

Mathematical Optimization Modelling for Fast-Switched and Delay Minimized Scheduling for Intra-Cell Communication in an AWGR-Based PON Data Center

Ali Hammadi

Faculty of Electronic Engineering Technology, College of Technological Studies, The Public Authority for Applied Education and Training (PAAET), Kuwait, The State of Kuwait

Email: aa.hammadi@paaet.edu.kw

How to cite this paper: Hammadi, A. (2017) Mathematical Optimization Modelling for Fast-Switched and Delay Minimized Scheduling for Intra-Cell Communication in an AWGR-Based PON Data Center. *Int. J. Communications, Network and System Sciences*, 10, 13-29.

<https://doi.org/10.4236/ijcns.2017.102002>

Received: January 25, 2017

Accepted: February 25, 2017

Published: February 28, 2017

Copyright © 2017 by author and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

As the internet traffic along with the processing power in data centers are exponentially growing, the need for the design of energy efficient with highly elastic networking infrastructure to support the different applications and cloud services that can be hosted in data centers have become a hot research area. A key departure from the norm is that conventional routers and switches in conventional data centers are replaced with high performance Passive Optical Networks (PONs) to take over the role of routing and traffic forwarding through efficient resource provisioning algorithms. In this paper, the different aspects of PONs in the design of energy efficient, high capacity, and highly elastic networking infrastructures to support the applications and services hosted by modern data centers are considered. In this work, a mathematical optimization model for energy-efficient and delay-minimized scheduling in AWGR based PON data center for PON cell fabric configuration will be presented. The performance of the proposed architecture in terms of efficient scheduling against average delay and power consumption for different traffic loads and patterns will be evaluated. Different scenarios of traffic; random and unbalanced with hotspots are examined to evaluate the average delay and power consumption with and without sleep mode. Results have shown that with sleep mode enabled, power savings for two evaluated objective functions have shown similar results when examining different traffic patterns. The power savings range between 8% and 55% during low and high load activities, respectively. However, minimization of delay model has shown improvement in reducing total average delay reaching up to 42% if compared with the model with objective of minimization of power consumption.

Keywords

PONs, Data Center Networks, AWGR, Energy Efficiency, Scheduling, Optimization

1. Introduction

In recent years, efforts for the re-design of data center architectures have been devoted to tackle two major issues appeared in conventional data center designs; power consumption in the first place which has main impact on the global warming, and high cost of the electricity bill resulted from high number of electronic devices used in the fabric networking interconnection [1]. The introduction of PONs in the design has shown a significant potential to not only reduce power consumption and cost but also to provide high per server data rate and improve bandwidth allocation methods. PONs are found as a promising mature solution for data centers as their performance has been historically proven in access network to facilitate services to hundreds of thousands of homes. PONs has also shown a great feature of dynamic provisioning through flexible protocols to efficiently assign resources to accommodate large or small flows [2].

In the last decade, the emergence of PONs in the design of data centers fabric interconnections have been introduced in several publications. In [3] authors have shown a study on the PON technology and its capability in terms of storage, processing and transmission speed to provide all requirement needed for cloud computing applications. In [4] authors proposed architecture based on passive Arrayed Waveguide Grating Router (AWGR) to eliminate the use of high power consuming aggregation switches. The design employs Orthogonal Frequency Division Multiplexing (OFDM) technology to manage inter-rack communication between access switches and the AWGRs. Unlike the design proposed in [4], authors in [5] have introduced a design that considers a co-existence of AWGR in addition to aggregation switches. The design aims to off-load the burden on access switches when processing and forwarding inter rack traffic. The proposed design as reported by the authors has shown a 10% of reduction in power consumption on access switches. Similarly, to design presented in [5], authors in [6] demonstrated a design based on AWGRs in addition to TOR access switches. However, authors in [6] proposed a design with multiple uplink ports at the access layer to facilitate inter-rack communication through either AWGRs or TOR switches.

Another partial PON implementation was presented in [7]. Authors in [7] proposed a novel Passive Optical Pod Interconnect (POPI). The design eliminates the use of access switches and employs passive couples, splitters, and AWGRs instead. POPI relies on flow aware dynamic bandwidth allocation (DBA) algorithm to control the sharing and assignments of wavelengths channels among the servers. In [8] authors introduced a software defined based passive optical intra-rack network SD-POIRN. The work proposes a MAC design to

schedule the intra-rack data transmissions based on Max-Min Fair share bandwidth allocation algorithm. The algorithm improved the network throughput in terms of utilization and showed acceptable measures with respect to average delay.

In our previous work, we investigated the feasibility of PONs deployment in future data centers by re-designing the current paradigm of PONs (used traditionally in residential access networks) to furnish scalable, low cost, energy-efficient, and high capacity interconnections infrastructure to accommodate the different traffic patterns in data centers. We have proposed, compared, and patented five novel designs relaying mostly on PONs to handle intra and inter rack communications [2]. In [9], 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 [10] and BCube [11] architectures respectively. In [12] a Software Defined Network (SDN) based AWGR based PON data center architecture was introduced. The SDN design has improved the energy efficiency by up to 90% while introducing no blocking for typical average data rates. In [13], the server centric PON data center design was briefly introduced. The work in [13] demonstrated the optimization mathematical results for minimizing the power consumption of a PON cell through efficient placement of Virtual Machines (VM). In [14], further investigation of the design presented in [9] and [12] is carried. The work tackled the problem of resource provisioning optimization for cloud applications addressing delay and non-delay sensitive applications.

In this paper, further investigation is carried on the AWGR PON architecture design presented in [9] to design a protocol through a centralized scheduler capable of coordinating channel access arbitration and bandwidth allocation for intra PON cell communication. The work is based on development of Mixed Integer Linear Programming (MILP) optimization mathematical model for energy-efficient and minimized delay scheduling examining different traffic patterns such as random and unbalanced traffic with hot spot.

The remainder of the paper is organized as follows. In Section 2, a re-visit with a brief description of the AWGR based PON data center architecture shall be given for completeness and through presentation of the work in this paper. In Section 3, the idea of PON cell fabric configuration through energy-efficiency and delay-minimized scheduling is described. In Section 4, a presentation of the MILP model is given. In Section 5, discussion of the results is presented. Section VI concludes the paper.

