Network Services SLAs over 5G Infrastructure Converging Disaggregated Network and Compute Resources

I. Mesogiti, E. Theodoropoulou, K. Filis, G. Lyberopoulos,
COSMOTE Mobile Telecommunications S.A.
Athens, Greece
{imesogiti, etheodorop, cfilis, glimperop} @cosmote.gr

Anna Tzanakaki
University of Bristol
Bristol, UK
anna.tzanakaki@bristol.ac.uk

R. Cantó Palancar, N.A. Serrano Linares
Telefonica,
Barcelona, Spain
{rafael.cantopalancar, nicolasangel.serranolinares} @telefonica.com

Daniel Camps-Mur
i2CAT Foundation
Barcelona, Spain
daniel.camps@i2cat.net

Jesús Gutiérrez
IHP Microelectronics
Frankfurt Oder, Germany
teran@ihp-microelectronics.com

Abstract— **5G networks will comprise multiple network and compute infrastructure setups at finest fragmentation consisting of multiple disaggregated pools of network, compute and storage resources. Thus, 5G shall provide the means to move from the currently strictly defined, simple, single service models of provisioning connectivity or cloud or software services separately, to more complex "service provisioning" models including a mix of infrastructure resources, network services and even applications, specifically tailored to the specific stakeholder/vertical/end-user requirements. In this context, new SLAs shall be defined along with the SLA monitoring and maintenance mechanisms to be deployed over the distributed/disaggregated 5G deployments resources. This paper provides insights on the 5G network services SLAs definition along with an SLA maintenance paradigm.**

Keywords—5G; SLAs; Transport Network services;

I. INTRODUCTION

In legacy network technologies (including 4G), Quality of Service (QoS) provisioning has been rather abstract and considered in an application/service agnostic manner, while on the contrary, in the expected 5G systems, the QoS provisioning shall be based on a more stakeholder/ application/service aware approach. Practically this means that besides the generic technical QoS Key Performance Indicators (KPIs) and target values to be achieved by the 5G technologies, the actual 5G network deployments and operation shall be tailored (automatically) to support the requirements of a range of stakeholders and services in a holistic manner. For this purpose, the development of 5G technologies is tightly coupled with the analysis of stakeholders' and their services' requirements so that the latter can be mapped with the 5G network capabilities, functionalities, deployment strategies, etc. The ultimate goal worldwide is to deliver "a stakeholder driven, holistic ecosystem for technical and business innovation integrating networking, computing and storage resources into one programmable and unified infrastructure" [2], "enabling the transport of software to the data rather than the other way round, i.e. executing software on the device where the data is produced instead of sending all data to a centralised data centre" [2], while satisfying versatile, on a per stakeholder and service basis defined Service Level Agreements (SLAs).

This paper provides insights and suggestions on the definition of new SLAs concerning the provisioning of various future 5G network services to various stakeholders over converged Fronthaul (FH) and Backhaul (BH) network architectures, and further over 5G networks comprising distributed, even disaggregated network and compute resources. It also provides a deployment paradigm of functionalities to support monitoring and maintenance of these SLAs.

The paper is organized as follows: at first the 5G network principles are summarized along with the technology flexibility in terms of technologies to be deployed and infrastructure owners variety. Next, the envisioned 5G network services provisioning is identified and the corresponding SLAs (Service Level Agreements) are elaborated. The maintenance of these SLAs is then discussed on the basis of the 5G-XHaul architecture [14]. 5G-XHaul [1] is a European project working on the definition of converged Fronthaul (FH) and Backhaul (BH) networks for future 5G mobile networks, defining a logical transport architecture that integrates various wireless and optical technologies under a common SDN control plane [8].

II. 5G NETWORKS PRINCIPLES OVERVIEW

To date, the standardization of the 5G network architectures has not been finalized. However, we can identify a set of principles underpinning the 5G network architectures which are common in the currently proposed options by various SDOs (Standards Developing Organizations) [4], 5G-PPP projects ([1],[5],[6]), as well as vendors' and operators' roadmaps. In general terms, 5G networks will comprise multiple network and compute infrastructure setups [7]; at finest fragmentation comprising multiple disaggregated pools of network, compute and storage resources as proposed by the 5G-PICTURE project [5], which will:

- employ different wireless and optical network technologies to provide connectivity and incorporate hardware programmability, software definition at various layers, resources virtualization and new hardware implementations;
- make use of spatially distributed (local, in-network and/or remote) compute and storage resources which are often denoted as “cloud” or “edge” datacenters – and which can be versatile in terms of technologies and management platforms, and
- be orchestrated so as to provide a service abstraction layer over multiple domains,

towards delivering a complete environment to satisfy the IMT-2020 service and system KPIs and support existing and future (5G/5G ready) applications.

