Reliable Slicing in 5G Networks

Luca Valcarenghi, Alessio Giorgetti, Barbara Martini, Koteswararao Kondepu, Molka Gharbaoui, Piero Castoldi

Abstract In 5G, the term slicing refers, in general, to the possibility for different customers (usually called tenant) to share the same physical network. Thanks to the *softwarization* of networks according to the Network Function Virtualization (NFV) concept and the *programmability* of network connectivity through Software Defined Networking (SDN), new network and service capabilities can be envisioned by integrating networking, computing and storage resources while serving a multitude of tenants. Each tenant is assigned a logical network that can satisfy its requirements. Survivability is one of the most important requirements especially for vertical applications requesting Ultra Reliable Low Latency Communications (URLLC). In this chapter the concept of slice is introduced and the some use cases for providing reliability in a slice.

1 5G Slicing

With the advent of NFV and SDN a novel network scenario is envisioned enabled by network deployments into the cloud also extended to the network edge and by programmability of network connectivity through network controllers. This trend known as *softwarization* is enabling new unique network and service capabilities by integrating networking, computing and storage resources into one programmable and unified infrastructure while serving a multitude of distributed smart devices and applications (e.g., robots, drones, smart vehicles). As result, current communications network scenario is moving from having a separate network for each application (e.g., fixed telephone network, mobile telephone networks, Internet access) to a single network shared by different applications or verticals. Network Slicing is a key feature of the 5G System that allows Operators to flexibly structure the network resources to match the services offered to subscribers, third-party customers, including the roaming scenario. The concept of slicing emerged as a way of setting up several logical networks for different verticals on the same physical network. Each vertical is then assigned to the logical network that guarantees the required QoS. Such setup potentially allows communication providers to save capital and operating expenditures (CAPEX and OPEX). However, as for any shared medium, guaranteeing the required QoS to network slices sharing the same physical network is not a trivial task and remains an open issue. In particular, slice control and management planes shall be designed for slice provisioning and dynamic reconfiguration and the data plane shall guarantee each slice requirements (e.g., QoS requirements, slice isolation, etc.).

This chapter overviews how the concept of slice is defined in different Standard Developing Organizations (SDOs) and research projects.

1.1 The Concept of Slice in SDOs and Research Project

Several SDOs are focusing on the network slicing concept 1. The Next Generation Mobile Networks (**NGMN**) alliance defines a Network Slice Instance (NSI) as *"... a set of network functions, and resources to run these network functions, forming a complete instantiated logical network to meet certain network characteristics required by the Service Instance(s)"* 2. In 2 the network slicing concept consists of three layers depicted in **Figure 1**: Service Instance Layer, Network Slice Instance Layer, and Resource layer. The Service Instance Layer represents the services (i.e., end-user or business services) which must be supported. The Network Slice Instance Layer provides the network slice instances with specific network characteristics that are required by the related Service Instances (e.g., Enhanced MBB, M2M, Enterprise and Industry). The Resource Layer provides the physical or virtual resources for slice deployment.

