Network Slicing in Networks Beyond 5G – NECOS[1] White Paper

Alex Galis (*), Christian Esteve Rothenberg (**), Joan Serrat (***)

(*) University College London - a.galis@ucl.ac.uk
(**) University of Campinas - chesteve@dca.fee.unicamp.br
(***) Polytechnic University of Catalonia - serrat@tsc.upc.edu

1. Introduction and Context

1.1. 5G Networks

5G networks are conceived as extremely flexible and highly programmable end-to-end connect-and-compute infrastructures that are both application- and service-aware, as well as being time-, location- and context-aware. These 5G networks represent: (i) an evolution, over 4G networks, in terms of capacity, performance, and spectrum access in radio network segments and (ii) an evolution of native flexibility and programmability conversion in all radio and non-radio 5G network segments: Radio Access Network, Fronthaul and Backhaul Networks, Access Networks, Aggregation Networks, Core Networks, Mobile Edge Networks and Clouds, Software Networks, Cloud Networks, Satellite Networks, and Edge IoT Networks.

5G networks represent a shift in networking paradigms: namely a transition from today's "network of entities" toward a "network of functions." Indeed, this "network of functions," and its most likely manifestation, a "network of virtual functions," will, in some cases, result in the decomposition of current monolithic network entities into network functions. These functions will constitute the unit of networking for next-generation systems and should be able to be composed in an "on-demand," "on-the-fly" basis. One of the current research challenges consists of devising the architecture and design of solutions that identify the set of elementary functions or blocks from which to compose network functions, which are today implemented as monolithic elements.

New concepts and techniques leveraged by 5G networks include: (i) Realizing network slicing in a cost-efficient way; (ii) Addressing both end-user and operational services; (iii) Supporting softwarization natively; (iv) Integrating communication and computation, and (v) Integrating heterogeneous technologies (including fixed and wireless technologies).

5G networks are expected to present several advantages. One, in particular, is a high degree of flexibility. They enforce the necessary degree of flexibility, where and when needed, with regards to capability, capacity, security, elasticity, and adaptability. These networks will serve highly diverse types of communication – for example, between humans, machines, devices and sensors – with different performance attributes.

5G communication technologies are making rapid progress towards commercial deployment, including URLLC – Ultra-Reliable and Low Latency Communication, mMTC – massive Machine Type Communication, eMBB – enhanced Mobile Broadband, and UDN – Ultra-Dense Networking and IoT trends. Emerging vertical industry applications, such as autonomous/cooperative driving, drone surveillance, remote sensing, and data analytic, are causing the conventional cloud computing model to evolve toward edge computing by utilizing these advanced communications to deliver ultra-low latency, high broadband, and customized data transport service.

[1] This paper summarizes the results of work conducted by the eleven partners of the NECOS project consortium in the context of 5G networks and beyond. The project was funded by the European Union (H2020-777067) and the Rede Nacional de Ensino e Pesquisa under the EU-Brazil Joint Call EUB-01-2017

1.2. Networks Beyond 5G

Figure 1 represents the evolution trends of 5G networks. A Network Beyond 5G will have to go well beyond the characteristics that have been identified for 5G infrastructures so far and will be exemplified by the design and implementation of new platforms, whose characteristics and capabilities are expected to go well beyond those that have been described so far in 5G including:

