5G & SLAs Automated proposition and management of agreements towards QoS enforcement

Conference Paper · February 2018
DOI: 10.1109/ICIN.2018.8401587

4 authors:

Evgenia Kapassa
University of Piraeus

Marios Touloupou
University of Piraeus

Argyro Mavrogiorgou
University of Piraeus

Dimosthenis Kyriazis
University of Piraeus

5G & SLAs: Automated proposition and management of agreements towards QoS enforcement

Evgenia Kapassa, Marios Touloupou, Argyro Mavrogiorgou, Dimosthenis Kyriazis
Department of Digital Systems
University of Piraeus
Piraeus, Greece
{ekapassa, mtouloup,margy, dimos}@unipi.gr

***Abstract*—Efficient Service Level Agreements (SLA) management and anticipation of Service Level Objectives (SLO) breaches become mandatory to guarantee the required service quality in software-defined and 5G networks. To create an operational Network Service, it is highly envisaged to associate it with their network-related parameters that reflect the corresponding quality levels. These are included in policies but while SLAs target usually business users, there is a challenge for mechanisms that bridge this abstraction gap. In this paper, a generic black box approach is used to map high-level requirements expressed by users in SLAs to low-level network parameters included in policies, enabling Quality of Service (QoS) enforcement by triggering the required policies and manage the infrastructure accordingly. In addition, a mechanism for determining the importance of different QoS parameters is presented, mainly used for "relevant" QoS metrics recommendation in the SLA templates.**

Keywords: *service level agreement, artificial neural networks, network workload prediction, network parameters forecasting, SLA translation*

I. INTRODUCTION

In recent years, Software Defined and 5G Networks - a combination of Software-Defined Networking (SDN) and Network Function Virtualization (NFV) - is a significant area of research and innovation, that tend to make cloud and network services more agile. Major network and service providers predict that, by 2020, 70% of deployed networks will rely on cloud infrastructures, virtual network functions, as well as multi-domain SDN controllers [1,2]. These technologies are essential to support many aspects of the anticipated functionality offered by the 5G networks. The current vision and research challenges in 5G go beyond the focus on the underlying infrastructure, towards the stakeholders of the ecosystem, including SLA management [3]. The management of such a virtualized network focuses on maintaining acceptable quality. Therefore, Quality of Service (QoS) expectations are driving end-users to negotiate specific QoS levels with their service providers. This is increasingly done through service level agreements (SLAs). Considering that the quality requirements submitted by the end-user are stated into the SLA, the thing that becomes essential for the service providers prior to signing the specific SLA, refers to an estimation of the resources needed to fulfill the user requirements. Service providers need to determine the resources to fulfill the QoS requirements of the service, while at the same time, resource utilization must be maximized [4].

One of the most significant part, is the role of the Service Platform, in mapping the end-user's defined service requirements (i.e. workload parameters), and policies expressions [5], to the resource level attributes. However, the exchange of information between the different entities that exist is extremely difficult due to technical and business reasons. On the one hand, each entity focuses on a particular set of parameters, and on the other hand exchange of such information is not always technically obtainable.

From the service provider's perspective, the benefits of SDN include cutting down the time and cost to deploy a new service. From a customer's perspective, they get an on-demand virtual environment and the IT services they need in an instant [6].

Considering the aforementioned problem, in this paper, we present an approach for mapping the high-level end-user requirements to the low-level policy parameters, as briefly described in [7]. What is more, we propose a mechanism for suggesting the most important QoS parameters to the Service/Infrastructure Provider, in order to achieve better QoS assurance as part of the 5G Development and Validation Platform for global Industry-specific Network Services and Apps [8]. Data Analytics allow automated identification of relationships and dependencies between different parameters of the datasets and thus such approaches are exploited in the context of SLA mapping given the need to correlate input parameters (i.e. SLA high level requirements) with output parameters (i.e. policy-related low-level network attributes). To that end, we considered an Artificial Neural Network (ANN) [9] as a black box approach.