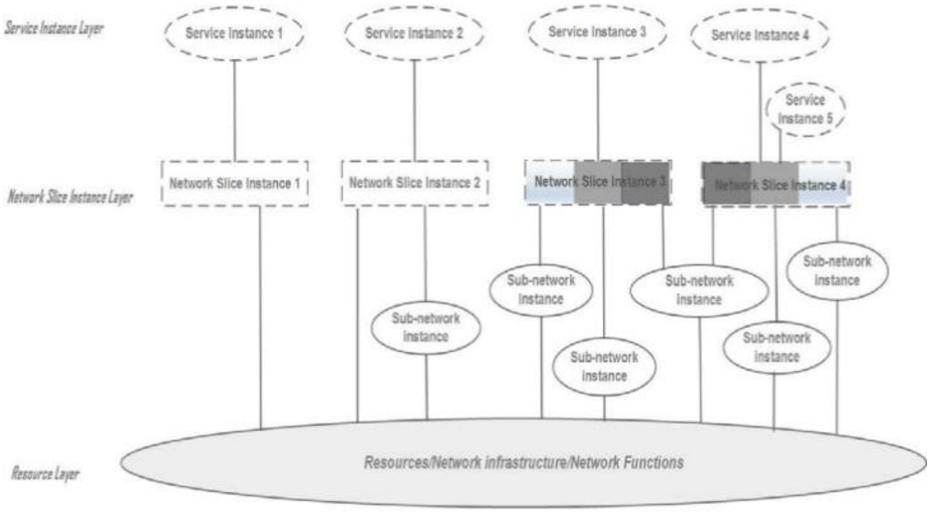

Figure 1 Slice concept in 2

3GPP in TR 28.801 3 defines, in accordance with NGMN, a network slice instance (NSI) as *"... a set of network functions and the resources for these network functions which are arranged and configured, forming a complete logical network to meet certain network characteristics…"*. In addition, it defines the following phases of a network slice lifecycle: preparation phase, instantiation, configuration and activation phase, run-time phase, decommissioning phase. Moreover, it introduces three management functions to manage the NSIs to support communication services: Communication Service Management Function (CSMF), responsible for translating the communication service related requirements to network slice related requirements; the Network Slice Management Function (NSMF), responsible for management and orchestration of NSI; and the Network Slice Subnet Management Function (NSSMF), responsible for management and orchestration of a network slice subnet instance (NSSI). Finally, it defines the different roles of the actors (e.g., costumers, providers, operators, etc.) involved in slice provisioning as depicted in **Figure 2**. However, TR 28.801 does not specify how to implement such functions and their relationship with respect to the ETSI NFV architectural framework.

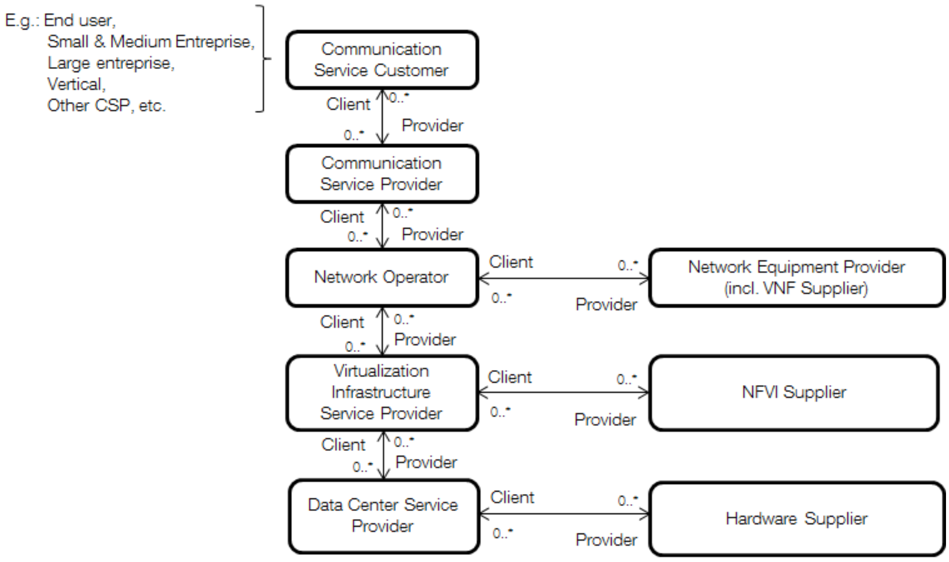

Figure 2 High level function of roles in 3

ETSI NFV EVE012 4 establishes the correspondence between a network slice (3GPP) and a network service (ETSI NFV). There, ETSI describes that an NFV Network Service (NFV-NS) can be regarded as a resource-centric view of a network slice, for the cases where a NSI would contain at least one virtualized network function. Moreover, ETSI NFV EVE012 proposes that 3GPP slice management functions interact with ETSI NFV Architecture through the Os-Ma-Nfvo reference point as depicted in **Figure 3**. However, in 4 it is stated that *"...3GPP slice-related management functions are still under definition in 3GPP SA5 and future updates might require further analysis about the interaction between 3GPP slicing related management functions and NFV-MANO..."*.

