Ref. Ares(2017)2619970 - 23/05/2017

Grant Agreement No. 619572

COSIGN

Combining Optics and SDN In next Generation data centre Networks

Programme: Information and Communication Technologies

Funding scheme: Collaborative Project – Large-Scale Integrating Project

Deliverable D2.6

Report on Successful Co-Integration of Data Plane Network Including 2nd Generation TOR Switch and POLATIS Optical Circuit Switch

Due date of deliverable: December 31, 2016
Actual submission date: December 20, 2016

Start date of project: January 1, 2014 Duration: 39 months

Lead contractor for this deliverable: Polatis

Project co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Executive Summary

This document summarises the progress made in WP2 relating to the integration of one of the two optical switching technologies included in the project with the TU/e TOR switch. This technology is optical circuit switching (OCS) realised by Polatis optical switches which are based on low insertion loss beam steering technology. The Polatis switches used in the COSIGN project have been augmented by the addition of an on-board OpenFlow agent to facilitate control of the switches by SDN controllers (such as OpenDaylight) via the OpenFlow wire protocol. Since this was last reported in D2.1 [14], the agent has been much improved and this is summarised in this document and reported in more detail in D3.5 [16].

The TU/e TOR switches have been upgraded since the 1st generation TORs described in D2.2 [7]. A new ASIC supporting 128 channels at 10Gbps has been included in the 2nd generation TOR and real mid-board optical engines have been integrated in close proximity to the ASIC. The result of this additional effort is a further size reduction of the single switch and further potential power reduction. In section 2 some of the impact of using the 2nd generation TOR for large scale deployment in datacentres is discussed. The TORs also fully support SDN control via an integrated OpenFlow Agent provided by Broadcom and adapted to the ODL controller by PhotonX Networks. Since the TORs have been designed using mid-board optical engines that only support multi-mode fibre transmission while the other optical switching solutions rely on single-mode fibre transmission systems, additional HW adaptors converting MM to SM optical links have been realized.

This integration has allowed the COSIGN project to realise an SDN controlled, highly performant, hybrid packet-OCS demonstrator data plane that has been used to great effect by two of the COSIGN application use cases (VApp and VDC) as reported in other COSIGN deliverables. This data plane is now being augmented with new optical technologies such as multiple and hollow core fibres and fast optical TDM switching in preparation to support the COSIGN long term demonstrator.

Document Information

Status and Version:	0.6	
Date of Issue:	20/12/2016	
Dissemination level:	Public	
Author(s):	**Name**	**Partner**
	Michael Enrico	POLATIS
	Oded Raz	TU/e
	Chris Jackson	University of Bristol
	Michael Galili	DTU
Edited by:	Michael Enrico	POLATIS
Reviewed by:	Michael Galili	DTU
	Karel Van de Plassche	PHOTONX
Checked by :	Sarah Ruepp	DTU

Table of Contents

1 Introduction

Realizing an integrated data plane with its associated control plane elements such as OpenFlow agents and SDN controller plugins is key to the COSIGN project being able to demonstrate the advances it has made in many of its work packages and tasks. This is achieved through the demonstration of applications that are pertinent to data centre networking and infrastructure management that can also make good use of the optical technologies being further developed in COSIGN.

This report explains how the top of rack (TOR) packet switching technology from TU/e and PhotonX has been further developed during the project and integrated with the optical circuit switching technology from Polatis to realise a data plane that has then formed the heart of the medium term testbed and demonstrator platform. This has then been used to support the development and subsequent demonstration of two applications – VApp and VDC. The control plane software elements, although equally important in realizing the COSIGN goals are only summarised in this deliverable and reported in much more detail in other deliverables ([16], [17]).

1.1 Reference Material

1.1.1 Reference Documents

[1] Abts and J. Kim, "High Performance Datacenter Networks: Architectures, Algorithms, and Opportunities," Synth. Lect. Comput. Archit., vol. 6, no. 1, pp. 1–115, Mar. 2011
[2] "Annex A6: 120 Gb/s 12x Small Form-factor Pluggable (CXP)"
[3] "Finisar - 10G Board-mount Optical Assembly." [Online]. Available: https://www.finisar.com/optical-engines/fbotd10sl1c00
[4] "Broadcom - BCM56850 Series." [Online]. Available: http://www.broadcom.com/products/ethernet-communication-and-switching/switching/bcm56850-series
[5] "Finisar - 25G Board-mount Optical Assembly." [Online]. Available: https://www.finisar.com/optical-engines/fbotd25fl2c00
[6] Broadcom, "Broadcom - BCM56950 Series." [Online]. Available: https://www.broadcom.com/products/Switching/Data-Center/BCM56960-Series
[7] https://www.broadcom.com/products/ethernet-connectivity/software/of-dpa
[8] Public COSIGN Project Deliverable D2.2 "Report on 1st generation TOR switch operation" http://www.fp7-cosign.eu/wp-content/uploads/2015/07/cosign_d2_2_final1_submit.pdf
[9] "Open Rack." [Online]. Available: http://www.opencompute.org/projects/open-rack/
[10] "Open Compute Project." [Online]. Available: http://www.opencompute.org/
[11] "COMSOL Multiphysics." [Online]. Available: https://www.comsol.com/
[12] M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data center network architecture," in Proceedings of the ACM SIGCOMM 2008 conference on Data communication - SIGCOMM '08, 2008, vol. 38, no. 4, p. 63
[13] B. Heller, S. Seetharaman, and P. Mahadevan, "ElasticTree : Saving Energy in Data Center Networks," 2006
[14] http://archive.openflow.org/wk/images/8/81/OpenFlow_Circuit_Switch_Specification_v0.3.pdf
[15] Public COSIGN Project Deliverable D2.1 "Integration of OpenFlow SW Interface with POLATIS System and Venture 4x4 OXS" http://www.fp7-cosign.eu/wp-content/uploads/2015/09/COSIGN_Deliverable2.1_V2.0_Final.pdf
[16] **Consortium-only** COSIGN Project Deliverable D3.4 "SDN Controller for DC Optical Network Virtualization and Provisioning Software Prototypes: Final Release" https://share.dtu.dk/sites/COSIGN_91600/Shared%20Documents/Final%20Deliverables/WP3/COSIGN-D3.4-v1.0_Final_Submit.docx?Web=1

[17] **Consortium-only** COSIGN Project Deliverable D3.5 “Integrated SDN Framework for Optical DCNs” https://share.dtu.dk/sites/COSIGN_91600/Shared%20Documents/Final%20Deliverables/WP3/COSIGN-D3.5_final.docx?Web=1

[18] Public COSIGN Project Deliverable D5.2 “Demonstrator Result of Data Plane and SDN Environment Integration” (currently in preparation)

[19] Open Networking Foundation “OpenFlow Switch Specification Version 1.4.1 (Protocol version 0x05)” (ONF TS-024) https://www.opennetworking.org/images/stories/downloads/sdn-resources/onf-specifications/openflow/openflow-switch-v1.4.1.pdf

[20] https://datatracker.ietf.org/doc/draft-ietf-netconf-restconf

[21] https://github.com/Broadcom-Switch/OpenNSL

1.1.2 Acronyms and Abbreviations

Most frequently used acronyms in the Deliverable are listed below. Additional acronyms can be specified and used throughout the text.

