Proceedings Paper:
Hammadi, A, Musa, M, El-Gorashi, TEH et al. (1 more author) (2016) Resource provisioning for cloud PON AWGR-based data center architecture. In: Proceedings. 21st European Conference on Network and Optical Communications (NOC), 01-03 Jun 2016, Lisbon, Portugal. IEEE , pp. 178-182. ISBN 978-1-5090-0296-2

https://doi.org/10.1109/NOC.2016.7507009

Resource Provisioning for Cloud PON AWGR-Based Data Center Architecture

Ali Hammadi, Mohamad Musa, Taisir E.H. El-Gorashi, and Jaafar M.H. Elmirghani
School of Electronic & Electrical Engineering, University of Leeds, LS2 9JT, United Kingdom
{elaah, el10moi, t.e.h.elgorashi, j.m.h.elmirghani}@leeds.ac.uk

***Abstract*—Recent years have witnessed an ever-increasing growth for cloud computing services and applications housed by data centers. PON based optical interconnects for data center networks is a promising technology to offer high bandwidth, efficient utilization of resources, reduced latency and reduced energy consumption compared to current data center networks based on electronic switches. This paper presents our proposed scheme for data center interconnection to manage intra/inter communication traffic based on readily available low cost and power PON components. In this work, we tackle the problem of resource provisioning optimization for cloud applications in our proposed PON data centers architecture. We use Mixed Integer Linear Programming (MILP) to optimize the power consumption and delay for different cloud applications. The results show that delay can be decreased by 62% for delay-sensitive applications and power consumption can be decreased by 22% for non-delay sensitive applications.**

Keywords— Passive Optical Network (PON); data center; energy efficiency; arrayed waveguide grating routers (AWGRs); resource provisioning.

I. Introduction

The use of Passive Optical Networking (PON) technology in data centers and the useful functionalities provided by devices like Arrayed Waveguide Routers (AWGR), fiber Bragg grating (FBG), and star couplers/splitters have attracted much attention from the research community in the last few years. As PON technology performance has been proven in access networks and has shown its capability in provisioning low cost, high capacity, low latency, scalable, and energy efficient networks, it has become more attractive to be adopted to provide fabric interconnection in modern data centers.

PONs can resolve many issues in current electronic and optical data center architectures such as high cost and high power consumption resulting from the large number of access and aggregation switches needed to interconnect hundreds of thousands of servers [1]. PONs can also overcome the problems of switch oversubscription and unbalanced traffic in data centers where PON architectures and protocols have historically been optimized to deal with these problems and handle bursty traffic efficiently through flexible protocols.

PON solutions are scalable. This is readily proven in the combination of core and access networks that are able to connect easily 20 million homes in the UK, 5 times that in the US. PONs achieves scalability due to their cellular architecture. A PON cell may connect 256 servers and many cells can then provide coverage of a small data center or a large data center with a possible 1 million servers. PON solutions enable efficient wavelength/bandwidth utilisation as PONs can assign a wavelength to large “elephant” flows between servers and can also allocate a time slot in their TDM-WDM structure to accommodate “mice” flows.

In our previous work we investigated energy efficiency for core networks with data centers and clouds [2-8]. In [9], we proposed and compared five novel designs for PON deployment in future cloud data centers to handle intra and inter-rack communications. In [10], we have shown that the AWGR-based PON architecture can be scaled up efficiently to hundreds of thousands of servers and have shown energy savings of 45% and 80% compared to the Fat-Tree [11] and BCube [12] architectures, respectively. In this paper, we further investigate the proposed AWGR PON architecture for cloud resource provisioning applications. We present our results through a developed Mixed Integer Linear Programming (MILP) optimization model for power consumption and delay minimization for different applications that can be hosted in data centers.

The remainder of this paper is organized as follows: In Section II, we present our proposed PON data center architecture. In Section III, we present our results for resource provisioning through our optimization model. Finally we conclude the paper in Section IV.

