SLAs in 5G: A complete framework facilitating VNF- and NS- tailored SLAs management

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

***Abstract*— As the 5G technology is expected to impact the mobile network and associated ecosystems, appropriate guarantees for the service quality can maximize the ability of Virtual Network Functions (VNF) and Network Services (NS). This implies the utilization of Service Level Agreements (SLA) as a means to ensure that the NSs are provided in an efficient and controllable way. However, the complexity of determining resource provision policies in such multimodal environments as well as the characteristics and properties of various VNFs and NSs, results to custom SLAs that do not consider all aspects of the 5G environment. In this paper we propose an SLA Management framework to map high-level requirements expressed by users, to low-level resource network parameters to improve the service provider's ability to meet the corresponding SLA commitments. In addition, we consider a mechanism for dynamic SLA Templates generation with initial Service Level Objectives (SLO).**

***Keywords* — 5G Networks, Service Level Agreement; Quality of Service; network resources prediction; artificial neural networks;**

I. INTRODUCTION

The technology revolution of the 21st Century has opened the avenues towards many possibilities. The move into the future would require an evolution in networks to meet the enhanced needs of the future. The evolution to 5G networks would mean that the usage of the data will be met by an increasing network quality [1]. 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 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 [2]. In the 5G context, everything tends to be offered as a service, whether it is infrastructures, platforms or software, following the cloud software-platform-infrastructure (SPI) model [3]

The current vision and research challenges in 5G goes beyond the focus of the underlying infrastructure, towards the stakeholders of the ecosystem, including SLA Management, as highlighted in [4]. 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. Therefore, QoS provisioning and management is a very important and challenging field in distributed environments. 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, refers to an estimation of the resources needed to fulfill the user's requirements for every application-as-service. Once the network is virtualized, the SDN controller can configure network devices as quickly as new VMs [5]. As a result, SDNs, enable SLAs to cover the new dynamic nature of the data center and reduce the time and costs to deploy new compute resources. From the service provider's perspective, the benefits include cutting down the time to deploy a new service, lower costs and as a result to increase profitability and improve customer's satisfaction. From a customer's perspective, they get an on-demand virtual environment and the Information Technology services they need in an instance [6].

Considering the problem mentioned above, in this paper we describe the architecture of an SLA Management Framework under development in the EU-funded project 5GTANGO [7]. The SLA Framework incorporates an approach for mapping the high-level requirements (expressed by end users) to low-level policy (i.e. resource) parameters, as described in [8]. Machine learning allows the automated identification of relationships and dependencies between different parameters of datasets. For that reason, 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) [9]. To that end, the proposed approach considers an Artificial Neural Network (ANN) [10] as a black box approach with the major advantages. These refer to the limited amount of information needed to be trained and the generic and flexible nature of the ANNs, which makes them easily adaptable to different software components (as in this case VNFs). Moreover, the proposed framework includes a mechanism that generates SLA Templates in a dynamic way, by analyzing Service Provider's historical data and suggesting the most important QoS parameters for specific VNFs and NSs.

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 SLA life-cycle in 5G networks. Section IV introduces the overall architecture of the presented SLA framework. In Section V we present the functionality of each component included in the SLA Management Framework and at the Section VI we describe in detail the workflow of the framework. Finally, the paper concludes with some thoughts for future work and potentials for the current research.

II. RELATED WORK

Active and advanced work has been conducted on SLA management for Cloud infrastructure and as implied for 5G infrastructure. In this direction, there are several approaches and solutions for mechanisms that handle QoS parameters, in order to optimize the selection process within a workflow. A framework for incorporating QoS in Grid applications is discussed in [11]. 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. In recent cloud-related we have witnessed the emergence of a new category which is the mapping of SLA between Service Provider, Infrastructure Provider and end users [12]. A brief description, similar to our work, is presented in [8]. In this case, the authors presented an approach for mapping the high-level end-user requirements to the low-level policy parameters, as briefly described, and also proposed a mechanism for suggesting the most important QoS parameters to the Service/Infrastructure Provider, in order to achieve better QoS assurance. Moreover, a baseline for our work, was derived through the [13], where an architecture design of life cycle based SLA Management is described. Another interesting work is presented in [14], 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. Furthermore, in [15], 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 parameters to high-level requirements that are used in cloud SLAs is achieved. Furthermore, more details regarding the requirements mapping can be found in [16-17]. A generic black box approach, based on ANNs is used to perform the aforementioned translation between low and high-level requirements.