2. The Architecture of the AWGR Based PON Data Center Network

In this section, the modeled architecture for the PON data center is revisited and briefly described. The PON architecture is designed employing mainly passive optical devices to manage connectivity for traffic routing among servers located in the same rack or in different racks. Intra-rack communication is the term

used for traffic exchanged among servers within the same rack, while inter-rack communication is the server to server traffic forwarded through different racks.

The architecture is comprised of the three main parts. The first main segment of the network is the rack and it is defined as a PON group. In a PON group a number of servers say 16 - 32 are connected together via one of the three technologies depicted in **Figure 1(c)** and described in details in [2] and [14].

The use of active electronic or optical switches is completely avoided for the intra-rack communication. All servers are equipped with tunable ONU like

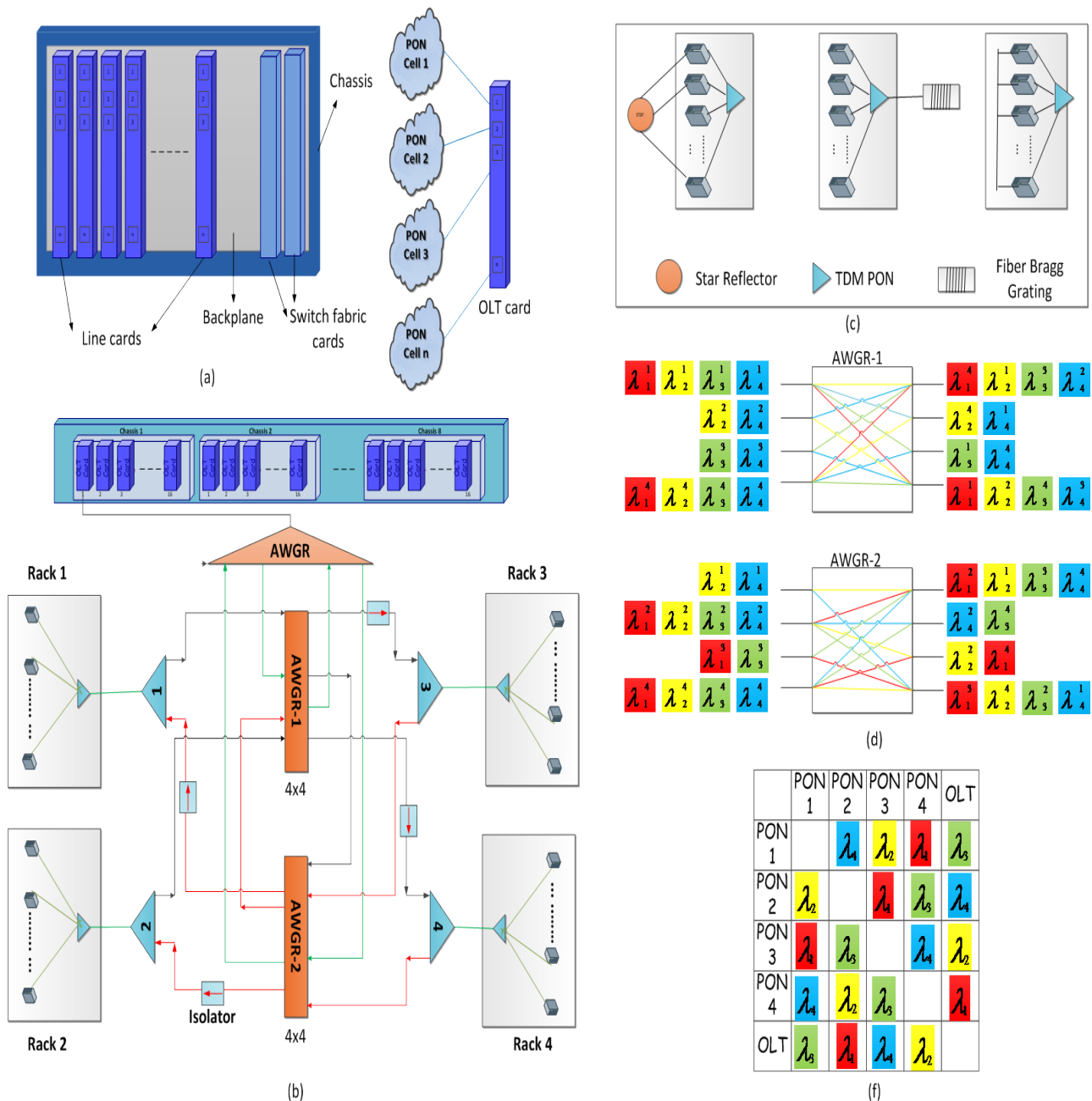


Figure 1. A possible PON deployment in a data center. (a) An OLT chassis with 16 AM, (b) AWG-based PON cell, (c) Technologies for intra-rack communication, (d) Fabric interconnection for AWGRs, (e) Routing table for inter-rack communication.

transceiver with a PCI-e interface that connects directly to the servers. The second part of the network is called the PON Cell and shown in **Figure 1(b)**. In a PON Cell, a number of PON groups (racks) are connected through intermediate AWGRs. The interconnection in a PON cell is given in a full mesh structure to provision all-to-all with bi-directional communication links through the two intermediate AWGRs. The AWGRs fabric interconnection and the wavelengths routing and assignment table for the intra PON cell communications are shown in **Figure 1(d)** and **Figure 1(f)**, respectively. For more details on how the fabric interconnection is optimized and how results are obtained from the MILP mathematical model, refer to [9] and [14].

The PON cell fabric interconnection is designed to be flexible to accommodate any number of PON groups through mathematical model that solves for topology physical interconnection and the wavelength routing and assignment within the cell [9]. The third part of the network is the optical switch (OLT). All PON groups with all the PON cells are physically connected to the OLT ports via the upper AWGRs that connect all the intermediate AWGRs located in each of the PON cells.

A network can be designed to accommodate multiple OLT switches and can be scaled up to increase the number of servers that can be hosted within the PON data center by adding more switches. Each OLT switch consists of a chassis that can host 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 [15]. An OLT chassis therefore can accommodate a total of 256 AM ports and hence connect a maximum of 256 PON group. Depending on servers' activity and the intra and inter rack traffic ratios, the number of OLT ports per PON group/PON cell are dynamically assigned.

The architecture is designed in fashion that is similar to the cellular network. The rationale behind this topology design is to have wavelengths that can be reused for other PON groups connected to different OLT ports connecting different PON cells. The cellular based architecture 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 different cells connected to different ports either in the same or different OLTs.