II. Architecture of the Proposed AWGR-Based PON Data Center

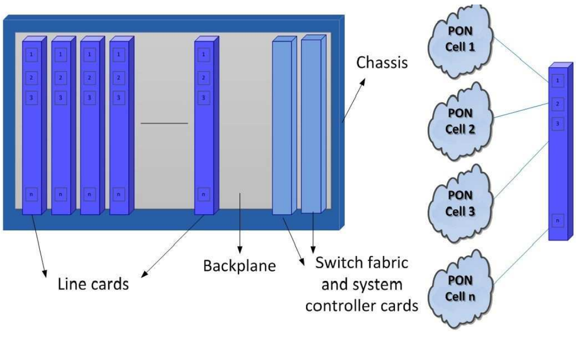
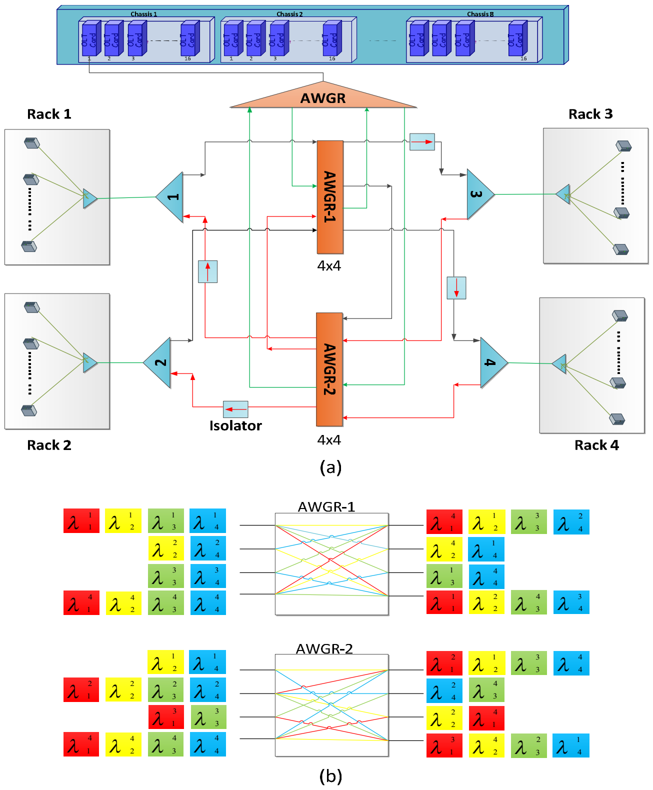


Fig. 1. A possible PON deployment in a data center (a) An OLT chassis with 16 AM, each of which with 16 ports, (b) AM OLT card connecting number of PON cells

In this section, we consider PONs for deployment in data centers. Figure 1 shows the proposed connectivity created by PON deployment. The OLT chassis hosts 16 access module cards (AM) where each AM has the capacity to connect 16 ports, each of which provides a transmission rate up to 10 Gb/s (e.g., XGPON2). A single card port can connect up to 128 servers, therefore, one card can connect 2048 servers and one chassis can provide connections to 32,768 servers. The architecture can be scaled up to host hundreds of thousands of servers by adding more chassis.

The architecture is similar to a cellular network in that wavelengths are reused for other racks connected to different OLT ports where each port connect a PON cell. The cellular based architecture using PONs improves scalability to allow such architectures to host millions of servers without having limitation on number of wavelengths as these wavelengths are reused in all cells.



	PON 1	PON 2	PON 3	PON 4	OLT
PON 1		λ_4	λ_2	λ_1	λ_3
PON 2	λ_2		λ_1	λ_3	λ_4
PON 3	λ_1	λ_3		λ_4	λ_2
PON 4	λ_4	λ_2	λ_3		λ_1
OLT	λ_3	λ_1	λ_4	λ_2	

(c)

Fig. 2. (a) Architecutre of proposed PON cell with servers equipped with tuneable lasers (b) Obtained MILP configuration for 4x4 AWGRs interconnection for wavelenght routing (c) MILP obtained wavelength routing table for intra-cell communication

Figure 2(a) depicts the architecture of a PON cell relying only on optical passive devices. The PON cell is equipped with two intermediate AWGRs to provision full interconnection among the PON cell racks. Each PON group hosts 16-32 servers connected by passive splitters/couplers. The connection between the racks and the OLT is established via a 1: N AWGR. Each server is equipped with a tuneable laser and needs to tune to a particular wavelength to communicate with other servers in different racks. The fabric interconnection and wavelength routing table shown in Figures 2(b) and (c) are obtained through a mathematical MILP optimization model.