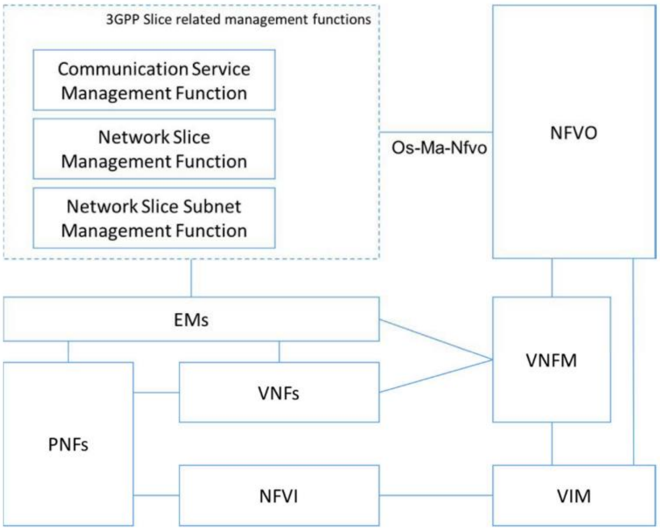
3GPP Slice related management functions
Communication Service Management Function
Network Slice Management Function
Network Slice Subnet Management Function
Os-Ma-Nfvo
NFVO
EMs
VNFM
VNFs
PNFs
NFVI
VIM

Figure 3 Interaction between slice management functions (3GPP) and ETSI NFV Architecture from 4

Additional definitions of slices have been proposed by the following organizations:

- The Internet Engineering Task Force (**IETF)** in 5 6 7
- The Broadband Forum (**BBF**) in 8
- The Optical Networking Forum (**ONF)** in 9
- **ITU-T** through the Focus Group on IMT (International Mobile Telecommunication)-2020 (FG IMT-2020) in 10.
- Metro Ethernet Forum (**MEF**) in [11][12]

Within the research community, the 5G-Transformer project 14 envisions three functional layers for providing verticals with slices: a Vertical Slicer as the logical entry point for verticals to support the creation of their respective transport slices in a short time-scale (in the order of minutes), a Service Orchestrator to orchestrate the federation of transport networking and computing resources from multiple domains and manage their allocation to slices, and a Mobile Transport and Computing Platform (5GT-MTP), that provides and manages the virtual and physical IT and network resources on which slices are deployed. Such architecture implementation in under development.

The 5G!Pagoda research project in 15 overviews most of the ongoing network slicing-related activities. In addition, it proposes a functional architecture of the slicing system and it delineates key implementation elements. The functional architecture of the slicing system for the single domain is based on a Domain Specific Slice Orchestrator (DSSO), a Slice Operations Support (SOS), a Slice Management Plane (SMP), a Slice Software Layer (SSL), a Slice Resource Layer (SRL), a Virtual Computing/Storage/Connectivity Infrastructure Layer (VCSCI), and a Physical Computing/Storage/Connectivity Infrastructure Layer (PCSCI).

Within the general architecture proposed by the research project SONATA 16, a “Slice Management Functional Block” is defined and integrated in the NFV Management and Orchestration (NFV MANO) functional block, detailed in the ETSI NFV architectural framework. In addition, the implementation of such functions has been developed within the project.

2 Reliability in 5G Slices

The envisioned 5G network architecture, including the Next Generation Core (NG Core, i.e., the new Evolved Packet Core --- EPC --- for 5G) and the New Radio Access Network (New RAN), will be heavily based on virtual network functions (VNFs) [17]. Network function virtualization (NFV) enables an easy introduction of new network services by adding dynamic programmability to network devices (e.g., as routers, switches, and applications servers) that, in turn, empowers fast, flexible, and dynamic deployment of new network and management services. Moreover, network function virtualization also enables network slicing by providing multiple instances of the same network function. In this context, the dynamic service chaining allows the delivery of a new breed of applications (e.g., cloud robotics, smart cities) by dynamically selecting and composing computational and network services deployed as virtual functions (VFs) in distributed micro-clouds located at the network Edge closer to the users [25]. The exploitation of network function virtualization is foreseen also in the NG core [18] and the

New RAN technology [19]. In the NG Core, the different network functions (e.g., Access and Mobility Function (AMF), Session Management Function (SMF), Policy Control Function (PCF), Application Function (AF), Authentication Server Function (AUSF), User Plane Function (UPF), and User Data Management (UDM)) can be virtualized, as it has been proposed for LTE-A [18], and placed in different virtual machines (VMs) or run as a single bundle in one VM.

As specified in [19] "*...Service continuity is not only a customer expectation, but often a regulatory requirement ...*". Thus, in a scenario where network functions are virtualized, both hardware and software failures assume the same importance, and their reliability shall be guaranteed. Similarly, reliability at service chain level is important to assure proper service availability features to application service platforms deployed by verticals [25][26].