3. PON Cell Fabric Configuration

In high performance routers and switches, schedulers are responsible of configuring the backplane crossbar to match a conflict-free pair of inputs and outputs with unmatched demands [16]. In routers, variable length packets are segmented into fixed sized cells before transmitting them across the backplane [17]. Segmentation of packets into fixed sized cells improves the performance in terms of throughput and delay as all transmitting packets traverse the crossbar and reach the destination outputs at the same time. Hence, schedulers can plan for subsequent transmission cells for queued packets at the inputs in advance. However, the choice is not straightforward as there can be multiple inputs waiting for

transmission to the same output destination. Efficient crossbar scheduling algorithm is developed for routers to improve throughput, overcome or reduce blocking, and minimize delay. Such implementations are proposed and discussed in [18] [19] [20] [21].

For illustration purposes, **Figure 2** presents a bipartite matching example for 16x16 ports network. Nodes in the left column present the inputs, column in the right are for the outputs, and the lines connecting the inputs with the outputs are the queued demands waiting for transmission from inputs to outputs. For maximum throughput in such scenario, three cell times are needed as shown in **Figure 2** to serve all queued demands. Maximum matching is produced in cell time 1 and 2 and rest of unmatched demands is scheduled for the third cell time. The delay experienced by a request waiting to be transmitted is a function of the number of requests in the queue. The average delay is defined as the interval of time of request's arrival and its departure. Then average delay is calculated by averaging out delay of all successfully transmitted requests. Scheduled transmission in first cell time will not experience delay; however, requests scheduled for second and third cell times will experience one and two cell time delays respectively. Example shown in **Figure 2** will result in 0.6 average cell delay.

Adoption of the same concept in the proposed PON design is taken into consideration; a local scheduler at the OLT is responsible for configuring the PON cell fabric for intra-cell traffic scheduling. Upon receiving requests from demanding servers, the scheduler assigns resources and plans the scheduling of accessing the channels by means of maximum matching for efficient utilization of resources and minimization of delay as possible. For every pair of demanding servers, time slots are assigned and a pair of wavelengths that transceivers need to tune to for transmission/reception are assigned. Scheduler in the proposed architecture is responsible for scheduling transmission for all requesting servers; some demanding servers will take part in the first TDM frame time, and other take part in subsequent frame times. In a given frame time a conflict free pairing between sources and destinations is selected by the scheduler so each destination is paired with at most one source and each source is matched with one destination. Configuration of the PON cell fabric is carried by assigning tuning wave-

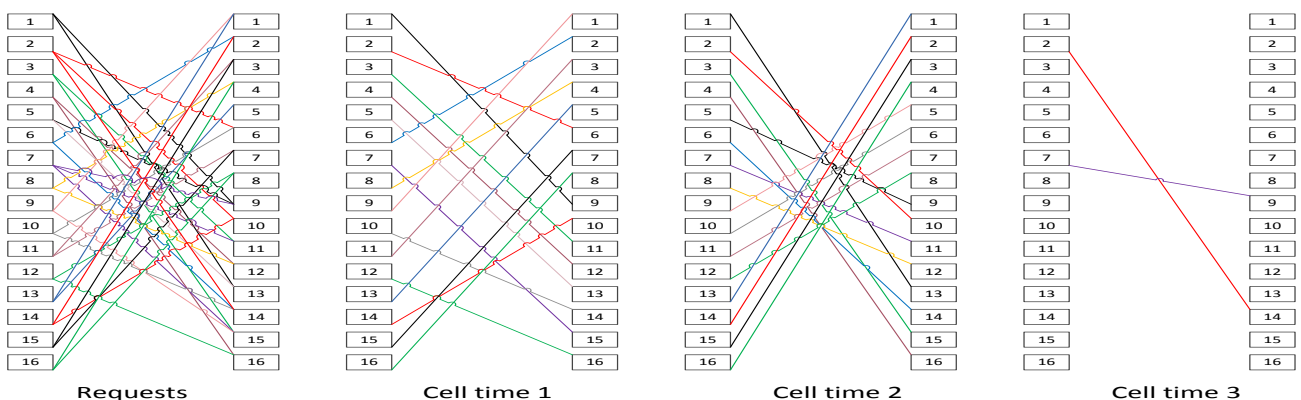


Figure 2. Sample of optimum scheduling with maximum matching assigned for transmission in three cell times.

lengths for transceivers of demanding servers. A scheduler performs a maximum size bipartite matching to maximize flows and connections made in each frame time for achieving efficient resource utilization and maximum throughput.

All intra cell communications are coordinated locally via the OLT. Servers with demands send a request control message containing destination address and resources requirements to the OLT. If OLT grants the request, it replies with control messages to the source and the destination ONUs connecting the pair of servers with information about the assignment of time slots with resources in the designated wavelength which both servers' ONUs need to tune to. At initial state, all ONUs transmitters and receivers are tuned to designated wavelength connecting them with OLT switch. Based on gate message received from the OLT, ONUs tunes their transmitters and receivers to assigned pair of wavelengths. ONUs needs to retune to other wavelengths in subsequent frame times for communication with different servers located at different PON groups. The switching time of converting wavelengths at tunable transceivers is in nanoseconds scale and has been demonstrated by experiments in Bell Labs [22].

Figure 3 depicts a sample of demand map requests received by the scheduler and output of the scheduler decision for frame times' assignments with Tx/Rx tuning map for the designated frame times. In **Figure 3**, server 2 in PON-1 (PON_2^1) has two queued requests for server 12 in PON-4 (PON_{12}^4) and server 10 in PON-3 (PON_{10}^3) for the architecture depicted in **Figure 1**. A control message is sent to the OLT using wavelength 3 routed through AWGR-1 input port 1 to output port 3 that connects with the OLT switch. Scheduler looks at the full list of demands received and match conflict free pairs to produce a maximum match. Hence, scheduler decides to grant demand between server 2 and 12 in frame time 1 and replies with a control message to PON-1 and PON-4 with wavelengths 3 and 2 respectively. Destination server at PON-4 tunes its receiver with wavelength 1 to receive data from source server in PON-1. Second queued demand at server 2 is scheduled for transmission in the second TDM frame time and will experience one frame time delay. Idle servers if not taking part in the frame time transmission should be tuned to wavelengths connecting them with the OLT switch or can be switched off for further energy savings.

