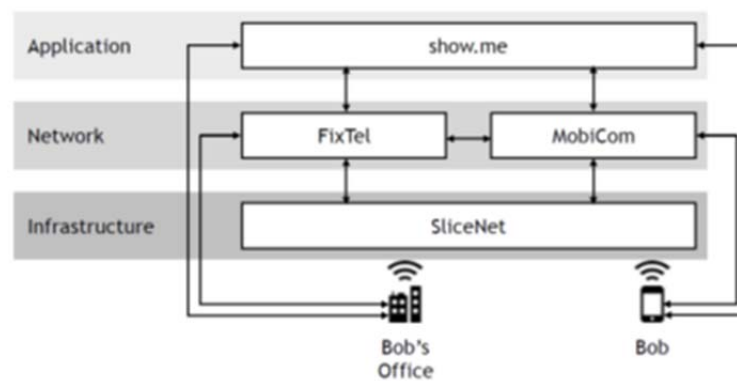


Figure 4-2: Virtual Network Operators sharing the infrastructure of an access network (including an edge cloud). The inter-domain application on top is video broadcasting

Base line scenario

In a baseline scenario of the envisioned use case, Bob starts streaming video footage from the construction during a visit to a site. A few minutes later, Alice joins the show.me application to view her colleague’s video broadcast, while on her way to the site. As a problem seems to arise in the constructions, a team of colleagues located at another office is notified to view the video. Bob’s manager is also later notified for the solution to be fixed. As a result, a series of viewers from different locations start receiving the video broadcast asynchronously.

Extended scenario

Further extending the baseline scenario, Bob cumulatively streams/uploads a series of videos and photos from the aforementioned construction site, during the construction process. Alice also streams/uploads videos/photos from subsequent stages of the construction. At a later point, a team of colleagues at the office inspects the available content to find material for a promotional video of the office. This results in a high volume of requests towards FixTel’s caches; however the corresponding content has been purged due to more recent requests.

⁵ In the simplest case, the application back-end always establishes a connection to the video source (user). As such mechanisms reside on the application level, they are out of CHARISMA’s scope.

4.2.3. Network setup and operation

An example illustration of the envisioned use case is provided in Figure 4-3. MobiCom and FixTel are instantiated on top of SliceNet's virtualized resources. The show.me application back-end resides at some external network and can be reached over the Internet. CHARISMA's Converged Aggregation Levels (CALs) are utilized in the considered setup. The established CAL hierarchy is followed by a corresponding cache hierarchy, where lower level vCaches are configured with *child-parent relationships*⁶ with the vCache on the immediate higher level (within the same VNO), e.g. the vCache instance on the μ BS (CAL1) acts as a parent cache to the cache instance at the office CPE (CAL0), and as a child cache to the vCache at the EPC (CAL3). All video viewing requests pass through the available vCaches on their way to the application back-end or the streaming end user in the case of live video; in the latter case the requests allow the crossed vCaches to cache the delivered content. vCaches intercept requests and examine whether the desired video is cached and thus can be immediately served to the user. This is accomplished through standard means of transparent caching. Upon a cache miss (e.g. the video is live, so no cached copy can exist yet), a vCache forwards the request to the final recipient (i.e. the streaming end user for live video, or the application back-end), caching the video stream sent in response. This means that if a request cannot be served from any of the caches in the established hierarchy, it will end up at the show.me application back-end, where a copy is always maintained. In addition to these hierarchical caching relationships, *peering relationships*⁷ are established between vCaches co-located on the same (μ)DC. This feature leverages on multi-tenancy, taking advantage of the co-location of the peering vCaches, i.e., the network proximity of the vCaches allows for low latency and high bandwidth availability in the exchange of both cache queries and content itself. However, the exchange of content over the established peering link obviously results in the consumption of the corresponding compute/storage and network resources of the peering VNO (consumption due to cache lookup and delivery of the content from storage). Facing the risk of overconsumption of own resources, with an impact on local services, VNOs monitor and police the traffic between the peering caches. . In addition to caching, a routing solution is further employed for the support of low latencies. The TrustNode router optimization allows traffic flows between source-destination UE pairs to be efficiently diverted towards their destination at the lowest common ancestor of the source-destination nodes in the network hierarchy. As video streaming scenarios may result in multiple concurrent recipients, issues emerge regarding the forwarding of multiple 1-to-1 video streaming flows, i.e. despite the route optimization sender, UEs would have to deliver one stream per recipient, over-consuming energy and bandwidth, while further compromising low latencies. A concept of a TrustNode based solution addressing this issue will be presented in the deliverable D1.3 at month M21 of the CHARISMA project.

