Gutiérrez-Estévez, D.M., et al. 5G-MoNArch use case for ETSI ENI: elastic resource management and orchestration, in: *2018 IEEE Conference on Standards for Communications and Networking (CSCN), 29-31 October 2018, Paris, France [Proceedings].* IEEE, 2018, 5 p.

DOI: https://doi.org/10.1109/CSCN.2018.8581789

5G-MoNArch Use Case for ETSI ENI: Elastic Resource Management and Orchestration

David M. Gutierrez-Estevez[1], N. di Pietro[2], A. de Domenico[2], M. Gramaglia[3], U. Elzur[4] and Y. Wang[1]
[1]Samsung Electronics R&D Institute UK, United Kingdom
[2]CEA-LETI, MINATEC, France
[3]Universidad Carlos III of Madrid, Spain
[4]Intel, USA

***Abstract*—5G networks will grant spectacular improvements of the most relevant Key Performance Indicators (KPIs) while allowing resource multi-tenancy through network slicing. However, the other side of the coin is represented by the humongous increase of the management complexity and the need for efficient algorithms for resource orchestration. Therefore, the management and orchestration of the network through Artificial Intelligence (AI) and Machine Learning (ML) algorithms is considered a promising solution, as it allows to reduce the human interaction (usually expensive and error-prone) and scale to large scenario composed by thousands of slice in heterogeneous environments. In this paper, we provide a review of the current standardization efforts in this field, mostly due to the work performed by the Experiential Network Intelligence (ENI) industry specification group (ISG) within the European Telecommunications Standards Institute (ETSI). Then, we thoroughly describe an exemplary use case on elastic network management and orchestration through learning solutions proposed by the 5GPPP project 5G-MoNArch and recently approved at ETSI ENI.**

I. INTRODUCTION

The 5th generation (5G) of cellular systems will change the access to technology for users, vertical markets and industries. Thanks to the 5G-enabled technical capabilities, they will experience a drastic transformation that will trigger the development of cost-effective new products and services. A large number of use cases and corresponding requirements for representative vertical markets such as automotive, health, factories of the future, energy, media, and entertainment will need agile access to network support functionalities [1], [2]. This will require a fundamental rethinking of the mobile network architecture and interfaces. In particular, the expected diversity of services, use cases, and applications in 5G requires a flexible, adjustable, and programmable network architecture. To this end, the latter must shift from the current network of entities to a network of capabilities.

In order to achieve the Key Performance Indicators (KPIs) envisioned by 5G, the most relevant standardization bodies have already defined the fundamental structure of the architecture and its building blocks [3], [4]. By leveraging on the novel concepts of Software Defined Networking (SDN), Network Function Virtualization (NFV) and modularization, the new architecture proposed by relevant organizations such as the 3rd Generation Partnership Project (3GPP) or the European Telecommunications Standards Institute (ETSI) will natively support the service diversity targeted by the future commercial ecosystem [4], [5].

Moreover, in sharp contrast with present and past generations of cellular communication systems, 5G will be able to exploit the cutting-edge tools and technological solutions made available by Artificial Intelligence (AI) and Machine Learning (ML), which have gained a lot of momentum in recent times. In response to the industry demand for AI-driven intelligent networks, ETSI has created the Experiential Network Intelligence (ENI) industry specification group (ISG) [6]. ENI's goal is to improve operators' experience and add value to the Telco provided services. In ENI's perspective, this can be done by means of automation and intelligence, assisting in complex decision making to deliver OPEX reduction and to enable 5G deployment. In particular, ENI aims to define an architecture that uses AI techniques and context-aware, metadata-driven policies. In this way, offered services can be adjusted based on changes in user needs, environmental conditions and business goals, according to the "observe-orient-decide-act" control loop model [7]. The adaptive capabilities of AI-enhanced network management and orchestration systems seem especially suitable to provide the dynamism required by the new 5G use cases, which are characterized by the need for preempting or reacting "on-the-fly" to substantial changes in user demands, service requirements, and resource availability.