III. SLAS IN 5G

The SLA lifecycle is an important part of the provision of services, especially in 5G networks. SLA Management is a dynamic process that consists of three main phases: a) Selection, b) Operation and c) Termination [18].

Selection of different network services is triggered by business aspects, being the basis to different QoS parameters, which can also be described as requirements of final SLA. The selection of a network service provider (SP) depends on the desired Network Service (NS), its properties, budget constraints etc. QoS expectations are driving end-users to negotiate specific QoS levels with their service providers. After a successful negotiation process, an SLA is formed to describe the agreed QoS parameters. The final SLA is being monitored with real-time usage data, in order to avoid and manage any unexpected violations. The whole process of defining the requirements, negotiation and management of the SLA, is described as the SLA life cycle [19-20].

The aforementioned SLAs lifecycle in the 5G domain is managed by the corresponding service platforms of 5G environments. Such a platform is developed by 5GTANGO, which is an EU-funded Innovation Action [21] that enables the flexible programmability of 5G networks with:

- an NFV-enabled Service Development Kit (SDK)
- a Catalogue with advanced validation and verification mechanisms for VNFs/Network Services qualification (including 3rd party contributions)
- a modular Service Platform with an innovative orchestrator so it can bridge the gap between business needs and network operational management systems.

The service platform of 5GTANGO, provides the service and function orchestration features, plus all the needed complementary and supporting features, like slice management, policy and SLA management, user access management and infrastructure adapter, as well as monitoring functionalities.

IV. SLA MANAGEMENT FRAMEWORK

The proposed SLA Management Framework is part of the 5GTANGO Service Platform and interacts with the Service/Infrastructure Provider, the end-user -that may be a developer of VNFs and NSs, or/and the service consumer- the Policy Management Framework and the Monitoring Framework. The framework should consider two phases: pre-deployment phase (SLA Template

Management) and post-deployment phase (SLA Information Management).

1) Internal Architecture

Fig.1 presents the architecture of the Framework. The major architectural components are going to be described in detail at the following sections. Briefly, the SLA Template Management component is responsible for receiving all the appropriate information from different stakeholders, and formulate a Service Level Agreement in an efficient way. In addition, it is responsible for the mapping of high-level requirements to low-level resource attributes, through a Mapping Mechanism. On the contrary, the SLA Information Management takes place during the NS runtime, and is responsible for obtaining historical monitoring data, analyze them and make any adaption if necessary to the SLA mapping results.

B. *Related Entities*

As mentioned above, the SLA Management Framework that is presented in this paper is part of the 5GTANGO service platform. The roles and entities that are expected to take part in the SLA management process are the service/infrastructure provider, and the end-user. Moreover, key entities for the SLA Management Framework are the Policies Framework and Monitoring Framework of the 5GTANGO SP, as shown in Fig.2. 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 [22]. The role of the policy rules and expressions are vital to the mapping mechanism as they give specific low-level parameters regarding the NS/VNF. Finally, the SLA Management Framework interacts with a Monitoring Management Framework [23], in order to be optimized.

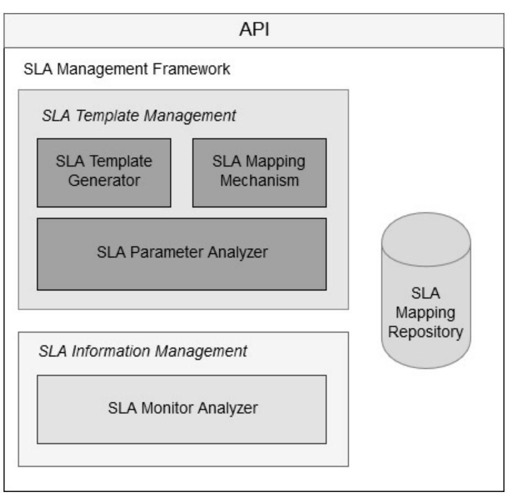

Fig. 1 - SLA Management Framework internal architecture

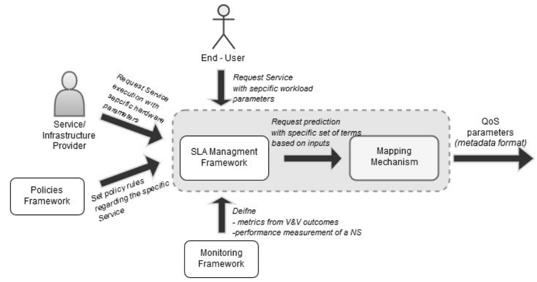

Fig. 2 – Related roles in the SLA Management Framework

V. COMPONENTS

A. *SLA Template Management*

The SLA Template Management component, is responsible for receiving the appropriate dataset from different stakeholders (e.g. service provider, end-user), analyze them, map them and formulate a SLA Template and b) a signed (final) SLA, in an automated and efficient way