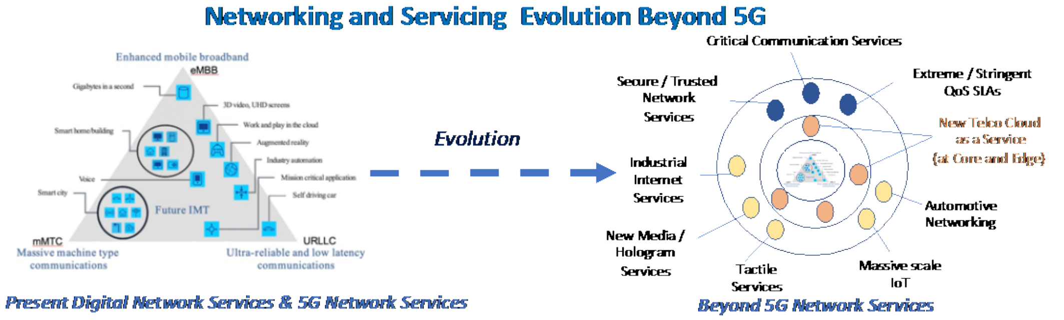

Figure 1- Networking and Servicing Evolution Beyond 5G

- Evolution of network services towards very low latency, very high bandwidth, very high reliability / resilience, trustworthiness and privacy, delivering stringent non-functional requirements / KPIs / SLAs per slice needed for future precision services (e.g., Immersive services: Hologram as a service, Augmented Reality and Virtual Reality services, 360-degree video, Industrial Internet services).
- Network service delivery evolution towards fully integrated, highly automated, greatly autonomic and substantially intelligent infrastructure and S/W platform, i.e., New Telco Clouds, where integration is based on a high dynamic interaction between all groups of communication, compute, storage and network services/applications resources/devices in all network segments with high-precision service assurance per slice.
- Hyper scale integration of elements and many networks (i.e., network devices, network (virtual) functions, edge elements and digital objects) in new telco clouds.
- Gradual replacement of IP best effort service provision with service information considered with network conditions together to achieve guarantees for KPIs or QoS as required by future precision services and applications per slice.
- In-network computation as an evolution of telecom clouds to support computation for third parties in both the core and edge.
- Network programmability as a key architectural principle and design choice to enable soft network evolution as requirements change.
- Anonymity and security support for all network operations with guarantees and monitoring abilities for mission critical services per slice.

2. Network Slicing

Slicing is a move toward advanced segmentation of resources and deployment of networking and servicing elements for the purpose of enhanced services and applications on a shared infrastructure in both 5G and Beyond 5G.

Slices are expected to considerably transform the networking perspective and enhance network architectures by

(i) Abstracting away the lower-level elements, in various ways;

(ii) Isolating connectivity at a subnetwork level;

(iii) Separating logical network behaviors from the underlying physical network resources;

(iv) Allowing dynamic management of network resources by managing resource-relevant slice configuration;

(v) Simplifying and reducing the expenditure of operations;

(vi) Supporting for rapid service provisioning; and

(vii) Supporting for network services deployment.

We refer to as a Cloud Network Slice a set of infrastructures (i.e., network, cloud, data centre - DC), components/network functions, infrastructure resources (i.e., connectivity, compute, and storage manageable resources), and service functions that have attributes specifically designed to meet the needs of an industry vertical or a service. As such, a network slice is a managed group of subsets of resources, network functions/network virtual functions at the data, control, management/orchestration, and service planes at any given time. The behavior of the network slice is realized via network slice instances (NSIs) (i.e. activated slices, dynamically and nondisruptive reprovisioned).

A management mechanism is required for allocating dedicated infrastructure resources and service functions to users.

The Cloud Network Slice key elements are provided below:

- A cloud network slice supports at least one type of service.
- A cloud network slice may consist of cross-domain components from separate domains in the same or different administrations, or components applicable to the infrastructure.
- A resource-only partition is one of the components of a Cloud Network Slice, however on its own does not fully represent a Network Slice.
- A collection of cloud slice parts from separate domains is combined, connected through network slices, and finally aggregated to form an end-to-end cloud network slice.
- Underlays / overlays supporting all services equally (with ‘best effort” support) are not fully representing a Network Slice.

The key characteristics of Cloud network slicing include:

1) The concurrent deployment of multiple logical, self-contained, and independently shared or partitioned slices on a common infrastructure platform.

2) Dynamic multiservice support, multitenancy, and integration meant for vertical market players.

3) The separation of functions, simplifying the provisioning of services, the manageability of networks, and integration and operational challenges, especially for supporting communication services.

4) The means for network/cloud operators/ISP and infrastructure owners to reduce operations expenditure, allowing programmability and innovation necessary to enrich the offered services, for providing tailored services, and allowing network programmability to over-the-top (OTT) providers and other market players without changing the physical infrastructure.