Next, a mathematical model for configuring and optimizing the PON cell fabric interconnection in the AWGR based PON cell for efficient scheduling examining minimization of delay and power consumption shall be presented.

4. MILP for Efficient Scheduling for Intra-Cell Communication

In this section, a MILP model is presented with two objective functions, minimization of delay and minimization of power consumption. The objective function with the minimization of the total delay aims to minimize total delay through minimization of total TDM frame times needed to serve all queued demand requests received by the scheduler. The model shall maximize the connection made in each frame time to minimize number of frame times and the num-

ber of transceivers' retuning needed and hence minimize the total delay. Minimization of the number of frames will not only minimize total delay but also will result in maximization of throughput through efficient utilization of uplink and downlink capacities.

The second objective function with the minimization of total power consumption aims to minimize the power consumption by minimizing the total number of active ONUs needed to be switched on in the designated frame times while serving all demands.

The parameters and variables used in the model are as follows:

Parameter:	
S	Set of ONUs
PG	Set of PON Groups (Racks).
T	Set of TDM time frames
$P\{PG\}$	Set of ONUs in each PON group PG.
Cu	Wavelength uplink capacity
Cd	Wavelength down capacity
α^{ij}	Demand matrix between every source and destination servers (i, j) .
PC_ONU	Power consumption of an ONU

Variables:	
X_{ijt}	Binary equivalent equals one if server i and server j are matched in frame time t , otherwise equals zero.
Z_{ijt}	Binary equivalent equals one if a request from server i (source) to server j (destination) is granted in frame time t
Δ_{ijt}	Traffic scheduled for transmission for the granted pair of servers (i, j) .
Λ_{ijt}	Traffic to be scheduled for transmission in subsequent frame times following frame time t
Ψ_{ijt}	Binary equivalent for frame time t to equal 1 if a demand exists between (i, j) servers, otherwise equals zero.
S_{it}	Set of source ONUs granted for transmission in frame time t
S_{jt}	Set of destination ONUs granted for transmission in frame time t
ϕ_{ijt}	Binary equivalent equals one if either of i or j , or both are matched for transmission in frame time t , otherwise equal zero
ξ_{ijt}	Total number of ONUs scheduled for transmission in frame time t

Objective:

Minimize Delay:

$$\sum_{t \in T} \left((t-2) \sum_{i \in S} \sum_{j \in S} Z_{ijt} \right) \tag{1}$$

The mathematical expression in equation (1) gives the model objective which is to minimize the total delay by minimizing the waiting time for requests to be served. The reduction of required number of frame times improves the throughput by the maximization size of granted requests in each frame time.

Minimize Power Consumption:

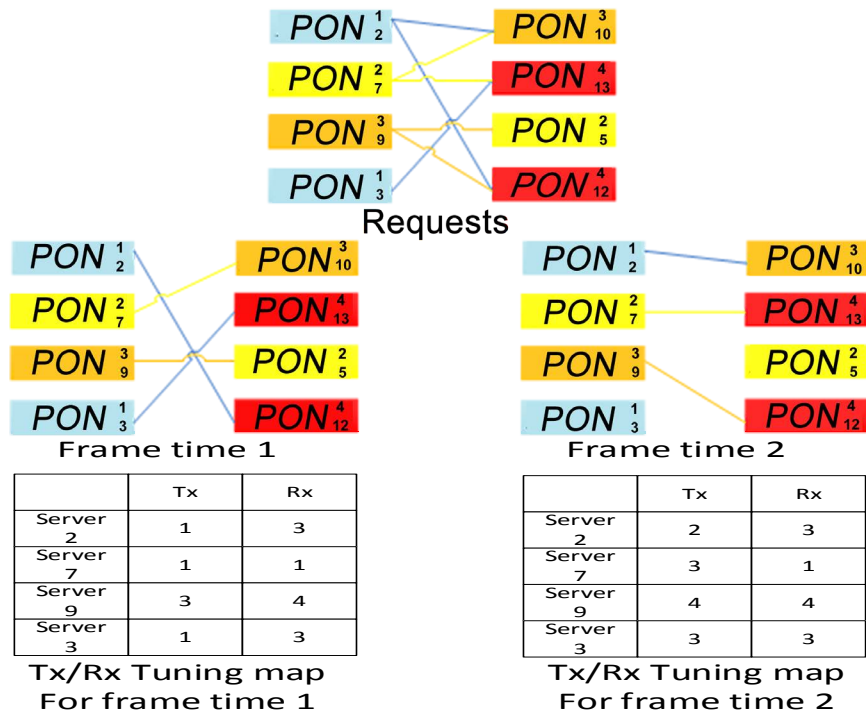


Figure 3. Sample of demand requests and cells assignments with Tx/Rx wavelengths tuning map for AWG-based PON cell.

$$\sum_{t \in T} \sum_{i \in S} \sum_{j \in S} \phi_{ijt} PC_{-ONU} \tag{3}$$

The mathematical Equation (2) gives the model objective which is to minimize the power consumption by minimizing the total number of ONUs needed to be switched on in the designated frame times while serving all demands.

Subject to:

$$\sum_{\substack{j \in S \\ i \neq j}} X_{ijt} \leq 1 \tag{3}$$

$$\forall i \in S, \forall t \in T \quad j \text{ and } t \neq 1$$

Constraint (3) ensures that any requesting server with a queued demand should be matched with at most one destination in frame time t .

$$\sum_{\substack{i \in S \\ i \neq j}} X_{ijt} \leq 1 \tag{4}$$

$$\forall j \in S, \forall t \in T \quad j \text{ and } t \neq 1$$

Constraint (4) ensures that each destination server should be matched to at most one source server in frame time t .

$$\Lambda_{ijt} \geq \psi_{ijt} \tag{5}$$

$$\Lambda_{ijt} \leq M \psi_{ijt} \tag{6}$$

$$\forall i, j \in S, \forall t \in T \quad j \text{ and } t \neq 1$$

Constraints (5) and (6) are binary equivalent to indicate that a request exist between (i, j) servers in frame time t .

$$Z_{ijt} \leq X_{ijt} \tag{7}$$

$$Z_{ijt} \leq \psi_{ij(t-1)} \tag{8}$$

$$\begin{aligned} Z_{ijt} &\geq \psi_{ij(t-1)} + X_{ijt} - 1 \\ \forall i, j \in S, \forall t \in T \\ i &\neq j \text{ and } t \neq 1 \end{aligned} \tag{9}$$

