Title*: 5G-MoNArch Elastic Network Slice Management

from **Source***: Chairman
Contact: Marco Gramaglia

input for **Committee***: ENI

Contribution **For***:

Decision	X
Discussion	
Information	

Submission date*: 2018-11-06

Meeting & Allocation: **Rapporteur's call#82:Use cases/Requirements/Terminology/Categorization** -

Relevant WI(s), or deliverable(s):

Decision/action requested: Please approve proposed to be approved with minor additions PoC To start on 6 November 2018

ABSTRACT: *This PoC is meant to showcase the network elasticity mechanisms proposed and developed by the European research project 5G-MoNArch, with a special attention to the AI/ML aspects. These procedures contribute to endow the network with the elasticity features required by ETSI ENI Use Case #2-6 [1], "Elastic Resource Management and Orchestration". In particular, the proposed mechanisms are compliant with its triggering conditions, operational flow, and post-conditions, as defined in ETSI GS ENI 001,*

PoC Project Details

1.1 PoC Project

PoC Number (assigned by ETSI):
PoC Project Name: **5G-MoNArch Elastic Network Slice Management**
PoC Project Host: Telecom Italia
Short Description: this PoC is meant to showcase the network elasticity mechanisms proposed and developed by the European research project 5G-MoNArch, with a special attention to the Artificial Intelligence / Mechine Learning (AI/ML) aspects. These procedures contribute to endow the network with the elasticity features required by the new use cases under discussion in ETSI ENI 001 [1] Use Case #2-6, "Elastic Resource Management and Orchestration". In particular, the proposed mechanisms are compliant with its triggering conditions, operational flow, and post-conditions, as defined in [1].

1.2 PoC Team Members

Table 1.1

	Organization name	ISG ENI participant (yes/no)	Contact (Email)	PoC Point of Contact (see note 1)	Role (see note 2)	PoC Components
1	Universidad Carlos III de Madrid	Yes	Marco Gramaglia (mgramagl@it.uc3m.es)	X	University	Storyline/ Implementation of the baseline architecture / Implementation of the specific algorithms
2	CEA-Leti	Yes	Nicola di Pietro (nicola.dipietro@cea.fr)		Research body / Manufacturer	Design of the admission control algorithm
3	Samsung R&D Institute UK	Yes	David Gutierrez Estevez (d.estevez@samsung.com)		Manufacturer	Design of the Horizontal and Vertical scaling algorithms
4	Telecom Italia S.p.A.	Yes	Luca Pesando (luca.pesando@telecomitalia.it)		Network Operator	Testbed logistic and setup
5	Huawei	Yes	Dr. Wang Yali (wangyali11@huawei.com)		Manufacturer	Help with the architecture design and discussion about solution related intelligent

NOTE 1: Identify the PoC Point of Contact with an X.
NOTE 2: The Role will be network operator/service provider, infrastructure provider, application provider or other as given in the Definitions of ETSI Classes of membership.

All the PoC Team members listed above declare that the information in this proposal is conformant to their plans at this date and commit to inform ETSI timely in case of changes in the PoC Team, scope or timeline.

1.3 PoC Project Scope

1.3.1 PoC Goals

This PoC contributes to demonstrate ENI as a viable technology for the improvement of telecommunication networks. We intend to prove the applicability of the techniques described in ETSI GS ENI 001 [1] Use Case #2-6, "Elastic Resource Management and Orchestration".
The main goal of this PoC consists in showing the feasibility and the benefits of an AI-assisted "elastic" management and orchestration of the network, which entails an improvement of the network efficiency and its capability to smoothly adapt the resource allocation and utilization. The public demonstration of these ENI concepts, based on an architectural framework compatible with ENI Reference Architecture ETSI GS ENI 005 [2], will help to build commercial awareness and confidence in the ENI approach.

The PoC will demonstrate in a testbed environment how ENI can be used to integrate and improve the resource and network function management and orchestration, including slice admission control and deployment, horizontal and vertical scaling of virtual network functions, and their migration between different nodes of the network.

1.3.2 PoC Topics

PoC Topics identified in this clause need to be taken for the PoC Topic List identified by ISG ENI and publicly available, i.e. the three topics identified in clause 4.5 of the ENI PoC Framework ETSI GS ENI 006 [3]. PoC Teams addressing these topics commit to submit the expected contributions in a timely manner.