In this paper, we focus on a use case that fits in ENI's framework and is issued from the combination of an AI-endowed network management and orchestration system with some of the technological innovations introduced and developed by the H2020 European project 5G-MoNArch [8]. The focus of 5G-MoNArch is about a versatile, adaptable, and programmable architecture for 5G, based on network slicing, advanced inter-slice control and cross-domain management, and Virtual Network Function (VNF) scalability [9], [10]. The use case that we present in this paper addresses mechanisms to exploit the flexibility of a 5G system by means of resource and network *elasticity*. This can be understood as the ability to gracefully adapt to load changes in an automatic manner such that at each point in time the available resources match the demand as closely and efficiently as possible. Elasticity strictly depends on the design of the communication and computational resource orchestration mechanisms of the network and on the automated handling of its virtualized and cloudified components. These automation mechanisms can greatly benefit from the employment of AI techniques in general and the integration of an ENI system in particular, which would allow optimized decisions to be made based on real data. The elastic management and orchestration of 5G networks through AI techniques eventually allows to increase the resource utilization efficiency without sacrificing performance.

The remainder of this paper is structured as follows. In Section 2, we provide a description of a prominent architecture for the use of AI in the management and orchestration of future networks proposed by ETSI ENI. Section 3 presents the use case proposed by the 5G-MoNArch project. Finally, conclusions are drawn in Section 4.

II. ETSI ENI BACKGROUND

In response to the industry demand for AI-driven intelligent networks, ETSI has created the ENI workgroup [3]. ENI's goal is to improve operator's experience and add value to the Telco provided services, by assisting in decision making to deliver OPEX reduction and to enable 5G deployment with automation and intelligence. In particular, ENI aims to define an architecture that uses AI techniques and context-aware, metadata-driven policies, to adjust service configuration and control based on changes in user needs, environmental conditions and business goals, according to the "observe-orient-decide-act" control loop model [4].

Network slicing for 5G can serve as a prime example to demonstrate ENI's architecture and the operator's benefits it provides, especially around VNF's computational resources efficiencies, while preserving the user requested SLA.

The Telco industry's evolution towards standardization of ML/AI-assisted networks, requires various industry consensus, including grammar and syntax for service policy and associated domain specific language (DSL), as well as data ingestion format, to foster ability to interact with the broad variety of tools used for management and monitoring. A *normalized* format is required also to address the difficulty to harmonize the state of the divergent infrastructure, due to use of silo specific tools e.g., per compute, network and storage and due to the variety of "assisted systems", each with different capabilities and different exposed API and varying degrees of ability to interact with ML/AI system, like ENI. It is therefore essential for ENI to define architecture components such as data ingestion and normalization, to provide a common base for ENI's inter-modular interaction as well as for transforming the external assisted system (e.g., a 3GPP/5G implementation) inputs to a format that is understood by ENI.

To date, ENI has defined a modularized system architecture, as shown in Figure 1. Having a modularized system architecture, facilitates the flexibility and generalization in the system design, as well as increase vendor neutrality. A brief description of each module, according to [4], is given below.

- The *Policy Management* module provides decisions to ensure that the Operator goals and regulator policies are met.
- The *Context Awareness* module describes the state and environment in which a set of the assisted system entities exists or has existed. For example, an operator may have a business rule that prevents 5G from a specific type of a network slice in a given location.
- The *Situational Awareness* module enables ENI to understand how information, events, and recommended commands that it may provide to the assisted system, may impact its next state, actions and ability to meet its operational goals.
- The *Cognition Management* module operates at the higher level and enables ENI as a whole to consult and meet its end to end goals.
- The *Knowledge Management* is used to represent information about ENI and the assisted system, differentiating between known facts, axioms, and inferences.