The technical flexibility of the 5G environment implies that the various wireless, optical network, and cloud infrastructures may be owned by different stakeholders, whilst being orchestrated by a single layer to deliver the requested services with the guaranteed QoS. For this purpose, various SLAs will be needed, underpinning the business relationships mainly between different infrastructure providers, telecom operators and verticals [8].

III. 5G Network Services Provisioning

Moving from the currently defined service models of provisioning connectivity services or cloud services or software services separately, in the 5G ecosystem, "service provisioning" can include a mix of infrastructure resources, network services and even application/services, specifically tailored to the requirements of the stakeholder/vertical/end-user to be served.

Considering a telecom operator responsible for a 5G infrastructure deployment and providing 5G services to tenants, namely corporate users, such as Virtual Network Operators (VNOs) - Network as a Service (NaaS) providers using the 3GPP terminology - and verticals, we can identify the provisioning of the following three (3) major 5G network service categories [8]:

- TYPE A: Transport Network Services
- TYPE B: Transport Network and Cloud Services
- TYPE C: End-to-End Telecommunication Services and Cloud Services.

IV. SLAs for 5G Network Services Provisioning

Considering the expected 5G network services, new SLAs/SLA templates for 5G-network services shall be generated leveraging on existing connectivity and cloud services SLAs provided by network operators. In these new SLAs, the following 5G infrastructure aspects/capabilities should be reflected:

- The versatility of provided services; being one or a combination of the following:
    - Backhaul (BH) network connectivity services, of various types.
    - Fronthaul (FH) network connectivity services, of various types (various Functional Splits).
    - Access network connectivity services.
    - Cloud services, e.g. for Cloud-RAN deployments (hosting of vBBUs (virtual Baseband Units)), for hosting core network elements or even for application services.
    - Auxiliary services such as synchronization.
- The dynamic, spatio-temporal provisioning of resources to tenants/users.
- The support of dynamically (time) changing QoS for each service.
- The capability of scaling resources based on specific triggering events (e.g. performance measurements, monitored events, external requests).
- The evaluation of services provisioning on the basis of a number of KPIs such as:
    - Availability.
    - Reliability.
    - Performance: in terms of guaranteed and acceptable values of a wide range of monitored parameters such as: Bit Rates, Latency between two end points, Jitter, Packet Loss, Cloud Resources, such as storage space, vCPUs, RAM, etc.

To this end three (3) SLA categories have been identified [8] corresponding to the aforementioned 5G service provisioning categories.

A. *TYPE A SLA: Transport Network Services*

The users of this type of services could be NaaS providers, namely MNOs (Mobile Network Operators) or other tenants aiming at providing their own customers with telecommunication services. Many market players can be considered as tenants, including the commonly identified verticals (in the 5G ecosystem [9], [10], such as: major infrastructure owners (stadium/ airport/shopping malls, etc.), transportation industry, etc. These MNOs/tenants are considered to own an access and core network infrastructure while lacking a transport network. For instance:

- A MNO/tenant owning a small cell infrastructure and a remote core network may ask for BH services, of specific data rates, latency, jitter, packet loss, etc. depending on the services to be offered.
- A MNO/tenant owning a Remote Radio Head (RRH) infrastructure to be connected to a pool of remote BBUs (virtual or physical), and to a core network, may ask for FH services for the RRH-BBUs links and for BH/backbone (BB) services from that point to the core network end.

The various parameters and the values of the SLAs will correspond to the specificities of the MNOs'/tenants' infrastructure and services per case. The information to be included in a SLA template of this type is presented in Table I.