Constraints (7), (8) and (9) are binary equivalent to indicate a request from server i (source) to server j (destination) is granted for transmission in frame time t .

$$\begin{aligned} \Lambda_{ijt} &= \alpha^{ij} \\ \forall i, j \in S \text{ and } t = 1 \end{aligned} \tag{10}$$

Equation (10) defines the demand map requests received by the scheduler at $t=1$ is equal to the traffic demand.

$$\begin{aligned} \Delta_{ijt} &= Z_{ijt} \alpha^{ij} \\ \forall i, j \in S, \forall t \in T \\ i &\neq j \text{ and } t \neq 1 \end{aligned} \tag{11}$$

Equation (11) presents the size of accepted demand between (i, j) servers scheduled for transmission in frame time t .

$$\begin{aligned} \Lambda_{ijt} &= \Lambda_{ij(t-1)} - \Delta_{ijt} \\ \forall i, j \in S, \forall t \in T \text{ and } t \neq 1 \\ i &\neq j \end{aligned} \tag{12}$$

Equation (12) present the not served traffic to be scheduled for transmission in subsequent frame times following frame time t .

$$\sum_{\substack{i \in P[k] \\ i \neq j}} \sum_{j \in P[q]} \Delta_{jit} \leq Cd \tag{13}$$

$$\sum_{\substack{i \in P[k] \\ i \neq j}} \sum_{j \in P[q]} \Delta_{ijt} \leq Cu \tag{14}$$

$$\forall k, q \in PG, \forall t \in T \text{ and } t \neq 1$$

Constraint (13) and (14) ensures that uplink and downlink wavelengths' capacities (10 Gb/s) for each PON group are not exceeded while assigning resources to demanding servers in the designated frame times.

$$S_{it} = \sum_{\substack{j \in S \\ i \neq j}} Z_{ijt} \tag{15}$$

$$\forall i \in S, \forall t \in T \text{ and } t \neq 1$$

$$S_{jt} = \sum_{\substack{i \in S \\ i \neq j}} Z_{ijt} \tag{16}$$

$$\forall j \in S, \forall t \in T \text{ and } t \neq 1$$

Equations (15) and (16) allocate the set of granted source and destination ONUs for transmission in frame time t respectively.

$$\phi_{ijt} \geq S_{it} \tag{17}$$

$$\phi_{ijt} \geq S_{jt} \tag{18}$$

$$\begin{aligned}
\phi_{ijt} &\leq S_{it} + S_{jt} \\
\forall i, j \in S, \forall t \in T \\
i = j \text{ and } t \neq 1
\end{aligned} \tag{19}$$

Constraints (17-19) are binary equivalent to indicate one if either of i or j , or both i and j are granted for transmission in frame time t , otherwise equal zero.

$$\begin{aligned}
\xi_{ijt} &= \sum_{i \in S} \sum_{j \in S} \phi_{ijt} \\
&\quad i=j \\
\forall t \in T \text{ and } t \neq 1
\end{aligned} \tag{20}$$

Equation (20) calculates the total number of ONUs in TDM frame time t scheduled for transmission.

$$\sum_{t \in T} \sum_{i \in S} \sum_{j \in S} \Delta_{ijt} = \sum_{i \in S} \sum_{j \in S} \alpha^{ij} \tag{21}$$

$t \neq 1$ $i \neq j$ $i \neq j$

Constraint (21) is the demand satisfaction constraint to ensure granted demand is equal to the total demand.

5. Results and Discussions

In this section we evaluate the performance of the described architecture depicted in **Figure 1** in terms of scheduling against average delay and power consumption for servers hosted in the PON cell.

The model with the objective to minimize total delay aims for minimizing total number of frame times needed and hence minimize the overall average waiting time to service all queued demands. Minimization of frame times maximizes the number of granted requests in each frame time by finding the largest possible conflict-free matching between source and destination servers. Maximizing the matching can efficiently reduce number of iterations/frame times to successfully grant all demands and hence reduce the waiting time for queued requests. While minimizing total delay, the model ensures that uplink and downlink bandwidth assignment from and to each ONU in the designated PON group is not exceeding the capacities of the assigned wavelengths. Hence, while attempting to achieve best maximum match for a frame time, the output of the matching graph is mastered and constrained by the size of flows to be received and transmitted by each ONU.

The modeled PON cell is depicted in **Figure 1** with **Figure 4** PON groups each of which is hosting 4 servers. In the evaluation different considered loads reflect the size of queued traffic at each node arrived at the same time. The examined traffic patterns are based on random selection of communicating pairs of nodes to reflect the overall loads. Scenarios with hotspot nodes are also considered in the evaluation. A hotspot node is a server where all other servers within the PON cell having a queued request for. Average waiting time of a request is governed by the size of load (number of queued requests) and the total time needed for serving all demands. Average delay is calculated using Equation (22) [16].

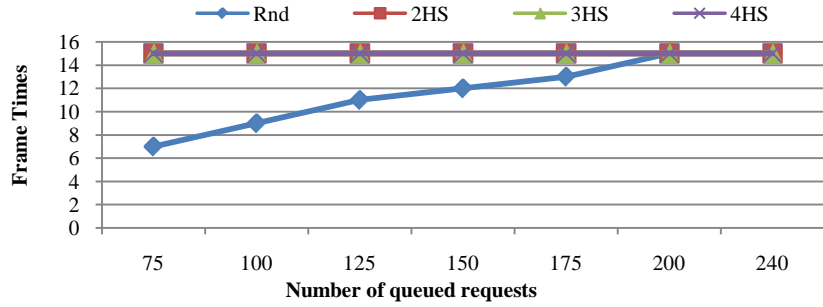


Figure 4. Number of frame times versus load for random traffic, random with 2, 3, and 4 hotspots for the delay-minimized objective.

$$\text{Average Delay} = \frac{\sum_{t \in T} \left((t-2) \sum_{i \in S} \sum_{j \in S} Z_{ijt} \right)}{\text{Tot Number of Req}} \quad (22)$$

Power consumption is governed by the total number of ONUs needed in all TDM frame times and is calculated using Equation (2). For comparison purposes, two objective functions are evaluated; one with minimization of delay and the second with minimization of power consumption. Minimization of delay approach target the minimization of average waiting time per request to be scheduled and serviced and can provide the optimum number of TDM frame times needed. For the model with minimization of power consumption, number of TDM frame times is obtained from the minimization of delay objective and will be given as an input parameter for the minimization of power consumption model. **Figure 4** presents the optimum number of TDM frame times for the different examined loads.

