Programmable Photonics in Data Centers: Architectures and Algorithms

Carla Raffaelli
DEI, University of Bologna, Italy

Abstract— Flexible networking is a challenge to design future Internet scenarios and service convergence. The main aspects related to data center interconnection based on emerging software defined photonics and networking concepts are here introduced and discussed. The programmable optical backplane is described and applied as a mean to provide enhanced network functionalities in a scalable and reconfigurable perspective.

1. INTRODUCTION

The recent evolution of the Internet is characterized by the need to dynamically share physical resources like storage, computing capacity and even networks themselves, thus enabling modern applications, like social networking, cloud computing, video streaming and others to efficiently perform their services.

Data centers are emerging as aggregates of increasingly powerful physical facilities which generate high amount of traffic and consume high power, accordingly. In spite of the continue increase of peak performance, the allowable power consumption is required to increase at a much slower rate [1]. As a consequence, data center interconnection design will result one of the most challenging networking problem in the next future which can be solved in a massive adoption of new technological solutions, possibly based on photonic solutions.

On the network service side, the evolving characteristics of content providers call for enhanced dynamic network reconfiguration capability. Emerging photonic technology can provide interconnection at extremely high rates with enhanced flexibility in spectrum sharing [2]. At the same time network control and management technologies should be able to exploit this flexibility in relation to cloud-based application dynamics.

Solutions based on combinations of Network Function Virtualization (NFV) and Software Defined Networking (SDN) can achieve the extreme flexibility required to support current and future development of cloud computing paradigms. Programmable features of emerging photonic technology can be fruitfully exploited in this context thus defining new roles and chances for optical interconnection both in capacity exploitation and energy efficiency perspective [3]. In any case the joint deployment of photonics and SDN to fulfill future dynamic network requirements, needs algorithms and architectures to harmonize and optimize several functionalities [4], which are addressed in this presentation.

The application of standard protocols like Open Flow [5] must be studied to map network management functionalities on programmable photonics. Different approaches can be adopted based on protocol extensions or on protocol compliance. Architectures and algorithms to support the management of flexible grid channels are here considered through possible virtualization of photonic network elements based on emerging SDN protocols. Solutions to optimize the usage of wavelength channels while supporting quality of service will be also discussed in terms of procedures and performance.

The reference data center architecture consists of several servers organized in racks. Traffic generated by servers is aggregated by top of the rack switches (ToRs) which possibly implement quality of service mechanisms. They are called hybrid switches because they implement both circuit and packet services and interface to server with elecrical interfaces and to the optical core switch with optical WDM interfaces. Example of such ToR switches are described in [6].

The paper is organized as follows. In Section 2 the concept of SDN and its evolution are presented. The introduction of optical programmable backplane in design of flexible interconnection in data center is discussed in Section 3. An example of application of the programmable architecture to support quality of service is described in Section 4. Concluding remarks are given in Section 5.

2. THE ROLE OF SDN

The SDN concept can be considered as firstly introduced in [5] with the definition of the Open Flow protocol, even though it is a quite broader concept. Since then it has evolved through many important steps while gaining increasing attention of the major players in the Internet world, like Google, Cisco, Juniper, Hewlett'Packerdand VMWare, just to cite a few. The basic concept behind SDN is not completely new, being it characterized by the separation of control and data planes [7], but what's more is the notion that the network is a programmable system. This property is achieved by the introduction of the network operating system as an analogous of the operating system of a standalone computer, distributed instead throughout the network to jointly work with software abstractions, i.e., virtualization, of available network hardware and infrastructure.

This appealing scenario requires a deep change in conceiving and designing network components and functionalities. In the SDN perspective the network intelligence is implemented in controlling entities set outside from the forwarding engines which results in simpler boxes with respect to today network nodes. Controlling entities become parts of the network operating system through which network services are configured and operated. Network elements instead provide forwarding services and standard interfaces to the network operating system.

As explained in [8] the reference SDN architecture includes control, management and data plane functionalities. The control plane, also called the network operating system, interfaces with the network application layer through a northbound interface which offer an open API (Application Programming Interface) to configure and define any kind of feasible network service with the desired characteristics. At the same time it interfaces with the forwarding plane through standard protocols, like Open Flow, at the southband interface to configure and exploit physical resources. A significant part of the SDN architecture is represented by network management that operates at all the abstraction layers which contribute to SDN implementation. Basically, the management plane supports network programmability in terms of infrastructure configuration and monitoring, while the control plane is in charge of dynamically configuring communication services on demand, typically based on flow-oriented operations. The management plane operations take place on longer time scale with respect to logical connections.