Table 1.2

PoC Topic Description	Related WI	Expected Contribution	Target Date
Network Operations -> Elastic Resource Management and Orchestration	ENI-005, (GS ENI 005 system architecture) – Early draft ENI-007 GS ENI 002 Requirements) Release 1 version 2.0.x or above, ENI-008 (GS ENI 001 Use Cases) Release 1 version 2.0.x or above	Horizontal and Vertical Virtual Ne5twork Function (VNF) Scaling, Admission Control	30/09/2019

1.4 PoC Project Stages/Milestones

Table 1.4

PoC Milestone	Stages/Milestone description	Target Date	Additional Info
P.S	PoC Project Submission	09/10/2018	
P.TP.1	PoC User Story finalization	ENI#8	Finalization of the high-level description of the use case for the three algorithms described In Section 2
P.TP.1	PoC Test Plan 1	31/01/2019	Initial testbed up and running
P.TP.1	PoC Test Plan 1	31/03/2019	Algorithm testbed
P.D1	PoC Demo 1	30/05/2019	Testbed showcased in Turin, Italy
P.D2	PoC Demo 2	30/06/2019	Testbed showcased in Valencia, Spain
P.D3	PoC Demo 3	30/07/2019	Webinar ETSI
P.C1	PoC Expected Contribution 1	30/09/2019	Contribution on the feasibility of Horizontal and Vertical VNF Scaling, in ENI 005 [2]
P.C2	PoC Expected Contribution 2	30/09/2019	Contribution on the feasibility of admission control, in ENI 005 [2]
P.C3	PoC Expected Contribution 3	30/09/2019	Reporting on the Tesbed showcase
P.R	PoC Report	30/09/2019	PoC-Project-End Feedback
P.E	PoC Project End	30/09/2019	presented to ISG ENI for information

1.5 Additional Details

The dates for the demos are tentative and will be confirmed. More demos may be provided, if the PoC team finds the right venue.

2 PoC Technical Details

2.1 PoC Overview

The proposed PoC features two different network slices: i) an enhanced Mobile Broadband (eMBB) slice that serves 360° videos to a virtual reality device that ii) also uses an Ultra-Reliable Low Latency Communications (URLLC) slice to provide voice connectivity and haptic interaction with other avatars in the virtual scenario. From the physical infrastructure point of view, the testbed includes a set of Physical Network Functions (PNFs) that implement the radio lower layers and a cloud infrastructure (composed by a larger but farther to the radio central cloud and a closer but less capable edge cloud). In particular, besides the virtual reality application, the testbed will be composed by a set of functionalities that are summarized as follows and detailed next:

- Network Slice Blueprinting and Onboarding;
- VNF Relocation due to latency;
- Horizontal and Vertical VNF Scaling (AI);
- Admission Control (AI).