Based on this setup, this use case builds on the available CALs as follows:

- CAL0: in this case aggregation takes place at the Customer Premises Equipment (CPE) CAL0 (degenerating as necessary to User Equipment (UE)). In our baseline scenario, a single CPE interfacing FixTel at Bob/Alice's office is equipped with a vCache (or even a physical cache device) localising traffic within the customer premises. This cache is configured with a parent-child relationship with the vCache

⁶In such relationships, a parent cache is queried by a child cache upon a cache miss on the latter. The process typically continues recursively until either a cached copy is found or the content origin is reached by the top most parent cache.

⁷In such relationships, caches either proactively or reactively query their peers for content requested but not locally available. When a content item is not found in a peering cache, a failure message is returned and the requesting cache contacts the content origin.

residing at FixTel's slice of the μ DC located at the μ BS. On the office side, Bob's colleagues issue a request for the video through the show.me application which initially reaches the CPE (v)Cache. Given that no previous request has been issued for this video, the request results in a cache miss which subsequently triggers a request towards the next vCache in the hierarchy, i.e. at CAL1/2. Upon reception of the video, the CPE cache stores a local copy for future reference. Indeed, when, later on, Bob's manager uses show.me to see the video, his request hits the CPE cache, which delivers the video directly. In another scenario, the source of the video broadcast is co-located with the recipients of the broadcast in the same 5G cell; in this case the video broadcast is directly diverted to the local recipients.

- CAL1/2: in this case, aggregation takes place at the (small-, micro-, macro) base station level. In our baseline scenario, Bob's video stream is cached at a vCache co-located with the base station, through which Bob accesses MobiCom's (virtual) network (i.e. μ BS in Figure 4-3); while a copy also reaches the application back-end. Alice's request for Bob's video hits this vCache, as Alice is attached to MobiCom through the same μ BS as Bob. Aggregation at this level is further facilitated by multi-tenancy: Bob's video needs to traverse the borders of the MobiCom domain, so as to enter the FixTel domain and reach Bob's office: as explained above, Bob's colleagues request the video through the show.me application. Their request reaches the CPE cache, which at the moment does not hold this video. The request is subsequently forwarded to FixTel's vCache which also does not hold a copy; this triggers a query from FixTel's vCache to MobiCom's vCache; the latter responds with the video, which is also subsequently cached at FixTel's vCache. This traversal from the FixTel domain to the MobiCom domain happens at the μ DC level, over the established peering link between the two corresponding vCaches. It is noted that in the absence of the introduced peering links, the video request from Bob's office would have to reach the show.me application back-end to be served. This would obviously result in increased latency for the video to start streaming, as well as unnecessary traffic inside the network. The same procedure is followed in the extended scenario, where the office requests lead to a high volume of requests towards MobiCom's vCache. Though the content may still be available there, a limit is imposed on the amount of traffic/requests to be served by the peering vCache. For the remainder of requests FixTel's cache further propagates the request, through the cache hierarchy, towards the show.me application back-end.
- CAL3: in this case aggregation takes place at the Central Office (CO) / Evolved Packet Core (EPC). A peering cache may also reside at this level, as shown in Figure 4-3.