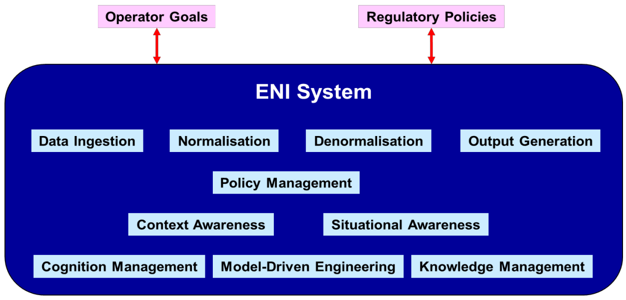

Figure 1. ETSI ENI modularized system architecture

The interaction and interoperability of ENI with an assisted system is determined by the latter's support of the ENI Reference Points. Specifically for the use of compute resources elasticity and efficiency, as presented in this paper, few elements, determined by relevant ENI Reference Points are needed. As depicted in Figure 2 below, the current NFVI Information allows ENI to be aware of the computational resources' capabilities (e.g., type of CPU, memory, data plane and accelerators) and availability (status and utilization level), while in turn this enables ENI to influence and optimize placement decisions made by the VIM, while ensuring that 3GPP policies, resources allocation and SLA are adhered too. Moreover, by using this information, ENI can further optimize resource utilization by i) enabling higher density for a given set of workloads under associated SLA, ii) anticipating and reacting to changing loads in different slices and assisting the VIM in avoiding resource conflicts, and/or iii) timely triggering of up/down scaling or in/out scaling of associated resources.

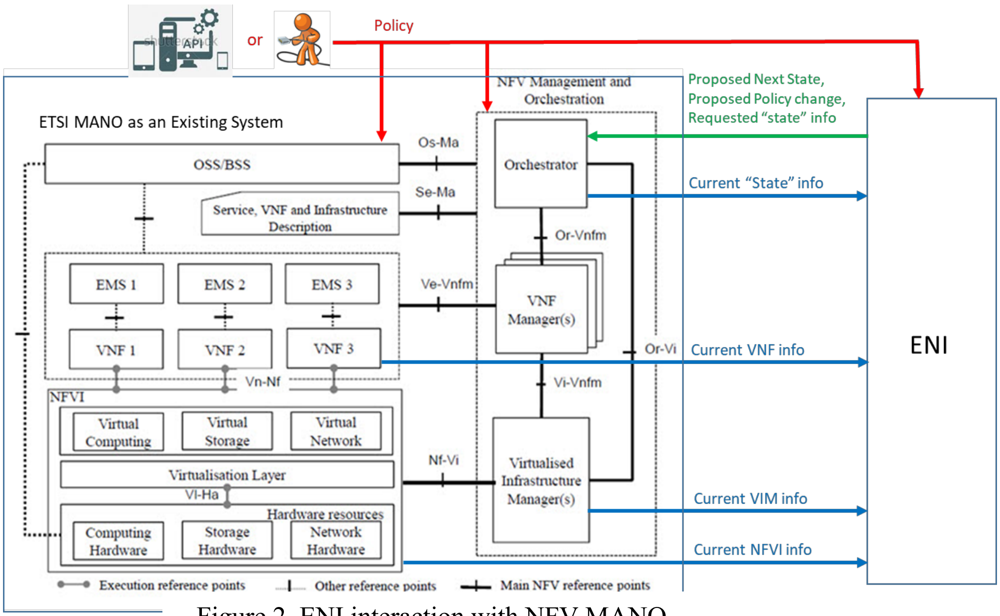

Figure 2. ENI interaction with NFV MANO

III. New ETSI ENI Use Case: Elastic Resource Management and Orchestration

A. *Use Case Context*

There is a need to design mechanisms that allow the 5G network infrastructure to be flexible enough to successfully host vertical services with very diverse requirements. An elastic resource management and orchestration increases the flexibility of the network by gracefully adapting the system configuration to the load and the available resources at every time. Furthermore, future networks to support network virtualization and network slicing need to be dimensioned by considering computational requirements and how these change with the network load. Therefore, autonomous and intelligent self-dimensioning of the network is targeted, along with a smart usage of the computational resources.

The elastic management and orchestration of resources can be achieved in different ways. Three different sets of elasticity mechanisms can be identified, each of them addressing a specific challenge.

- The computational aspects of network functions have not been taken into account in their original design, hence new computationally elastic VNFs should be redesigned to enable efficient network virtualization.
- Flexible mechanisms for orchestration and placement of NFs across central and edge clouds should be designed, considering source and destination hosts resources, migration costs, and services' requirements. In particular, latency requirement are a key driver for placement of VNFs, in addition to the computational requirements (see Figure 3).
- Multiplexing gains due to the sharing of the infrastructure and physical resources across different slices need to be fully exploited to enhance the system resource utilization. Moreover, an efficient network management has to capitalize on the possibility of sharing and re-using the same virtual resources for network slices with similar or identical requirements and shared VNFs.