2.1 Virtualised EPC reliability

In the technical specification document (TS) 23.007 [20], 3GPP specified different failure detection and recovery mechanisms for EPC components, including detection of path failure with the help of Echo Request/Echo Response timer messages. Moreover, approaches for recovering failures in a scenario where a mobile network function is virtualised can stem from schemes already proposed for grid and cloud networking [21]. Furthermore, scalable architectures for reliability management are being defined by ETSI NFV [22] and implemented in current open source orchestration frameworks such as Openstack [23]. However, the performance of resilience schemes based on the aforementioned approaches once applied to 5G mobile network have not been fully evaluated so far.

This section demonstrates the capability of *recovering vEPC failures* by means of a vEPC in "hot backup". Both working vEPC and backup vEPC are deployed in multiple Network Function Virtual Infrastructure Points of Presence (NFVI-PoPs) made available by the federated testbeds belonging to the SoftFIRE project [24]. The demo is designed to evaluate the *Service Recovery Time (SRT)*, that is the time required to regain user equipment (UE) connectivity, when the proposed resilient scheme is deployed in different NFVI-PoPs.

The considered scenario and the proposed resilience scheme are depicted in Fig. 4 and Fig. 5 by referring to functional elements of the Long Term Evolution-Advanced (LTE-A) architecture.

The proposed resilience scheme considers a scenario where the vEPC fails (e.g., a virtual machine where the vEPC runs crashes). Fig. 4 shows the two considered vEPC resilience schemes based on vEPCs hot backup deployed in federated NFVI-PoPs. The one on the left features two co-located vEPCs (i.e., vEPCa and vEPCb deployed in Surrey 5GIC testbed) while the one on the right features a remote hot backup vEPC (i.e., vEPCr) deployed in a different compute resource available in another testbed (i.e., Fokus). In the latter case two testbeds will be contemporarily utilized to implement the resilience scheme. In the Surrey 5GIC testbed two different VNF functions (vOAISIM and vEPC) will be implemented by exploiting open source mobile platforms (i.e., OpenAirInterface-OAI). Here, vOAISIM VNF provides emulation of virtual user equipment (vUE) and evolved NodeB (eNB) while vEPC will be used to emulate the core network.

Fig. 5 shows the considered scenario and lifecycle event when vEPC VNF fails. Here, when VNFs are deployed, vOAISIM connects with vEPCa, and Zabbix server start monitors the VNFs that are associated corresponding Zabbix agent. Note that each vEPC VNF and vOAISIM VNF deployement contain also Zabbix agent. If the Zabbix server detects an anomaly activity in vEPCa (e.g., overload) or does not receive any status report from vEPCa (i.e., vEPCa crashed) for a pre-defined period of time (i.e., time to trigger the activity), the Zabbix server check the status of the hot backup vEPC to initiate a recovery procedure. The receovery procedure consist in reconfiguring vOAISIM to connect to the hot backup vEPCr. Upon reconfiguration vOAISIM is able to communicate hot backup vEPCr. Similarly, the experiment also demonstrate to the recovery based on the local vEPCb deployed in 5GIC testbed.

