This is a post-print version of the following article: R. S. Montero, E. Rojas, A. A. Carrillo, I. M. Llorente, "Extending the Cloud to the Network Edge", IEEE Computer, pp: 91-95, Issue No. 04 – April, 2017 vol. 50

Extending the Cloud to the Network Edge

June 2017

Rubén S. Montero, OpenNebula, Elisa Rojas, Telcaria Ideas, Alfonso A. Carrillo, Telefónica, and Ignacio M. Llorente, OpenNebula

Abstract

Telefónica's OnLife project aims to virtualize the access network and give third-party Internet of Things application developers and content providers cloud-computing capabilities at the network edge.

1. Introduction

The Internet of Things (IoT) is drastically changing our world by connecting every kind of device to the Internet, from doorbells and sprinkler systems to health sensors and traffic lights. Ideally, these devices will interconnect with other devices or services to perform their tasks in smarter ways, forming the basis of an optimal environment that reacts to our needs and moods.

To realize this vision, we need a new computing infrastructure that can cope with massive device connectivity and is flexible enough to address the requirements of a diverse set of devices and their associated applications.[1] Reducing and managing communication latency will define the future of IoT applications like video streaming, gaming, and many mobile apps.[2] For example, voice-controlled smart-home systems benefit from content caching, health devices require low latency to respond to emergencies in real time, connected cars might rely on the collective processing of nearby vehicles' sensor data, and industrial robotics demand more computing capabilities with steady latency.

The geographical distance between IoT service providers and users from a centralized cloud infrastructure turns out to be an important issue.[3] Centralized clouds are appropriate for services with limited data communication—such as web services—or for batch processing, but not for applications that require moving large amounts of distributed data or those with interactive users that require low latency and real-time processing. Meeting these latency demands requires bringing resources as close to IoT devices as physically possible, as the response delay introduced by intercountry—or intercontinental—round trips would make IoT applications unfeasible. Moreover, if we consider the number of connected IoT devices, centralized processing of their generated data doesn't scale, and such processing needs to be distributed among resources close to the devices. These close-to-the-device resources also need to be provisioned in a cloud-like manner to support the various connected IoT devices, their applications, and service providers.

2. The Need for Extending the Cloud

Most IoT service providers across different industries have acknowledged the latency issue and are building or using distributed clouds to collocate their services across different geographical areas to provide the required quality of service (QoS) and functionality. Telecommunications companies are in a unique position to solve this problem because central offices (COs) are usually located close to their customers' premises and thus close to IoT devices. In addition, COs can be transformed into clouds. Similar to the cloudlet concept, in which small-scale cloud datacenters at the edge of the Internet are used to support resource-intensive and interactive mobile applications,[2] this can extend IoT service providers' computing facilities to the network edge.

There are several initiatives to achieve this goal, mainly built around mobile-edge computing (MEC)—a network architecture concept that enables cloud-computing capabilities and an IT service environment at the edge of the cellular network.[4] Among these, the Central Office Re-architected as a Datacenter (CORD) initiative seems to be better suited for convergent telecoms, as it integrates network function virtualization (NFV) and software-defined networking (SDN).[5] CORD aims to reduce costs while bringing agility and refined control to the network. CORD's reference architecture is based on three pillars: commodity hardware, an SDN kernel to control the underlying switching fabric, and a virtualization management platform to create and control the virtualized functions.

However, we believe that to support the various IoT devices and applications of future cities and homes, the CO must be further re-architected as a cloud at the edge of the access network. This will transform the CO into a multitenant environment where IoT service providers can deploy elastic applications with a great degree of control.

Telefónica, one of the world's largest telecoms, is exploring this approach through the OnLife project, whose main goal is to design a future-proof technology stack that could bring the benefits of cloud computing and network programmability to the access network. OnLife's technological core is the CO datacenter (COdc), which builds on some of CORD's principles but takes its disruptive approach a step further by simplifying the implementation and introducing an open framework to deploy edge applications.

3. OnLife

The COdc's functional goals are twofold: first, it must support current residential services, such as Internet access, voice calls, and Internet Protocol Television (IPTV); second, it must allow the deployment of third-party edge solutions. While designing the COdc, we adhered to the following principles: use open source software and open hardware specifications, greenfield to avoid constraining new applications with current protocols, and maintain simplicity by not over-engineering an intrinsically complex system.