The remaining of the paper is structured as follows. In Section II, similar approaches in the related field are presented, while in Section III a preliminary reference is made on the main elements of a SLA. Section IV describes the overall architecture of the 5GTANGO platform, while the core of the proposed approach appears is Section V. Finally, in Section VI, the paper concludes with some thoughts for future research and potentials for the current study.

II. Related Work

Active and advanced work has been conducted on SLA management for Cloud infrastructure and Grid Computing. A framework for incorporating QoS in Grid applications is discussed in [10]. In this paper, a performance model to estimate the response time and a pricing model for determining the price of a job execution is used. With the recent work of SLA in cloud computing we have witnessed the emergence of a new category which is the mapping of SLA between Service Provider, Infrastructure Provider and end-users [11]. A very interesting approach, similar to our work, is [12]. In this case, the authors study the performance interference of combinations of elementary applications when running inside collocated Virtual Machines (VMs). Another interesting work is presented in [13], where the importance of applying ANNs in the service-oriented field is demonstrated. In this approach, the main goal of the ANN is to set up technical targets of design attributes of web service systems, in terms of quality goals. Moreover, in [14], the LoM2HiS framework is presented. In this case, the authors provide a framework that implements the reverse process of the one that we suggest, where the translation of low level metrics to high-level terms that are used in Cloud Service Level Agreements is achieved. Furthermore, more details regarding the requirements mapping can be found in [15-16] A generic black box approach, based on ANNs is used in order to perform the aforementioned translation between low and high-level requirements.

The approach presented in this paper aims at providing a high to low level requirements mapping in a virtualized environment (i.e. a software defined network). For this reason, we propose an ANN, optimized with a Genetic Algorithm (GA) as a method for combining high level parameters with low, while at the same time decomposing service level objectives to associated policies. What is more, we present a mechanism for determining the importance of different parameters and include the "relevant" ones at the SLA for better QoS assurance at a-priori level. The overall architecture of the proposed framework is presented in Fig, 1.

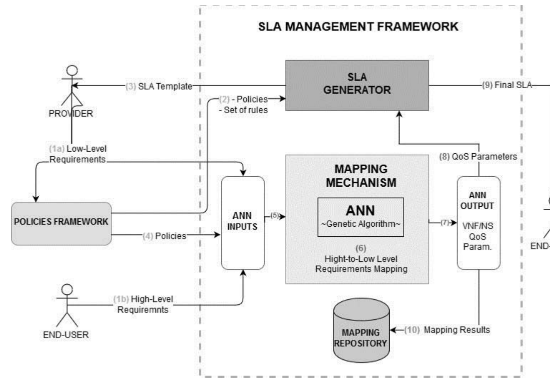

Fig. 1 SLA Management Framework overall process

III. Structure of a Service Level Aggreement

A. Definition

An SLA is a contract between a service provider and its internal or external end-users that documents what services the provider will accord, and defines the performance standards that are obligated to be met by the provider [17,18].

Over the years, SLAs set the expectations for a service provider performance, and established penalties for missing the targets and, in some cases, bonuses for exceeding them [20]. In that concept, SDN evolution could not leave unaffected the evolution of SLA models, and their flexibility in adapting more demanding and adapting parameters.

IV. 5G Tango Architecture Overview

5GTANGO is a 5GPPP Phase2 Innovation Action [21] that enables the flexible programmability of 5G networks with [8]: a) an NFV-enabled Service Development Kit (SDK),b) a Store platform with advanced validation and verification mechanisms for VNFs/Network Services qualification (including 3rd party contributions), c) a modular Service Platform with an innovative orchestrator in order to bridge the gap between the business needs and the network operational management systems.

The proposed SLA Management Framework is part of the Service Platform and interacts with the Service/Infrastructure Provider, the end-user (that may be developer of functions and services or/and the end-user), and the policies framework.