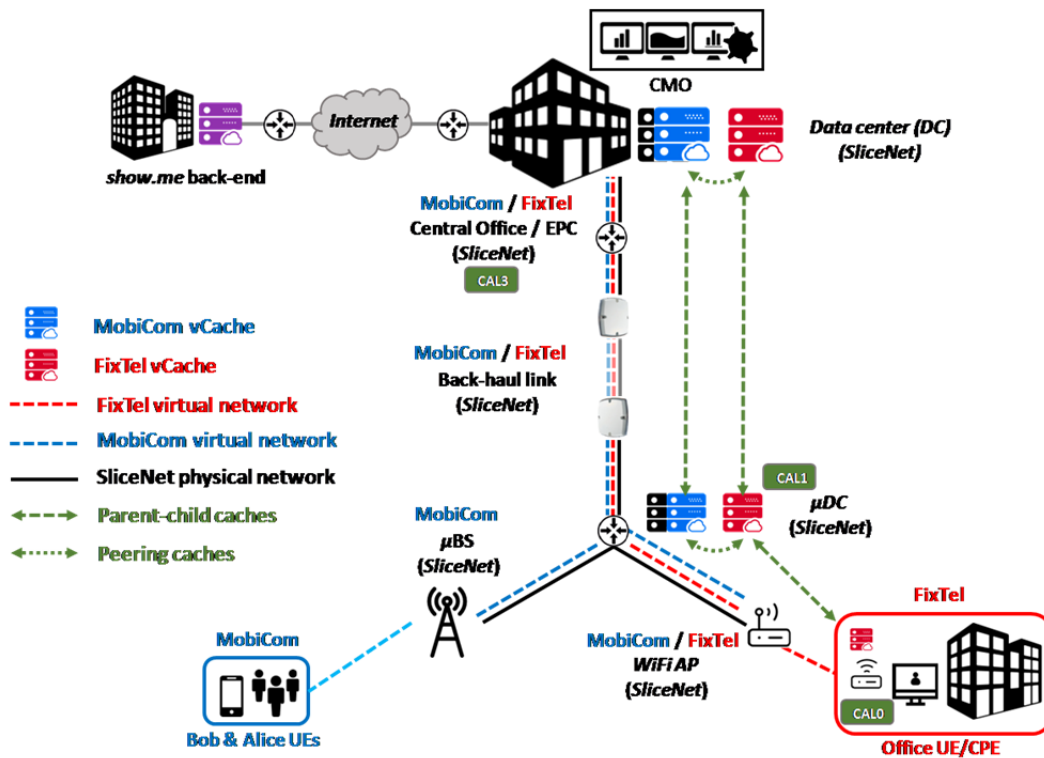


Figure 4-3: Multi-tenancy in a video streaming application

The setup described in the previous paragraphs for providing multi-tenant access and video broadcasting services involves several security risks. The infrastructure and services provided can be considered to be the targets of traditional security risks, such as Denial of Service (DoS) attacks, coming from outside. Even more, cloud computing and multi-tenancy factors add in complexity and introduce additional security concerns that demand conventional or new security techniques to ensure service availability, SLA compliance, and data confidentiality and integrity. In a multi-tenant environment, these risks might also come from the inside.

Multi-tenancy introduces several security risks that need to be addressed. The fundamental issue introduced arises from the premise upon which multi-tenancy is based on: the fact that multiple tenants are sharing common infrastructure. Tenants might be competitors, and if traffic isolation is not carefully addressed, one tenant might try to access another tenant's information or interfere or even disrupt the services run by other tenants. It is the infrastructure provider's responsibility to provide a secure infrastructure by isolating tenant's virtual machines, as well as the various networks within the infrastructure. Best practice suggests that the management networks, storage networks, and customer (tenant) networks all be isolated. Isolation is usually achieved by using virtual switches for each of the networks, utilizing VLAN segmentation (IEEE 802.1q) or a combination of both.

Side-channel attacks are a common type of attack that occur in multi-tenant environments and need to be addressed as well. In a side-channel attack, tenant information can be obtained by other tenants, by using bandwidth monitoring or similar techniques. Usually, this type of attacks occurs when authorisation mechanisms to monitoring information have not been foreseen or are not adequate, and information related to resources of other tenants or to all tenants is revealed, providing competitors with information through which conclusions on the usage or services of other tenants can be drawn. Lack of enforcement of appropriate authorization policies cause this interference between tenants.

Another possible issue arising in multi-tenant setups is interference between tenants caused by increased overload. For example, when the overload of one of the tenants (caused by using large amounts of CPU or memory resources) negatively impacts another tenant that has deployed its services on the same physical infrastructure as the first tenant. Applications that are sensitive to latency, disk I/O or CPU utilization might be impacted and performance might become unpredictable. Standard monitoring tools, capable of providing measurements of monitoring metrics for both physical and virtual resources of the infrastructure, such as CPU utilisation, memory utilisation, packet rate and bit rate of network interfaces, can help identify which tenant applications are causing such problems. The infrastructure provider should carefully decide actions on such incidents, and provide appropriate mitigation plans to ensure agreed SLAs with all tenants using its infrastructure. Scenarios that demonstrate interference between tenants due to overload, although common in multi-tenant setups will not be examined in CHARISMA.

Apart from standard monitoring tools, the infrastructure provider might use attack monitoring tools, such as Intrusion Detection Systems (IDS). Virtual instances of IDSs might be deployed as well to provide monitoring to tenants' networks and services. However, the placement of these security elements in a virtual multi-tenant environment is different from traditional environments. In-line placement of virtual IDSs might significantly impact performance, consuming resources from the resource pool of each tenant. For this reason, off-path placement of the virtual IDS is a more appropriate solution. Connecting to a spanning port of a switch, attaching to a hub, or attachment by using a network tap are techniques that allow off-path placement of virtual IDSs.

In a similar way to virtual IDSs, incident responses to identified threats might happen through virtual firewalls, enforcing security policies that block access to specific hosts. Instances of virtual firewalls should be connected in-line to customer services to protect them. However, given the complexity of the shared infrastructure, a platform for security policy management should be provided by the infrastructure provider to its customers (tenants), allowing them to define their own security profiles through policies. Obviously, different tenants and applications will demand different security policies and configurations.

Several of the above-mentioned challenges are within the scope of CHARISMA's work in the area of security. Specifically, in this use case scenario we will address:

1. Tenant isolation;
2. Infrastructure monitoring, including physical and virtual resource monitoring;
3. Automated provisioning and management of security services.

An example illustration of the security issues we plan to examine in CHARISMA is provided in Figure 4-4. First, as shown in the previous figures describing the envisioned use case, SliceNet is offering slices of its infrastructure to different VNOs, such as MobiCom and Fixel operators. For both VNOs a slice has been created, providing each tenant with their own, private, isolated and secure virtual network infrastructure. Services of each tenant will be isolated as well, even if deployed in the same cloud infrastructure, and the different configuration per each tenant's needs should be demonstrated.