5) Slicing capability exposure: through a utilization model, the providers can offer Application Programming Interfaces (APIs) to the vertical business customers for granting the capability of managing their own slices. Such management actions can include dimensioning and configuration.

6) Integration at customer premises: complementary network segments, in some cases pertaining to the vertical business customer, become an integral part of the solution, requiring a truly convergent network including the integration in existing business processes as defined by the vertical customer.

7) Hosting applications, offering the capability of hosting virtualized versions of network functions or applications, including the activation of the necessary monitoring information for those functions.

8) Hosting on-demand third parties/OTTs, empowering partners (third parties/OTTs) to directly make offers to the end customers augmenting operator network or other value creation capabilities.

3. Lightweight Slice Defined Cloud (LSDC) – NECOS Project

The NECOS project addresses relevant limitations of current cloud computing infrastructures to respond to the new services demands, as presented in two use-cases that are instantiated in a set of six scenarios. The first use-case is telco service provider focused and is oriented towards the adoption of cloud computing in large networks. The second use-case is targeting the use of edge clouds to support devices with limited computation and storage capacity. The envisaged solution is based on a new concept – ***Lightweight Slice Defined Cloud (LSDC)*** – as an approach that extends the virtualization to all the resources in the participating networks and data centres, providing uniform management with advanced levels of orchestration.

The NECOS approach is manifested in a newly developed platform whose main distinguishing features are:

- The Slice as a Service –- a new deployment model. A slice is a grouping of resources managed as a whole, which can accommodate service components, independent of other slices.
- Embedded algorithms for an optimal allocation of resources to slices in the cloud and networking infrastructure to respond to the dynamic changes of the various service demands.
- A management and orchestration approach making use of artificial intelligence techniques in order to tackle the complexity of large-scale virtualized infrastructure environments.
- Making reality the lightweight principle in terms of small footprint components deployable on a large number of small network and cloud devices at the edges of the network.

3.1. Use Cases

The NECOS project is driven by two use cases, namely the Telco Cloud and the Multi-Access / Mobile Edge Computing (MEC). These use cases are of paramount importance for the industry today and act as target platforms on top of what a number of relevant scenarios are presumed to be supported, controlled and managed through the NECOS platform, leveraging the concept of slicing as a form of segregating multiple services in the same (federated) cloud networking substrate. Specifically, these scenarios are the *virtual Radio Access Network* (vRAN) scenario, the *5G Services* scenario, the *Virtual Customer Premises Equipment* (vCPE) scenario, the *Network Slicing for Touristic Content Distribution* scenario, the *Multi-Domain Network Slicing for Next Generation Touristic Applications* scenario, the *Network Slicing for Smart Cities Data Content Distribution* scenario and the *Network Slicing for Metropolitan Integrated Monitoring* scenario. These scenarios have been specified in order to identify the functional and non-functional requirements that NECOS platform should satisfy as well as the critical success factors and key performance indicators.

The priority and relevance of the above-mentioned requirements were also analysed. This analysis was based on the Quality Function Deployment (QFD) methodology developed by Y. Akao. For each scenario, we evaluated correlations between the identified requirements and NECOS Critical Success Factors/NECOS Key Performance Indicators/NECOS Expected Differentiated Factors (NECOS Characteristics). In other words, we evaluated how each requirement is contributing to solve/enable each Project Critical Success Factors/Project Performance Indicators/Project Expected Differentiated Characteristics as seen from each scenario.

Efforts have also been devoted to the analysis of the ecosystem enabled by NECOS as a first step to study the viability of business models associated to it. Three different perspectives were considered: one for the cloud provider receiving the slice demands from the tenant; one from the federated cloud providers; and one from the overall federation. Simulations and analytical studies have been performed, deriving from them business guidelines related to NECOS-based solutions.

3.2. NECOS Architecture

The NECOS architecture is one of the main project outputs towards infrastructure slicing to provide Slice as a Service as established in the project objectives. The Lightweight Slice Defined Cloud (LSDC) stands as the core embodiment of this architecture, automating the configuration process of clouds and their interconnecting networks aimed to substantiate the infrastructure slicing. This is layered on the partitioning and virtualization of resources in data centres and wide area networks participating in the slice lifecycle and by providing uniform management with a high level of automaticity for the currently separated realms of computing, connectivity, and storage resources.