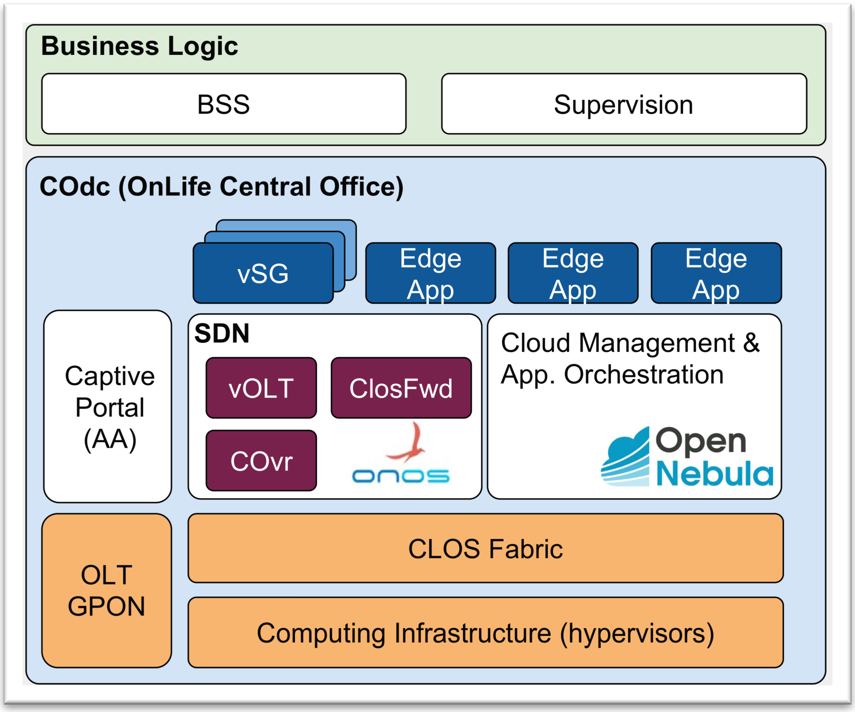

Figure 1. OnLife architecture. The business support system (BSS) in the upper layer provides the central office (CO) datacenter (COdc) with basic user authentication, authorization, and accounting capabilities. The software-defined networking controller, based on the Open Network Operating System (ONOS), is responsible for executing the networking logic that controls the Clos switching fabric in the CO. The cloud manager, based on OpenNebula, is responsible for managing the virtualized resources that implement the different NFVs and edge applications.

Figure 1 shows the main components of OnLife's architecture. In the upper layer, the business support system (BSS) provides the COdc with basic user authentication, authorization, and accounting capabilities. Interaction with Telefónica's business logic is performed through a custom captive portal that offers available edge applications (for example, remotely controlling the lighting in a house), connectivity, and additional services.

The SDN controller, based on the Open Network Operating System (ONOS; onosproject.org) and responsible for executing the networking logic that controls the switching fabric in the CO, is in the lower layer. There are two main network applications running in ONOS: vOLT and ClosFwd. The vOLT application reproduces the behavior of an optical line terminal (OLT) by redirecting traffic to the captive portal by default (where clients can consult with and hire different services), and switching the inbound traffic to the CO once the client is subscribed to the network. The ClosFwd application is in charge of internally forwarding the CO and creates the paths between the client, the virtual subscriber gateway (vSG; the virtual replacement of the customer premises equipment (CPE) that runs in a virtual machine (VM) and provides basic routing and filtering), and the various services. Both applications provide a RESTful API that is dynamically controlled by the cloud management platform (CMP), which is located at the same logical level of the SDN controller and built with OpenNebula (opennebula.org).[6]

OpenNebula, a lightweight and powerful CMP, is responsible for managing the virtualized resources that implement the different NFVs and edge applications. OpenNebula also interacts with the ONOS components to establish the network connectivity for each VM. Additionally, it provides the orchestration functionality needed to manage multiple-VM applications that might include interdependencies and elasticity rules to dynamically adjust the number of VMs based on the application load.