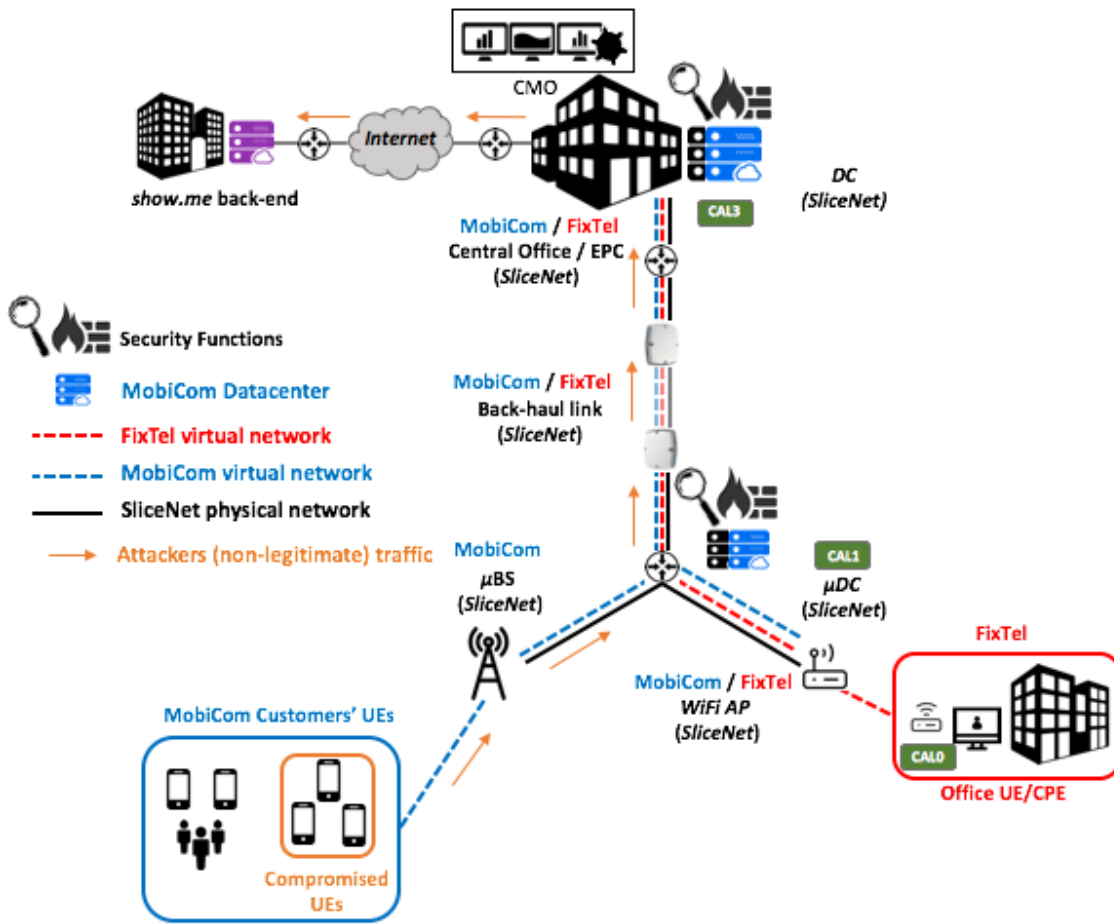


Figure 4-4: Security scenarios on the multi-tenancy and video streaming use case

Monitoring services that provide monitoring information and metrics acquisition from the physical and virtual resources comprising the infrastructure should be deployed. This information should be provided to both the infrastructure provider and the VNOs. In CHARISMA, we assume that the monitoring services belong to the infrastructure provider. Access to the monitoring information should be given as well to the VNOs that share the infrastructure. However, each tenant should be able to access only the information that concerns its slice (virtual resources) and services.

Additionally, each tenant must be able to deploy its desired security services. Examples of security services include virtual firewalls and virtual intrusion detection systems. The type of security services and their placement will be decided through a set of security policies defined by each VNO. As shown in Figure 4-4, the deployment of such security service can provide protection to the VNO's infrastructure and offered services. In the illustrated scenario, we assume that due to vulnerability some Mobicom customers' UEs have been compromised. The compromised devices perform an attack to the show.me application. Another possible case would be an attack towards components of the infrastructure of significant importance, such as the NFVO. Our work will focus on demonstrating how we can mitigate such security attacks through the automated provisioning and management of virtual security services deployed at the available CALs (CAL1 and CAL3 in this scenario).