A. *Inter-Rack Communication within the PON Cell*

Inter-rack communication within the cell can be provisioned either via the OLT switch or directly through the intermediate AWGR where a wavelength is selected for transmission based on the location of the destination server. Alternative routes facilitate multi-path routing and load balancing at high traffic loads, however, forwarding traffic through the OLT switch should be avoided if possible to reduce delay and power consumption. A server can reach servers in other racks by tuning its transceiver to the proper wavelength that matches the AWGR wavelength routing map.

This design is a Wavelength Routing Network (WRN) with N+1entities (N racks and the OLT) to communicate with each other. In a WRN for N+ 1 entity to communicate with N entities, we require either N fibers with N^2 wavelengths or N^2 fibers with N wavelengths. In our designs, we select N wavelengths and employ the 2 AWGRs to represents the N^2 fibers (connections). For the architecture depicted in Figure 2 with N=4, 4 wavelengths are needed.

According to the wavelength routing table shown in Figure 2(c), if a demand exists between servers A and B located in rack 1 and 4 respectively. A control request message is sent to the OLT switch using wavelength 3 routed through AWGR-1 input port 1 to output port 3. If the OLT decides to grant the request, the OLT then replies with a control messages to the two servers A and B using wavelengths 3 and 2 for racks 1 and 4 respectively. The control messages shall contain information about the wavelengths both servers need to tune to and assigned resources. Upon reception of the control information from the OLT switch, servers A and B tune their transceivers to wavelength 1. Idle servers by default should be tuned to wavelengths connecting them with the OLT.

B. *Intra-Rack Communication*

Intra-rack communication within the PON cell can be provisioned using one of the described techniques in Figure 3. The first proposed design uses a passive star reflector to connect servers within a rack allowing each server to broadcast to other servers using an additional transceiver. The main limitation of such a design is the complexity of the MAC protocol needed to coordinate and arbitrate channel access.

Another solution to support intra-rack connectivity is to deploy a Fiber Brag Grating (FBG) after the star coupler connecting the servers in the rack to reflect a dedicated wavelength assigned for intra-rack traffic communication. To facilitate the use of the FBG for the intra-rack communication, each server can be equipped with a second multi-wavelength (MW) transceiver. OFDM technology can be used to allow a single transceiver to generate multiple carriers, one for intra-rack communication and another for connections to the OLT or other racks. However, the expensive OFDM transceivers will increase the deployment cost of the PON design.

A third alternative which we find more practical for intra-rack communication is the Passive Polymer Backplane developed in [13]. This technology employs a passive backplane with multimode polymer waveguides and can provide non-blocking full mesh connectivity with 10 Gb/s rates per waveguide, exhibiting a total capacity of 1 Tb/s.

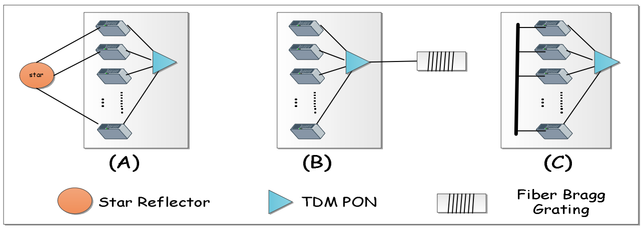


Fig. 3. PON based technologies for intra-rack communication

C. *Inter-Cell Communication*

The depicted design in Figure 4 shows a schematic of proposed upper level connectivity for inter-cell communication using only passive devices. For simplicity, we only show uplink connections from PON cells to the OLT switches. This facilitate multi path routing and also enhance the bandwidth allocation mechanism by introducing 2-tiers of optical passive AWGRs for connectivity with multiple OLT switches instead of having each PON cell connected to a single OLT port. This allows efficient utilization of resources in case of low activity in some PON cell by allowing servers at heavily loaded cells to join multiple OLT ports where resources are available. We employ SDN control and management system to coordinate and arbitrate the channel access for communication through the OLT links for uplink and downlink transmissions.