DC	Data Centre
DCN	Data Centre Network
DCIM	Data Centre Infrastructure Management
DMZ	Demilitarized Zone
IaaS	Infrastructure as a Service
IETF	Internet Engineering Task Force
ISO	International Organization for Standardization
IPS	Intrusion Prevention System
MOOC	Massive Open Online Course
MPLS	Multi-Protocol Label Switching
NaaS	Network as a Service
NIST	National Institute of Standards and Technology
OSS	Operational Support Services
PaaS	Platform as a Service
PCI DSS	Payment Card Industry’s Data Security Standard
PE	Provider Edge
PoD	Point of Delivery
RAS	Reliability, Availability, Serviceability
RFC	Request For Comments
SaaS	Software as a Service
SDE	Software Defined Environments
SDN	Software Defined Networking
SLA	Service Level Agreement
TCAM	Ternary Content-Addressable Memory
TCO	Total Cost of Ownership
ToR	Top of the Rack
VDC	Virtual Data Centre
VLAN	Virtual LAN
VPN	Virtual Private Network
VRF	Virtual Routing/Forwarding

1.2 Document History

Version	Date	Authors	Comment
01	24/11/2016	See the list of authors	TOC first draft
02	09/12/2016		First draft (incomplete)
03	13/12/2016		First draft (almost complete)
04	15/12/2016		First draft for internal review
05	18/12/2016		Second draft for internal review
06	20/12/2016		Version 1 for submission

2 Hardware Development

2.1 2nd generation TOR with mid-board optical engines

2.1.1 Design work on TOR with mid-board optical engines

In order to minimize the amount of optical transceivers required to implement a mega data center[1] optical network, optical transceivers with the largest possible amount of ports are desirable. Highest port count pluggable transceivers available have 12 ports [1],[2]. The same is true for the highest port count on board optics (OBO) transceivers [3]. Figure 1 is a picture of the 2nd generation TOR designed and fabricated by the TU/e. It includes eleven 12x10G OBO transceivers surrounding the 128x10G state-of-the-art electronic switching ASIC [4] with a total of 1.28 Tbps, covered by a fan in the image. The control plane processor, attached in the top right corner of the board, manages the data plane switching ASIC thanks to the decoupling between control and data planes provided by Software Defined Networking (SDN).

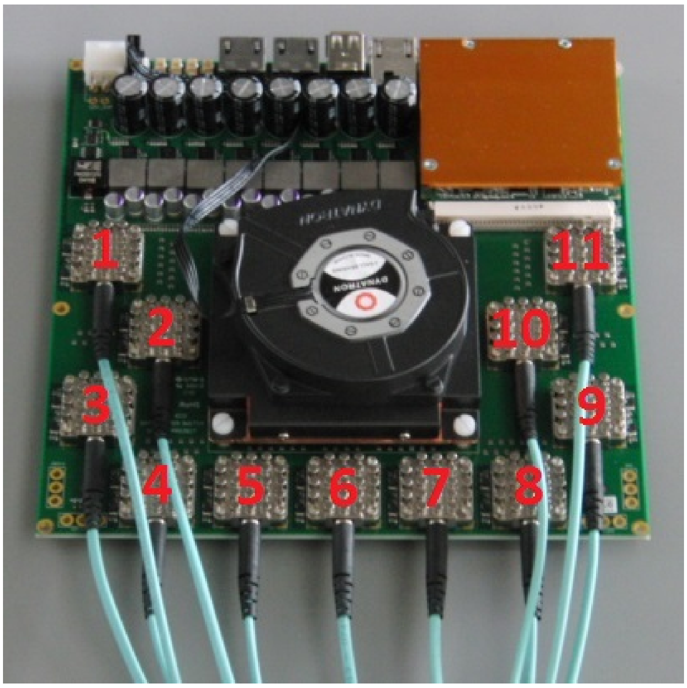

Figure 1: Implemented hybrid switch

The rest of the board, in the top left section, includes the required interface connectors, and the voltage converters generating the internal voltages. Our research shows that the reduced amount and small size of the OBO transceivers, closely placed to the switching ASIC, allows very compact designs, like our proof-of-concept implementation achieving a 20cm x 20cm size. A similar approach could be followed with the same-size 25G version of those devices [5],[6] to reach 3.2 Tbps. The ASICs and optical modules needed to achieve this target where not available at the time the design of this 2nd generation TOR was launched (see MS9).

[1] The term “mega data center” is in general associated with data centers operated by the likes of Google, Facebook, Microsoft and Amazon. These are usually based on the concept of Warehouse Scale Computers and often make use of compute pods (roughly the size of a ship container with several racks installed in them).

2.1.2 Multi-mode to single mode adaptors

In order to support data plane connectivity between the MMF based TOR and the SMF based Polatis switch, we have designed and built special optical convertors between multimode and single mode fibre based optical modules. Such a module, which can convert up to 8 multimode channels into 8 single mode channels, can be seen in Figure 2. The 8 channels are currently only supported in the O-band (1270-1330nm). In addition, we have also created a convertor from CXP to CFP2 modules which are being prototyped by a start-up company from the TU/e (Effect Photonics).

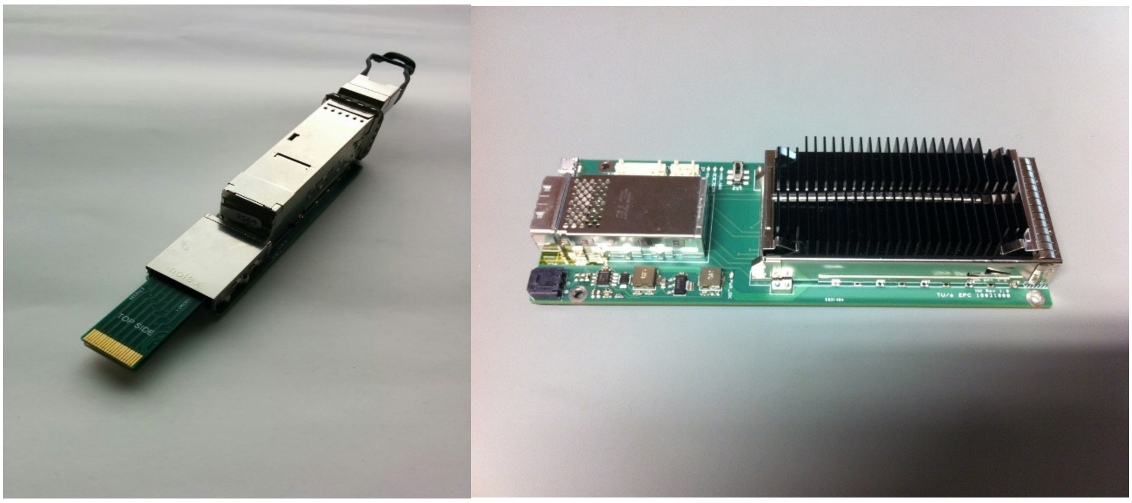

Figure 2: (Left) MMF to SMF conversion using QSFP;(Right) MMF to SMF convertor using CFP2

2.1.3 Software support for data plane integration

This has not changed over that previously reported in D2.2. In summary, the switch uses the Broadcom OF-DPA software [7] as agent. OF-DPA provides an Open-Flow 1.3.4-complient agent, as well as a convenient API for debugging. The agent has been slightly enhanced with better syntax parsing for the two COSIGN use-cases, but is mostly identical to the software reported in D2.2 [8].