The elements in the forwarding plane are called 'network elements' and their internal organization aims at layering hardware specific functionalities and logical flow oriented operations [3].

3. PROGRAMMABLE OPTICAL ARCHITECTURES

Programmable optical architectures can play a key role in future data center networking. The concept of programmable architecture has been recently introduced and demonstrated in practical experiments [9]. Basically, a programmable architecture consists of a switching matrix that can be reconfigured via management software. Reconfiguration times depend on optical switching technology, typical values being in the order of tens of milliseconds for most popular implementations [10]. Several applications of programmable architecture can be foreseen, ranging from reconfigurable interconnections of physical devices to programmable architectures of networks devices [9]. Key network elements such as Reconfigurable Optical Add Drop multiplexers (ROADM) have been shown to improve scalability and power efficiency aspects with costs comparable with traditional architectures when adopting programmable interconnections [11].

Programmable architectures can be efficiently applied in the interconnection of software defined photonics elements for evolution of data center interconnection and related functionalites. In Figure 1 a general scheme which represents the programmable architecture functionalities is reported. The programmable optical backplane provides the interconnection paths among network elements, which can be reconfigured on a wider time scale with respect to connection operation. They are typically configured by operation on the management plane. When referring to SDN context, connection operation, instead, are typically flow-based and the mostly deployed protocol to control their configuration is represented by Open Flow [5].

The above architecture can be applied also to introduce new network functionalities, such quality of service support, which will be considered in the next section.

4. TRAFFIC MANAGEMENT AND QUALITY OF SERVICE SUPPORT

The introduction of optical WDM links in data center aims at meeting both high capacity and limited power consumption requirements [12]. Flexible grid WDM attracts particular interest for its flexibility in supporting optical channels with different capacity. This property, if suitably exploited by programmable hardware, can sensibly enhance variable capacity requirements of cloud computing and big data applications. In any case WDM offers communication capacity in the Gbit/s range that can accomodate aggregate Ethernet traffic coming from top of the rack switch for intra and inter data center data transfer. Traffic generated in data centers is typically classified

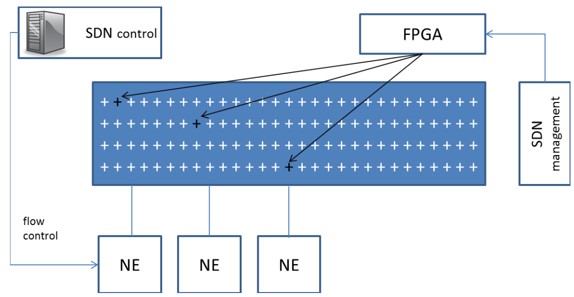

Figure 1: Programmable optical backplane in SDN context with software defined network elements.

in elephant and mice flows and mapped onto optical channel accordingly. Huge elephant flows are typically assigned circuit switched connections while light elephant flow or mice flows are switched on a lower time scale by fast packet switching [6]. A wavelength channel is able to provide very high transmission capacity up to the range of 100 Gbit/s or even beyond. Circuit and packet switched information are carried in most proposals by separate wavelength [13]. Efficient usage of this resource is highly important because even small unused intervals lead to huge waste of transport capacity.

In [14] a scheme to support three different levels of service has been proposed and in [15, 16] performance of the aggregation stage and of a core switch have been evaluated. Basically the proposed approach consists of a delayed guaranteed service traffic (GST) that experience no loss and no variable delay when traversing the WDM interface. Each GST flow is assigned to one wavelength and GTS traffic engineering is performed on a call blocking basis [17]. The delay applied to the GST traffic is represeted by the maximum acceptable duration of real time packets, which is typically very low, in the range of tens of microseconds. Real time traffic (RT) has no variable delay, being it dropped when no WDM resources are available. RT traffic engineering is performed with reference to blocking probability at a statistical multiplexer without queuing [15]. RT blocking takes place in case a GST burst is already waiting for transmisison in the system or no wavelength channels are available upon RT packet arrival. Finally best effort traffic (BE) is admitted to the free WDM channels of an interface. A pre-emptive policy is applied, meaning that a BE packet is interrupted upon arrival of GST or RT traffic.

The above quality of service mechanism can be implemented on a programmable architecture, which allow flexibility in scalability, implementation and possible upgrade of the algorithm. The proposed scheme, thought as an optical core switch for a data center, is represented in Figure 2. Basically, the scheme includes the functional blocks described in [14] which are interconnected by a prograbble backplane instead than wired interconnections, with consequent enhanced flexibility.