Power consumption of ONUs (PC_ONUs) is evaluated with respect to traffic load and activity. For further energy savings, ONUs can be switched off if not scheduled by OLT for transmitting or receiving in the designated frame time. The power consumptions of the ONU during the on and off times are 2.5W and 0 W [23]. The approach is to compare the power consumption when ONUs are always on if it was active or not and when ONUs are off or on sleep mode during inactive state. For the different traffic scenarios, as the requirement of frame times increases, more power is consumed. If no energy awareness of inactivity is enforced to shut off idle ONUs, overall power consumption tends to increase as ONUs need to be kept on in all frame times until all requests are served. If energy awareness is enforced, traffic scenarios requiring higher frame times will result in considerable savings of power.

Figure 4 and **Figure 5** present the number of frame times and the saving percentages of power consumption for the different examined traffic patterns, respectively. Savings can reach 55% at low loads when many servers are inactive allowing longer times for ONUs to be put in switch-off or sleep mode. Traffic with hot spots scenario requires higher number of frame times, therefore by enforcing sleep mode for inactive ONUs results in more power savings. **Figure 6** and **Figure 7** present the power consumption of the two examined objective

functions. The two objective functions produce the almost same power consumption results for the different examined loads as the same number of ONUs need to be switched on for both cases to serve all demands regardless of the number of frame times needed. The power consumption increases linearly as the

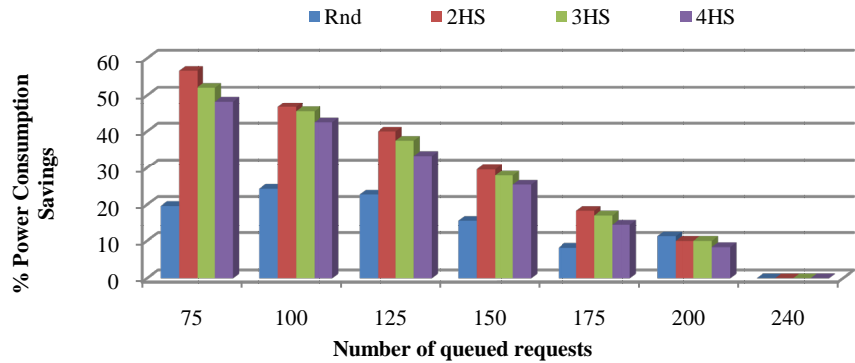


Figure 5. ONUs sleep mode enabled power consumption savings versus load for random traffic, random with 2, 3, and 4 hotspots for the delay-minimized objective.

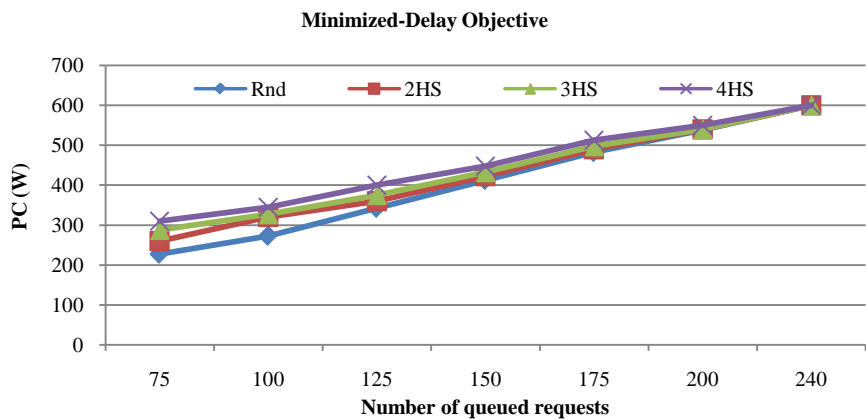


Figure 6. Power consumption versus load for random traffic, random with 2, 3, and 4 hotspots for the delay minimization objective with ONUs sleep mode enabled.

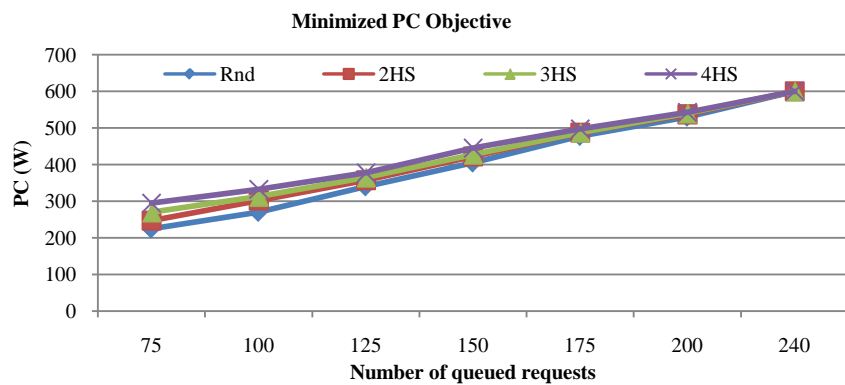


Figure 7. Power consumption versus load for random traffic, random with 2, 3, and 4 hotspots for the power consumption minimization objective with ONUs sleep mode enabled.

load increases for both cases. Introduction of hot spots causes an increase in the total power consumption for each load applied. For low loads such as the case with 75 requests (30% load), if we compare random traffic with the traffic with 4 hot spots, power consumption increases by 26.6%. This is mainly due to the selection of the granted servers in each frame time as with random scenario more options for granting are available, unlike the case with hotspots where more frame times are needed and less number of demanding servers competing for a grant in each frame time.

Evaluation results for power consumption against the two objective functions presented in **Figure 8**, which have shown negligible difference. Maximum of 6% savings can be achieved with the model of minimization of power consumption if compared with the minimized-delay model.

Random selection of demanding pairs of servers shows an increase of request average waiting time as the load increases. Results shows average delay ranges from 2 to 7 frame times for 30% (75 requests) to 100% loads (240 requests) respectively. Introducing hotspots result in unbalanced load assignments; where some nodes are targeted heavily. Hotspots increases average cell delay and degrade the throughput especially at lower loads as it introduces more frame times for conflict free pair matching of demanding servers.