TABLE I. SLA INFORMATION FOR 5G BACKHAUL/FRONTHAUL SERVICES ([8])

Backhaul / Fronthaul Service	
SLA Parameters	***Definition***
Time Period	The actual time over which the service is provided, incl. the start and end date (e.g. 1/1/2020 - 30/4/2020).
Periodicity	The periodicity of the service offering; e.g. "Continuous", "Every weekend", 8:00am – 17:00pm, etc.
Location of the Access Network Nodes (ANNs) to be connected	The location of the ANNs to be connected (e.g. at a specific stadium area/ municipality/ building block(s)/campus/ hotspot (shopping mall, park, etc.), nationwide, etc.
Technical details & QoS	The transport service technical details or restrictions in terms of e.g. functional split, protocol ([11],[12]), wireless/optical technology, as well as the required QoS, e.g. including Data Rates, Latency, Jitter, Packet Loss.
Availability	The time within a specific time-period in which the service is up and running, or inversely the max. time within a specific time-period in which the service is unavailable.
Reliability	Frequency (number of times) of non-availability of the service per specific time period.
Monthly Availability	The percentage (%) of time within a specific time-period in which the service is up and running.
Scaling Rules	The set of pre-defined rules based on which the service can be scaled up/down.
Operational Rules	The set of pre-defined rules based on which specific operations are triggered, e.g. events' warning, monitoring initiation, etc.

B. *TYPE B SLA: Transport Network and Cloud Services*

The customers of this type of services mix could be MNOs or other tenants -as in TYPE B case- owning an access network while lacking a transport network and cloud infrastructure, and aiming at providing their own customers with telecommunication services. For instance:

- A MNO/tenant owning a Small Cell infrastructure may ask for backhaul services of specific data rates, latency, jitter, packet etc., depending on the services to be offered, and for cloud services to host the virtual instances of its core network and/or services/applications.
- A MNO/tenant owning a RRH infrastructure to be connected to a pool of remote vBBUs (virtual BBUs) ([12]), and to a core network, may ask for FH services for the RRH-BBUs links and for BH/BB services towards the core network end, as well as for cloud services to host the vBBUs and/or the virtual instances of their core network elements and possibly services/applications. In case of an application comprising more than one distributed elements, it is necessary to define service function chaining with SLAs.

The various parameters and the values of the SLAs will correspond to the specificities of the MNOs'/tenants' infrastructure and services per case. The information to be included in a SLA template of this type is presented in Table II.

TABLE II. SLA INFORMATION FOR 5G TRANSPORT NETWORK AND CLOUD SERVICES ([8])

Transport Network and Cloud Services SLA Parameters	
Backhaul Service OR Fronthaul Service	
SLA Parameters	***Definition***
As in Type A SLA for Transport Network Services (see TABLE I.)	
Cloud Services (for virtual instances of Telecom Network Components)	
SLA Parameters	***Definition***
Time Period	The actual time over which the service is provided, incl. the start and end date (e.g. 1/1/2020 - 30/4/2020)
Periodicity	The periodicity of the service offering; e.g. "Continuous", "Every weekend", 8:00am – 17:00pm, etc.
Location	The location of the ANNs to be connected (where the end-users are expected to reside) (e.g. at a specific stadium area/ municipality/ building block(s)/campus/ hotspot (shopping mall, park, etc.), Nationwide, etc.
Cloud Service Resources	The Cloud Service Resources in terms of vCPUs, Memory-RAM, Storage Space, etc.
Availability	The time within a specific time-period in which the service is up and running, or inversely the max. time within a specific time-period in which the service is unavailable.
Reliability	Frequency (number of times) of non-availability of the service per specific time period.
Monthly Availability	The percentage (%) of time within a specific time-period in which the service is up and running.
Scaling Rules	The set of pre-defined rules based on which the service can be scaled up.
Operational Rules	The set of pre-defined rules based on which specific operations are triggered, e.g. events' warning, monitoring initiation, etc.

C. *TYPE C SLA: End-to-End Connectivity and Cloud Services*

The customers of this type of services mix could be MVNOs or other tenants (as in the previous types) including all vertical categories ([9],[10]) such as factories of the future (Industry 4.0 operators), eHealth and emergency services providers, energy/ security/ surveillance services providers (focusing on mIoT (massive Internet of Things) services), media and Entertainment (multimedia content providers/broadcasters provisioning UHD (Ultra High Definition) content, Crowdsourced/ multi-user created content, highly interactive services etc.), leasing end-to-end telecommunication services and possibly cloud services for hosting their applications.