2.1.4 Multiplying front panel bandwidth

Figure 3 shows the packaging of four 2nd generation TORs into a 19” rack case. Being only limited by the case size, our proof-of-concept design may allow also the packaging of six switches in the bigger OpenRack [9] unit of OpenCompute [10].

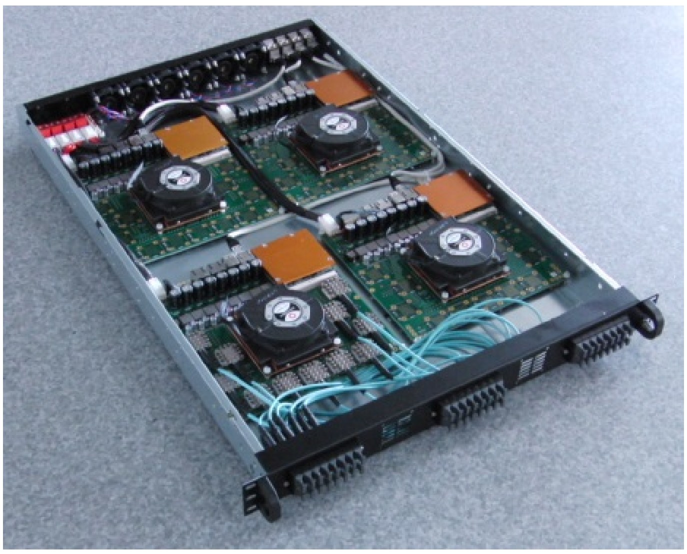

Figure 3: Four switches packaged into a 19” rack unit

This approach of packaging several hybrid switches into a single rack unit effectively decouples switching ASIC and front-panel optical bandwidth, resulting in a front panel port and bandwidth density several times higher than the underlying switching ASIC characteristics. With our proof-of-concept design and packaging, the optical front panel density scales up to 512 / 768 ports and 5.12 / 7.68 Tbps bandwidth, accessible through the MPO connectors provided by the 4 / 6 switching ASIC’s and 44 / 66 OBO transceivers packaged into the 19” / OpenRack case. The number of ASIC’s, transceivers and ports would be the same when implementing the 25G version of the hybrid switch, but in that case the bandwidth would scale up to 12.8 Tbps / 19.2 Tbps.

2.1.5 Thermal management of 4-in-1 switch box

By placing several hybrid switches into a single rack unit, thermal management becomes more critical, especially because the performance and reliability of passively cooled optical transceivers is highly dependent on temperature. Figure 4 includes the thermal simulation [11] results, including temperature and air velocity of the case with a six fans configuration, predicting a maximum temperature of 35.5° C assuming an air in-take temperature of 20° C. This estimated low temperature is the results of several key changes made possible thanks to the packaging concept and use of OBO. The main modification consists of removing the power supplies from the rack case, and thereby getting rid of the excess heat they generate. This decoupling, suggested by OpenCompute, allows power distribution network designers to better scale the data center power distribution system by sharing power supplies units between several devices.

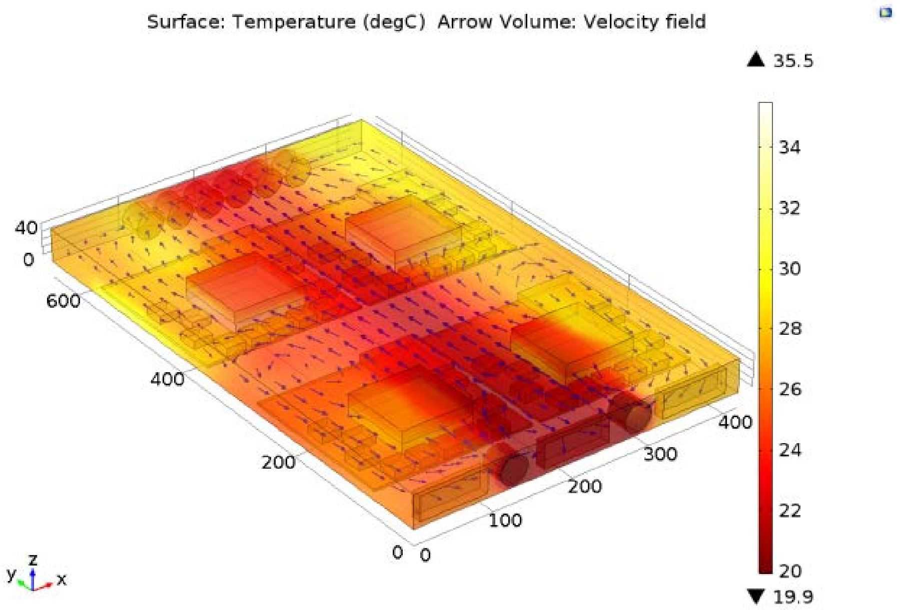

Figure 4: Thermal simulation results

In addition, without the power supplies, it is possible to install more fans in the back panel for extra forced cooling. In our system we chose to place six fans. To improve things even further the front panel has some free area for extra ventilation due to the shift to OBOs. All of these changes mean that good airflow and heat removal of the rack unit are obtained. Finally, another important advantage is that the OBO modules are distributed all around the rack case, and not concentrated in the front panel (as standard pluggable modules would), so they receive additional benefit from the improved airflow around them in the case.

The simulation results are verified experimentally by direct reading of the temperature sensor included in the transceivers during a set of experiments. This set of experiments includes several steps: from an initial state where only the control plane boards are running, to a fully loaded state where all ports are enabled and transmitting or receiving packets. Figure 5 presents the minimum and maximum registered operational temperatures for each one of the eleven optical transceivers of the top left board, numbered counterclockwise as shown previously in Figure 1.

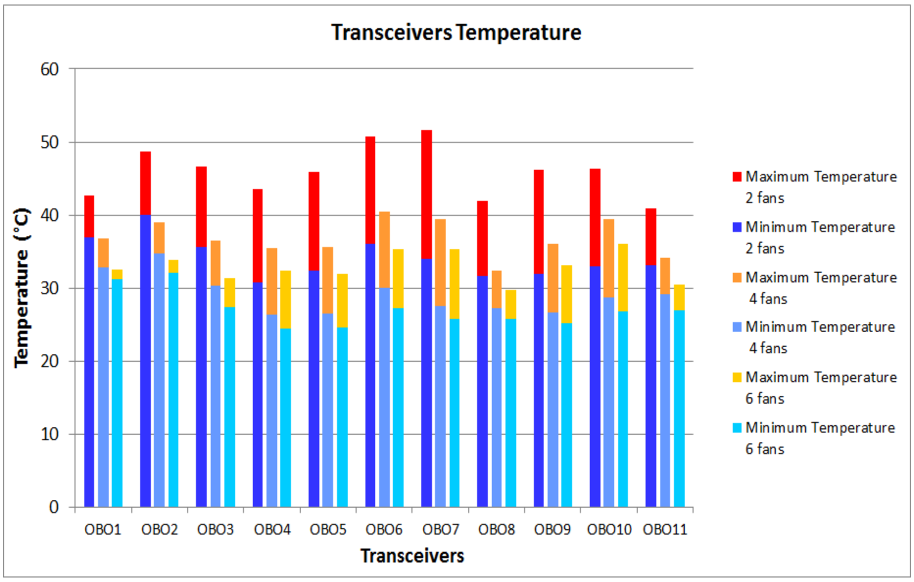

Figure 5: Measured temperature of the top left board transceivers

In the chart, the first, second, and third column of each transceiver correspond to experiments with two, four, and six fans, respectively. These results confirm that every extra pair of fans reduces the thermal stress in the transceivers, not only by reducing their maximum operating temperature, but also by reducing the temperature range variation. For instance, transceiver six has a maximum temperature of 50.7° C / 40.5° C / 35.2° C, a minimum temperature of 36° C / 30° C / 27.3° C with 2 / 4 / 6 fans, resulting in operating ranges of 14.7° C / 10.5° C / 7.9° C respectively. The measured maximum operating temperature of the transceivers is 51.6° C with 2 fans, 40.5° C with 4 fans, and 36° C with 6 fans. Finally transceiver six and seven are placed in the worst thermal places of the board: first, because the fan covering the ASIC interrupts the cooling air flow going from the openings in the front panel to the hot-air extracting fans of the back panel; and second, because they are in the hot air path extracted by the fan covering the ASIC's of the bottom boards.