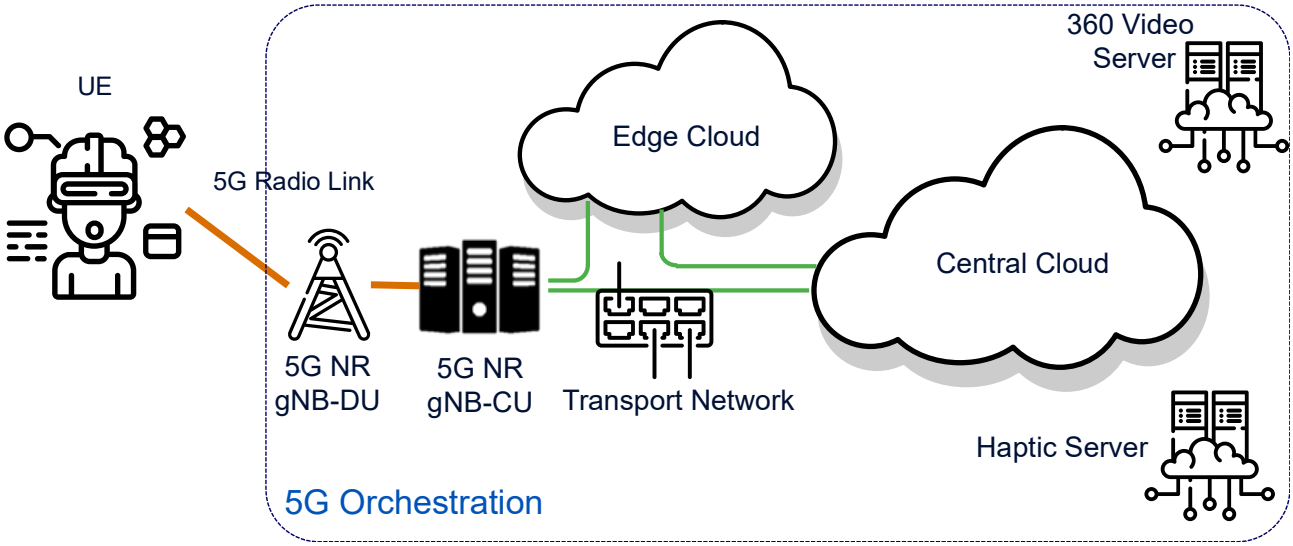

Figure 1: PoC hardware and the software setup

2.2 PoC Architecture

The diagram represented below shows the framework of the PoC. We will rely on the well know ETSI NFV and 3GPP Management and Orchestration architectures that are also building the overall 5G MoNArch architecture. We plan to populate the specific modules of the architecture with our algorithms, as specified next.

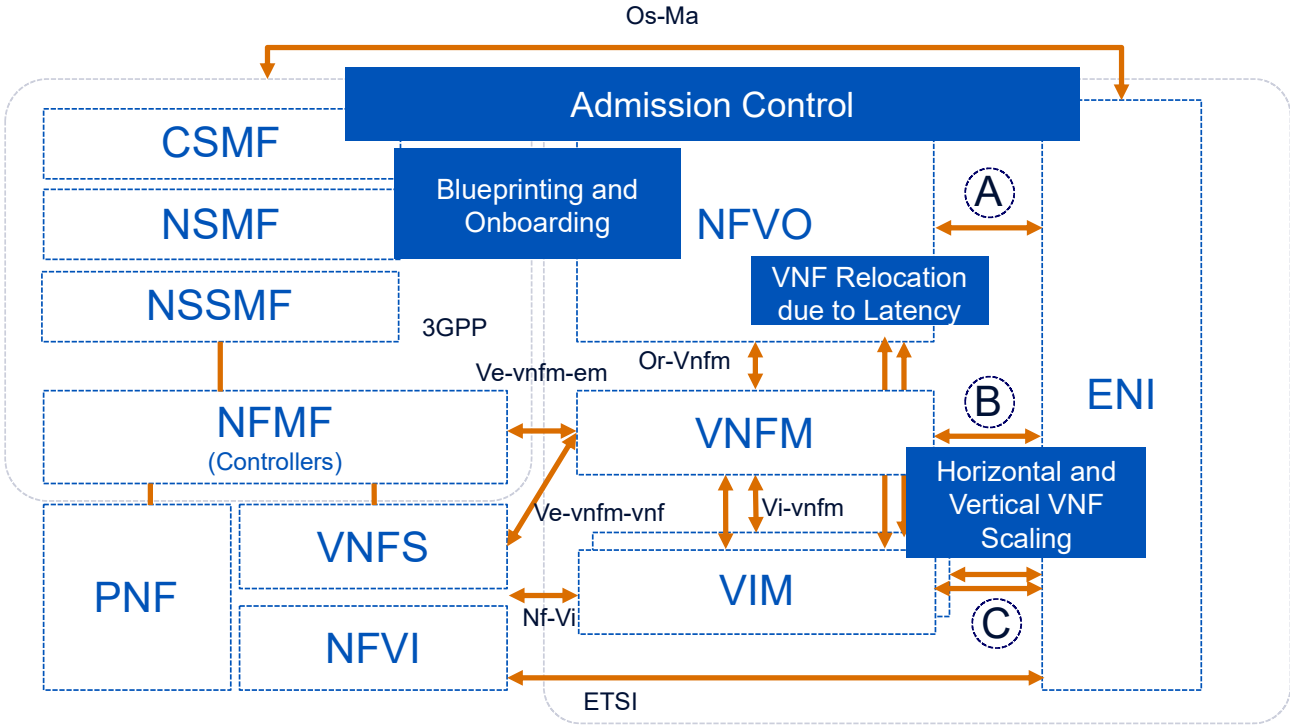

Figure 2: PoC architecture [8]

