Zero-Touch Provisioning of Distributed Video Analytics in a Software-Defined Metro-Haul Network with P4 Processing

Bogdan-Mihai Andrus[1], Silviu Adrian Sasu[1], Thomas Szyrkowiec[1], Achim Autenrieth[1], Mohit Chamania[2], Johannes K. Fischer[3], Stephan Rasp[4]

[1] ADVA Optical Networking SE, Fraunhoferstr. 9a, 82152 Martinsried, Germany; [2] ADVA Optical Networking SE, Justus-von-Liebig-Str. 7, 12489 Berlin, Germany; [3] Fraunhofer Institute for Telecommunications Heinrich Hertz Institute, Einsteinufer 37, 10587 Berlin, Germany;
[4] SeeTec AG, Werner-von-Siemens-Str. 2–6, 76646 Bruchsal, Germany
Author e-mail address: BAndrus@advaoptical.com

Abstract: We demonstrate automated network service provisioning and virtual network function orchestration with P4-based VNF acceleration. Zero-touch provisioning of distributed computing resources at the edge and central office is validated with a video analytics use case.

OCIS codes: (060.4250) Networks; (060.4510) Optical communications.

1. Introduction and Application

The increasing need to dynamically provision services in a cost-effective way, spanning multiple knowledge domains, technologies and administrative boundaries has driven the evolution of architectures and protocols for the operation of networks. Such services have grown in scale and complexity from conceptually simple voice and data connections in homogeneous networks within the scope and control of a single administrative entity. Future services require the allocation of heterogeneous resources with complex placement constraints and highly dynamic usage patterns. To address these new challenges, a new Metro-Haul architecture based on Software-Defined Networking (SDN) and Network Fuction Virtualization (NFV) has been introduced and described [1]. It complies with future service requirements which are derived from emerging trends in recent technology and market outlook surveys, network evolution and traffic forecasts.

Due to the system's complexity, we incorporate a Zero Touch Provisioning (ZTP) functionality in our proposed solution such that new hardware can be installed directly into the environment by provisioning and configuring it automatically [2]. This significantly eases arising scalability constraints when adopting a novel architecture. Offloading Virtual Network Function (VNF) processing [3] is another objective of this demonstration, which seeks to optimize computing resource utilization by moving stateless L2-L4 packet processing into dedicated programmable switching fabric platforms using the programming language P4 [4].

2. Overview and Low-Latency Video Analytics Use Case

While low latency has always been a desirable attribute, it is now becoming a must-have for next- generation networks especially on the edge segment, where a growing number of applications, (e.g., virtual and augmented reality, autonomous driving, drone control, tele surgery) require low latency to be viable. As a relevant example, video analytics can have a great impact on conventional Video Management Systems (VMS) by incorporating detection and alert algorithms. However, with areas like London, UK where the number of CCTV cameras is estimated at roughly 500 thousand, deployment, scaling and control of such extensive VMS is non-trivial.

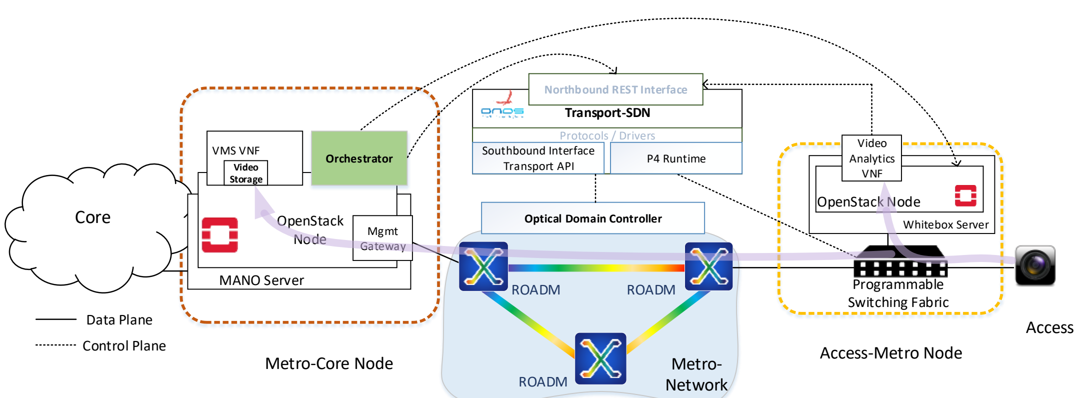

Fig. 1: Demo and architecture overview (VMS: Video Management System, VNF: Virtual Network Function).

