Computer Communications 91–92 (2016) 62–75

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

A PCE-based architecture for green management of virtual infrastructures

Giovanni B. Fioccola [a], Pasquale Donadio [b], Roberto Canonico [a,*], Giorgio Ventre [a]

[a] Department of Electrical Engineering and Information Technology, University of Napoli “Federico II”, Via Claudio 21, 80125 Napoli, Italy
[b] Consorzio Interuniversitario Nazionale per l'Informatica (CINI), Via Cinthia, 80126 Napoli, Italy

ARTICLE INFO

Article history:
Received 23 December 2015
Revised 31 May 2016
Accepted 7 July 2016
Available online 9 July 2016

Keywords:
Cloud computing
Optical networks
Path computation element

ABSTRACT

Recent evolutions of virtualization technologies allow carriers to optimize and monetize their network infrastructures in new ways, acting as virtual infrastructure providers. By assuming an underlying GMPLS-enabled network infrastructure connecting a number of geographically dispersed data centers, in this paper we define an architectural framework that allows infrastructure providers to optimally use their resources to provide Virtual Infrastructures on demand. The architecture we propose is designed as an extension of the standard Path Computation Element (PCE) architecture. A centralized entity, named VRO, is responsible for optimally allocating the physical resources needed to deploy a requested Virtual Infrastructure. In the paper, we also present how it is possible to apply our framework to pursue green management objectives so that OPEX expenditures can be reduced, while preserving contractual SLAs. We also describe a prototype of our framework that is able to configure GMPLS-enabled network nodes and Cloud-enabled data centers in order to create Virtual Infrastructures.

1. Introduction

The emergence of the Network Function Virtualization (NFV) paradigm allows network service providers to extend their business models by combining the traditional service portfolio with innovative Cloud computing services [1]. By relying on the virtualization of resources, service providers can automate the highly dynamic delivery of virtualized network services and create on-demand multiple isolated virtual infrastructures for their customers. This concept, previously known as network virtualization [2], has recently been further extended to include both communication and computational resources. In this context, the term Virtual Infrastructure (VI) is used to denote a set of computational resources (i.e., virtual machines and virtual disk volumes) deployed in a number of distributed data centers and connected by guaranteed-bandwidth virtual links [3]. The aim of creating such Virtual Infrastructures is to provide a given service to a known and variable population of end users. Network Function Virtualization is the best paradigm to implement the idea of virtual and distributed service infrastructures. For this paradigm to be effective, infrastructure providers need more powerful management platforms to efficiently combine management procedures for both communication and IT resources. As a matter of fact, the European Telecommunications Standards Institute (ETSI) has identified the necessity of “a consistent management and orchestration architecture” as one of the challenges to be addressed for successfully implementing NFV.

Hereinafter, we assume that the underlying networking infrastructure supports Generalized Multi-Protocol Label Switching (GMPLS) [4] and conforms to the Path Computation Element (PCE) architecture [5]. GMPLS enables dynamic topology reconfiguration, while PCE establishes Label Switched Paths (LSPs) as virtual Traffic Engineering (TE) links for the allocated virtual GMPLS control plane. Such an infrastructure is able to automatically provide guaranteed-bandwidth network connections between given endpoints (e.g., the WAN core routers of distributed data centers [6]). Based on this assumption, we propose an architecture that assigns to a central management entity, named Virtual Resource Orchestrator (VRO), the responsibility for optimally allocating the resources needed to deploy a requested Virtual Infrastructure. Fig. 1 shows how a VI request, issued by a Virtual Service User (VSU), is taken by a Virtual Infrastructure Provider (VIP), which tasks VRO with deploying virtual machines, virtual volumes and virtual links needed to build the Virtual Infrastructure. The resulting VI must satisfy a set of requirements expressed in the initial VI request. Proper mechanisms (e.g., based on transparent migration of VMs across data centers [7]) to provide recovery mechanisms for the

* Corresponding author.
E-mail addresses: giovannibattista.fioccola@unina.it (G.B. Fioccola), padonadio@gmail.com (P. Donadio), roberto.canonico@unina.it (R. Canonico), giorgio.ventre@unina.it (G. Ventre).

http://dx.doi.org/10.1016/j.comcom.2016.07.003

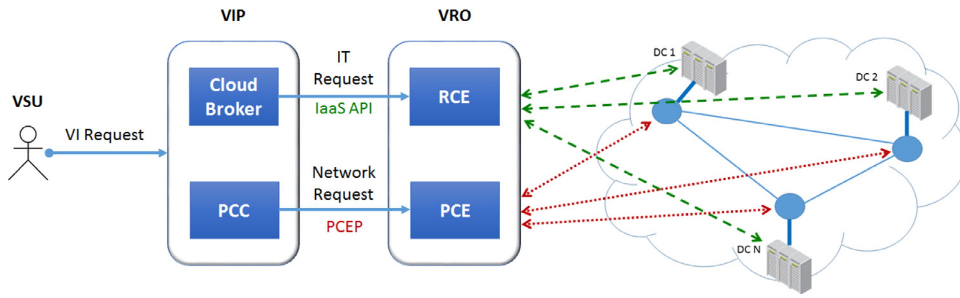

Fig. 1. Creation of a virtual infrastructure on a VRO-based infrastructure.

created infrastructure could also be implemented, as proposed in [8].