V. Proposed Mechanisms

A. SLA Template Generator

In a scenario like 5GTANGO a vital part of the SLA Management Framework is the SLA Generator that firstly creates the SLA templates for the Service/Infrastructure Provider, and secondly creates the final SLA itself. This Generator will be able to obtain a set of policies for the specific NS, in order to create an SLA template in an automated way. In more detail, SLA Management Framework should have access both to the NS-Descriptors and its associated set of policies, in order to be able to analyze them and generate in a dynamic way a structured and at the same time generic SLA template. The generated template should be in human readable format and immediately available to the Service/Infrastructure Provider.

Another critical issue, is the negotiation process. SLA negotiation is an important mechanism that guarantees the service performance and enhances the trust between service end-user and service providers. Through negotiation providers can customize the SLA before signing. To this end, it should be mentioned that customization of a SLA refers to the modification of the SLA template which is defined and offered by the service provider, as an indication of the acceptable guarantees that may be included in the contract content. This process is necessary but sometimes may also be time consuming. For the negotiation process to become more efficient we suggest the implementation of a mechanism that determines the importance of different QoS parameters by correlating and analyzing the: a) Predefined SLA template parameters, b) Associated policies parameters and c) Provider's negotiation historic data.

Thus, this mechanism will be able to compile importance weight factors and identify dependencies between the parameters in order to recommend and include in the SLA templates the "relevant" QoS parameters for each different case [22].

B. Mapping Mechanism

As mentioned above, the SLA Management Framework that presented in this paper is part of the 5GTANGO platform. In order for NS to be deployed in the service provider's infrastructure, several steps must be implemented with regards to the performance estimation and the required quality, that should be agreed and signed in the SLA. The roles that are foreseen in this process include the service provider, the infrastructure provider – in our scenario is the same entity, and the end-user as shown in Fig. 2.

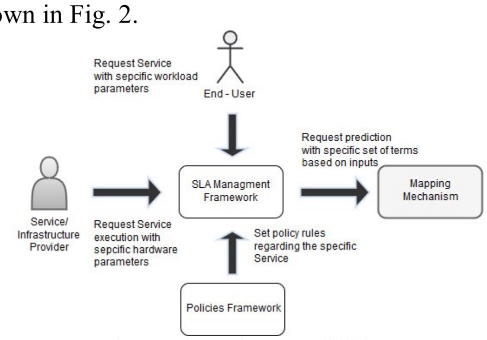

Fig. 2 Roles and Responsibilities

Apart from these roles, an important component for the SLA Management Framework is the Policies Framework. To create an operational NS service, it is highly envisaged to associate it with a set of deployment and operations-specific policies, supporting a set of orchestration aspects. The role of the policy rules has a huge part to the mapping mechanism as they give specific low-level requirements regarding the NS/VNF. The benefits for the involved parties are [23]: a) The Service Provider does not have to reveal sensitive information regarding the internal structure of the NS, b) The Infrastructure Provider may utilize the mapping mechanism to create black box models for the components in an automated way, for driving the allocation of resources for a specific service instance requested by the end-user.

The mechanism is based on unsupervised learning, using an Artificial Neural Network (ANN) [24]. ANNs, since they represent a black box approach, are perfect for usage in an environment where information is not easily relayed from one entity to the other, mainly for confidentiality purposes. Also, they do not need any knowledge regarding the internal structure of the services. They only need inputs and outputs of the model, which are available as they are the prior terms that are used between the roles described above [25]. In addition, they are perfect candidates for mapping, as the influence of high-level network service requirements (i.e. workload parameters from the end-user), and the changes in low-level requirements are reflecting at the changes in the ANN output. By having such a correlating function, it becomes very easy to dynamically map the one parameter to the other as shown in Fig 3.