Considerable savings reaching up to 42% for average delay are achieved with the minimization of delay objective compared to the minimization of power consumption model. This is justified by the efficient scheduling through minimization of average waiting time per request achieved by the minimized-delay objective. **Figure 9** and **Figure 10** show the average delay in frame times for the minimized delay and minimized power objectives respectively. Percentage of delay reductions between the objectives is shown in **Figure 11** for the different evaluated traffic scenarios against the different loads. The reduction in the delay as appears in **Figure 11** is a result of the minimization of the waiting time for the queued requests to be served through the efficient utilization of the resources.

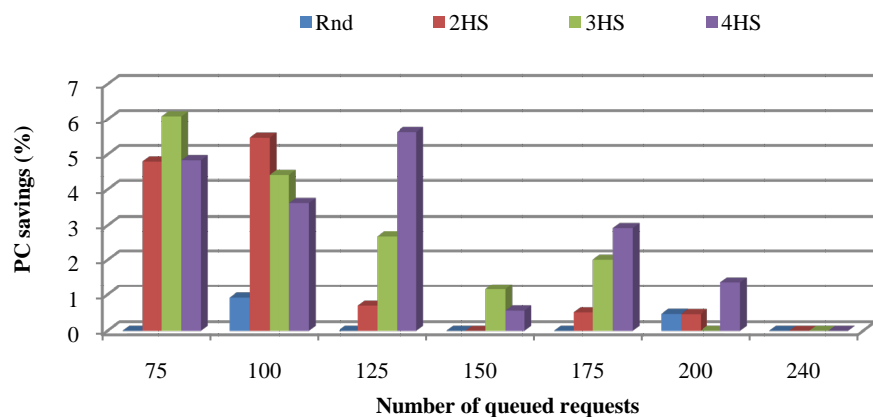


Figure 8. Power consumption savings for the minimized power consumption objective against the minimized delay objective for random traffic, random with 2, 3, and 4 hot-spots.

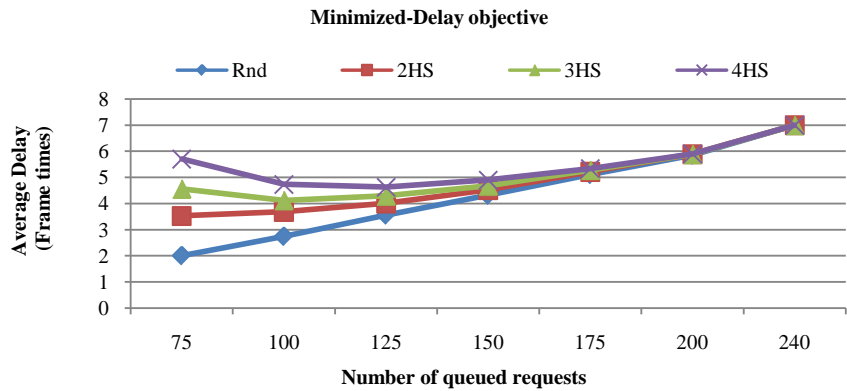


Figure 9. Average delay versus load for random traffic, random with 2, 3, and 4 hotspots with ONUs sleep mode enabled.

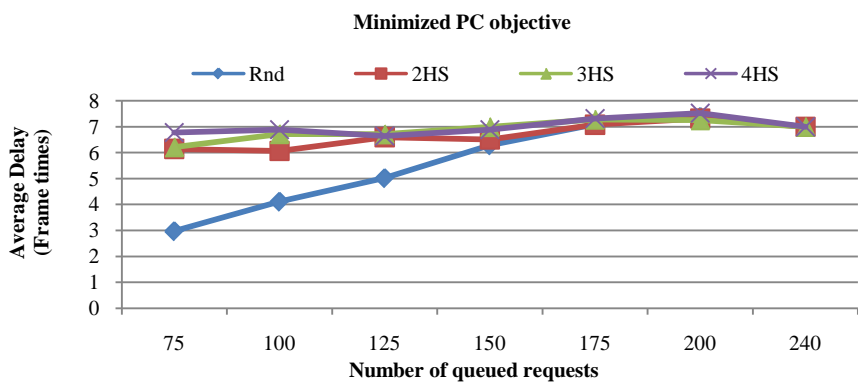


Figure 10. Average delay versus load for random traffic, random with 2, 3, and 4 hotspots with ONUs sleep mode enabled.

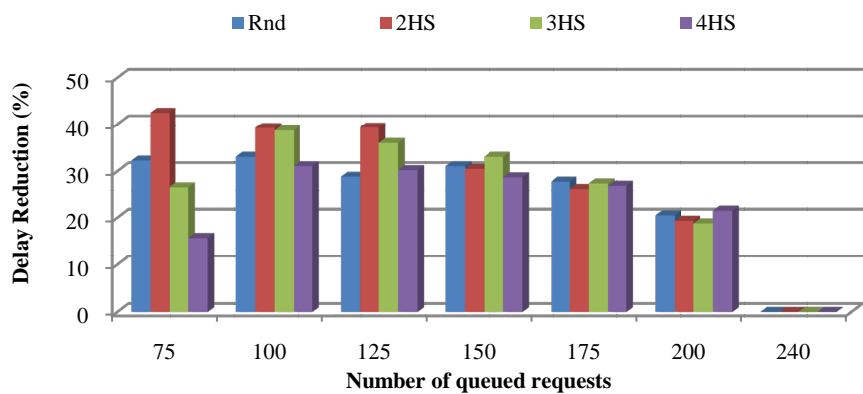


Figure 11. Average delay reduction for the minimized power delay objective against the minimized power consumption objective for random traffic, random with 2, 3, and 4 hotspots.

The less number of transmission frames needed, the lower is the total time required to serve queued demands. Therefore; the objective with the minimization of total delay targets the minimization of overall frames needed and hence minimize the overall queuing delay of the network. While the objective with power consumption minimization targets the reduction of the total number of switched

on devices when serving queued demands regardless of the overall time needed to serve the demands.

6. Conclusions

In this paper, PONs is considered in the design of the fast-switched and delay minimized fabric interconnection for future data centers networks. A mathematical optimization model using MILP was developed and presented to minimize the overall network delay through efficient scheduling and resource assignments. Different loads with random unbalanced traffic representing the nature of traffic in data centers have been evaluated. Results have shown that power savings with sleep mode enabled for the different examined traffic patterns range between 8% and 55% during low and high load activities respectively. Model with delay minimized scheduling has shown reduction in average delay reaching 42% with a trade-off of negligible maximum increase in power reaching 6% if compared with the model with minimized power consumption. The work presented in this article has been limited to mathematical modeling. However, the design of real time heuristics for hardware implementation and also to validate and verify the results obtained from the MILP mathematical model is part of the planned future work.