1) SLA Template Generator

In a scenario like 5GTANGO a valuable part of the SLA Management Framework is the SLA Generator that firstly creates the SLA templates for the Service/Infrastructure Provider, and secondly formulates the final SLA itself. The Generator, as described in Fig. 3, will be able to obtain a set of policies for the specific NS by the Policy Framework, as well as service/infrastructure provider's historical information by the Monitoring Framework (i.e. NS performance data, resource parameters preferences). The purpose is to be able to analyze and correlate them, in a way that can formulate a generalized SLA Template with initial objectives about the NS. The analysis mentioned above is planned to use Artificial Intelligence (AI). Firstly, clustering algorithms should be applied, to assign a set of observations (i.e. a dataset as described before) into subsets (i.e. clusters) so that observations in the same cluster are similar in some sense [25-26]. Then, a classification model should be developed, in order to come up to a conclusion from the observed subsets [26].

2) SLA Mapping Mechanism

The SLA Mapping Mechanism (MM) is responsible for the mapping between the high-level requirements described by the end-user and the low-level requirements described by the service provider.

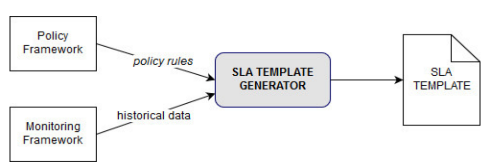

Fig. 3 – SLA Template Generator

The mechanism is also responsible for decomposing the service level objectives to associated policies for runtime QoS enforcement. More specifically, the MM obtains a set of policies from the Policy Manager, a set of low-level requirements from the Service/Infrastructure Provider and a set of high-level requirements from the end-user, based on the services being selected. The datasets are used to compose, aggregate, or convert the low-level metrics to form the high-level SLA parameters. As a result, the outputs of the mapping mechanism are explicit SLA parameters and metrics. The output is categorized between simple and complex mapping results. A simple mapping result maps "end-to-end", from low-level to high-level. For instance, mapping low-level metric *"disk space"* to high-level SLA parameter *"storage"*. Complex mapping results contain predefined formulations for the calculation of specific SLA parameters, using the low-level resource metrics. Table 1, presents a complex mapping result.

The MM 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 another, mainly for confidentiality purposes. Besides, ANNs 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. 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, one parameter can be dynamically mapped to another as depicted in Fig.4.

3) SLA Parameter Analyzer

The SLA Parameter Analyzer is responsible to decide whether the process of the mapping mechanism should take place or not. The decision is taken after the component searches into the mapping repository (see next section) and check if there is already a correlation between the input parameters and the stored mapping results.

B. SLA Information Management

The SLA Information Management, as previously mentioned, takes place after the successful deployment of the NS and it is mainly responsible to optimize the mapping results. In detail, the component obtains historical monitoring

Table 1: Complex Mapping Result

Low-Level Metric	SLA Parameter	Mapping Formulation
downtime, uptime	Availability (A)	$A = 1 - \frac{downtime}{uptime}$

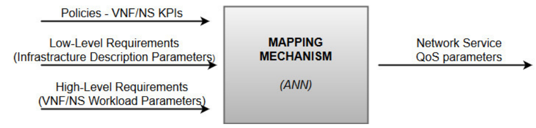

Fig. 4 – Prediction of QoS parameters for given requirements

data for the deployed NS, as well as resource parameters preferences. After that, it has to decide whether there is any breach of contract between the mapping results stored in the Mapping Repository, and the actual monitoring data. Specifically, the SLA Monitor Analyzer component, compares the QoS parameters from the Mapping Repository, with the computed monitoring data, and then calculates the delta between them. If delta is more than "0", the MM will obtain the monitoring feedback as an additional dataset and re-do the mapping process, as described in the section "SLA Mapping Mechanism".