Finally, several virtualized components of the architecture are implemented as either VMs or SDN applications: a vSG; a CO virtual router (COvr; in charge of routing traffic to Telefónica's transport networks), and the edge applications to support IoT devices that are deployed in independent VMs in an isolated network. The idea behind edge applications is that third-party companies can develop their own appliances to implement the associated edge logic for the devices.

The OnLife architecture has been implemented in a proof-of-concept CO, based on compute nodes with 8 CPUs and 32 Gbytes of RAM, using a virtualized Clos fabric consisting of 4 x 2 leaf-spine OpenFlow switches and an emulated OLT. This setup allowed us to showcase a complete workflow from customer authentication to the deployment of associated edge applications. As test cases, we implemented basic connectivity applications (Internet access and video on demand) and a content delivery network (CDN).

4. Edge Applications

One of OnLife's main challenges is to make the CO available third-party edge computing applications, similar to the infrastructure as a service (IaaS) model, which opens the datacenter to external workloads. The ability to provide this edge-computing platform in a pay-as-you-go model (similar to IaaS) opens up avenues in both innovative use cases and business models.

However, given the CO's specific characteristics in terms of computational and storage resources—in addition to the environment's security constraints—a well-defined framework to develop such edge applications is required. For example, an application deployed at the network edge has to be rapidly reallocated when the user moves across the access network (for instance, from home to office). Therefore, we require edge applications to not store any state or persistency information at the edge. This includes the application logic itself, so edge applications also need to be able to autoconfigure. The autoconfiguration process is performed using specific information passed to the edge application upon bootup. The context could include user data, configuration parameters, or additional resources to install the application.

We envision a wide range of edge applications that will work with OnLife, from single-component instances to applications that require the deployment of multiple VMs. An edge application in the COdc provides this capability and includes deployment dependencies between the VMs. The interconnection of the VMs for each edge application happens in a separate private network. Figure 2 depicts the deployment of several applications in the COdc. IoT devices use the residential network to connect to the edge applications deployed in the COdc. Within the COdc, the IoT traffic is then forwarded to the target application through specific switching circuits in the Clos, which eventually may send the data to the ISP backbone network. Apart from IoT applications, basic services apps are also deployed for each customer and accessed in the same way (e.g. vSG for Internet access).

Edge applications are tied to the environmental conditions where IoT devices operate. A problem such as a traffic jam, a large event, or an emergency in a neighborhood might require allocating additional computational resources to the associated edge application. The COdc can increase (or decrease) the number of VMs considering application-specific performance metrics; for example, to add more VMs at specific times and dates or when the number of requests are above a given threshold.

The COdc also provides a well-defined API to manage edge applications. This API resembles the classical IaaS API to control a VM's lifecycle. The COdc uses the functionality exposed by OpenNebula and ONOS to deploy the edge applications and provide them with the features mentioned earlier.