2.1.6 Software controlled power management

By taking advantage of the control plane processor included in the switches, it is possible to configure and operate the data plane hybrid switches included in the rack unit in different operational modes, adjusting the optical front panel capacity and power consumption of the rack unit according to the utilization scheme of the data center optical network. Figure 6 compares the total power consumption of the 512 optical ports 19" rack unit under different operating conditions. For completeness this analysis is repeated for 3 different fan utilization modes. Every line is composed of eight segments, each one of them corresponding to a different operational mode. The first segment corresponds to the lowest power consumption of the rack unit, and it is achieved when only the four control plane boards are running to maintain communication with the data center central controller, but all the switching ASIC's and the optical transceivers are powered off. The power consumption for the whole system in this state is as low as 25.8W. This mode is mainly interesting for software defined data center networks where it is desired to keep the control plane network always on, and under certain conditions, to power down sections of the data plane network to save in power.

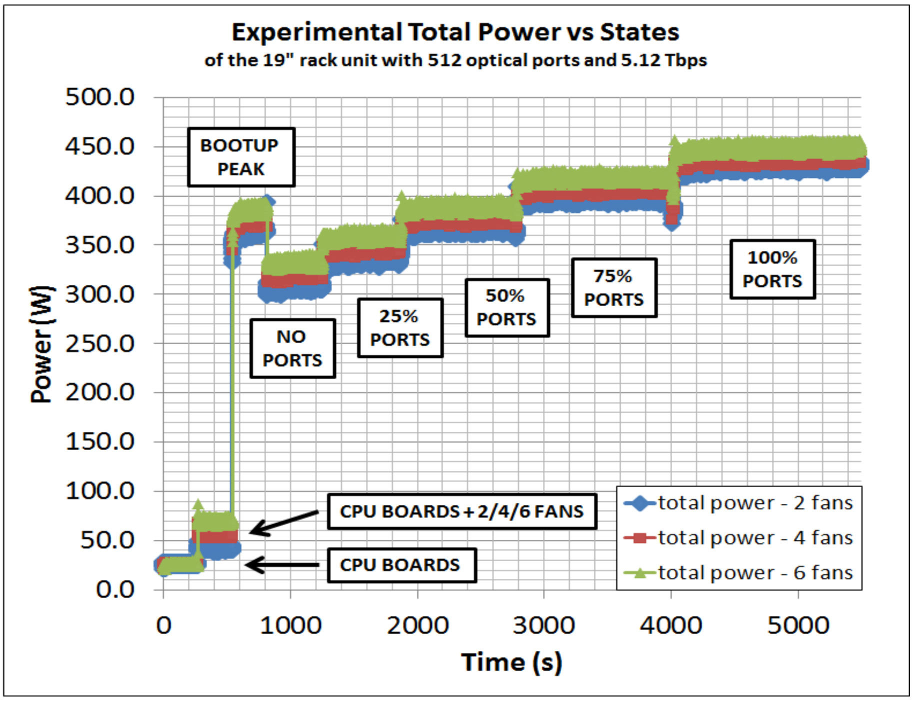

Figure 6: Measured power consumption of the rack unit under different operational conditions

The second segment of the three lines shows the power increase when enabling two (43.4W), four (63.9W) or six (83.0W) fans respectively. The third section shows the boot up peak produced when powering up ASIC and transceivers. Finally, the fourth to eighth segments display the power consumption of the system when the OpenFlow controller is running, for configurations with no ports, 25%, 50%, 75% and 100% enabled (and increasingly loaded) ports. For instance, with six fans, the total power consumption increases from 341.3W with no ports, to 367.4W with 25% ports, 400.6W with 50% ports, to 427.3W with 75% ports, and to 457.2W with 100% ports. As expected, the six fans configuration produces the highest power consumption of the system, but also, as shown before during the thermal analysis of the system, it provides the best thermal behavior of the system.

2.2 Impact at data centre level

Using the compact switch and the integrated 4 in 1 1RU box we now examine the impact on the entire data center network. The analysis is based on a 3-level fat-tree network topology built with 128-port switches. Fat-tree topology is widely used in data centers since, among other relevant features, it allows to connect any number of servers using switches of any number of ports by adding layers of switches [1],[12]. Point A of Figure 7 represents a data center with 131072 servers, requiring 5120 switches.

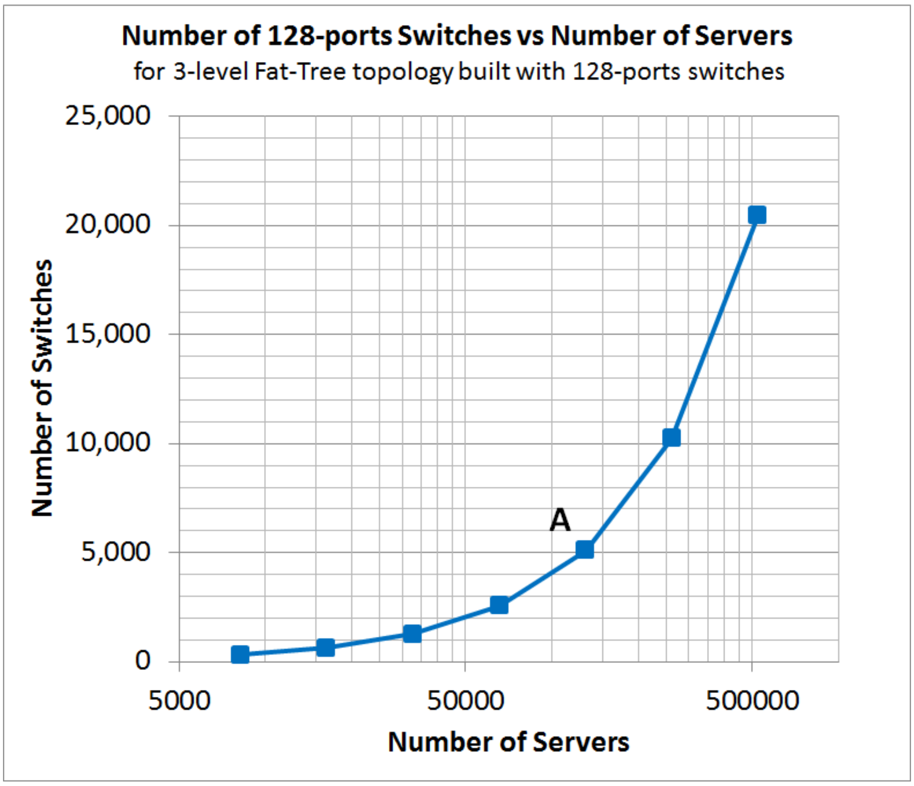

Figure 7: Number of switches for a 3-level, fat-tree topology implemented with 128-ports switches

Fig. 8 shows the possible savings in transceivers, the most costly part of the switch boxes, when moving to OBO transceivers which often, as is the case for our design, offer 12 channels in one transceiver module. The number of optical transceivers needed to implement the topology represented by point A of Figure 7 is computed for QSFP/QSFP28 transceivers, with four channels, and for OBO transceivers with 12 channels. By using 12-port transceivers, the number of optical transceivers is reduced by 65%, from 163840 QSFP / QSFP28 (point B) to 56320 OBO transceivers (point C). This is due to the higher port density of OBO transceivers.