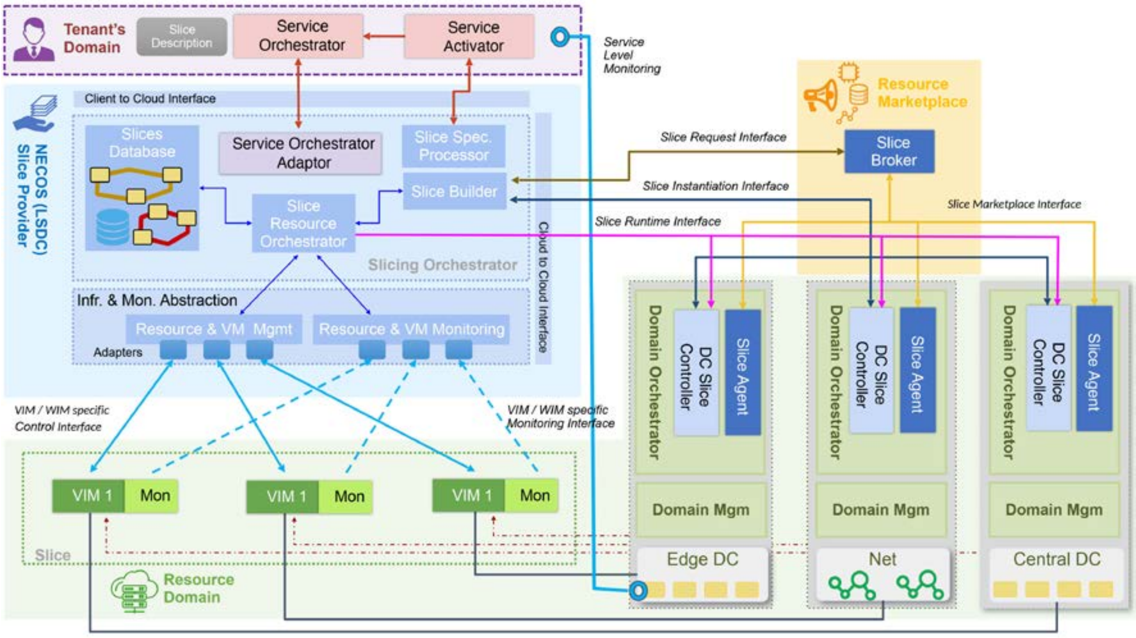

Figure 2- NECOS Functional Architecture

This architecture has been devised to support the Slice as a Service approach. The NECOS architecture provides novel artifacts to instantiate the LSDC, and for cloud network slicing in general, based on three main sub-systems:

(1) the NECOS (LSDC) Slice Provider subsystem (colored in blue),

(2) the Resource Marketplace subsystem (colored in yellow), and

(3) the Resource Providers subsystem (colored in green).

Figure 2 presents these main elements, how they are grouped, and the way they interact with the tenants (colored in red) within the scope of NECOS.

The NECOS Slice Provider (LSDC) is the sub-system that allows for the creation of full end-to-end slices from a set of constituent slice parts. In NECOS, a slice looks the same as the full set of federated resources, but with domains looking a lot smaller. The Resource Marketplace provides the way for the NECOS Slice Providers (LSDC) to find the slice parts to build up a slice. Rather than having a pre-determined set of providers that have been configured in a federation, we chose to follow a more flexible model of a marketplace from which slice parts with specific characteristics can be found and provisioned. The Resource Providers are organizations that can provide the resources required for the slice parts – namely, DC resources in the form of servers, storage, and network resources in the form of connectivity and network functions. Further resources can be provided by organizations that have mobile edge, sensor networks, wireless access networks, etc. Each resource provider will be capable of providing slice parts of a full end-to-end slice.