As of today, the creation of a Virtual Infrastructure would be accomplished by statically allocating resources. This process is lengthy, costly, error-prone, and in the case of long-term contracts there is a heavy underutilization of over-committed resources. The purpose of the VRO is to translate a VI request into a number of coordinated network planning and provisioning actions. Being solicited only when creating VIs, the VRO is not subject to heavy load, as we expect only a few VI requests per hour to be processed. Besides assuring contractual SLAs, the purpose of the VRO is to pursue optimization objectives that are compatible with the requirements negotiated in the SLAs. The VRO design has been derived by extending the standard PCE architecture. One of the most important goals for the network management procedures is to minimize energy consumption of IT and networking infrastructures. In fact, it has been estimated that data centers accounted for 1.3% of worldwide electricity use in 2010, as they are one of the major sources of energy consumption of the whole ICT sector [9]. Hence, finding a compromise between power consumption and the perceived Quality of Service (QoS) is one of the main objectives in Cloud data centers. Equipment and network topologies must be chosen by taking into account energy consumptions, as well as performance and costs. Indeed, energy costs are a significant amount of the OPerating EXpenditure (OPEX) in current Cloud infrastructures. Therefore, the trade-off between CAPital EXpenditure (CAPEX) (for example infrastructure costs) and OPEX has to be considered according to an energy-efficiency point of view. In this paper, a framework that allows network management procedures to be used by infrastructure providers has been described. User's resource deployment requests are processed with the aim of saving energy, so as to reduce OPEX, while preserving SLAs negotiated with customers. Moreover, we suppose that the overall network performance cannot be improved through over-designing after an initial investment in hardware and software acquisition, hence CAPEX costs are given. Finally, the VRO and its prototype implementation are described. A preliminary description of the VRO and its functions was already presented in [10]. This paper extends that work with an experimental evaluation that shows how the VRO can be configured to pursue significant energy savings, by combining green management procedures in data centers with a green management of the geographical networking infrastructure.

The remainder of this paper is organized as follows. In Section 2 we discuss related work. Section 3 presents the Path Computation Element, which is the key component of the control plane architecture. Section 4 presents the Virtual Resource Orchestrator (VRO) and its interactions with network and IT resources. In Section 5, we present the results of a simulation study that shows how the VRO can be configured to pursue the green management of the infrastructure. Section 6 describes a prototype of the VRO, which has been implemented on a local testbed. Section 7 concludes the paper by drawing our final remarks.

2. Related work

The concept of Virtual Infrastructures provided as a service is the basis of the emerging Cloud computing paradigm. In this paper, we refer to a scenario in which users have control over the geographical location of their virtual machines and are provided with contractual guarantees regarding the geographical connectivity among VMs located in different data centers. Such a scenario has been assumed in other recent works already, e.g. [3]. For example, Virtual Infrastructures created on-demand might be Content Delivery Networks, as proposed in recent papers [11,12]. We claim that NFV may play a significant role in this context. As a matter of fact, ETSI envisions dynamic creation of on-demand Content Delivery Networks (CDNs) as a relevant use case for NFV in [13].

Network and Cloud resource orchestration require complex management procedures to guarantee performance, energy efficiency, security, robustness and reliability. Therefore, the problem of resource management in virtualized data centers has been largely investigated in the last few years [14]. Defining the mapping of virtual resources to physical ones is commonly known as embedding; this problem has been extensively investigated in the context of network virtualization in wide area networks [15,16]. More recently, the same formulation has been applied to data center context [17]. Within a single data center, resource management procedures typically rely on live migration techniques for dynamic relocation of VMs [18]. In [19] authors propose an effective sequencing technique (named CQNCR) for determining the execution order of massive VM migrations within data centers. Specifically, given an initial and a target resource configuration, CQNCR sequences VM migrations to efficiently reach the final configuration with minimal time and impact on performance. Experiments show that CQNCR can significantly reduce total migration time by up to 35 percent and service downtime by up to 60 percent. Both solutions are very promising, but they do not provide any support for network resources orchestration.

In [20] the authors address the role of high-performance dynamic optical networks in Cloud computing environments. The use case described in their work is completely analogous to our scenario. A new architecture based on a Data Center as a Service (DCaaS) solution is introduced, so that customers are able to create Cloud platforms. Central to the proposed architecture is the coordinated virtualization of optical networks and IT resources allocated in distributed data centers, in order to deploy virtual infrastructures with guaranteed QoS. Their paper also provides a detailed description of possible approaches for the optical network virtualization. Finally, authors present three different Virtual Optical Network (VON) composition methods with two distinct customized objectives (i.e., cost-aware or resource-aware). However, none of the proposed methods takes into account other management objectives, such as reduction of energy consumption. Moreover, the opportunities deriving from the dynamic consolidation of virtual machines within data centers are not considered.