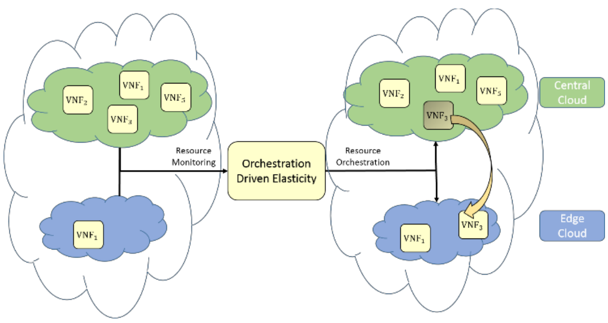

Figure 3. Dynamic VNF placement through and elastic Orchestrator.

The three above described challenges are the target of the proposed elastic management and orchestration of resources. To that aim, AI and the ENI system may play an important role as a tool to enhance the performance of elasticity algorithms. Prominent examples of performance-boosting capabilities that could be provided by the ENI system are the following: i) speeding the service deployment process by realizing an AI-based, automatic, accurate, and reliable mapping from service requests to network slice instantiations, ii) identification of similarities (in terms of requirements or shared VNFs) across slices to facilitate resource sharing, thus increasing the system resource utilization efficiency, iii) learning and profiling the computational utilization patterns of VNFs, thus relating performance and resource availability, iv) traffic prediction models for proactive resource allocation and relocation, v) optimized VNF migration mechanisms for orchestration using multiple resource utilization data (CPU, RAM, storage, bandwidth), and vi) optimized elastic resource provisioning to network slices based on data analytics.

B. Use Case Description

In the following, details are provided to the description of this use case.

Actors and roles: The AI-assisted "elastic" network management and orchestration is enabled by a predisposition to elasticity of the whole network infrastructure that provides end-to-end services through network slicing. Note, however, that this predisposition can be achieved with the standard 3GPP and ETSI NFV architecture, where management and orchestration functionalities of several architectural elements would be enhanced with elastic capabilities. In particular, the following architectural elements and elements play an active role in the current Use Case:

- *Management and Orchestration System*: it is composed of the functions from different network, technology, and administration domains (such as 3GPP public mobile network management, ETSI ISG NFV Orchestration) that manage network slices and related communications services across multiple management and orchestration domains in a seamless manner.
- *Network Slice Management Function (NSMF)*: it is part of the Management and Orchestration System (such as 3GPP public mobile network management, ETSI ISG NFV Orchestration) and would use AI to extend the 3GPP NSMF/NSSMF [3GPP TS 28.530] functionalities, in order to support the elastic intra-slice (or cross-domain) orchestration and the elastic cross-slice orchestration. The former deals with the orchestration of the different VNFs part of the same slice across multiple domains, while the latter addresses the joint orchestration of the multiple slices deployed on a common architecture. The NSMF also includes functions related to performance monitoring, measurement, and alarm. It is also in charge of defining and instantiating elastic slices, creating first the slice blueprint based on the service-related resource requirements and then defining the appropriated Network Slice Instance.
- *Elastic Slice*: a set of VNFs and the associated resources to support a mobile service with elastic (non-stringent) requirements that admit graceful performance degradation. This allows e.g. more flexibility in the allocation of resources and in the deployment of the associated VNFs.
- *Elastic VNFs*: they can be (re-)designed with elastic principles in mind such that the computational resources available for its execution are taken into account, or its temporal and/or spatial interdependencies with other VNFs are mitigated.
- *ENI System*: system solution that provides a set of AI methods (e.g. supervised/unsupervised and reinforcement learning schemes) to the Elastic Network Slice Management Function.