4.2.4. Requirements & KPIs

The functional and performance requirements associated with this use case are the following:

Index	Description	Importance	KPIs and verification
1	The infrastructure owner has to be able to offer its virtualised resources in a way that multiple virtual network operators can be easily instantiated and become fully functional. This means that virtualised resources should be easily (e.g., through automated MANO procedures) bundled together into slices of the physical infrastructure.	Mandatory	VNO slice instantiation delay , i.e. the time required to instantiate VNFs as well as apply all network, compute and storage resource configurations. Adopting 0, a target value of 90 minutes is set for this KPI.
2	VNO resource availability should remain unaffected by the instantiation, resource reconfiguration and operation of other VNOs on the same infrastructure (except through explicit peering).	Optional	VNO service availability and quality should remain unaffected by the instantiation of a new VNO(s) on the same infrastructure, e.g., instantiation of FixTel while MobiCom is already in operation. <ul style="list-style-type: none"> • Availability/reliability will be assessed in terms of packet loss on a running service of a VNO during the instantiation of a second VNO on the same infrastructure. CHARISMA will target high availability, which according to [8] corresponds to <1% of losses during observation time. • Quality will be assessed in terms of latency variation. Latency will be measured as one trip time latency, i.e. the time it takes from when a data packet is sent from the transmitting end (e.g. Bob in the considered baseline scenario) to when it is received at the receiving end (e.g. Alice in the considered baseline scenario) [8]. Variation will be measured by comparing the observed values before, during and after the instantiation of a second VNO, e.g. FixTel.
3	The created network slices must support complete traffic isolation between VNOs, i.e. no traffic from one VNO should reach the other without explicit consent.	Mandatory	No specific performance metric suitable for this requirement. Tests will be contacted to verify that no traffic crosses administrative (VNO) borders apart from the case of vCache peering. These tests will include traffic originating from clients residing at a foreign VNO.

4	Throughput is a key parameter for a large part of existing and envisaged applications, as is in this case, video broadcasting. High-definition video is already the norm (with higher resolutions coming up) especially for receivers with large screens, so bandwidth requirements are considerable and will become more so in the near future. However, device capabilities need also to be taken into account.	Mandatory	Throughput (Mb/s) will be measured as the user data rate [8] perceived by the receiving end user device. In the considered baseline scenario throughput will be measured at both Alice's (mobile) end device (subscriber of the same VNO with Bob, i.e. MobiCom), and the office's fixed end device (subscriber of FixTel). A low user data rate value, i.e. up to 50 Mbps [8] will be targeted. It is noted that this range of values corresponds to the current highest quality Video Bitrates for YouTube (Standard Frame Rate) ⁸ .
5	Traffic should be routed as close to the edge as possible to minimize hops and hence traffic impairments.	Mandatory,	<p>Path length (hop count): this will be compared against the available topology. This is considered as a means of verification. For the baseline scenario, the target values will be:</p> <ul style="list-style-type: none"> • Alice's end device: 2, i.e. MobiCom vCache (μDC) ->μBS-> Alice's UE • Bob's office colleagues: 4, i.e. MobiCom vCache (μDC) ->FixTelvCache (μDC)->WiFi AP ->CPE cache->UE • Bob's manager: 1 CPE cache->UE <p>Latency (ms). Here, latency will be measured as round trip time latency, i.e. the time it takes from when a video request is sent from the receiving end to when the first data packet is received. Corresponding to the baseline scenario, it will be measured for Alice, Bob's office colleagues and Bob's manager cases. Target values are in the Low (1-10 ms) to Medium (10-50 ms) ranges [8]. It is noted that video streaming is not considered as a mission critical application requiring Low range latency values. However, the selected use case aims to deliver low to medium latencies as a demonstration and validation of the ability of integrated, NFV-enabled caching capabilities to significantly lower perceived latencies. To this end, and corresponding to the extended use case scenario, the project will also measure the end-to-end latency for the delivery of a video from the</p>

⁸<https://support.google.com/youtube/answer/1722171?hl=en>

			application back-end.
6	The inter-domain, cache peering communication mechanisms should allow peering VNOs to impose limits on the resource consumption due to peering. Subsequently, this calls for the monitoring of exchanged data and request rate, as well as for corresponding request/data traffic policing mechanisms.	Optional	Request and data rates. Focus is on the ability to monitor and police the corresponding traffic, rather than on the specific threshold values.
7	The exchange of control and data plane traffic between peering vCaches must reduce the perceived latency at the end user side. This is the essential motivation for this technological innovation.	Mandatory	Latency (ms). It will be measured similarly to Requirement #5. Measurements will be contacted against a comparison scenario where no peering links are established across the vCaches of the different VNOs. The measurements will allow the validation and quantification of the intuitive latency reduction expectancy in this scenario.
8	The use of vCaches must reduce the perceived latency at the end user side. This is the essential motivation for the development of the vCaches.	Mandatory	Latency (ms). It will be measured similarly to Requirements #5 and #7. Measurements will be contacted against a comparison scenario where no vCaches are instantiated across the different VNOs. The measurements will allow the validation and quantification of the intuitive latency reduction expectancy associated with caching.
9	Cache peering should allow for controlled and identifiable resource consumption overheads, i.e. VNOs should be able to identify/measure the burden of peering links.	Optional	Typical IT resource metrics such as CPU and RAM utilization will be employed to quantify overheads per peering request. Measurements will be contacted against a comparison scenario where no peering links are established across the vCaches of the different VNOs.
10	Each tenant should be able to define its own security policies, deciding the deployment of desired security services (e.g. virtual IDS, firewall) and their configuration without affecting the other tenant's services.	Mandatory	There are no specific KPIs to validate this requirement. However, final demonstrations performed at the end of the project will showcase the ability of each tenant to define its security policies and that different policies of tenants for security services that run on the same infrastructure do not cause any conflicts due to the provided tenant isolation.