The same reduction in number of transceivers could be achieved by using the CXP [2] front panel transceiver, also with 12 ports at 10G. However, since a fully populated front panel allows a maximum of 24 CXP transceivers, only two switches and 256 ports can be packaged into a single rack case.

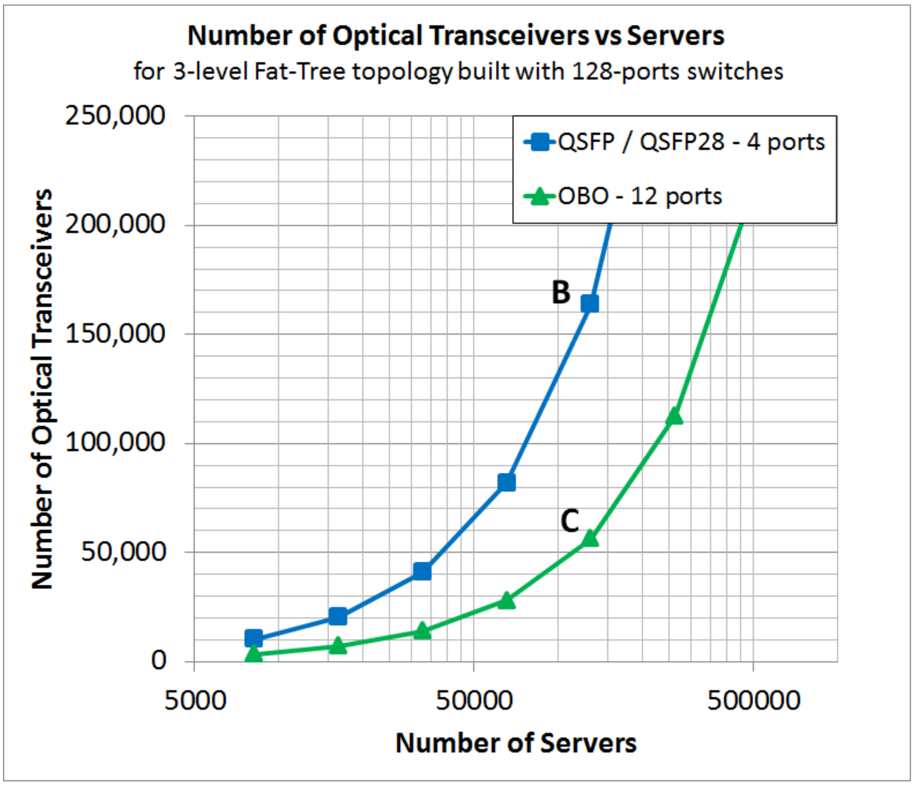

Figure 8: Number of optical transceivers for

a 3-level, fat-tree topology implemented with 128-ports switches

Figure 9 depicts the number of racks required to package the 5120 switches of point A, assuming 40 rack units per rack. By having more compact switches using OBO it is possible to pack more switches into each rack leading to a large reduction in the number of racks. For instance, the number of racks is reduced by 75%, from 128 racks (point D) to 32 racks (point E), when packaging 4 switches in a 19” rack unit. In case of packaging 6 switches in an OpenRack unit, only 22 racks are required, achieving around 83% reduction (point F). This is due to the higher switch density per rack unit.

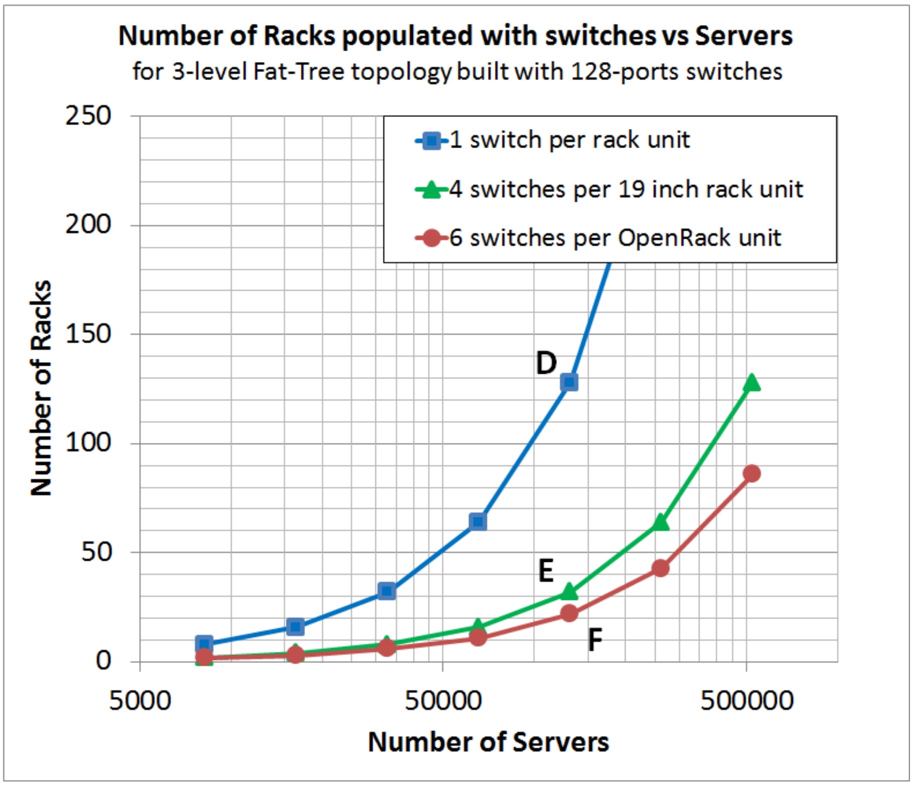

Figure 9: Number of racks populated with switches for

a 3-level, fat-tree topology implemented with 128-ports switches

Finally, Fig. 10 compares the total power consumption of the network fabric considering three different approaches. In the first approach, all switches are permanently fully on, so the consumption of the switches is always maximum, regardless the actual utilization of the data center. Point G represents this situation for the mentioned example, where 5120 switches consume a total of 585.2kW, according to the experimental measurements of Fig. 6. The second and third approaches consider power saving schemes where the data center central controller commands the control plane boards of the switches to disable (total or partially) the data plane infrastructure according to the utilization of the data center. In the second approach, all the switches are also permanently on, but the amount of enabled ports in switching ASIC and transceivers is configured according to the utilization of the data center. For instance, point H represents the data center with 25% utilization where 5120 switches are running with 25% of the ports enabled, reducing the power consumption by 19.6% to 470.4kW. Finally, in the third approach, some switches are kept in the lowest power operational mode, with only the control plane boards running. The remaining switches are fully powered on, with 100% enabled ports. For example, point J represents a data center with 25% utilization where 3840 switches are in low power mode, and 1280 switches have 100% of the ports enabled. This approach reduces the power consumption by 70.7% to 171.2kW. we can conclude that in case that the data center topology and application are allowed to operate in different modes according to the actual utilization of the infrastructure, the third approach provides the best power savings, since the effect of disabling ports in ASIC and transceivers of the second approach is limited. It must be noted that the idea of powering off sections of the data center to save power is not novel [1],[13] , although we follow a different approach here since the devices are not completely powered off, which requires long boot up times. With our approach, the control plane network remains always on, and only a fraction of the data plane is disabled on demand.