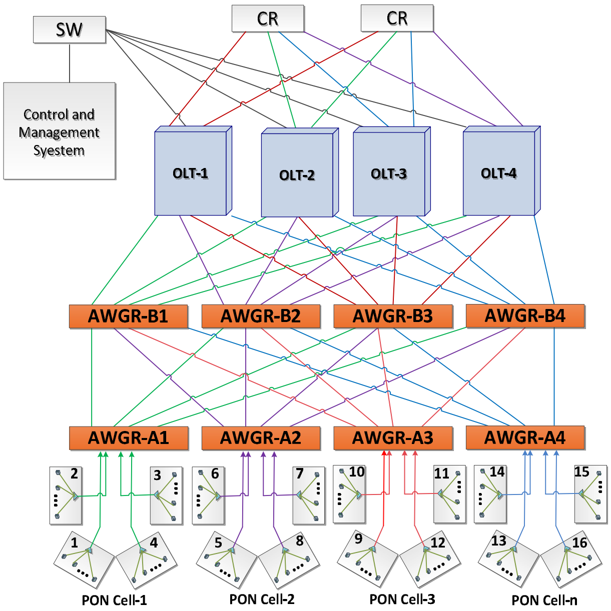


Fig. 4. Upper-level connectivity for inter-cell communication

Depending on the activity ratio of servers in different PON cells, the PON protocol through SDN can reconfigure the network by retuning the servers' transceivers to distribute and perform load balancing to different OLT ports. Alternatively the SDN can also consolidate loads at fewer PON cells to save power in response to the variation of the daily load. The main advantage of the design is its flexibility which allows servers to join different OLT ports based on availability of resources in order to reduce oversubscription and improve resources provisioning mechanism through energy efficient grooming and reconfiguration.

III. Optimization of Resources Provisioning in PON Data Center

In this section, we report the MILP optimization results with the objectives of power and delay minimization for efficient resource provisioning in a PON data center. We examined a different number of VMs (20, 40, and 60), where each VM has a requirement for CPU, memory, and communication traffic with other VMs selected randomly following uniform distribution. Table I presents the input parameters used for the model.

Table I: Input Data for the Model

Link capacity	10Gb/s
Power consumption for idle servers	201 W [14]
Maximum power consumption for servers	301 W [14]
Clients' processing requirements of CPU cycles in MHz	500-2000
Clients' memory requirements in MB	500-2000
Server's processing capacity in GHz	2.5
Server's memory (RAM) in GB	8
ONU power consumption	2.5W [15]
VMs traffic	40-200 Mb/s

The model ensures that the ONU link capacity along with the physical machines' CPUs and memory capacities are not exceeded when assigning VMs. We also ensure that servers in each PON group do not exceed the shared capacity of assigned wavelengths while communicating with other servers in different PON groups. A wavelength continuity constraint is used to ensure that the wavelength going into a node is the same wavelength leaving it for all nodes except the source and destination. Furthermore, we ensure the PONs' directionality property is satisfied by ensuring flows are only directed from inputs to outputs of the optical passive devices.

The objective of minimizing the servers' power consumption results in the allocation of CPU and RAM resources requested by clients to the minimum possible number of servers by means of consolidation (a packing optimization problem). This objective does not take into account the communication demands among the VMs to decide the location of each VM. We will consider such communication requirements in extensions to our work. The objective that addresses the minimization of delay aims to minimize traffic flow on communication links and does not take into account the number of servers used to provision resources to the clients