4.3. Updated Mandatory CHARISMA Requirements and KPIs

A first attempt for defining CHARISMA requirements and KPIs was reported in deliverable D1.1. However, some of these KPIs were either too tightly defined, or not specific enough to allow verification of the project results. In this section we define the updated CHARISMA requirements and KPIs that have resulted from the two updated use cases described above, the work performed in WP3, WP4 and WP5, as well as from the results of our three initial intermediate demos. These requirements and KPIs are the parameters that will be supported by the CHARISMA architecture, and will be used to verify the project results through the project demos, and final demonstration and field trials.

Table 4-2: Updated mandatory CHARISMA requirements and KPIs

Number	Requirement	CHARISMA support
1	CHARISMA shall offer low latency services	CHARISMA’s architecture shall support low latency services via: <ul style="list-style-type: none"> • Routing of data at the lowest common aggregation point; • Devolved offload strategies for device-to-device, device-to-remote-radio, device-to- baseband, device-to-central office/metro, cloud-to-cloud/cellular, etc.; • Mobile distributed caching; • Trust Node enabled secure hierarchical and ID routing. CHARISMA shall support a latency of 10ms or less.
2	The system shall support advanced end-to-end security	CHARISMA’s architecture shall support distributed (decentralized) security, as opposed to centralized security in 4G, as well as physical layer security. The CHARISMA virtualized OpenNaaS-based architecture-level design provides a C&M plane offering improved security. A holistic security approach is proposed where the underlying infrastructure is virtualized and shared among several SPs, who operate simultaneously the same physical resources.
3	CHARISMA shall support open access in a way that multiple virtual network operators can be easily instantiated and become fully functional.	This means that virtualised resources should be easily (e.g., through automated MANO procedures) bundled together into slices of the physical infrastructure. VNO slice instantiation delay (i.e. the time required to instantiate VNFs as well as apply all network, compute and storage resource configurations) shall not exceed 90min.

4	The created network slices must support complete traffic isolation between VNOs, i.e. no traffic from one VNO should reach the other without explicit consent.	No specific performance metric is defined for this requirement. During the field trials, tests will be conducted to verify that no traffic crosses administrative (VNO) borders, apart from the case of vCache peering. These tests will include traffic originating from clients residing at a different VNO.
5	Each tenant should be able to define its own security policies, deciding the deployment of desired security services (e.g. virtual IDS, firewall) and their configuration without affecting the other tenant's services.	No specific KPI to validate this requirement. However, final demonstrations performed at the end of the project will showcase the ability of each tenant to define its security policies and that different policies of tenants for security services that run on the same infrastructure do not cause any conflicts due to the provided tenant isolation.
6	A high throughput will be supported, which is a key parameter for a large part of existing and envisaged 5G applications, including high bandwidth video streaming.	CHARISMA's architecture shall support data-rates up to 10 Gbps for SMEs and residential users, and up to 1 Gbps for mobile end-users, through the use of a hierarchical intelligent data processing approach at the C-RAN and RRH, where statistical multiplexing, aggregation, and caching allow access data volumes to be significantly increased. In addition, CHARISMA's architecture shall incorporate mm-wave (60 GHz and E-band technologies), as well as optical LoS and non-LoS (NLoS) final-drop technologies, including converged wireline (FTTH) connections from the RRH and/or the C-RAN to end-user premises. The typical value of data-rate for video streaming shall be at least 50Mbps per user.
7	Low packet loss rate	The 5G system proposed by CHARISMA shall provide packet loss rate of 10^{-5} or less.

5. Conclusions

In this deliverable D1.2 “Refined architecture definitions and specifications” we have provided an update of the CHARISMA architecture, which we have first described in deliverable D1.1. A key feature of the CHARISMA architecture is the converged aggregation level (CAL) node, which integrates networking and computational resources at *all* aggregation levels of the access network. It has been designed to feature three particular features of future 5G networking: low latency, open access, and security. The computational and networking resources at each CAL node establish a distributed cloud architecture for 5G, which, in addition, requires a sophisticated management and control architecture. A more detailed description of the data plane as well as the control and management plane architectures has been given in chapter 2.