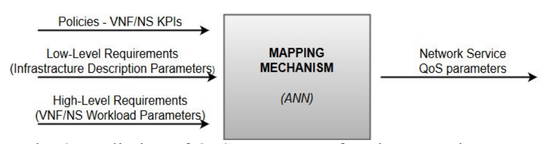

Fig. 3 Prediction of QoS parameters for given requirements

The inputs and outputs of the ANN are determined through proper information provided by the Policy Framework, the Service/Infrastructure Provider and the End-User. In more details, each one of them is responsible for the following inputs (Fig. 2).:

- Policy Framework defines a set of policy rules associated to the VNF/NS.
- Service/Infrastructure Provider defines a set of low-level requirements – Hardware description parameters.
- End-User defines a set of high-level requirements –VNF/NS workload parameters.

As a result, the mapping between the high-level requirements described by the end-user and the low-level requirements described by the provider (also taken from the policy rules) takes place. To this end, the result of the mapping outputs the QoS parameters that are needed to be expressed in the SLA. This information is available to the SLA Generator, being the basis of the final SLA, through the process described at subsection A, of section V.

The ANN will be optimized by a GA [26] which should select the fittest parameters of each network, after an extensive process, to adapt it to each NS component's needs (i.e. more suitable transfer functions, number of neurons per layer, total hidden layers etc.) [27]. A potential ANN model is pictured in Fig. 4.

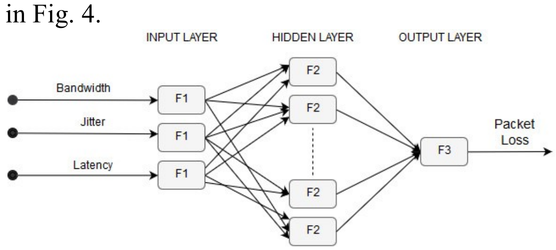

Fig. 4 Example structure of potential ANN model

C. SLA Template Compilation

At the current subsection, a combination of all the described mechanisms take place, in order to present step-by-step the creation of a SLA. Following the 5GTANGO approach it is assumed that district VNFs/NS are accessible in a Catalogue. The Catalogue provides optimal selection of VNFs/NS by the service/infrastructure provider, based on her requirements and constraints. The sequence diagram for the compilation appears in Fig. 5.

With that in mind, the following steps are followed:

1. The Service/Infrastructure Provider selects VNF/NS from 5GTANGO Catalogue in order to be deployed.
2. The requirements needed as an input for the Mapping ANN are defined: a) Service/Infrastructure Provider defines a set of low-level requirements – Hardware description parameters, b)End-User defines a set of high-level requirements – VNF/NS workload parameters.
3. The Policy Framework creates policy rules using the low-level requirements as described from the Service/Infrastructure Provider.
4. The SLA Templates are generated through the SLA Generator, being accessible to the provider.
5. The ANN gets the appropriate dataset that consists of the high and low-level requirements as well as the policies.

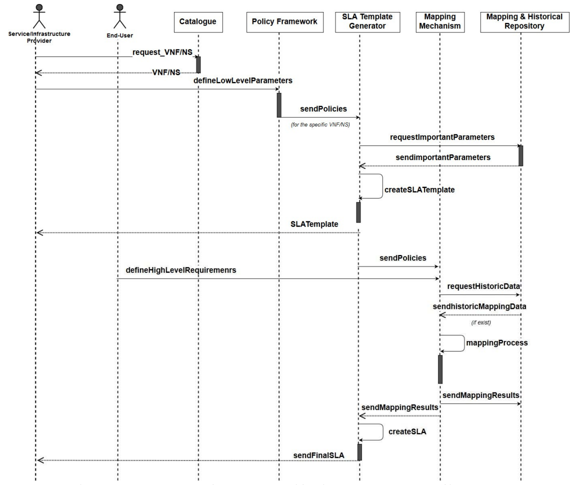

Fig. 5 SLA template compilation sequence diagram

6. The ANN correlates the inputs so as to predict the QoS KPIs (i.e. output).
7. The SLA Generator gets the produced QoS parameters and creates the final SLA.

