Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 97 (2016) 114 – 117

Procedia Computer Science

CLOUD FORWARD: From Distributed to Complete Computing, CF2016, 18-20 October 2016, Madrid, Spain

In-Network Programmability for next-generation personal cloUd service support (INPUT)

Roberto Bruschi[a], Paolo Lago[a,b], Chiara Lombardo[a,b,]*

[a]*CNIT - University of Genoa Research Unit - Genoa, Italy*
[b]*DITEN - University of Genoa - Genoa, Italy*

Abstract

In order to overcome the cloud service performance limits, the INPUT Project aims to go beyond the typical IaaS-based service models by moving computing and storage capabilities from the datacenters to the edge network, and consequently moving cloud services closer to the end users. This approach, which is compatible with the concept of fog computing, will exploit Network Functions Virtualization (NFV) and Software Defined Networking (SDN) to support personal cloud services in a more scalable and sustainable way and with innovative added-value capabilities.

Peer-review under responsibility of organizing committee of the international conference on cloud forward:
From Distributed to Complete Computing

Keywords: SDN; Cloud services; Personal Networks; Virtualization

1. Introduction

The traditional Internet infrastructure operates most of the times on top of proprietary and specialized firmware/hardware components, making the support of new features impossible. In addition, as underlined by Nokia[1], next-generation cloud services will have very challenging requirements for the maximum end-to-end response time, but today's network latencies are much higher (around two orders of magnitude) than the values

* Corresponding author. Tel.: +39 010 353 2075.
E-mail address: chiara@tnt-lab.unige.it

Peer-review under responsibility of organizing committee of the international conference on cloud forward:
From Distributed to Complete Computing
doi:10.1016/j.procs.2016.08.286

required by next-generation cloud services. Even if upcoming 5G network technologies are envisaged to provide much better performance in terms of bandwidth and network latency, traffic exchange inside the access/edge portions of telecom operator networks remains a big concern. The presence of multiple stakeholders involved in the services deployment, and often with competing objectives, also represents a significant change in accessing the network resources. Telecom operators rent portions of their physical and computational resources to cloud service providers, so the former see costs and performance at infrastructure level, while the latter only perceive abstracted/virtual computing platform with no energy consumption indications. Moreover, the presence of multiple service providers calls for refined mechanisms to guarantee isolation and connectivity of each service, which can further hinder the scalability of already unsustainable datacenters[2].

In order to respond to the challenges of increased device/object connectivity and data-intensive applications, deploying storage, computing and configuration features in the edge network, hence closer to the end-user, instead of inside datacenters can improve the control on the allocation of physical and logical resources with the possibility of applying economies of scale and power management schemes, and improved privacy.

In this respect, the European Commission-funded INPUT Project[3] aims to go beyond the typical IaaS-based service models by moving computing and storage capabilities from the datacenters to the edge network, and consequently moving cloud services closer to the end users. This approach, which is compatible with the concept of fog computing, will exploit Network Functions Virtualization (NFV) and Software Defined Networking (SDN) to support personal cloud services in a more scalable and sustainable way and with innovative added-value capabilities.

2. The INPUT Solutions for Supporting Personal Cloud Services

The INPUT Project aims at designing novel infrastructures and paradigms to support Future Internet personal cloud services in a more scalable and sustainable way. The INPUT technologies will enable next-generation cloud applications to go beyond classical service models (i.e., IaaS, PaaS, and SaaS), and even replace physical Smart Devices (SD), usually placed in users' homes (e.g., network-attached storage servers, set-top boxes, video recorders, home automation control units) or deployed around for monitoring purposes (e.g., sensors), with their "virtual images," providing them to users "as a Service" (SD as a Service – SDaaS).

In this way, virtual and physical SDs will be made available to users at any time and at any place by means of virtual cloud-powered Personal Networks, which will constitute an underlying secure and trusted service model (Personal Network as a Service – PNaaS). These Personal Networks will provide users with the perception of always being in their home Local Area Network with their own (virtual and physical) SDs, independent of their location.