The roles of the two main actors in the CHARISMA control and management system, namely the network operator and the virtual network operators have been detailed for various services. The provisioning of network slices, caching and security services are described in chapter 3. These refer again to the three defining characteristics of CHARISMA: low latency, open access, and security.

While in D1.1 a rather large number of use cases were described, in this deliverable two main and updated use case scenarios, based on the transportation vertical sector, and the support of VNOs in a multi-tenancy environment have been described. The selected use case scenarios have also been used to enable verification of project results through the specific and updated KPIs resulting from them. These KPIs will be verified through the project demos, and final demonstration and field trials in WP4 during the final year of CHARISMA.

References

- [1] CHARISMA D1.1, “CHARISMA intelligent, distributed low-latency security C-RAN/RRH architecture”, 2016
- [2] CHARISMA D2.1, “CHARISMA Initial Architecture Design and Interfaces”, 2016
- [3] CHARISMA D2.2, “CHARISMA PHY Design and Interfacing”, 2016
- [4] CHARISMA D3.1, “V-Security Management Plane Design and Definition”, 2016
- [5] CHARISMA D3.2, “Initial 5G multi-provider v-security realization: Orchestration and Management”, 2016
- [6] CHARISMA D3.3, “Initial Content Caching and Traffic Handling at SW regex integration”, 2016
- [7] 5G-PPP White Paper on Media & Entertainment Vertical Sector – January 2016
- [8] 5G PPP White Paper on Use cases and Performance Evaluation Models – April 2016
- [9] R3-161687, Draft TR 38.801 (v0.3.0) “Study on New Radio Access Technology: Radio Access Architecture and Interfaces”, DOCOMO Communications Lab, September 2016
- [10] iCIRRUS D3.2, “Preliminary Fronthaul Architecture Proposal”, 2016
- [11] Strategic Research and Innovation Agenda (SRIA) document on “Pervasive Mobile Virtual Services”, Expert Advisory Group of the European Technology Platform Networkworld 2020, July 2016
- [12] Strategic Research and Innovation Agenda (SRIA) document on “Service Level Awareness and Open Multi-Service Internetworking”, September 2016
- [13] 5G-PPP white paper “5G Automotive Vision”, October 2015

Acronyms

5G	5 th generation mobile network
ACE	Accelerated
ADC	Analog-to-Digital-Converter
ALG	Application Level Gateway
AP	Access Point
APD	Avalanche Photodiode
API	Application Programming Interface
AR	Access Router
ARN	Active Remote Node
ARP	Address Resolution Protocol
ASE	Amplified Spontaneous Emission
ATX	Advanced Technology eXtended
AWG	Arbitrary Waveform Generator/ Arrayed Waveguide Grating
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
BNG	Broadband Network Gateway
BOSA	Bi-directional Optical Sub-assembly
BRAS	Broadband Remote Access Server
BS	Base Station
CAL	Converged Aggregation Layer
CCD	Charge-Coupled Device
CDN	Content Delivery Network
CE	Coexistence Element
CGH	Computer-Generated Hologram
CO	Central Office
CP	Cyclic Prefix
CPE	Customer Premise Equipment
CPU	Central Processing Unit
CRC	Cyclic Redundancy Check
CST	Computer Simulation Technology (AG)
D2D	Device to Device
D2I	Device to Infrastructure
DAC	Digital-to-Analog-Converter

DC	Data Centre
DDM	Digital Diagnostic Monitoring
DDoS	Distributed DoS
DDR	Double Data Rate
DDS	Direct Digital Synthesis
DFB	Distributed Feedback Lasers
DFT	Discrete Fourier Transform
DHCP	Dynamic Host Configuration Protocol
DL	Downlink
DOP	Degree of Polarisation
DPDK	Data Plane Development Kit
DPI	Deep Packet Inspection
DoS	Denial of Service
DoW	Description of Work
DP	Distribution Point
DQPSK	Differential Quadrature Phase Shift Keying
DRAM	Dynamic RAM
DSCP	Differentiated services code point
DSP	Digital Signal Processing
EC	European Commission
ECMA	European Computer Manufacturers Association
EDFA	Erbium Doped Fibre Amplifier
EIRP	Equivalent isotropically radiated power
EM	Element Manager
EML	Externally Modulated Laser
ETH	Ethernet
EVM	Error-Vector-Magnitude
FEC	Forward Error Correction
FFT	Fast Fourier Transform
FMC	FPGA Mezzanine Card
FPGA	Field-Programmable Gate Array
FSAN	Full Service Access Network
FSO	Free Space Optics
FW	Firewall
GbE	Gigabit Ethernet
GPON	Gigabit Passive Optical Network
GUI	Graphical User Interface
HD	High-Definition
HDMI	High-Definition Multimedia Interface

