Autonomous Resource Assignment for Optimal Utilization in Optical Data Centre Infrastructures

Albert Pagès[(1),*], Jordi Biosca Caro[(1)], Fernando Agraz[(1)], Salvatore Spadaro[(1)]
(1) Optical Communications Group (GCO), Universitat Politècnica de Catalunya (UPC), Barcelona (08034), Spain
** e-mail: albert.pages-cruz@upc.edu*

Abstract: We present an architectural solution based on data analytics for self-organizing optical data centers. Thanks to a reinforcement learning-based cognitive layer, an adaptive and autonomous resource assignment to deployed services is achieved. © 2021 The Authors

1. Introduction

As telecom infrastructures move towards 5G and beyond architectures, it is no longer sufficient to automate their control and management loops, but it is also required to provide the means for autonomous operation of the infrastructures, adapting to dynamic workloads while anticipating future states. The concept of self-organizing infrastructures (SOIs) has risen as an answer to these needs [1]. SOIs refer to those infrastructures that have the capacity to autonomously handle their control and management operations. In such cases, data-analytics/machine learning (ML) frameworks [2] usually aid the management/control layers on gaining insights about the infrastructure and steer its behavior towards the desired state. The governance entity may specify the necessary operations, named actuations, to achieve a specific goal. Among the several approaches, declarative solutions have gained interest: actuations are defined as a high-level intent, leaving the concrete (re-)configuration operations to the discretion of the governed systems, separating thus the "what" from the "how". Given such a framework, in this work, we focus on an intra-data center (DC) scenario with an optical DC network (DCN), tasked with guaranteeing the performance of deployed services. The presented work relays on an ML-based governance layer which, starting from monitoring the status of both computing and networking resources, it determines the most suitable actuations to maintain an optimized infrastructure performance.

2. Architecture for self-organizing optical datacenters

Fig. 1 depicts the presented architecture for a self-organized optical DC infrastructure. The specific technologies employed for each one of the components are also indicated, which will be exercised during the Proof-of-Concept (PoC) experiment at the following section. The DC infrastructure consists of several servers arranged in racks, interconnected thanks to the capabilities of an optical DCN fabric composed by opto-electronic Top of the Rack (ToR) and all-optical switches, interconnected by fiber links. Network services (NSs) are deployed as per the requests of customers, by means of the service/slice and resource orchestrators, and a Software Defined Networking (SDN) controller. An NS requests for several Virtual Network Function (VNF) instances, which will be deployed as Virtual Machines (VMs), interconnected following a specific topology, materialized through optical connections.

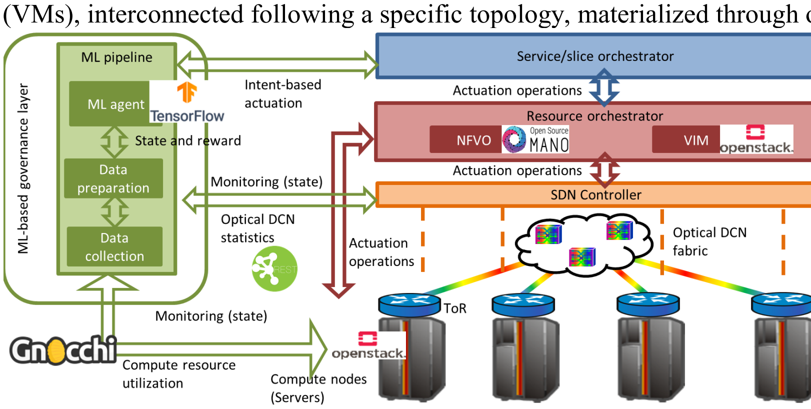

Fig. 1. Architecture of the intent-based self-organizing framework in optical DC infrastructures.