2.1. Personal Networks

A Personal Network (PN) is a secure and trusted virtual overlay network capable of interconnecting the SDs of a user with standard L2 protocols and operations equivalent to the ones presently available in the user's home network, independent of their location (inside/outside the user's home) or nature (physical/virtual). The correct routing of the L2 data and signaling packets is guaranteed by the OpenFlow's matching/action rules, and is dynamically provided by the control processes as described in Section 3.

PNs are realized by virtualizing typical Network Functions provided by the user's home gateway, and transferring them into software instances (*Net_Functions*) running in commodity computing facilities deployed in the telecom operator edge network. A *Net_Function* can replace a single data- or control-plane network functionality (e.g., IP forwarding/routing, firewall, deep packet inspection, NAT, DHCP), so a chain of *Net_Functions* provides all the L3 signaling and data-plane operations on behalf of the user's physical home gateway.

Physical SDs typically connected to the user's LAN are fully or partially virtualized through software instances, named Service Applications (*Service_Apps*), running at different levels of the edge network infrastructure. Additionally, *Service_Apps* and *Net_Functions* can be dynamically migrated from one computing facility to another, while guaranteeing service continuity. The purpose of the migration process is twofold: on the one hand, it can be used to put under-utilized servers in low power idle or standby states, and hence reduce the carbon footprint produced by the telecom operators, on the other hand placement of the service chain physically "closer" to the user position (e.g., migration of the service chain to a server facility closer to the current user position) can reduce end-to-end latency and thus improve the overall QoE.

Service_Apps are meant to cooperate with applications residing in the users' smart devices (*User_Apps*) and optionally in datacenters (*DC_Apps*) to realize innovative personal cloud services. The presence of such *Service_Apps* allows user requests to be directly satisfied by completely offloading datacenters and/or to be pre-manipulated before crossing the network and arriving at datacenters in ways that enhance performance.

In order to correctly manage the INPUT infrastructure resources, both *Service_Apps* and *Net_Functions* must be specified by their computational and memory/storage requirements and by the expected levels of Quality of Service (QoS).

3. The INPUT Edge Network Architecture: Building Blocks and Interfaces

As illustrated in Fig. 1, the edge network control plane includes the Network and Service Management (NS-MAN) and the Network and Service Operating System (NS-OS). The NS-MAN is responsible for the long-term configuration of the network, the administrative configuration of the infrastructure, the overlaying cloud services and PNs, and for the monitoring of the resources usage and power consumption of the overall INPUT infrastructure. The NS-OS, on the other hand, drives the real-time configuration of the programmable resources and the dynamic instantiation and migration of *Service_Apps* and *Net_Functions* according to users' locations. Three tasks are used to this goal: *Consolidation* is in charge of calculating the optimal re-configuration of the infrastructure, *Orchestration* exploits the set-up coming from *Consolidation* to instantiate/migrate *Service_Apps* and *Net_Functions* to the identified subset of devices/hardware resources, and *Monitoring* collects performance measurements and alerts, including infrastructure- and device-level power consumption, end-to-end latency, and user mobility statistics.

The data path is composed of a number of SDN switches and NVF servers geographically distributed across the edge network to host the service chains allocated according to the directives coming from the NS-OS. Hypervisors are used for the proper management of the Virtual Machines (VMs) composing the service chains.

Fig. 1 also shows the North- and South-bound interfaces (NBIs and SBIs) used for monitoring and communication purposes within the edge network. Some are well-known interfaces widely used in network and cloud infrastructures, like SNMP and OpenFlow, while others stem from achievements of the ECONET Project, like the Green Abstraction Layer (GAL), a recently approved ETSI standard[4].

The INPUT SBIs will be realized by extending current SDN and NFV protocols beyond stateless network operations (as in today's version of the OpenFlow architecture) and the Network-as-a-Service (NaaS) paradigms typical of current NFV approaches. Power management primitives will be abstracted and mapped onto the SBI protocols as well. telecom operators will expose a standard and simple environment to cloud service providers by means of the NBIs, which will allow the definition or updating of the *Service_Apps*, the type of data to be maintained, and their operative requirements and features (e.g., Service Layer Agreement – SLA). This kind of interfaces will be provided by the telecom operators to the cloud service providers through a PaaS paradigm (since the infrastructure will be directly managed and consolidated by NS-MAN and NS-OS).