3.3. NECOS Information Model

The NECOS information model provides the appropriate views of slices and resources for the Tenant, the Slice Service Provider, and the Infrastructure Provider needs. This information model enables the interaction between the different stakeholders in the NECOS ecosystem, and therefore, it is also a main achievement of the project. In addition, as these interactions happen through well-defined APIs, the project has devoted efforts in specifying a set of cloud API methods, which are either invoked by the Tenant during slice request, configuration, run-time management (i.e., client-to-cloud APIs), or by the NECOS system for slice provisioning, control, and resource orchestration (i.e., cloud-to-cloud APIs). In that context, proper methods have been designed to materialise these two types of APIs.

3.4. NECOS Demonstrators

The NECOS demonstrators and validation prototypes are based on the state of the art open software tools, which have been carefully selected. These tools have been used to design, integrate, and deploy the management and orchestration design choices, systems, and the APIs that have constituted the main research activity of the project. The NECOS approach has been validated in the context of implementations, using experimental environments like FIBRE, 5TONIC, and Emulab. Five different demonstrators were produced. The functionality of each one is described hereafter:

- The Multi-Slice/Tenant/Service (***MUSTS***) demonstrator is meant to exercise the following key features of NECOS: slice creation, slice decommission, slice monitoring, service deployment, service update, VIM heterogeneity, and elasticity upgrade (both vertical and horizontal). The KPIs obtained with this demonstrator are the average slice provisioning time, the average service provisioning time, CPU isolation, average elasticity response time, and monitoring data availability.
- The marketplace (***MARK***) demonstrator is meant to demonstrate the marketplace concept introduced in NECOS as a dynamic resource discovery mechanism that can cope with slices of significant size and multiple geographically distributed resource providers.
- The Experiments with Large-scale Lightweight Service Slices (***ELSA***) demonstrator is meant to show the deployment of end-to-end Slices that will be utilised by a Tenant in order to host services consisting of a vast number of lightweight elements (i.e., Virtual Network Functions (VNFs) and vLinks) deployed at the Edge of the infrastructure.
- The Machine-learning based orchestration of slices (***MLO***) demonstrator was created to show how machine learning algorithms can add value to slice orchestration.
- Finally, the Wireless Slicing Services (***WISE***) demonstrator shows the NECOS LSDC capabilities in expanding the cloud-network slinging concept towards wireless network domains.

The five above described demonstrators have been released by means of an open software license and constitute one of the instruments by means of which the NECOS will sustain its impact beyond the project lifetime.

4. Additional Research Tracks

Several research tracks have been conducted by NECOS project to design and implement supporting mechanisms and algorithms towards the realization of the “Slice-as-a-Service” concept, the foundation of the LSDC proposal. All they had ended in several contributions beyond state of the art.

WISE (WLAN Slicing Service) proposal [CARMO, 2018][2], comprising a heuristic based on an auxiliary graph that allows active/stand-by instances, request routings, and state update paths to be jointly considered to evaluate the availability of DC computing and accumulative bandwidth resources of DC's inbound links following a Software-Defined Networking (SDN) approach. The evaluation outcomes suggest that WISE algorithms significantly improve the request admission rate while reducing DCs' cost overtime. At the same time, they outperform existing solutions that separately consider placements, routing, and path updates.

[KIBALYA, 2018][3] has proposed a practical RAN slicing framework with a dynamic and intelligent two-level resource sharing algorithm capable of capturing dynamic properties of slice conditions based on a dynamic online resource sharing policy for sharing resources across multiple domains within heterogeneous network environments.

Our work on Unikernel-based CDN proposes a CDN platform, which places micro-content proxies near the users [VALSAMAS, 2018][4]. Proof-of-concept experimental results show UNIC most relevant features: modular orchestration of VM hosting replicas of Internet content, a Change-Point Detection (CPD)-based mechanism to detect content popularity changes, a DNS-based dynamic load balancing, and real-time monitoring of server resource utilization and end-user performance.

[YANG, 2018][5] has conceived an efficient solution to the fault-tolerant VNF placement problems, a heuristic algorithm that jointly computes the placement of active and standby instances of stateful VNFs. This solution maximizes the request admission rate and further evaluated outperforming the existing solutions that separately determine placements, routings, and state update paths, in terms of DC resource utilization, cost and runtime.