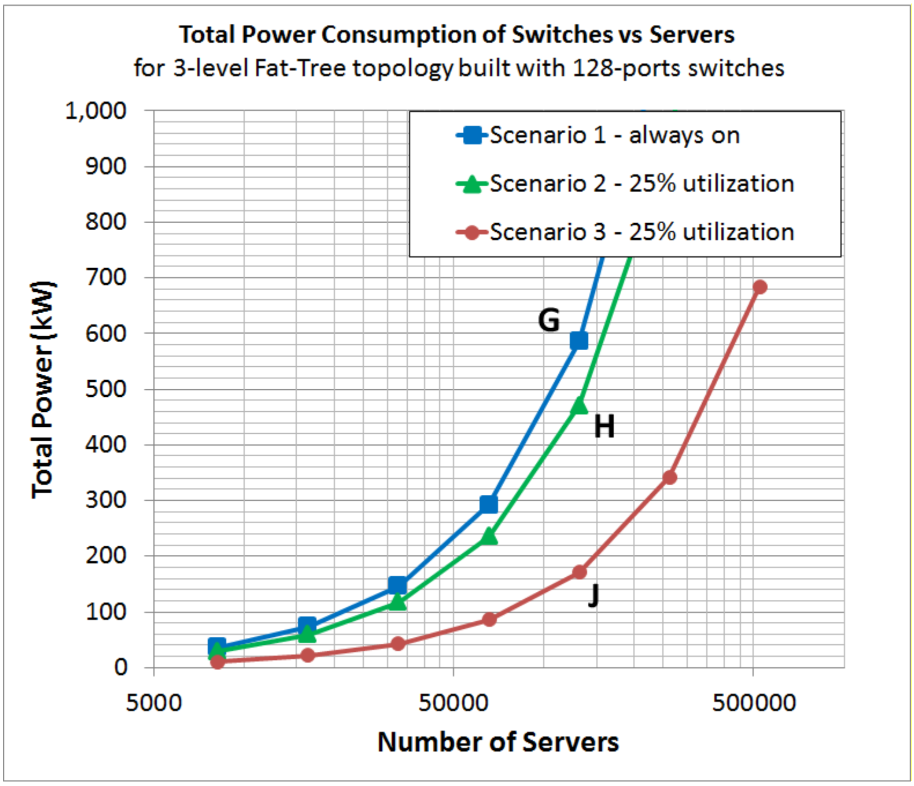

Figure 10: Total power consumption of the hybrid switches for a 3-level, fat-tree topology implemented with 128-ports switches

All those results imply an important cost reduction for data center operators through reduction in capital expenditure (CAPEX) and operational costs (OPEX). The significant decrease in the floor space occupied by switches enables smaller data centers. Besides, the reduced number of transceivers, which are the main drivers of the hybrid switches cost, and the reduced size of the boards, also bring the cost down of those devices. Finally, if the data center application and topology allows to power down (totally or partially) some devices depending on utilization in order to save power, best results are achieved by powering down the data plane in a fraction of the switches, and keeping the remaining switches fully on.

2.3 Polatis Optical Circuit Switches in COSIGN

A number of optical circuit switches have been supplied by Polatis to COSIGN partners for the purposes of developing a number of elements of the project - namely: (i) development of the OpenFlow agent for the Polatis switch as used in COSIGN, (ii) integration into the OpenDaylight (ODL) SDN controller, (iii) creation of the testbed and demonstrator setups, (iv) development and testing of the medium and long term demonstrators.

Independently of the COSIGN project, the University of Bristol have a number of 192x192 Series 6000 Polatis switches which have been used to develop the improved OpenFlow V1.0 + V0.3 circuit extensions and the ODL integration. An example of such a switch is shown in Figure 11. This particular switch uses discrete LC/APC connectors on the front panel. The rear view shows the redundant network interface cards (through which the management and control plane interfaces are realized), fans and power supplies.

Figure 11: Polatis Series 6000 192x192 port switch (front and rear panel views)

Two Series 6000 switches have been provided by Polatis to COSIGN partners DTU and TU/e for the purposes of constructing the medium and long term testbed and demonstrator hardware platforms. DTU have a 24x24 port switch as shown in Figure 12

Figure 12: Polatis Series 6000 24x24 port optical circuit switch with LC/UPC connectors (front and rear panel views)

TU/e have a 48x48 port switch such that as shown in Figure 13 (which shows two variants – one with high-density LC/UPC connectors and one with MTP connectors):

Figure 13: Polatis Series 6000 48x48 optical circuit switches with LC/UPC connectors (top) and MTP connectors (bottom)

3 Control plane integration

3.1 OpenFlow Agent Development

3.1.1 Polatis OpenFlow V1.0 + Circuit Extensions

To enable the Polatis device to be controlled and monitored by the SDN controller, an embedded OpenFlow agent is used. This agent uses OpenFlow V1.0 + V0.3 circuit extensions [14]. In addition, it supports proprietary optical circuit monitoring statistics.

Since the last reporting in D2.1 [15], the OpenFlow agent has evolved quite substantially. Originally the agent communicated internally with the switch control systems via the existing TL1 interface. As was planned at the time, the University of Bristol team improved the agent making it faster and more efficient by integrating it directly with the Polatis user services API (Figure 14) and eliminating redundant code (reducing the size of the agent by a factor of about three).

This agent is described in detail in D3.4 [16] and D3.5 [17]. The V0.3 circuit extensions and those required to support the optical circuit monitoring statistics had to be reflected in the OpenFlow support of the SDN controller as we will describe below.