The INPUT Project will aim at designing algorithms and criteria for autonomously and dynamically driving the configuration of network infrastructures to find the best compromise between the minimization of energy consumption (and consequently of operating expenses of telecom operators) and the maximization of the perceived performance levels of cloud services/applications (which brings benefits to the cloud service providers), thus promoting interactions and collaboration among the stakeholders in a way that would not hinder their independence: for example, the telecom operators may encourage the service providers to adopt a fairer approach in the use of their rented resources through economic incentives. To this goal, an important role will be played by the design of consolidation/orchestration algorithms for the minimization of energy consumption and the maximization of the perceived performance levels of cloud services/applications through proper service placement.

4. Conclusions

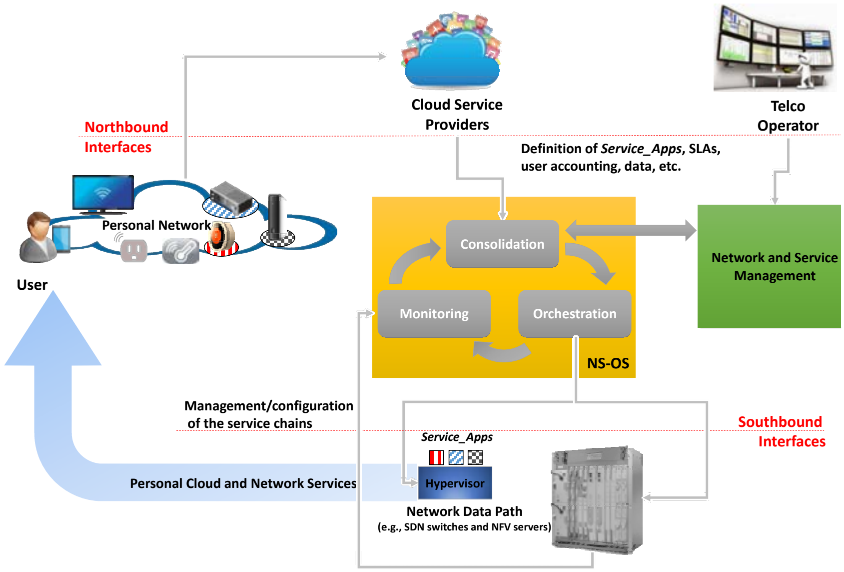

Fig. 1. Functional blocks and interfaces of the INPUT architecture.

In this article, we have introduced the INPUT Project and the edge network architecture we are defining, with particular attention to the control plane and interfaces used for internal and external communication. As the Internet scenario has become more heterogeneous and demanding throughout the years, with multiple stakeholders characterized by competing objectives, the traditional TCP/IP paradigm requires enhancements to provide flexibility, upgradability, and integration with IT services. In order to improve scalability and user-experience, the INPUT vision consists of drastically reducing end-to-end latency by exploiting the SDN control capabilities to introduce in-network programmability into edge network devices, and make them available to host cloud applications and place them according to algorithms that enhance performance and energy efficiency.

Acknowledgment

This work has been supported by the INPUT (In-Network Programmability for next-generation personal cloUd service supporT) project, funded by the European Commission under the Horizon 2020 Programme (Grant no. 644672).

References

1. “Technology Vision 2020 update – Personalizing the Network Experience,” Nokia Networks, July 2015, URL: https://blog.networks.nokia.com/customer-experience/2015/07/30/12537/
2. Cisco, “Cisco Visual Networking Index: Forecast and Methodology, 2012-2017,” White paper, May 2013, URL: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html
3. The In-Network Programmability for next-generation personal cloUd service supporT (INPUT) Project, URL: http://www.input-project.eu/.
4. ETSI Environmental Engineering (EE), “Green Abstraction Layer (GAL) - Power management capabilities of the future energy telecommunication fixed network nodes,” ES 203 237 V1.1.1, URL: http://www.etsi.org/news-events/news/822-2014-09-news-green-abstraction-layer-standard-to-manage-energy-consumption-of-telecom-networks.