To this end, it should be noted that the mapping results are due to be stored into a repository. The goal is the mapping results to be available for the SLA framework for a future selection of the appropriate SLA terms without unnecessary processing. The overall procedure is presented in Fig. 1.

VI. CONCLUSIONS AND FUTURE WORK

SDN brings several important benefits to the enterprise datacenter. One such benefit is the advancement and optimization of the traditional Service Level Agreement (SLA). An effective SLA is the key to ensure that a service/infrastructure provider delivers the agreed terms of services to the SDN end-user. In SDNs, end-users with clear definition of SLA parameters and flexible negotiation methods can increase the reliability and trust level of network provider – end-user relationship.

In this paper, we presented an SLA Management Framework, which is used for mapping high-level workload parameters to low-level resource attributes. This framework is embedded into 5GTANGO service platform for achieving autonomic SLA management. We considered a generic approach that is based on ANNs to efficiently be used as a mediator for the network provider and the end-user. In addition, we presented a mechanism to determine the importance of different QoS parameters so as to include the "relevant" ones at the SLAs, for better QoS assurance.

The next step in our work is to complete the design and implementation of all the components of the proposed SLA Management Framework, while we aim to streamline and enhance the interfaces between the different components and the entities. What is more, we envision the need for a deeper investigation in QoS enforcement using the autonomous SLA mechanisms described in Section V, as well as the real-time adoption based on monitoring feedback and the Quality of Experience parameters.

ACKNOWLEDGMENT

This work has been partially supported by the 5GTANGO project, funded by the European Commission under Grant number H2020ICT-2016-2 761493 through the Horizon 2020 and 5G-PPP programs (http://5gtango.eu).

REFERENCES

[1] LT. Park, JW Baek, JWK Hong, "Management of Service Level Agreements for Multimedia Internet Service Using a Utility Model", IEEE Communications Magazine vol.: 39 May 2001

[2] VC Emeakaroha, I. Brandic, M. Maurer, S. Dustdar , "Low Level Metrics to High Level SLAs - LoM2HiS Framework: Bridging the Gap Between Monitored Metrics and SLA Parameters in Cloud Environments" , International Conference on High Performance Computing and Simulation (HPCS) 2010

[3] P. Gouvas, E. Fotopoulou, A. Zafeiropoulos, C. Vassilakis, "A Context Model and Policies Management Framework for Reconfigurable-bydesign Distributed Applications" , Procedia Computer Science 2016

[4] G. Kousiouris, D. Kyriazis, K. Konstanteli, S. Gogouvitis, G. Katsaros, T. Varvarigou, "A Service-Oriented Framework for GNU Octave-Based Performance Prediction" , IEEE International Conference on Services Computing 2010

[5] A. Evangelidis, D. Parker, R. Bahsoon, "Performance Modelling and Verification of Cloud-based Auto-Scaling Policies", 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2017

[6] H. Basilier, L. Frid, G. Hall, G. Nilsson, D. Roeland, G. Rune, M. Stuempert, "A Vision of the 5G Core: Flexibility for new business opportunities", Charting the future of innovation vol.93 2016

[7] M. Zhao, FL. Gall, P. Cousin, R. Vilalta, R. Munoz, S. Castro, M. Peuster, S. Schneider, M. Siapera., E. Kapassa, D. Kyriazis, P. Hasselmeyer, G. Xilouris, C. Tranoris, S. Denazis, J. Martrat , "Verification and Validation Framework for 5G Network Services and Apps" , in press

[8] 5G Development and Validation Platform for global Industry-specific Network Services and Apps , http://5gtango.eu/

[9] Artificial Intelligence - Neural Networks, https://www.tutorialspoint.com/artificial_intelligence/artificial_intellige nce_neural_networks.htm

[10] S. Benkner, G. Engelbrecht, "A Generic QoS Infrastructure for Grid Web Services", Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services 2006