The performance parameters of this type of SLAs will be very similar to existing (legacy) network connectivity and cloud services SLAs. However, in case of applications comprising more than one distributed elements, the service function chaining shall be reflected in the SLAs. Moreover, taking advantage of the 5G network flexibility in resources provisioning, SLAs of this type can be dynamic in time and in terms of provided resources and QoS levels (as for SLAs of type A and B).

V. Support of SLAs by 5G Infrastructure Management Systems

The service management processes are responsible to guarantee the fulfillment of the agreed SLAs. In the legacy networks, the static –or actually of little flexibility- allocation of transport network resources to access network nodes determines the service performance, especially in terms of latency; thus the SLA guarantees. On the contrary, in 5G networks the allocation of transport network resources will be determined by the SLAs that shall be satisfied. Currently, performance monitoring is performed as a per network node health-check process, while in future 5G networks it will be carried out, additionally, on a per slice basis. For this purpose, tight coordination between service management and the multiple 5G infrastructure controllers is necessary for the definition and handling of monitoring alarms and the maintenance of the SLAs.

Considering the 5G-XHaul control plane (CP) architecture [13] (Fig. 1), as an indicative 5G transport network architecture (on which 5G-PICTURE leverages), it is necessary to determine the entities in charge of the SLA maintenance.

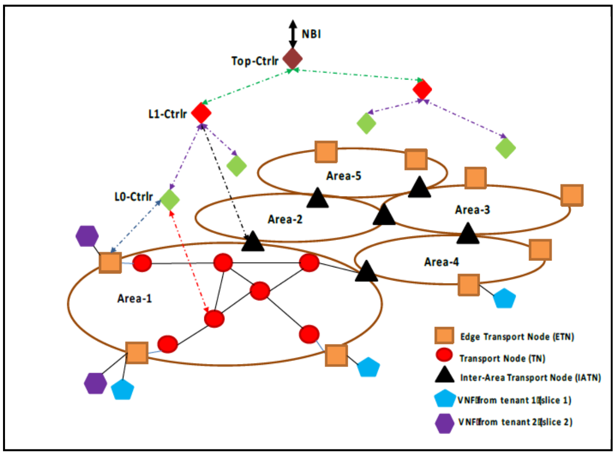

Fig. 1. 5G-XHaul: Control Plane Architecture [8]

The hierarchical control plane architecture proposed by 5G-XHaul consists of:

- Local Agents (LA), embedded in each transport node (i.e. Transport Node (TN), Edge Transport Node (ETN) and Inter Area Transport Node (IATN)).
- Level-0 controllers (L0-Ctrlr), ruling the control plane functionalities of a single local area. Level 0 control plane are technology specific, e.g. you may have a wireless L0-Ctrlr and an optical L0-Ctrlr.
- Level-1 controllers (L1-Ctrlr), in charge of maintaining connectivity between several ACs. L1-Ctrlr are technology agnostic.
- Top controllers (Top-Ctrlr), responsible for provisioning per-tenant slices and orchestrating connectivity across different 5G-XHaul areas.

With regard to the SLA monitoring and maintenance, LAs need to collect statistics and events related to their own utilization and health status, corresponding at least to the SLAs' KPIs on a per node and per configured slice basis. This information, besides being sent to the 5G deployment main orchestrator, shall be further delivered to Level-0 controllers, through the various south-bound interfaces developed in 5G-XHaul for the mmWave and Sub6 wireless technologies, and optical network nodes; to be used for SLA monitoring and maintenance purposes (Fig. 2).

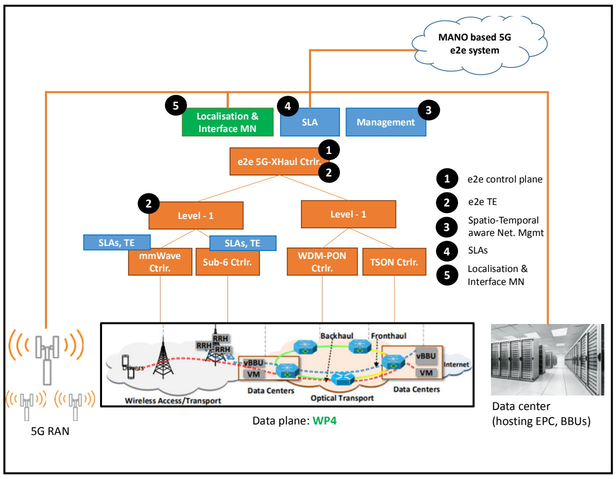

Fig. 2. 5G-XHaul: SLA Monitoring Functionalities [8]

Level-0 and Level-1 controllers need to have functions for setting up and adapting paths to network dynamics based on the SLA guaranteed performance per slice, using also the information delivered by LAs.