C. SLA Mapping Repository

The SLA Mapping Repository is responsible for keeping all the produced mapping results from the Mapping Mechanism component. More specific, it will store the correlations between the high-level and the low-level requirements.

VI. SLA MANAGEMENT WORKFLOW

In order for a network service to be deployed in the service provider's infrastructure, several steps should be done with regards to the performance estimation and the required quality, that should be agreed and signed in the final SLA, as shown in Fig. 5.

A. SLA Template Generation Process

During the SLA template timeline an initial SLA template will be formulated for the service provider. The SLA Generator will analyze the metadata that are attached in every given NS, and generate in a dynamic way a structured and at the same time generic SLA template.

B. Final SLA Formulation Process

- Part I: The SLA is made available by the service provider to the end-user. In sequel and based on the user's business requirements, the SLA might be accepted and signed as is or a new negotiation process (with updated QoS parameters) might be triggered
- Part II (a): Once the SLA Manager gather all the appropriate datasets, checks if there is already a combination in the Mapping Repository, between the latter and already existing mapping results, in order to decide whether the process of the mapping mechanism should be done again or not. In case the

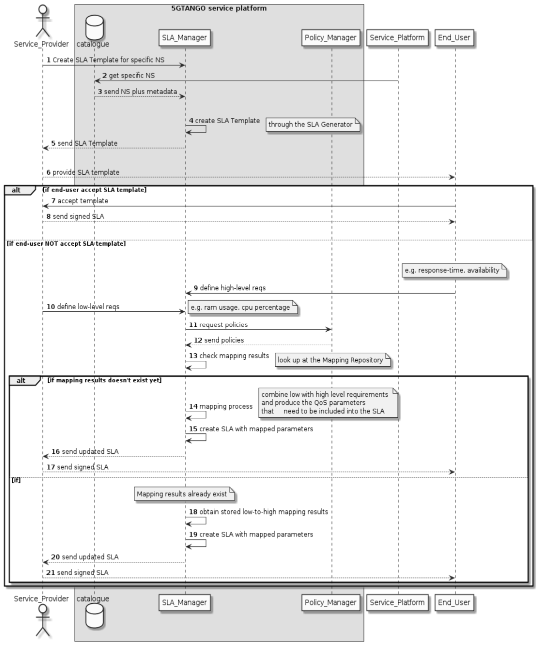

Fig.5 – SLA management workflow

provider's low-level requirements and policies takes place. The aim is to foresee the performance estimation and the required quality that should be agreed and signed in the SLA.

- Part II (b): In case there is already a combination between the input dataset (requirements obtained from the service provider and the End-User) and the stored mapping results, the SLA Manager bypasses the mapping process and dynamically creates the SLA.

C. Overall SLA Compilation

At the current subsection, a combination of all the aforementioned mechanisms is described, in order to present step-by-step, the formulation of an SLA. Given that the VNFs/NS will be accessible through a catalogue, the latter will propose the corresponding VNFs/NSs based on the service/ infrastructure providers' requirements and constraints.

Based on the above, the following steps are followed as shown in Fig.6:

1) The Service/Infrastructure Provider selects VNF/NS from 5GTANGO Catalogue in order to be deployed.
2) The required datasets are defined through the Parameter Analyzer:
    a) Service/Infrastructure Provider defines a set of low-level requirements –. resource parameters
    b) End-User defines a set of high-level requirements – VNF/NS workload parameters.
    c) The Policy Management Framework provide policy rules and expressions for the selected NS.
3) The SLA Template is generated and being accessible to the provider for further use.
4) The Mapping Mechanism correlates the datasets so as to predict the QoS parameters (i.e. output) in order to be included to the final SLA.
5) The mapping results are getting stored to the Mapping Repository, as well as the SLA Generator obtain them in order to formulate the final SLA.
6) With repetitive flow, and considering the SLA lifecycle, the SLA Management Framework obtains monitoring information to optimize the overall process.

CONCLUSIONS AND FUTURE WORK

The mapping problem between high-level requirements (workload and QoS) and low-level recourse attributes, is a challenging problem in virtualized environments as in the 5G domain. An effective SLA is the key to ensure that a service/infrastructure provider delivers the agreed QoS terms of Network Services to the end-user.