[11] N. Sfondrini, G. Motta, and L. You, "Service level agreement (SLA) in Public Cloud environments: A Survey on the current enterprises adoption," , 5th International Conference on Information Science and Technology (ICIST) 2015

[12] Y. Koh, R. C. Knauerhase, P. Brett, M. Bowman, Z. Wen, and C. Pu, "An analysis of performance interference effects in virtual environments", IEEE International Symposium on Performance Analysis of Systems & Software 2007

[13] L. Zhu, X.Liu, "Technical Target Setting in QFD for Web Service Systems Using an Artificial Neural Network," IEEE Transactions on Services Computing 2010

[14] C. Emeakaroha, I. Brandic , M. Maurer, S. Dustdar "Low level Metrics to High level SLAs - LoM2HiS framework: Bridging the gap between monitored metrics and SLA parameters in cloud environments", International Conference on High Performance Computing and Simulation (HPCS) 2010

[15] G. Kousiouris, D. Kyriazis, S. Gogouvitis, A. Menychtas, K. Konstanteli, T. Varvarigou, "Translation of Application-level Terms to Resourcelevel attributes across the Cloud Stack Layers" , IEEE Symposium on Computers and Communications (ISCC) 2011

[16] T. Cucinotta, F. Checconi, G. Kousiouris, K. Konstanteli, S.V. Gogouvitis, D. Kyriazis, T.A. Varvarigou, A. Mazzetti, Z. Zlatev, J. Papay, M. Boniface, S. Berger, D. Lamp, and T. Voith, "Manuel Stein: Virtualised E-Learning on the IRMOS Real-Time Cloud," Service Oriented Computing and Applications 2012

[17] M. Alhamad, T. Dillon, E.Chang, "Conceptual SLA Framework for Cloud Computing", 4th IEEE International Conference on Digital Ecosystems and Technologies 2010

[18] Y. Chen, S. Iyer, XLD. Milojicic, A. Sahai , "SLA Decomposition: Translating Service Level Objectives to System Level Thresholds", Fourth International Conference on Autonomic Computing 2007

[19] R. Buyya1, S. Kumar Garg, RN. Calheiros , "SLA-Oriented Resource Provisioning for Cloud Computing: Challenges, Architecture, and Solutions" , International Conference on Cloud and Service Computing 2011

[20] M. Comuzzia, C. Kotsokalisb, G. Spanoudakisa, R. Yahyapourb, "Establishing and Monitoring SLAs in complex Service Based Systems" , IEEE International Conference on Web Services 2009

[21] The 5G Infrastructure Public Private Parternship, https://5g-ppp.eu

[22] Z. Rehman, F.K. Hussain, O.K. Hussain, "Towards Multi-Criteria Cloud Service Selection", Fifth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing 2011

[23] G. Kousiouris, A. Menychtas, D. Kyriazis, SV. Gogouvitis, TA. Varvarigou, "Dynamic, behavioral-based estimation of recource provisioning based on high-level application terms in Cloud platforms", Future Generation Computer Systems 2014

[24] J. Bendriss, IGB. Yahia, P. Chemouil, D. Zeghlache, "AI for SLA Management in Programmable Networks", Design of Reliable Communication Networks; 13th International Conference 2017

[25] Artificial Intelligence - Neural Networks, https://www.tutorialspoint.com/artificial_intelligence/artificial_intellige nce_neural_networks.htm

[26] R.S.Sextona, R.E. Dorseyb, J.D. Johnson, "Optimization of Neural Networks: A Comparative Analysis of the Genetic Algorithm and Simulated Annealing" , European Journal of Operational Research, vol.114, no.3 1999

[27] G. Kousiouris, A. Menychtas, D. Kyriazis, K. Konstanteli, S.V. Gogouvitis, G. Katsaros, T.A. Varvarigou, "Parametric Design and Performance Analysis of a Decoupled Service-Oriented Prediction Framework Based on Embedded Numerical Software", IEEE Transactions on Services Computing, vol.6, no.4 2013