References

- [1] Hammadi, A. and Mhamdi, L. (2014) Review: A Survey on Architectures and Energy Efficiency in Data Center Networks. *Computer Communication*, **40**, 1-21. <https://doi.org/10.1016/j.comcom.2013.11.005>
- [2] Elmirghani, J., EL-Gorashi, T. and Hammadi, A. (2016) Passive Optical-Based Data Center Networks. Patent No. WO2016083812 A1.
- [3] Luo, Y.Q., Effenberger, F. and Meng, S. (2012) Cloud Computing Provisioning over Passive Optical Networks. *2012 1st IEEE International Conference on Communications in China (ICCC)*, Beijing, 15-17 August 2012, 255-259. <https://doi.org/10.1109/ICCCChina.2012.6356888>
- [4] Ji, P., Qian, D., Kanonakis, K., Kachris, C. and Tomkos, I. (2012) Design and Evaluation of a Flexible-Bandwidth OFDM-Based Intra Data Center Interconnect. *IEEE Journal of Selected Topics in Quantum Electronics*, **19**, 3700310. <https://doi.org/10.1109/JSTQE.2012.2209409>
- [5] Kachris, C. and Tomkos, I. (2011) Power Consumption Evaluation of Hybrid WDM PON Networks for Data Centers. *2011 16th European Conference on Networks and Optical Communications (NOC)*, United Kingdom, 20-22 July 2011, 118-121.
- [6] Wang, K., Zhao, L., Gu, H., Yu, X., Wu, G. and Cai, J. (2015) ADON: A Scalable AWG-Based Topology for Datacenter Optical Network, *Optical and Quantum Electronics*, **47**, 2541-2554. <https://doi.org/10.1007/s11082-015-0136-z>
- [7] Indre, R.M., Pesic, J. and Roberts, J. (2014) POPI: A Passive Optical Pod Interconnect for High Performance Data Centers. *2014 International Conference on Optical Network Design and Modeling*, Stockholm, 19-22 May 2014, 84-89.
- [8] Cai, Y., Luo, S. and Zhou, L. (2016) Software Defined Passive Optical Intra-Rack Networks in Data Centers. *2016 IEEE Globecom Workshops (GC Wkshps)*, Washington DC, 4-8 December 2016, 1-6.

- <https://doi.org/10.1109/GLOCOMW.2016.7849048>
- [9] Hammadi, A., El-Gorashi, T.E.H. and Elmirghani, J.M.H. (2015) High Performance AWGR PONs in Data Centre Networks. *17th International Conference on Transparent Optical Networks (ICTON)*, Budapest, 5-9 July 2015.
- [10] Al-Fares, M., Loukissas, A. and Vahdat, A. (2008) A Scalable, Commodity Data Center Network Architecture. *Proceedings of the ACM SIGCOMM 2008 Conference on Data Communication*, Seattle, 17-22 August, 63-74.
<https://doi.org/10.1145/1402958.1402967>
- [11] Guo, C., Lu, G., Li, D., Wu, H., Zhang, X. and Shi, Y. (2009) B Cube: A High Performance, Server-Centric Network Architecture for Modular Data Centers. *SIGCOMM Computer Communication*, **39**, 63-74.
<https://doi.org/10.1145/1594977.1592577>
- [12] Hammadi, A., El-Gorashi, T.E.H. and Elmirghani, J.M.H. (2016) Energy-Efficient Software-Defined AWGR-Based PON Data Center Network. *18th International Conference on Transparent Optical Networks*, Trento, 10-14 July 2016, 1-5.
- [13] Hammadi, A., El-Gorashi, T.E.H., Musa, M.O.I. and Elmirghani, J.M.H. (2016) Server-Centric PON Data Center Architecture. *18th International Conference on Transparent Optical Networks*, Trento, 10-14 July 2016, 1-4.
- [14] Hammadi, A., El-Gorashi, T.E.H., Musa, M.O.I. and Elmirghani, J.M.H. (2016) Resource Provisioning for Cloud PON AWGR-Based Data Center Architecture. *21st European Conference on Networks and Optical Communications*, Lisbon, 1-3 June 2016, 178-182.
- [15] Cisco: Data Sheet of Cisco-ME 4600 Series Optical Line Terminal Data Sheet.
- [16] Chao, H.J. and Liu, B. (2007) High Performance Switches and Routers. John Wiley & Sons, Inc., Hoboken. <https://doi.org/10.1002/0470113952>
- [17] McKeown, N. (1997) White Paper: A Fast Switch Backplane for a Gigabit Switched Router. *Business Communications Review*, **27**.
- [18] LaMaire, R.O. and Serpanos, D.N. (1994) Two-Dimensional Round-Robin Schedulers for Packet Switches with Multiple Input Queues. *Transactions on Networking*, **2**, 471-482. <https://doi.org/10.1109/90.336324>
- [19] Lund, C., Phillips, S. and Reingold, N. (1996) Fair Prioritized Scheduling in an Input-Buffered Switch. *Proceedings of the IFIP-IEEE Conference on Broadband Communications*, Montreal, 23-25 April 1996, 358-369.
https://doi.org/10.1007/978-0-387-34987-9_30
- [20] Hui, J. and Arthurs, E. (1987) A Broadband Packet Switch for Integrated Transport. *Journal of Selected Areas Communications*, **5**, 1264-1273.
- [21] Ali, M. and Nguyen, H. (1989) A Neural Network Implementation of an Input Access Scheme in a High-Speed Packet Switch. *Proceedings of GLOBECOM*, Dallas, 27-30 November 1989, 1192-1196. <https://doi.org/10.1109/glocom.1989.64143>
- [22] Chao, H.J., Kung-Li, D. and Jing, Z. (2003) PetaStar: A Petabit Photonic Packet Switch. *Selected Areas in Communications*, **21**, 1096-1112.
<https://doi.org/10.1109/JSAC.2003.815678>
- [23] Grobe, K., Roppelt, M., Autenrieth, A., Elbers, J.P. and Eiselt, M. (2011) Cost and Energy Consumption Analysis of Advanced WDM-PONs. *Communications Magazine*, **49**, s25-s32. <https://doi.org/10.1109/MCOM.2011.5706310>

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact ijcns@scirp.org