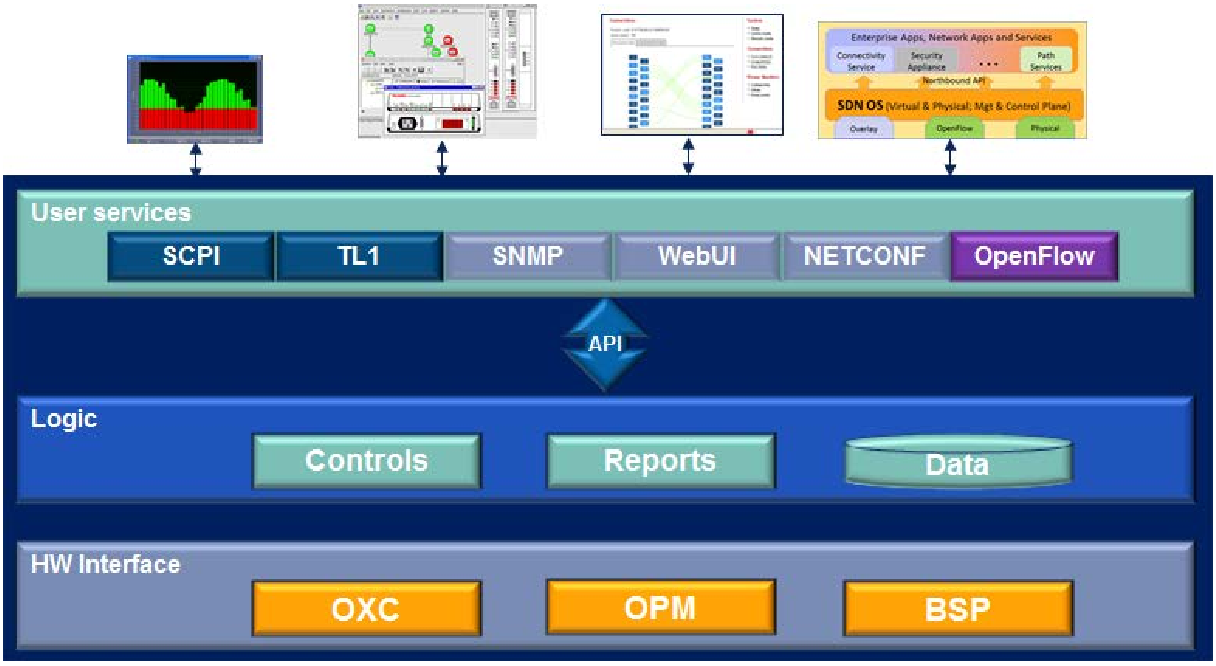

Figure 14: Polatis Optical Switch User Interface Structure

3.2 SDN Controller Integration

The OpenDaylight Lithium SDN controller (ODL) required a number of extensions in order enable control and monitoring of the Polatis device. This required the extension of ODL’s OpenFlow v1.0 protocol library with the v0.3 circuit extensions, with relevant extensions to the statistics monitoring modules. In addition, the agent available for the Polatis device underwent further development to enable statistics replies to ODL.

Further description of the extensions to both ODL and the Polatis agent can be found in D3.4 [16] and D3.5 [17]. Parts of the integration work were verified in the medium term demonstrator (ODL control of Polatis) which was present at ECOC 2016 executing the VDC and VApp use cases as will be described in the forthcoming D5.2 [18]. The statistics requests and responses will be verified by the long term demonstrators.

4 Physical Data Plane Integration

As a combined integration and dissemination activity the COSIGN project prepared and executed a live demonstration at the European Conference on Optical Communication (ECOC) in Düsseldorf in September of 2016. The COSIGN demonstrator was hosted by Polatis at their stand on the conference exhibition floor. The demonstrator setup was designed and implemented in a way to simultaneously showcase key features and benefits of the COSIGN mid-term architecture while at the same time being compact enough to allow transport and demonstration on the conference show-floor.

Consequently, the physical demonstrator was scaled to the minimum size still allowing for demonstration of the benefits of both VDC and vApp use cases. The main purpose of the demonstrator was to show network components and functionalities developed in the project. The demonstrator emulated three server racks in a datacentre by having three TOR switches each of them connected to a single high-performance server. A network was then built from these three TORs and a Polatis fibre switch. A fourth high-performance server was included to host controller and orchestrator software and a multi-channel Ethernet tester from Xena Networks was included to produce additional traffic in the network. The schematic in Figure 15 shows the data plane setup of the demonstrator.

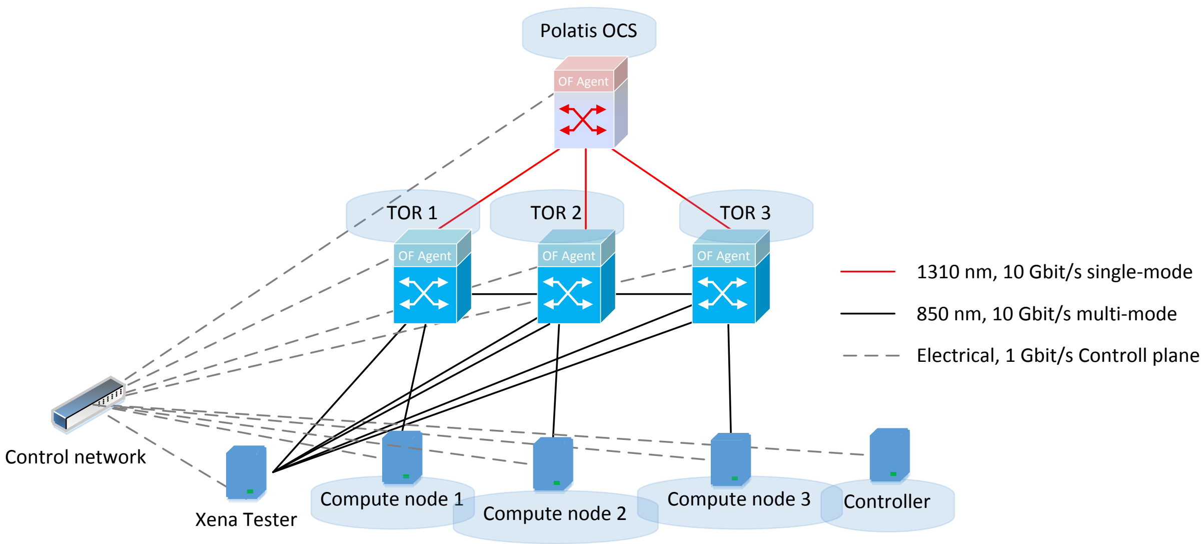

Figure 15: Data plane configuration of the ECOC demonstrator

Hardware configurations used in the lab setup and medium term demonstrator are as follows:

Component	Details
Servers	IBM X3690 X5 2xIntel Xeon E7-4807 6-core CPUs 256GB Memory dual 10G+dual 1G NICs
TORs	TU/e Generation 1 Ethernet switch 64 ports 850 nm multimode CXP interface

OCS	Polatis Series 6000 24 ports (48 fibres) Single mode fibre interfaces
Ethernet tester	Xena Bay 5x 10Gbit/s interfaces in use All with 850 nm multimode SFP+ transceivers

Table 1: Demonstrator hardware configurations

In order to provide connectivity between the TU/e TOR switches and the Polatis OCS a 10Gbit/s media converter is employed to convert between the multimode interfaces of the TOR switches and the single mode fibre in the OCS. The media converter (Fiber24 CONV) is equipped with pairs of 1310 nm single mode and 850 nm multimode transceivers performing transparent conversion between wavelengths and fibre types. This solution is employed in the three connections between the TORs and the OCS. It has no impact on the performance of the system.

The setup is mounted in two 20U rack cabinets to provide an easier overview of the demonstrator when placed on the exhibition floor. Figure 16 shows the two racks as they were displayed at the conference.

Figure 16: Demonstrator racks on the exhibition floor of ECOC 2016

The data plane integration of the mid-term demonstrator has been fully successful. Beyond the completed live demonstrations at ECOC the integrated testbed is also being used to produce scientific results related to both the vApp and VDC use cases and derived concepts.

5 Further Work

Following on from the success of the medium term demonstrator and the subset of that used as a demonstrator at ECOC 2016 and in accordance with the overall COSIGN description of work, at the time of writing this deliverable, work is underway to integrate the COSIGN long term demonstrator.