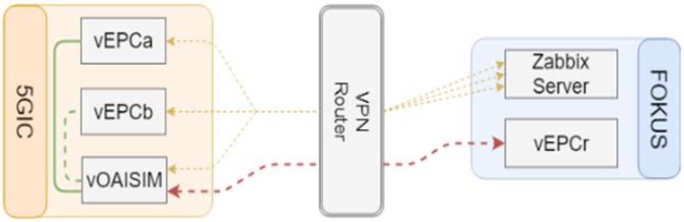

Figure 4: RAN and Core network deployment in federated environment

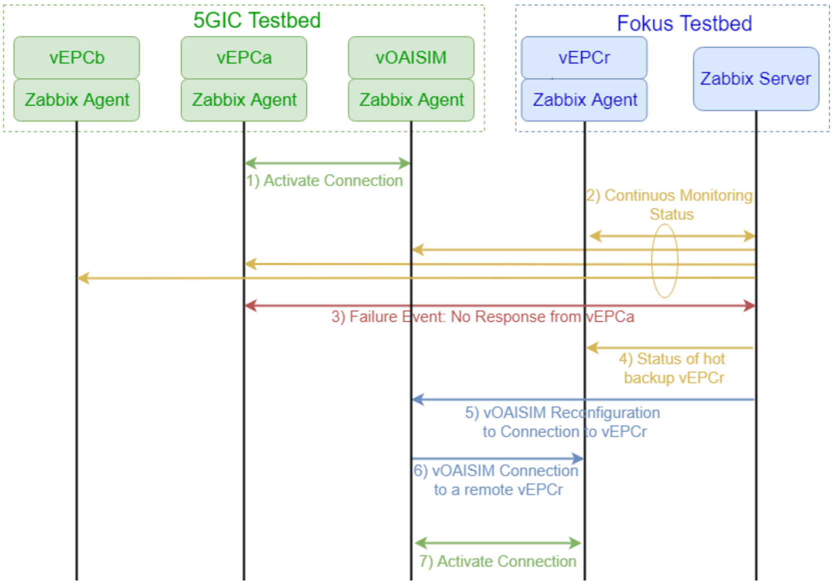
5GIC Testbed
Fokus Testbed
vEPCb
vEPCa
vOAISIM
vEPCr
Zabbix Agent
Zabbix Agent
Zabbix Agent
Zabbix Agent
Zabbix Server
1) Activate Connection
2) Continuos Monitoring Status
3) Failure Event: No Response from vEPCa
4) Status of hot backup vEPCr
5) vOAISIM Reconfiguration to Connection to vEPCr
6) vOAISIM Connection to a remote vEPCr
7) Activate Connection

Figure 5: Proposed scheme experimental evaluation setup

2.2 Service chaining reliability

The advent of SDN and NFV enables a convergent network-cloud ecosystem offering more effective and operative network and service deployments on top of virtual networking, computing and storage resources integrated into one programmable and unified infrastructure while serving a multitude of distributed smart devices and applications (e.g., robots, drones, smart vehicles) in turn being part of the infrastructure itself.

Thanks to *softwarization*, a scenario can be envisioned where service providers may offer not only communication services, but also virtualized computing and storage capabilities by elastically slicing the (cloud and network) infrastructure into partitions (i.e., *network slices*) offering customized network functions and services (e.g., NAT, firewall, deep packet inspection) tailored for specific applications. Moreover, with the softwarization of telecommunication infrastructures, a new breed of applications can be conceived (e.g., cloud robotics, smart cities) by dynamically composing (i.e., chaining) computational and network services deployed as virtual functions (VFs) in distributed micro-clouds located at the Edge of the current telecommunication infrastructure. Indeed, SDN can effectively provide programming abstractions that can be exploited for the dynamic enforcement and in-line steering of data traffic along the network path of service chains (i.e., service chain paths).

Through slicing and dynamic service chaining, service providers can deploy service infrastructures to serve many different verticals while saving capital and operating expenditures (CAPEX and OPEX). However, the concurrent usage of resources, the high dynamicity of services and the geographical distribution of VFs pose new challenges to service providers in terms of service lifecycle management and automation to address the QoS and service availability requirements of heterogeneous applications. To this purpose, close control loops and techniques are required towards providing automation, resource usage optimization and reliability eventually leveraging network analytics assisted decisions [27]. In this direction, ONAP is working on new solutions for providing automation, performance optimization and, in general, service lifecycle management capabilities [28]. On the other hand, the reliability of service chains is stated as a primary requirement to assure proper service availability [29][30]. However, the problem to address QoS and service chaining reliability is challenging due many different and heterogeneous application requirements. A way to effectively address reliability is to assure adaptive resource provisioning and protection mechanisms while service chains runs aiming at preventing service degradations due to the concurrent use of resources from different applications [31][32]. Moreover, application-oriented mechanisms are desirable that can be achieved through intent-based approach [33] and detection of service degradation on end-to-end basis [32].