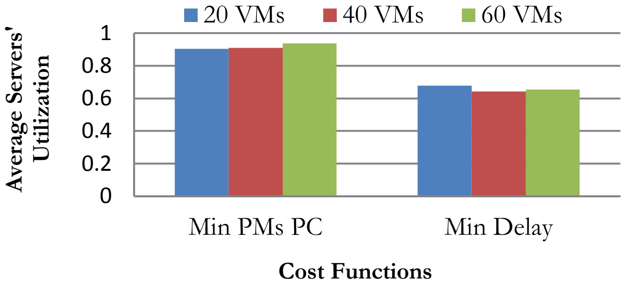


Fig. 5. Average server's utilization examining three sets of VMs; 20, 40, and 60 for the two objective functions; Minimization of physical machine power consumption and minimization of delay

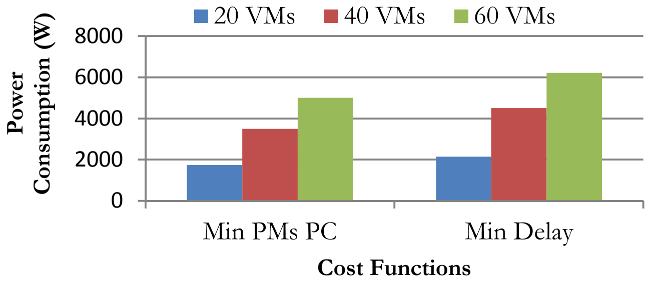


Fig. 6. Total power consumption examining three sets of VMs; 20, 40, and 60 for the two objective functions; Minimization of physical machine power consumption and minimization of delay

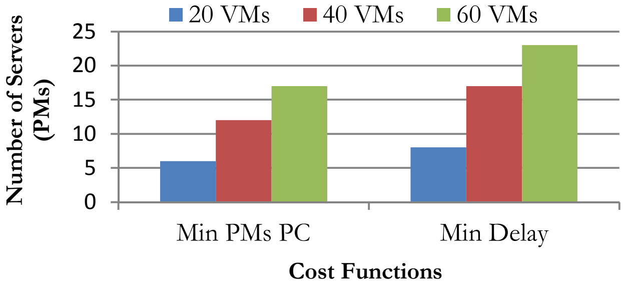


Fig. 7. Total number of switched on servers examining three sets of VMs; 20, 40, and 60 for the two objective functions; Minimization of physical machine power consumption and minimization of delay

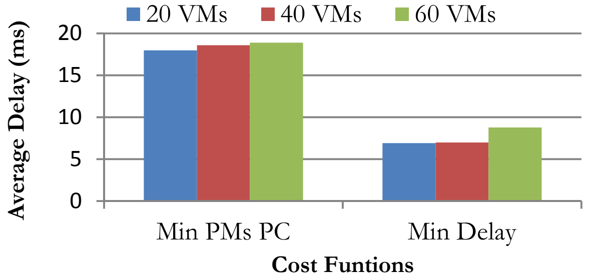


Fig. 8. Average delay examining three sets of VMs; 20, 40, and 60 for the two objective functions; Minimization of physical machine power consumption and minimization of delay

(the placement optimization problem). The number of servers can however be taken into account in a hybrid problem. We have therefore chosen to focus on the boundary problems to establish the limits on power saving and delay minimization through our proposed PON architecture and MILP optimization.

With minimization of delay, the model tries to allocate VMs that have mutual communication traffic in the same servers as much as possible. As a result, traffic flow among servers, ONU power consumption, and average delay are reduced. However, total power consumption is increased compared to the minimization of servers' power consumption model, as more servers are required to accommodate groups of VMs with mutual bandwidth. For the different number of VMs under examination, the model with the objective of delay minimization shows that the average delay can be decreased by 62%, while the power consumption minimization objective shows that power savings can reach 22%.

Minimization of delay also results in lower utilization of servers' resources as more servers are used to serve the same number of VMs to ensure that the maximum possible number of VMs are collocated in servers. Figures 5 and 7 present average servers' utilization and the number of activated servers for the different objectives for 20, 40, and 60 sets of VMs. Delay minimization results in lower average utilization of around 65%, while the minimization of servers' power consumption approach results in servers' utilization that is more efficient and approaches 90%.