In this paper, we have presented an SLA Management Framework enabling translation of service-level parameters (workload and QoS) to resource-level attributes. The framework exploits a generic approach that is based on ANNs, placing the framework as a mediator between Service Providers and end-users, with the intention to estimate how many recourses resources are needed for a Network Service instance with specific QoS levels. Moreover, the framework exploits a template generation mechanism to ensure that the proposed SLAs are tailored to the corresponding VNFs and NSs and their management is performed according to their specific parameters. The next step refers to the extension of the framework to enable QoS enforcement and triggering of actions to adapt the infrastructure accordingly in runtime.

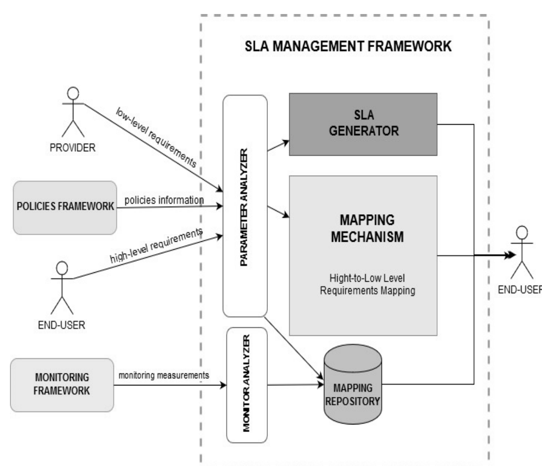

Fig. 6 - SLA Management Framework overall process

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] Explore IoT, 5G and VROpportunities, https://www.datasciencecentral.com/profiles/blogs/mobile-data-traffic-explosion-explore-iot-5g-amp-vr-opportunities

[2] 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

[3] National Institute of Standards and Technology, "The NIST Definition of Cloud Computing," http://csrc.nist.gov/ publications/nistpubs/800-145/SP800-145.pdf, 2013.

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

[5] SDN Architecture Overview, https://www.opennetworking.org/images/stories/downloads/sdn-resources/technical-reports/SDN-architecture-overview-1.0.pdf

[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 2016, vol.93

[7] 5G Development and Validation Platform for global Industry-specific Network Services and Apps , http://5gtango.eu/

[8] E. Kapassa, M. Touloupos, A. Mavrogiorgou, D. Kyriazis, "5G & SLAs: Automated proposition and management of agreements towards QoS enforcement", in press.

[9] An Introduction to Machine Learning Algorithms https://www.datascience.com/blog/introduction-to-machine-learning-algorithms

[10] Overview of Artificial Neural Networks an, https://hackernoon.com/overview-of-artificial-neural-networks-and-its-applications-2525c1addff7

[11] 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

[12] 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

[13] Z. Shu, S.Meina, "An Architecture Design of Life Cycle Based SLA Management ", he 12th International Conference on Advanced Communication Technology (ICACT), 2010

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

[15] 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

[16] 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

[17] 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, T. Voith, "Virtualised E-Learning on the IRMOS Real-Time Cloud," Service Oriented Computing and Applications 2012

[18] Cloud SLA lifecycle, http://www.sla-ready.eu/cloud-sla-lifecycle

[19] [19] L. Wu and R. Buyya, "Service level agreement (SLA) in utility computing systems", Performance and Dependability in Service Computing: Concepts, Techniques and Research Directions, IGI Global, pp. 1, 2011

[20] G. Conway, E. Curry, "Managing cloud computing-a life cycle approach", CLOSER, pp. 198 2012

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

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

[23] P. Trakadas, P. Karkazis, H.C. Leligou, T. Zahariadis, S. Kolometsos, "Implementation of a Monitoring Framework to Support the Management of 5G Services" , in press

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

[24] I. Hussain,"Clustering in Machine Learning", http://www.cad.zju.edu.cn/home/zhx/csmath/lib/exe/fetch.php?media=2011:presentation_ml_by_ibrar.pdf

[25] An Introduction to Clustering and different methods of clustering, https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/

[26] N. Bui, M. Cesana, S.A. Hosseini,Q. Liao, I. Malanchini, J. Widmer, " A Survey of Anticipatory Mobile Networking: Context-Based Classification, Prediction Methodologies, and Optimization Techniques", IEEE Communications Surveys & Tutorials, vol. 19, no. 3, third quarter 2017