Initial context configuration: Consumer-facing service descriptions are mapped to network slice "blueprints". Based on the slice blueprints, a running network slice instance (NSI) is selected or created. Once the NSI deployed, the AI schemes can be used to predict network loads, estimate resource usages, and react accordingly by activating Elastic Cross-slice (or Intra-slice) Orchestrator functions in order to optimize the resource usage across slices and prevent system faults.

Triggering conditions: The ENI System may recommend or enforce the application of one or more algorithms for an elastic (re-)orchestration of resources when at least one of the following events happens:

- A new service request arrives.
- The resource requirements of a new slice cannot be satisfied in the current system configuration.
- The amount of resources allocated to one instantiated slices exceeds a given "efficiency" threshold.
- The requirements of running services change (or is predicted to change) and become substantially more stringent.
- A risk of imminent resource shortage is detected.

Operational flow of actions: During the slice setup process, the ENI System may be used first to define the slice blueprint; then, based on the slice blueprint, to identify whether it exists one deployed NSI that can support the new service, with a minimum amount of additional resources. Based on this, the resource required are allocated, the slice is instantiated and managed during its lifecycle. If there are not enough resources available prior to the slice instantiation or an alarm notifies congestions, the ENI System may be used to support the following "elastic" system adaptation functions:

1) Elasticity solutions at the VNF level: VNF computational resource scaling and graceful degradation of performance.
2) Elasticity solutions at the intra-slice level: migration of VNFs to different clouds, to create room for other VNFs with tighter (latency or computational) requirements or enhance the performance of the migrated VNFs.
3) Elasticity solutions at the cross-slice level: cross-slice resource management to maximise resource sharing and optimise the resource utilization efficiency.

The three (families of) elasticity functions mentioned above can be jointly executed and are not mutually exclusive. Nonetheless, in general, they act at different time scales and involve different hierarchical elements of the network architecture (e.g. cross-domain or per-domain).

Post-conditions: The Elastic Network Slice Management Function entails an improvement in the exploitation of the network resources. On the one hand, less resources are employed to guarantee the same QoS. On the other hand, more service requests can be accepted and treated at the same time, improving the network efficiency and reducing redundancy in resource exploitation. Network slicing is re-organised still meeting non-elastic slice requirements.

IV. Conclusions

In this paper, introduced a use case recently proposed and accepted by the 5G-MoNArch project to ETSI ENI. In it, we propose the novel idea of utilizing AI techniques with the purpose of exploiting the resource elasticity of a 5G network, hence improving resource efficiency and the overall performance of its management and orchestration machinery. Using as basis the architectural work recently developed by ETSI ENI and the concept of resource elasticity, we have provided here the use case details in terms of actors and roles, initial context configuration, triggering conditions, operational flow of actions, and post-conditions.

Acknowledgment

This work has been performed within the 5GMoNArch project, part of the Phase II of the 5th Generation Public Private Partnership (5G-PPP) program partially funded by the European Commission within the Horizon 2020 Framework Program.

References

[1] 3GPP, "TR22.891 – Study on new services and markets technology enablers," Sept. 2016.Xxx

[2] 5GPPP, "5G PPP use cases and performance evaluation," Apr. 2016.

[3] 5GPPP, "View on 5G architecture," v. 2.0, Dec. 2017.

[4] 3GPP, "TS 23.501 – System architecture for the 5G system; Stage 2", June 2018.

[5] ETSI, "GS NFV-IFA 014 – Network functions virtualisation (NFV); Management and orchestration; Network service templates specification," Oct. 2016.

[6] Y. Wang, R. Forbes, C. Cavigioli, H. Wang, A. Gamelas, A. Wade, J. Strassner, S. Cai, S. Liu, "Network management and orchestration using artificial intelligence: Overview of ETSI ENI", to appear in IEEE Communications Standards Magazine.

[7] ETSI ENI, "Improved operator experience through experiential network intelligence", White Paper, Oct. 2017.

[8] 5G-MoNArch: 5G mobile network architecture for diverse services, use cases, and applications in 5G and beyond, https://www.5g-monarch.eu.

[9] 5G-MoNArch, "Architecture and mechanisms for resource elasticity provisioning," deliverable 4.1, May 2018.

[10] 5G-MoNArch, "Initial overall architecture and concepts for enabling innovations," deliverable 2.2, June 2018.