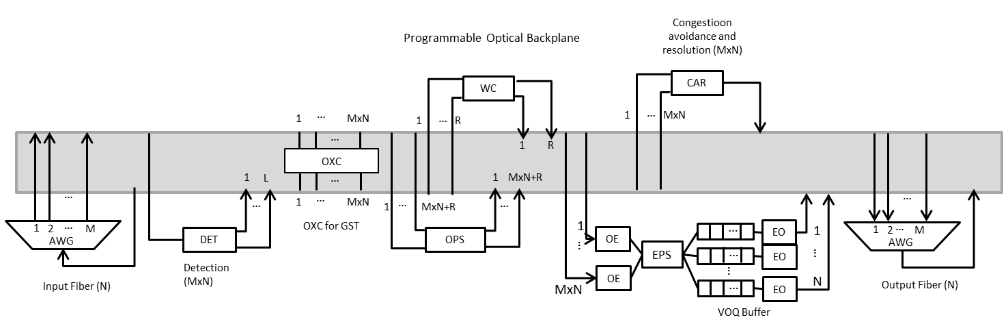

Figure 2: Programmable implementation of the optical core switch with 3-level quality of service support.

Let us assume to have N top of the rack (ToR) switches, with M WDM output channels, each. The optical core switch is configured with N input de-multiplexers with M output channels and N output demultiplexers with M channels to provide inter-rack interconnections. Additional interfaces should be provided for inter-data center interconnections that are not considered here.

Each de-multiplexed wavelength channel is sent to a detect packet type block (DET), which applies physical layer decoding to select the different kinds of traffic (GST, RT, BE). Each output of this block is connected to the corresponding required switch functionality. In case of packet switching, wavelength conversion can be applied to shift a packet from one channel to another one of the same output interface. OPS (Optical Packet Switching) switch size is increased with R interfaces in relation to the wavelength conversion approach and packet loss performance [18].

After switching, collecting blocks are introduced with the contention avoidance and resolution mechanisms (CAR). These blocks are $L:1$ devices, being L the number of QoS levels, M for each output interface. These blocks provide GST delay for guaranteed service and internal signaling capability. The status of these devices is properly set by the SDN controller. Each output is connected by the programmable backplane to the proper output multiplexer.

The programmable backplane introduces additional comlexity represented by the number of optical switches needed to implement the architecture. Let us introduce the following notation:

Table 1: Model parameters.

N	number of input and output fibre interfaces
M	number of wavelengths per interface
R	number of wavelength converters
L	number of QoS levels
C_{IO}	number of MEMS related to input and output interfaces
C_D	number of MEMS related to detection
C_D	number of MEMS related to switching modes
C_W	number of MEMS related to wavelength conversion
C_C	number of MEMS related to contention avoidance and resolution
C_I	total cost in MEMS of programmable interconnection

The following expressions can be obtained:

$$C_I = C_{IO} + C_D + C_S + C_W + C_C = LMN + 4MN + 2R + 3N \tag{1}$$

As shown in Equation (1), the total complexity function C_I results dominated by the product MN. At the same time the solution provides the flexibility to introduce additional level of functionality by increasing the detection module and possibly adding ne switching blocks.

In Figure 3 some numerical values are reported regarding programmable switch complexity, where the number of wavelength converters R is defined to obtain packet loss probability less then 10^{-4} with a shared per node saring scheme [18]. The figure shows the effect of the product MN which dominates the behaviour of the interconnection complexity function C_I.

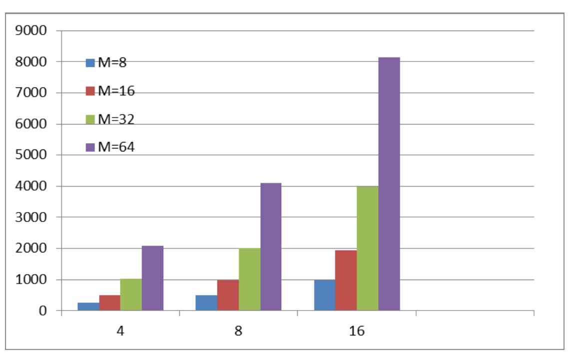

Figure 3: Number of MEMS switch in the programmable optical backplane as a function of the number of input and output interfaces, varying the number of wavelengths as a parameter, correponding to packet loss probability 10^{-4} with fully tuneable wavelength converters.