IV. CONCLUSIONS

This paper has proposed an optimization model and has introduced two objective functions: minimization of delay and minimization of power consumption, to cater for different applications that can be hosted in a PON cloud data center. Our results have shown the trade-off between minimization of power consumption and minimization of delay objectives. The minimization of delay model is best for real-time delay-sensitive applications and our PON architectures and optimization have shown that a reduction in delay of 62% is possible compared with the model and approach where the server's power consumption is minimized. On the other hand, minimization of server's power consumption can be used for non-delay sensitive applications with power savings that can reach 22%.

ACKNOWLEDGMENT

The authors would like to acknowledge funding from the Engineering and Physical Sciences Research Council (EPSRC), INTERNET (EP/H040536/1) and STAR (EP/K016873/1).

REFERENCES

[1] A. Hammadi and L. Mhamdi, "Review: A survey on architectures and energy efficiency in Data Center Networks," Computer Communication. vol. 40, pp. 1-21, 2014.

[2] X. Dong, T. El-Gorashi, J.M.H. Elmirghani,"IP Over WDM Networks Employing Renewable Energy Sources", IEEE Journal of Lightwave Technology, vol.29, no.1, pp. 3-14, Jan. 2011.

[3] X. Dong, T. El-Gorashi, J.M.H. Elmirghani, “Green IP Over WDM Networks With Data Centers”, IEEE Journal of Lightwave Technology, vol.29, no.12, pp.1861-1880, June 2011.

[4] X. Dong, T. El-Gorashi, J.M.H. Elmirghani, “On the Energy Efficiency of Physical Topology Design for IP Over WDM Networks”, IEEE Journal of Lightwave Technology, vol.30, no.12, pp.1931-1942, June 2012.

[5] A.Q. Lawey, T. El-Gorashi, J.M.H. Elmirghani, “ Distributed Energy Efficient Clouds Over Core Networks”, IEEE Journal of Lightwave Technology, vol.32, no.7, pp.1261-1281, April 2014.

[6] N.I Osman, et al. “Energy-Efficient Future High-Definition TV”, IEEE Journal of Lightwave Technology, vol.32, no.13, pp.2364,2381, July 2014.

[7] A.Q. Lawey, T. El-Gorashi, J.M.H. Elmirghani, “ BitTorrent Content Distribution in Optical Networks”, IEEE Journal of Lightwave Technology, vol.32, no.21, pp. 4209-4225, Nov 2014.

[8] L. Nonde, T. El-Gorashi, J.M.H. Elmirghani, "Energy Efficient Virtual Network Embedding for Cloud Networks”, IEEE Journal of Lightwave Technology, vol.33, no.9, pp. 1828-1849, May 2015.

[9] J. M. H. Elmirghani, Hammadi, A. and El-Gorashi, T.E., “ Data Center Networks”, UK patent, 26 November 2014.

[10] A. Hammadi, T. E. H. El-Gorashi, and J.M.H. Elmirghani,”High Performance AWGR PONs in Data Center Networks,” IEEE ICTON, Hungray, 2015.

[11] M. Al-Fares, A. Loukissas, and A. Vahdat, “A scalable, commodity data center network architecture”, in Title of Proceeding, ACM SIGCOMM, Seattle, WA, USA, 2008.

[12] C. Guo, G. Lu, D. Li, H. Wu, X. Zhang, Y. Shi, “BCube: a high performance, server-centric network architecture for modular data centers”, SIGCOMM Computer Communication, vol. 39, pp. 63-74, 2009.

[13] J. Beals IV, N. Bamiedakis, A. Wonfor, R. Penty, I. White, J. DeGroot Jr, et al.: A terabit capacity passive polymer optical backplane based on a novel meshed waveguide architecture, Applied Physics A, vol. 95, pp. 983-988, 2009.

[14] D. Kliazovich, P. Bouvry, and S. U. Khan, “GreenCloud: a packet-level simulator of energy-aware cloud computing data centers,” The Journal of Supercomputing, vol. 62, pp. 1263-1283, 2012.

[15] K. Grobe, M. Roppelt, A. Autenrieth, J. P. Elbers, and M. Eiselt, “Cost and energy consumption analysis of advanced WDM-PONs,” Communications Magazine, IEEE, vol. 49, pp. s25-s32, 2011.