The presented architecture follows the Monitoring, Analyze, Plan and Execute over a shared Knowledge (MAPE-K) loop [3] approach to achieve a holistic actuation framework. After the initial provisioning of NSs, the status of physical infrastructure resources is continuously monitored to potentially modify the provisioned resources through appropriate actuations. The heart of the actuation framework lays on a ML-based governance layer; in particular, Deep Reinforcement Learning (DRL) [4] is here used. The state space is a vector which contains the information needed to get an overview of the resources deployed (e.g., memory/CPU/storage usage, link bandwidth, assigned optical path, etc...). The reward function weights those metrics that form the state; computing resource metrics are weighted independently according to their relevance based on the computing resource policies of the DC operator, while the metrics related to the monitored optical network are rewarded to give higher values of bandwidth, low bit error rate and lower usage of optical resources. The DRL agent interacts with the environment (i.e. the DC orchestration system) by changing the status of DC resources in order to obtain a higher reward. The action space defines how and to who allocate resources, that is, it specifies an intent-based actuation to instruct the orchestration system with the necessary (re-)configurations to adjust the allocated resources to supervised NSs to keep with their loads. After training, the expected behavior of the agent is that, according to the state, it will predict which action is optimal, thus keeping the assigned resources to NSs at the optimal values in an autonomous way.

3. Experimental evaluation

In this section, we showcase the autonomous provisioning of resources to deployed services. As a starting point, an NS consisting of two interconnected VNFs has been deployed. Then, a DRL agent based on TensorFlow, is also deployed. The DRL agent focuses on monitoring the resources assigned to the deployed VNFs (VMs) and instruct scale-up/down/maintain actuations of the number of assigned CPU cores, memory and storage capacity to keep with their evolving utilization. In addition, it also monitors a normalized metric E, ranging from 100 to 0, exposed by the SDN controller, corresponding to the quality of the transmission of the inter-VNF connection, based on electrical packet statistics (packet drop) and optical channel metrics (optical SNR).

Fig. 2 (left) depicts the autonomous assignment of computational resources. Specifically, it depicts the memory assigned to one of the supervised VNFs and its utilization as a consequence of the actuations. For the particular case, the DRL agent seeks to adjust the assigned resources to meet medium utilization levels, after going through a Gaussian distribution reward function which punishes under- or over-provisioned states. To do so, it contacts the orchestrator specifying which of the VM resources should be up- or down-scaled, or kept the same. The solid line represents the currently assigned resources while the dashed line represents their usage. It can be seen how, in cases in which the usage is high, the agent instructs to increase the resources, while when the utilization lowers, resources are scaled-down. This ultimately leads to a better usage of DC resources, potentially allowing to host more services. Fig 2 (right) represents the evolution of the network related E metric. In such a case, we force situations in which the E value decreases over the time. As a response, the DRL agent learns that, after reaching a predefined threshold (represented by the solid horizontal line) it needs to instruct the network control to assign alternative optical resources (e.g. by means of re-routing), regaining optimal values. We can appreciate how as the agent learns, the difference (delta) between the E threshold and the value which triggers the actuation, lowers, tending to zero. This highlights that the actuation trigger of the agent improves over the time, executing the actuation only when needed.

Fig. 2. Autonomous computing resource assignment (left) and E factor evolution in response to actuations (right).

Acknowledgments

This work has been supported by the Spanish Government through project TRAINER-B (PID2020-118011GB-C22) with FEDER contribution.

References

[1] 3GPP, "Self-Organizing Networks (SON); Concepts and requirements (Release 16)", Technical Specification (TS) 32.500, July 2020.
[2] W. Kellerer et al., "Adaptable and Data-Driven Softwarized Networks: Review, Opportunities and Challenges", Proceedings of the IEEE, vol. 107, no. 4, pp. 711-731, April 2019.
[3] S. Spadaro et al., "Enabling Service Provisioning and Quality Maintenance in Sliceable Optical Networks", in Proc. Photonics West OPTO 2021, March 2021.
[4] M. Naeem et al., "A Gentle Introduction to Reinforcement Learning and its Application in Different Fields", IEEE Access, vol. 8, pp. 209320-209344, November 2020.