References

1. https://datatracker.ietf.org/meeting/99/materials/slides-99-netslicing-alex-galis-netslicing-terms-and-systems
2. NGMN Alliance, “Description of Network Slicing Concept”, NGMN 5G P1 Requirements & Architecture Work Stream End-to-End Architecture, v1.0.8, Sep. 14, 2016, https://www.ngmn.org/fileadmin/user_upload/161010_NGMN_Network_Slicing_framework_v1.0.8.pdf (accessed Apr. 30, 2018)
3. “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication management; Study on management and orchestration of network slicing for next generation network (Release 15)”, 3GPP TR 28.801 V15.1.0 (2018-01)
4. ETSI GR NFV-EVE 012, “Network Functions Virtualisation (NFV) Release 3; Evolution and Ecosystem; Report on Network Slicing Support with ETSI NFV Architecture Framework” v3.1.1, December 2017.
5. https://datatracker.ietf.org/wg/netslicing/about/ (accessed april 30, 2018)
6. L. Qiang *et al.*, “Technology Independent Information Model for Network Slicing draft-qiang-coms-netslicing-information-model-02, https://tools.ietf.org/html/draft-qiang-coms-netslicing-information-model-02, Expiration date: July 30, 2018 (accessed April 30, 2018)
7. “Applicability of Abstraction and Control of Traffic Engineered Networks (ACTN) to Network Slicing draft-king-teas-applicability-actn-slicing-02”, https://tools.ietf.org/html/draft-king-teas-applicability-actn-slicing-02, Expiration date: July (accessed April 30, 2018)
8. https://www.broadband-forum.org/5g
9. ONF, “TR-526 Applying SDN Architecture to 5G Slicing”, April 2016
10. Yoshinori Goto, “Activity Report of ITU-T Focus Group on IMT-2020”, NTT Technical Review, Vol. 15 No. 6 June 2017
11. http://www.mef.net/Assets/Technical_Specifications/PDF/MEF_55.pdf (accessed April 30, 2018)
12. https://www.mef.net/Assets/White_Papers/MEF_Third_Network_LSO_Vision_FINAL.pdf (accessed April 30, 2018)
13. https://www.gsma.com/futurenetworks/5g/introduction-to-5g-network-slicing/ (accessed April 30, 2018)
14. 5G-Transformer website, http://5g-transformer.eu/
15. 5G!Pagoda, “D2.3: Initial report on the overall system architecture definition”, https://goo.gl/2Z7PWu, June 30, 2017
16. SONATA, “D2.2 Architecture Design”,
17. 5G PPP, “View on 5g architecture,” White Paper, https://5g-ppp.eu/white-papers (accessed 10/07/2017).
18. V. G. Nguyen et al., "SDN/NFV-Based Mobile Packet Core Network Architectures: A Survey," in IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1567-1602, third quarter 2017.
19. 3GPP TR 38.801, “Study on new radio access technology: Radio access architecture and interfaces (Release 14)”, V14.0.0 (2017-03).
20. 3GPP TS 23.007, “Technical Specification Group Core Network and Terminals; Restoration procedures”, Dec 2017.
21. L. Valcarenghi, et al., “Quality-of-service-aware fault tolerance for grid-enabled applications”, Optical Switching and Networking, Volume 5, Issues 2–3, 2008.
22. ETSI GS NFV-REL 002 V1.1.1 (2015-09), “Network Functions Virtualisation (NFV); Reliability; Report on Scalable Architectures for Reliability Management”.
23. F. F. Moghaddam, et al., "Self-Healing Redundancy for OpenStack Applications through Fault-Tolerant Multi-Agent Task Scheduling," IEEE CloudCom 2016
24. EU SoftFIRE project, https://www.softfire.eu/
25. M. Gharbaoui, C. Contoli, G. Davoli, G. Cuffaro, B. Martini, F. Paganelli, W. Cerroni, P. Cappanera, P. Castoldi, “Experimenting latency-aware and reliable service chaining in Next Generation Internet testbed facility,” IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), 2018.

26. K. Antevski, et al., “Resource Orchestration of 5G Transport Networks for Vertical Industries,” arXiv preprint arXiv:1807.10430, 2018.
27. N. Sousa, et al., “Network Service Orchestration: A Survey, ” 2018.
28. http://github.com/onap/clamp
29. A. Hmaity, et al., "Virtual Network Function placement for resilient Service Chain provisioning," 8th International Workshop on Resilient Networks Design and Modeling (RNDM), Halmstad, 2016.
30. M. Gharbaoui, C. Contoli, G. Davoli, G. Cuffaro, B. Martini, F. Paganelli, W. Cerroni, P. Cappanera, P. Castoldi, “Demonstration of Latency-Aware and Self-Adaptive Service Chaining in 5G/SDN/NFV infrastructures,” IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), 2018.
31. AA. Mohammed, M. Gharbaoui, B. Martini, F. Paganelli, P. Castoldi, “SDN controller for network-aware adaptive orchestration in dynamic service chaining,” IEEE NetSoft Conference and Workshops (NetSoft), Seoul, 2016.
32. M. Gharbaoui, S. Fichera, P. Castoldi, B. Martini, “Network orchestrator for QoS-enabled service function chaining in reliable NFV/SDN infrastructure,” ," IEEE NetSoft Conference and Workshops (NetSoft), 2017.