HDTV	High Definition TV
HG	Hermite-Gaussian
HGI	Home Gateway Initiative
HGW	Home Gateway
HP	Holographic Plate
HW	Hardware
I/Q	In-phase/Quadrature
ICMP	Internet Control and Management Protocol
ICP	Internet Cache Protocol
ID	Identification Data
IDFT	Inverse Discrete Fourier Transform
IEEE	Institute of Electrical and Electronics Engineers
IETF	The Internet Engineering Task Force
Ifc	Interface
IFFT	Inverse Fast Fourier Transform
IMU	Intelligent Management Unit
IoT	Internet of things
IP	Internet Protocol
IPTV	IP Television
IPoE	IP over Ethernet
IPS	Intrusion Protection System
ISI	Inter-Symbol-Interference
ITU	International Telecommunication Union
JTAG	Joint Test Action Group
KPI	Key Performance Indicators
LG	Laguerre-Gaussian
LO	Local Oscillator
LOS	Line of Sight
LP	Low-Pass
LTE	Long Term Evolution
MAC	Media Access Control
MANO	management and organization
MB	MoBcache
MBH	Mobile Backhaul
MDC	Mobile Distributed Caching
MIMO	Multiple-Input Multiple-Output
mmW	millimeter Wave
MPLS	Multiprotocol Label Switching
MPPS	Million Packets per second

MPW	Multi Project Wafer
MU	Multi-User
MUX	Multiplexer
MVNO	Mobile Virtual Network Operator
NAPT	Network Address Port Translation
NAT	Network Address Translation
NFV	Network Function Virtualisation
NG-PON2	Next Generation Passive Optical Network 2
NIC	Network Interface Card
NLOS	None Line of Sight
O/E	Optical-to-Electrical
OAM	Orbital Angular Momentum
ODN	Optical distribution Network
OFDM	Orthogonal Frequency Division Multiplexing
OLT	Optical Line Terminal
OMCI	Optical Network Unit Management and Control Interface
ONT	Optical Network Termination
ONU	Optical Network Unit
OPEX	Operating Expense
OS	Operating System
OSNR	Optical Signal to Noise Ratio
OTT	Over-The-Top content
OVS	Open vSwitch
PC	Personal Computer
PCB	Printed Circuit Board
PCI	Peripheral Component Interconnect
PCIe	Peripheral Component Interconnect Express
pCPE	Physical CPE
PCS	Physical Coding Sublayer
PHY	Physical Layer
PIC	Photonic Integrated Circuits
PIN	Positive-intrinsic-Negative
PMA	Physical Medium Attachment
PNF	Physical Network Functions
PoSK	Polarisation Shift Keying
PON	Passive Optical Network
PPP	Public Private Partnership/
PPPoE	PPP over Ethernet
PtP	Point-to-Point

PTP	Precision Time Protocol
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
QKD	Quantum Key Distribution
QoS	Quality of Service
RAM	Random Access Memory
RF	Radio Frequency
RFC	Request For Comment
RGW	Residential Gateway
RN	Remote Node
RSSI	Received Signal Strength Indication
SC	Small Cell
SDN	Software Defined Networks
SFI	SerDes Framer Interface
SFP+	Small Form Factor Pluggable
SLA	Service level Agreement
SMA	Sub-Miniature version A (connector)
SMF	Single Mode Fibre
SNMP	Simple Network Management Protocol
SNR	Signal to Noise Ratio
SOA	Semiconductor Optical Amplifier
SP	Service Provider
SPP	Spiral Phase Plate
SRIOV	single root input/output virtualization
SSD	Solid State Disk
SSMF	Standard Single Mode Fibre
STB	Set Top Box
STP	Spanning Tree Protocol
syncE	Synchronous Ethernet
TCAM	Ternary Content Addressable Memory
Tcl	Tool command language
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
TIA	Trans-Impedance-Amplifier
TS	Training Symbol(s)
TWDM	Time and Wavelength Division Multiplex
Tx	Transmitter
UHDTV	Ultra HDTV
UNI	User Network Interfaces

UPnP	Universal Plug-and-Play
USB	Universal Serial Bus
VAS	Value-added Service
vCache	Virtual Cache
vCC	virtual Cache Controller
vCPE	Virtual CPE
vE-CPE	Virtual Enterprise CPE
Ve-VNFM	Virtual Network Function Manager
Vi-VNFM	Virtual Network Function Manager
vIT	Virtual IT
VLAN	Virtual LAN
VM	Virtual Machine
VNF	Virtual Network Function
VNFM	Virtual Network Function Manager
VNO	Virtual Network Operator
VPN	Virtual Private Network
VoIP	Voice over IP
WDM	Wavelength Division Multiplexing
WOC	WAN Optimization Controller
WP	Work Package

<END OF DOCUMENT>