The probabilistic monitoring structure, based on bitmap and counter-array [MARTINS, 2018][6], allows to reduce the computational cost associated with the detailed collection of information in the network and can get traffic statistics requiring a fixed size memory with controlled accuracy.

A Blockchain-based decentralized application into Multi-Administrative Service Orchestration (MdOs) is presented in [ROSA, 2018a][7], along with its feasible opportunities in three use case scenarios, pursued by ongoing works at SDOs, the SD-WAN, NFVIaaS, and Network Slicing.

A novel DC Slice Controller design is proposed to support a slice-as-a-service (SlaaS) and on-demand VIM models, introducing the concept of transformable resources [FREITAS, 2018][8].

[TUSA, 2018][9] focuses on the impact on the monitoring process of new scenarios, where Cloud Network Service Providers take advantage of using more flexible resource management and orchestration solutions in the form of dynamic virtualised compute, network and storage resources.

---

[2] [CARMO, 2018] M. Carmo, S. Jardim, A. Neto, R. Aguiar, D. Corujo, J. Rodrigues. Slicing WiFi WLAN-Sharing Access Infrastructures to Enhance Ultra-Dense 5G Networking. IEEE ICC 2018, Kansas City, MO, USA, 20-24 May 2018.

[3] [KIBALYA, 2018] Godfrey Kibalya, Joan Serrat and Juan-Luis Gorricho. RAN Slicing Framework and Resource Allocation in Multi-Domain Heterogeneous Networks. IFIP AIMS 2018 conference, June 2018.

[4] [VALSAMAS, 2018] P. Valsamas, S. Skaperas and L. Mamatas. Elastic Content Distribution Based on Unikernels and Change Point Analysis. IEEE Wireless 2018, Catania, Italy, May 2018.

[5] [YANG, 2018] B. Yang, Z. Xu, W. K. Chai, W. Liang, D. Tuncer, A. Galis, and G. Pavlou. Algorithms for Fault-Tolerant Placement of Stateful Virtualized Network Functions. IEEE ICC 2018, Kansas City, MO, USA, 20-24 May 2018.

[6] [MARTINS, 2018] Regis Martins, Luis Garcia, Rodolfo Villaça and Fábio L. Verdi. Using Probabilistic Data Structures for Monitoring of Multi-tenant P4-based Networks. IEEE ISCC 2018 Conference, June 2018.

[7] [ROSA, 2018a] Raphael Vicente Rosa, Christian Esteve Rothenberg. "Blockchain-based Decentralized Applications for Multiple Administrative Domain Networking, In IEEE Communications Standards Magazine, Sep. 2018.

[8] [FREITAS, 2018] L. Freitas, V. Braga, S. L. Correa, L. Mamatas, C. Esteve Rothenberg, S. Clayman, and K. Cardoso. Slicing and Allocation of Transformable Resources for the Deployment of Multiple Virtualized Infrastructure Managers (VIMs). In Workshop on advances in slicing for softwarized infrastructures (S4SI 2018), June 2018

[9] [TUSA, 2018] Tusa, F; Clayman, S; Galis, A; Real-Time Management and Control of Monitoring Elements In Dynamic Cloud Network Systems. In: 2018 IEEE 7th International Conference on Cloud Networking (CloudNet). IEEE, 2018

[CLAYMAN, 2018][10] shows how creating a VIM (Virtual Infrastructure Manager) on-demand and dynamically allocating a new VIM for each slice, rather than having one for the whole DC, which can be beneficial for various precision scenarios.

[TUSA, 2019][11] presents the deployment of a hierarchical multi-MANO environment for the instantiation and management of end-to-end 5G network services in a slice-enabled infrastructure.

[MEDEIROS, 2019][12] introduces the elasticity in cLOud - neTwork Slices (SLOTS) that aims to extend the horizontal elasticity control to multi-providers scenarios in an end-to-end fashion, as well as to provide a novel vertical elasticity mechanism to deal with critical insufficiency of resources by harvesting underused resources on other slices.