The original overall plan for the testbed to support the development and integration of the long term demonstrator is shown in Figure 17. As well as the SDN based control plane and orchestration elements controlling the integrated packet and optical circuit switching described in this deliverable, the long term demonstrator will also include transmission over multi-core optical fibre (MCF) and hollow core fibre (HCF), use of a fast optical time division multiplexing (TDM) switch (developed by the University of Bristol) and a MCF optical space switch (developed by the University of Southampton and Polatis) and, if it is ready in time, the Venture Photonics switch as described in D2.1 [15].

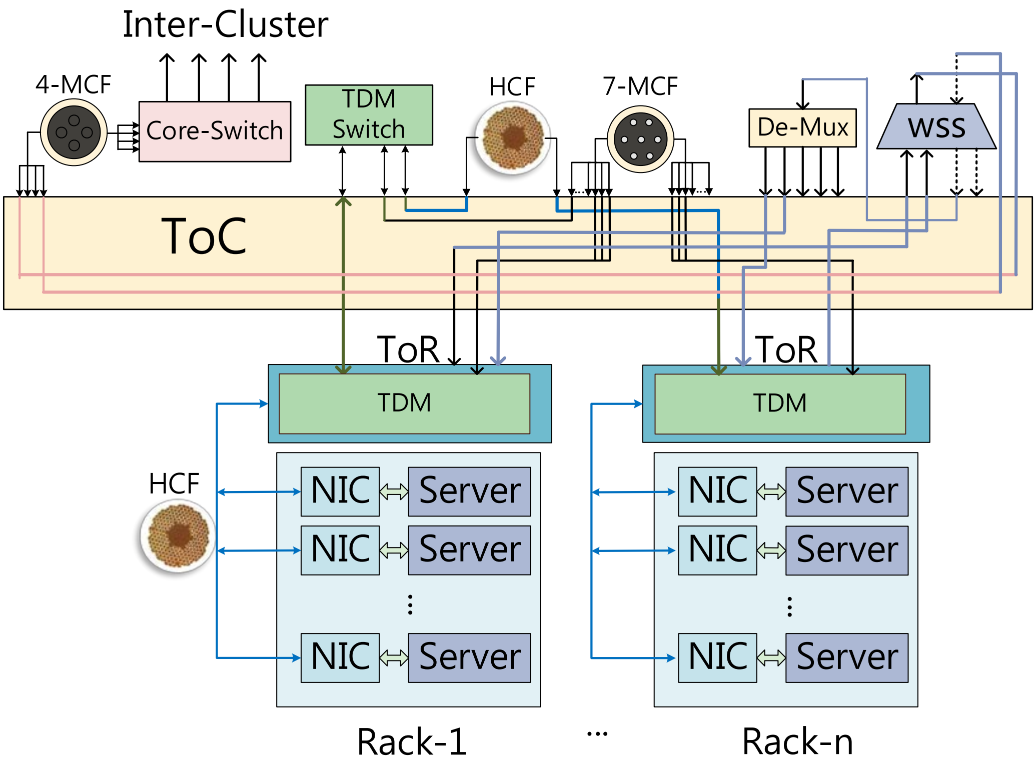

Figure 17: Overall setup envisaged for the COSIGN long term demonstrator

A subset of this demonstrator which is intended to be smaller and more portable, enabling it to be exhibited at some key industry events, is shown in Figure 18.

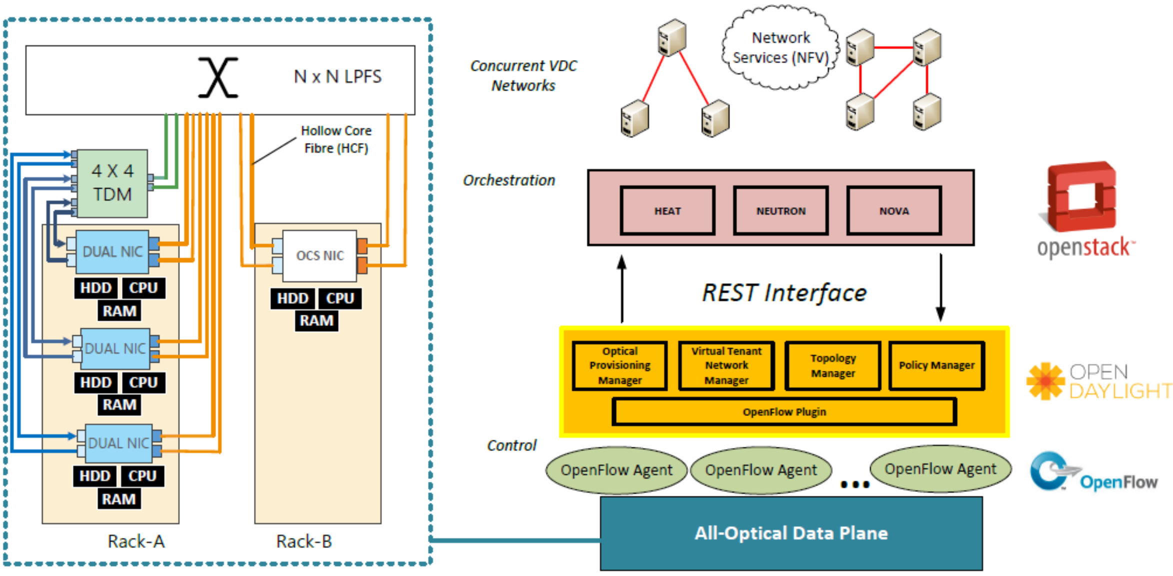

Figure 18: Illustrative setup of the COSIGN long term demonstrator

5.1 Polatis switch management agents

Inspired by participation in the COSIGN project but now independently of the project, Polatis is, at the time of writing, finalising the integration into their switches an OpenFlow agent based on the 1.4 release of the protocol (version 5) [19]. This will provide direct support for optical circuit switching, as opposed to the vendor/experimenter-specified extensions used with OF1.0+.

In addition, and again independently of the COSIGN project, Polatis has also added support in its switches for the NETCONF protocol. In the near future support for RESTCONF will be enabled. Although the RESTCONF standard is formally still being developed in a sequence of IETF drafts (the latest of which is version 18 dated October 27th 2016 [20]), it is now considered stable enough to enable the RESTCONF interface.

5.2 sFlow support in TOR switch

Although unlikely to be done under the auspices of the COSIGN project, PhotonX plan to use the open source OpenNSL software [21] that has been made available by Broadcom to facilitate export of packet flow statistics from hardware based on the Broadcom chipset using the sFlow protocol.

6 Conclusion

This deliverable describes the successful further development of the TOR switch by COSIGN partners TU/e and Photonx and its integration with the Polatis optical circuit switch, suitably augmented with an efficient on-board software agent to implement OpenFlow-mediated software defined network control. This has allowed the COSIGN project to realise an SDN controlled, highly performant, packet-OCS hybrid data plane that has been used to great effect by two of the COSIGN application use cases (VApp and VDC) as reported in other COSIGN deliverables.

Now work is under way to build on this success and realise a “long term demonstrator” testbed platform that will showcase most if not all the functionalities and technologies originally envisaged at the beginning of the project and, more importantly, how these can be used to increase the efficiency and performance of modern fibre-rich datacentre infrastructures.