5. CONCLUDING REMARKS

In this paper the main aspects in designing future dynamic SDN based networks are presented. The concept of software defined photonic network element has been introduced and the possibility of its interfacing with SDN framework is presented. The availability of commodity optics to implement optical programmable backplane is outline to achieve high flexibility, scalability and reconfigurability both for data center interconnections and advanced quality of service features.

REFERENCES

1. Gringeri, S., et al., "Technologies and protocols for data center and cloud networking," *IEEE Comunications Magazine*, Vol. 51, No. 9, 24–31, 2013.
2. Orlandi, P., et al., "Reconfigurable silicon filter with continuous bandwidth tunability," *Journal Optics Letters*, Vol. 37, No. 17, 3669–3671, 2012.
3. Channegowda, M., et al., "Software-defined optical networks technology and infrastructure: Enabling software-defined optical network operations," *J. Opt. Commun. Netw.*, Vol. 5, No. 10, A274–A282, 2013.
4. Raffaelli, C., et al., "SDN-controlled flexible-grid optical switch," *Proceedings EUCNC 2014*, Bologna, Italy, 2014.
5. McKeon, N., et al., "Open flow: Enabling innovation in campus networks," *ACM Sigcomm Computer Communications Review*, Vol. 38, No. 2, 69–84, April 2008.
6. Veisllari, R., et al., "Field-trial demonstration of cost efficient sub-wavelength service through integrated packet/circuit hybrid network," *IEEE/OSA Journal on Optical Communications and Networking*, March 2015.
7. Gude, N., et al., "NOX: Towards an operating system for networks," *ACM Sigcomm Computer Communication Review*, Vol. 38, No. 3,, 105–110 July 2008.
8. Araujo Wickboldt, J., W. Palm de Jesus, P. H. Isolani, C. Bonato Both, J. Rochol, and L. Zam-Benedetti Granville, "Software-defined networking: Managment requirements and challenges," *IEEE Communications Magazine*, 278–285, January 2015.
9. Peters, A., E. Hugues-Salas, G. Zervas, and D. Simeonidou, "Mesuring flexibility and design trade- offs of N×M SSS-based ROADMs and BVTs," *Proceedings of OFC*, paper W3J.5, 2015.
10. http://www.polatis.com/.
11. Garrich, M., N. Amaya, G. S. Zervas, J. R. F. Oliverira, P. Giaccone, A. Bianco, D. Simeonidou, and J. C. Oliveira, "Architecture on demand design for high-capacity optical SDM/TDM/FDM switching," *IEEE/OSA Journal of Optical Communications and Networks (JOCN)*, Vol. 7, No. 1, 21–35, January 2015.
12. "A survey on optical interconnects for data centers," *IEEE Communications Surveys & Tutorials*, Vol. 14, No. 4, 1021–1035, Fourth quarter 2012.
13. Guo, B., S. Peng, C. Jackson, Y. Yan, Y. Shu, W. Miao, H. Dorren, N. Calabretta, F. Agraz, S. Spadaro, J. Perell, G. Bernini, R. Monno, N. Ciulli, R. Nejabati, G. Zervas, and D. Simeonidou, "SDN-enabled programmable optical packet/circuit switched intra data centre network," *OFC 2015*, Los Angeles, CA, US, March 22–26, 2015.
14. Stol, N., C. Raffaelli, and M. Savi, "3-Level Integrated Hybrid Optical Network (3LIHON) to meet future QoS requirements," *IEEE Globecom 2011*, Houston, Texas, USA, ISSN: 1930-529X, 2011.
15. Cerroni, W. and C. Raffaelli, "Analytical model of quality of service scheduling for optical aggregation in data centers," *Photonic Network Communications*, Vol. 27 , No. 3, 1–12, June 2014.
16. Raffaelli, C. and G. Leli, "Performance of 3-level scheduling in optical core switch for data center interconnection," *Proc. NOC 2014*, Milan, Italy, June 2014.
17. Veisllari, R., N. Stol, S. Bjornstad, and C. Raffaelli, "Scalability analysis of SDN-controlled optical ring MAN with hybrid traffic," *ICC 2014*, Sidney, Australia, Giugno, 2014.
18. Savi, M., H. verby, Norvald Stol, and C. Raffaelli, "Cost evolution model to design optical switching fabrics with wavelength converters," *Elsevier Computer Networks*, Volu. 72, 99–112, October 29, 2014.