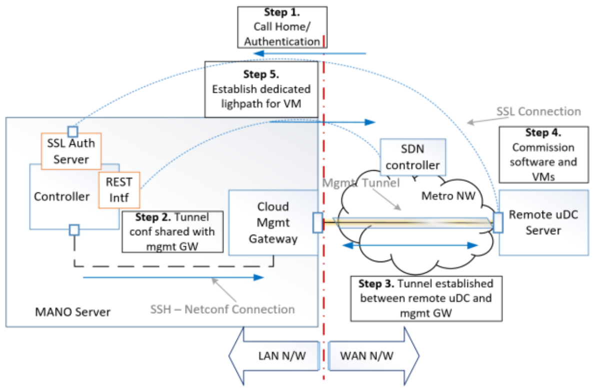

Fig. 2: Zero-Touch provisioning.

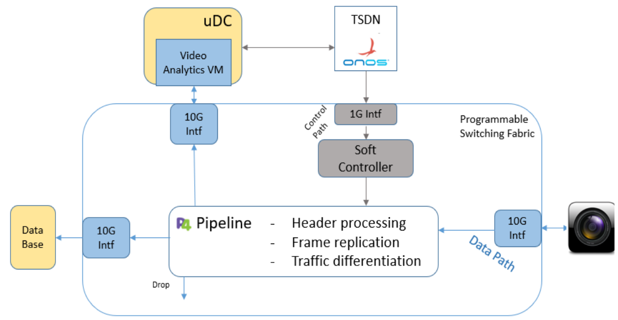

Fig. 3: Video stream processing.

The network architecture defined in the EU 5G-PPP project METRO-HAUL (Fig. 1) is designed to encompass distributed computing at the metro network edge to provide support for low latency services. Using generic Commercial off-the-shelf (COTS) servers and combining SDN and NFV solutions, ZTP deployment of micro Data Centers (µDCs) is achieved.

By focusing on a distributed video processing use case, as highlighted in Fig.1, we are able to offload part of the processing from a video analytics VNF using a programmable switching fabric by dynamically filtering and forwarding control traffic and video streams to desired destinations.

3. Component Description

Zero touch and authentication allow the operator to stage the deployment of new µDC servers and provision them in the network with minimum intervention. Fig.2 shows a simplified diagram of the zero-touch provisioning demonstration. Zero touch is triggered when the remote system is first powered on. Minimal config is present on the remote system such as secure management configuration. Using the management config, the system is capable of authenticating with the orchestrator. The cloud management gateway on the Management and Orchestration (MANO) server acts as a tunnel end-point termination for the remote system. Upon initial authentication, the controller VM is responsible for establishing the secure tunnel between the remote system and the management gateway. This is accomplished by sending the tunnel end point information to both the remote system and management gateway. Once the secure tunnel is established, the orchestrator can orchestrate the service VNFs, networks and resources inside the µDC. In the next step, when the video system has been established, the controller requests an additional lightpath through the optical network for the dedicated live video streaming to the metro-core node.

Fig. 3 visualizes the video stream processing that is offloaded from the VNF to the programmable networking hardware. Without the offloading, the video analytics VM would need to detect events and forward the video to the video storage, whenever an event occurs. The programmable switching fabric takes over the packet forwarding to the storage by replicating the video on demand and directing the stream towards the storage. This processing and output behavior of the pipeline is controlled by the video analytics VM through the open-source SDN controller ONOS.

The programmable switching fabric provides low latency packet processing and port mirroring functionality, which is activated whenever an event (e.g., object or motion detection) is detected by the analytics software. The configuration is controlled through the switch internal controller, which is part of the fabric, by changing entries of the match+action table in the P4 pipeline. In case of active frame replication, the P4 pipeline is additionally used for packet filtering and header processing. It classifies the incoming data from the camera into video and control packets. Based on the header information control packets are dropped and video packets are passed to the next match+action tables, in which the header processing takes place. At the end, these packets forwarded to the video storage.