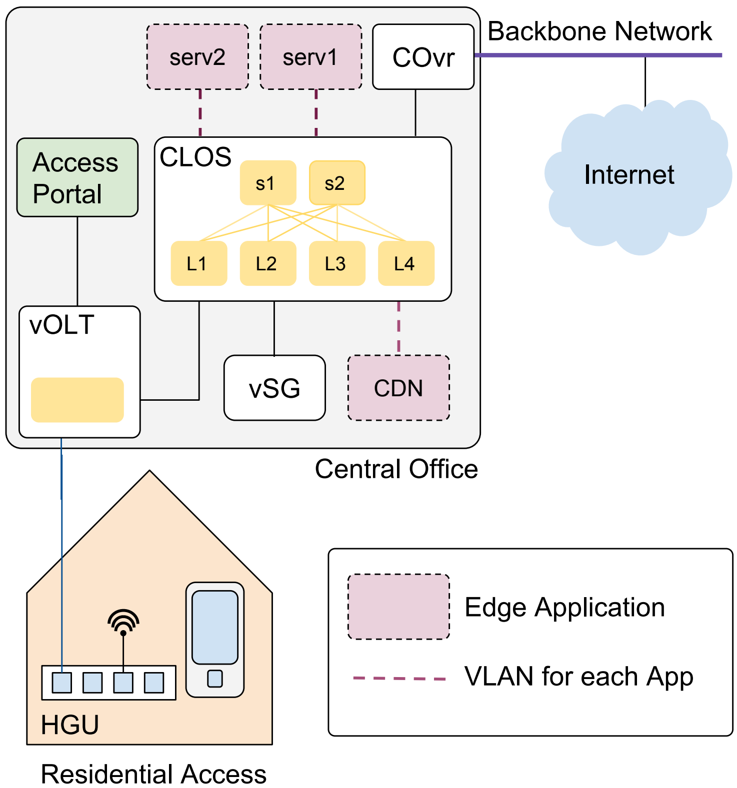

Figure 2. Service architecture for edge applications of the OnLife project. Applications are deployed as VMs by OpenNebula and interconnected in the Clos through specific switching circuits installed by ONOS. IoT devices connect to each application in the COdc through the residential access network. Together with the edge applications, standard services applications are also deployed in COdc (e.g. Internet access or VoIP) and accessed in the same way. CDN: content delivery network; HGU: home gateway unit

5. Moving Services from Customer Premises to Central Office

The initial functional and performance analyses made using the proof-of-concept and demo applications are very promising, and show us how to move other Telefónica solutions (currently deployed in the CPE or in expensive centralized locations) to the network edge. CPEs have limited capacity to host new IoT services such as internal security, access control, and energy management, which currently require the installation of additional physical equipment. The COdc allows us to host these services within a vSG built for the specific needs of the product offering. In particular, the following Telefónica solutions and services are being considered:

- Inmótica Hydra. This energy-efficiency enterprise solution helps customers manage and reduce their energy consumption and requires the installation of on-premise servers that occupy floor space and remote maintenance for each customer facility. Telefónica aims to remove this equipment and host all functionality within the COdc, leaving only the system's meters and sensors on premises.
- On the Spot. This retail commerce–oriented solution, which provides small businesses with in-store music, digital signage, and customer Wi-Fi, also requires the installation of on-premise servers that occupy floor space. On the Spot's maintenance cost and continuous software updates pose a challenge that the COdc is well suited to help resolve.
- FAAST Vulnerability. This residential solution, which provides protection against IoT threats, requires an agent in the CPE. However, most home CPEs don't have the capabilities to host it. The COdc is the only way to deliver this service without replacing or upgrading the residential CPE-installed base.

Making use of NFV and SDN, flexible datacenters built on commodity hardware can now be deployed in telecom Cos. Furthermore, it's been shown that the open source ONOS and OpenNebula projects can be adapted to different application scenarios and support new requirements, while allowing for fast and inexpensive prototyping.

Next steps for the OnLife project will consist of migrating and adapting the solution to a production-ready hardware infrastructure and replacing the emulated elements (such as vOLT) with actual equipment. We also aim to replace our current gigabit passive optical network (GPON) access technology with the latest XGS-PON and NG-PON2 technologies, without modifying the COdc software solution and at a reduced capital expenditure.

References

1. K. Xu, X. Wang, and W. Wei, "Toward Software Defined Smart Home," *IEEE Comm. Magazine*, vol. 54, no. 5, 2016, pp. 116–122.
2. M. Satyanarayanan et al., "The Case for VM-Based Cloudlets in Mobile Computing," *IEEE Pervasive Computing*, vol. 8, no. 4, 2009, pp. 14–23.
3. S. Clinch et al., "How Close Is Close Enough? Understanding the Role of Cloudlets in Supporting Display Appropriation by Mobile Users," *Proc. 2012 IEEE Int'l Conf. Pervasive Computing and Comm.* (PerCom 12), 2012; elijah.cs.cmu.edu/DOCS/clinch-percom-2012-CAMERA-READY.pdf.
4. M. Patel et al., *Mobile-Edge Computing*, white paper, ETSI, Sept. 2014; portal.etsi.org/portals/0/tbpages/mec/docs/mobile-edge_computing_-_introductory_technical_white_paper_v1%2018-09-14.pdf.
5. L. Peterson et al., "Central Office Re-Architected as a Data Center," *IEEE Comm. Magazine*, vol. 54, no. 10, 2016, pp. 96–101.
6. B. Sotomayor et al., "Virtual Infrastructure Management in Private and Hybrid Clouds," *IEEE Internet Computing*, vol. 13, no. 5, 2009, pp. 14–22.