Top-controllers need to provide an interface for receiving the application plane requests including the SLA parameters and to forward these to the underlying controllers. These nodes will interface also with the main orchestrator of the deployment.

Leveraging on this concept, the SLA monitoring and maintenance of the disaggregated resources of the 5G-PICTURE architecture, implies that LAs shall also be incorporated at the storage/processing domains (i.e. cloud infrastructures) along with those controllers that handle the lifecycle of resources' instantiation at these domains and respective SLA guarantees shall be met.

VI. Conclusions

In this paper, the future 5G network principles have been summarized, along with their capability to support complex "service provisioning" models including a mix of infrastructure resources, network services and even applications, specifically tailored to the stakeholder/ vertical/ end-user requirements; thus moving from the currently provided, simple, single service models of provisioning connectivity or cloud or software services separately.

In this context, SLAs need to be adjusted to address the versatility of provided 5G network services/ technologies/ performance as well as to reflect the dynamic, spatio-temporal availability and QoS fluctuations related to the provisioning of these services.

With this respect, such new SLA templates to be provided for 5G network services provisioning have been identified and proposed.

Finally, an indicative breakdown of SLA monitoring and maintenance functionalities to be deployed on various 5G network nodes is proposed on the basis of the 5G-XHAUL control plane architecture, along with insights of how this can be leveraged in the context of disaggregated network and compute resources of the 5G-PICTURE infrastructure.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union's Framework Programme Horizon 2020 under grant agreements (1) No. 671551 and project name "5G-XHaul: Dynamically Reconfigurable Optical-Wireless Backhaul/Fronthaul with Cognitive Control Plane for Small Cells and Cloud-RANs" and (2) Nº 762057 and project name "5G-PICTURE: "5G Programmable Infrastructure Converging disaggregated neTwork and compUte REsources".

REFERENCES

[1] 5G-XHaul Project, http://www.5g-xhaul-project.eu/

[2] 3GPP TS 22.261 V16.1.0 (2017-09) 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Service requirements for the 5G system; Stage 1 (R16).

[3] 3GPP, Technical Report, TS 28.530, "Management of 5G networks and network slicing; Concepts, use cases and requirements", Release 15.

[4] 3GPP, Technical Report, TS 23.501, "System Architecture for the 5G System; Stage 2", Release 15.

[5] 5G-PPP Project 5G-PICTURE, http://www.5g-picture-project.eu/index.html

[6] 5G-PPP Project MATILDA, http://www.matilda-5g.eu/

[7] 3GPP, Technical Report, TR 22.891, "Feasibility study on new services and markets technology enablers; Stage 1", R14.

[8] 5G-XHAUL, Deliverable D3.3, "D3.3 5G-XHaul algorithms and services Design and Evaluation", June 2018.

[9] 5G-PPP Program, European Commission, "5G empowering vertical industries", white paper, 2016/ URL: http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=14322 (visited on 5/12/2017).

[10] 5G-PPP Program, European Commission, Living document on "5G-PPP use cases and performance evaluation models".

[11] 5G-XHaul Deliverable D2.1 Requirements Specification and KPIs Document, November 2015.

[12] 5G-PICTURE, D2.2 System architecture and preliminary evaluations, May 2018.

[13] Mur, D. C., Flegkas, P., Syrivelis, D., Wei, Q., & Gutiérrez, J. (2016, December). 5G-XHaul: Enabling scalable virtualization for future 5g transport networks. IEEE, Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security (IUCC-CSS), International Conference on (pp. 173-180).

[14] I. Demirkol, D. Camps-Mur, J. Bartelt, J. Zou 5G Transport Network Blueprint and Dimensioning for a Dense Urban Scenario European Conference on Networks and Communications (EuCNC), Oulu, 2017.