5. Concluding Remarks

The NECOS approach features a set of innovative characteristics in the context of 5G and Beyond 5G. It allows for slicing to be performed either at the physical infrastructure level or at the VIM/WIM management level with relatively small changes in the software components of one alternative with respect to the other. The tenant's orchestrator interacts directly with VIM/WIM elements created on-demand, or with shim objects of the virtualized VIM/WIM at each local domain. From the tenant's point of view, these alternatives offer more control over the resources. From the service providers' point of view,these low-level slicing approaches can be considered lightweight because they do not require large slicing capable VIM/WIM or orchestrators in support of slicing and the providers can participate directly in a slice marketplace.

The slice may be built assembling parts of resources that belong to different administrative domains, as the NECOS architecture follows a multi-domain approach. To participate in the NECOS ecosystem, a given domain must implement the appropriate APIs to offer resources through a marketplace. The NECOS LSDC contacts the marketplace to decide, based on different criteria, which resources will constitute the slice parts. Once this is done, the LSDC takes control of the resources at each local domain through on-demand created VIM/WIM or shim objects. Such a process is one of the main distinguishing features of NECOS with respect to other slicing architectures, which require major software changes of large VIM/WIM platforms and/or peer-to-peer interaction between the (slicing-ready) resource orchestrators of participating domains.

The LSDC empowers a new service model – the Slice as a Service, by dynamically mapping service components to a slice. The enhanced management capabilities of the infrastructure create slices on-demand and slice management takes over the control of all the service components, virtualized network functions, and system programmability functions assigned to the slice, and (re)-configure them as appropriate to provide the end-to-end service.

The LSDC enables easy reconfiguration and adaptation of logical resources in a cloud networking infrastructure, to better accommodate the QoS demand of the Slice, through using software that can describe and manage various aspects that comprise the cloud environment. The LSDC allows each aspect of the cloud environment – from the networking between virtual machines to the SLAs of the hosted applications – to be managed via software. This approach reduces the complexity related to configuring and operating the infrastructure, which in turn eases the management of the cloud and network infrastructures.

---

[10] [CLAYMAN, 2018] Clayman, S; Tusa, F; Galis, A; Extending Slices into Data Centers: the VIM on-demand model. In: IEEE 9th International Conference on Network of the Future (NoF 2018). IEEE, 2018

[11] [TUSA, 2019] F. Tusa, S. Clayman, D. Valocchi, A. Galis. Hierarchical Service Providers on Sliced Infrastructure. Wiley Journal of Internet Letters. May, 2019

[12] [MEDEIROS, 2019] Medeiros, A, Neto, A, Sampaio, S, Pasquini, R, Baliosian, J. End-to-end elasticity control of cloud-network slices. Internet Technology Letters. May 2019

The LSDC platform offers the ability to a specific cloud provider to federate his own infrastructure with other cloud providers with different configurations in order to realize virtualized services using the Slice as a Service concept. The users of the LSDC APIs and platform will be able to create virtual services that can span the merged cloud infrastructure offered by different cloud providers. This concept is not purely technical, and it can also encompass business, cultural, and geographical among other domains.

6. Acronyms

API	Application Programming Interface
DC	Data Center
eMBB	enhanced Mobile BroadBand
GRE	Generic Routing Encapsulation
IMA	Infrastructure and Monitoring Abstraction
YAML	YAML Ain't Markup Language
KPI	Key Performance Indicator
LSDC	Lightweight Software Defined Cloud
MEC	Multi-Access Edge Computing
mMTC	massive Machine Type Communication
NFV	Network Functions Virtualization
QoS	Quality of Service
OTT	Over The Top
PoC	Proof of Concept
SDN	Software Defined Networking
SLA	Service Level Agreement
SLO	Service Level Objective
SRO	Slice Resource Orchestrator
URLLC	Ultra Reliable and Low Latency Communication
VNFM	Virtual Network Function Manager
VM	Virtual Machine
VNF	Virtual Network Function
VIM	Virtual Infrastructure Manager
WAN	Wide-Area Network
WIM	Wide-area network Infrastructure Manager