4. Contributions and Innovations

The demonstration acts as a proof of concept for the previously described Metro-Haul architecture. It covers one metro-core node and one access-metro nodes interconnected by a transport network. The ZTP procedure deploying all required software components to the bare metal servers and configuring the optical network between the edge nodes represents the first integral part of the demo. The P4 switching fabric enables the video analytics software to offload network processing and to control the handling through the controller in the case of an event. This enables the VNF to dynamically adapt the pipeline and activate the frame replication on demand. The intermediate ONOS controller translates the demands to the protocol-independent P4 runtime API and sends them to the switching fabric.

This second part offloads parts of the computational and network overhead from the VM to the programmable hardware and reduces the end-to-end delay for the replicated stream.

5. Demo Workflow

The live demo presented at the conference venue uses a remote connection to access the lab and all hardware. Web interfaces and GUIs visualize the demo flow, which covers two phases: the deployment of the VNFs including the optical connectivity and the event-triggered network control by the VNF adapting the configurable stream processing.

The hardware comprises a video camera, two servers (HP Proliant DL360 Gen9, with 2x 10Gbps NICS, 128GB RAM, and 2x Intel(R) Xeon(R) CPU E5-2630 v4) representing the two data centers on both sides of the ROADM ring, consisting of three ADVA FSP 3000s. A programmable switching fabric (see Fig. 3) acts as a direct data path interconnection between the access-metro node, the video camera and the ROADM ring.

The first phase covers a zero-touch provisioning of the network functions, which includes the setup of the optical lightpath. The demo starts out with a preinstalled Management and Organization (MANO) central server using an OpenStack cloud installation hosting the NFV controller and the VMS, a bare metal server deployed on the access side and an unprovisioned ROADM ring as shown in Fig. 1. The VMS contains a central video storage for stashing remote camera feeds.

Upon the initial boot, the bare metal server initiates a “Call-Home” and establishes a management tunnel to the controller VNF following an authentication procedure (see Fig. 2). The orchestrator discovers the new device and triggers its provisioning by installing the cloud operating system, the preconfigured virtual machine (i.e., video analytics function) and establishes the virtual network connections. An optical lightpath is established through the metro ROADM-based network through the ONOS controller. As a result, connectivity is established between the VMS, the distributed video processing VM and the video camera.

The second part is covered by the programmable switching fabric that provides an offloading of computational and network communication resources for the VNF and thereby reduces the load on the VM. The main functionality implemented by the switching fabric is divided into two subtasks. First, the packet stream is replicated when an event is discovered. Second, special P4 processing and filtering procedures are applied to the replicated packet stream.

The included P4-pipeline covers multiple stages. After the replication is triggered, the P4 pipeline applies a filtering function to the stream that only forwards the video and drops the control traffic. The next step of the P4 pipeline creates the correct headers by setting the destination for the packets that are sent to the video storage. The input for the header generation is supplied by the controller and contains information such as the IP address of the host running the VMS. Afterwards, the stream is sent through the optical network toward the metro-core node and the storage, while the original stream is forwarded (untouched) to the video analytics VM. We display the live video content captured by the camera, using a remote desktop connection to the VMS.

6. Conclusion and Relevance for OFC

The proposed demonstration combines software-defined control with VNF management and orchestration in a realistic use case of low-latency video analytics. With increasing complexity of the software architectures, it is important to validate the potential benefit of automation and zero-touch-provisioning of network services in a complex system architecture. This demo is relevant to OFC audience as it features a live proof-of-concept demonstration of a proposed high capacity optical Metro-Haul network architecture capable of P4-based VNF offloading to a programmable switching fabric.

Acknowledgement

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No. 761727 (METRO-HAUL) and No. 762057 (5G-PICTURE).

References

[1] R. Casellas et al., “Metro-Haul: SDN Control and Orchestration of Disaggregated Optical Networks with Model-Driven Development,” in 20th International Conference on Transparent Optical Networks (ICTON), Bucharest, Romania, Jul. 2018, Mo.C3.5.
[2] Y. Demchenko et al., “Enabling Automated Network Services Provisioning for Cloud Based Applications Using Zero Touch Provisioning,” 2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC), Limassol, 2015, pp. 458-464.

[3] T. Duan et al., “Separating VNF and Network Control for Hardware-Acceleration of SDN/NFV Architecture,” in ETRI Journal, Volume 39, Number 4, August, 2017.

[4] P. Bosshart et al., “P4: Programming Protocol-independent Packet Processors,” SIGCOMM Comput. Commun. Rev. 44, 3 (July 2014), pp. 87-95.