Network Slice Blueprinting and Onboarding: this functionality encompasses all the baseline operation available in a 5G MANO system. The automatic onboarding of the two network slices is a seamless operation that allows i) tenants to define the requirements associated with each slice (in this case, high bandwidth for the eMBB and very low latency for the URLLC) and ii) the deployment of the VNF in the cloud to fulfill the set of requirements. This part has specific elements of novelties for the blueprinting. Specifically, the network slice templates that are currently being investigated by 3GPP and NGMN, and their practical implementations with technologies such as TOSCA or OpenStack Heat are extended and integrated with specific elasticity fields.

VNF Relocation due to latency: after the first step, both network slices are orchestrated with all of their VNFs (namely the higher layers of the RAN stack, the c-gNB) instantiated and running in the central cloud. The MANO system continuously collects data about the network parameters of the virtual network (i.e. latency, throughput, available and used radio resources). This is especially important for the URLLC slice, which has very stringent requirements on the end-to-end latency between the UE and the VR server running in the cloud (i.e. less than 80ms for perfect operation, less then 200ms for a not ideal but still usable experience). For this reason, the delay is constantly monitored to avoid operational glitches caused by a sudden delay increase due to external factors, such as additional incoming network slices causing congestion in the radio and transport, or internal ones, like a high number of UEs connected to the VR application. In these cases, the orchestration framework (more specifically the NFVO, as depicted in Figure 2) triggers a relocation of the network functions and VR application to the edge cloud, to benefit from the reduced latency.

Horizontal and Vertical VNF Scaling: the AI algorithm (such as the one described in [5]) running in the ENI module and the VNFM constantly monitors the cloud resources usage through specific probes and takes orchestration decisions in case of new slices coming in. For example, when a new eMBB slice is coming in, the intelligent orchestration framework has to take a decision about where to run the associated VNFs and what resources assign to them. Due to the scarcity of resources at the edge, the algorithm may take one of the following decisions, based on the requirements of the incoming eMBB slice and the constraints imposed by other slices:

- Horizontally scaling the higher layer RAN stack (i.e. creating a new instantiation of the running VNF) to support the additional load, if it is permitted;
- Vertically scaling the VNF, if there are resources available on the same server;
- Relocating the NF to a new location. In the latter case, the URLLC-related functions could be kept at the edge and just the eMBB ones are moved.

Admission Control: throughout the network operation, the management system is constantly running an AI/ML based algorithm, implementing selected features of the algorithms described in [6], [7]. These algorithms take as input both the available physical resources in terms of spectrum and cloud and the current orchestration patterns. The goal is to optimize the monetization of the network and the overall performance, also through periodic re-orchestrations.

2.3 PoC Success Criteria

All goals are met when the described functionality is proved to be available.

Demonstration of Horizontal and Vertical VNF Scaling, Admission Control
Report of the results of Horizontal and Vertical VNF Scaling, Admission Control
Description of the results and parameters of elasticity in the Horizontal and Vertical VNF Scaling.

References

[1] ETSI RGS ENI 008 (GS ENI 001), “Experiential Networked Intelligence (ENI); ENI Use Cases” v.2.0.4 (under development, expected publication in 2019), Sec 5.3.6
[2] ETSI DGS/ENI-005 (GS ENI 005), “Experiential Networked Intelligence (ENI); System Architecture” v.0.0.13 (under development, expected publication in 2019)
[3] ETSI GS ENI 006, “Experiential Networked Intelligence (ENI); Proof of Concepts Framework” v1.1.1
[4] ETSI RGS ENI 007 (GS ENI 002), “Experiential Networked Intelligence (ENI); ENI requirements” v.2.0.2 (under development, expected publication in 2019)
[5] 5G MoNArch Deliverable D4.1, “Architecture and mechanisms for resource elasticity provisioning” Sec 4.2.2.2
[6] 5G MoNArch Deliverable D4.1, “Architecture and mechanisms for resource elasticity provisioning” Sec 4.3.2.4
[7] 5G MoNArch Deliverable D4.1, “Architecture and mechanisms for resource elasticity provisioning” Sec 4.3.2.6
[8] 5G MoNArch Deliverable D2.2, “Initial overall architecture and concepts for