Volume 7, Issue 2
April 2018

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

©2012-18International Journal of Information Technology and Electrical Engineering

A Critical Review of Next Generation Data Centers

[1]Pronaya Bhattacharya, [2]Amod Kumar Tiwari, [3]Rajiv Srivastava
[1]Dr. A.P.J. Abdul Kalam Technical University, Lucknow,
Uttar Pradesh, INDIA

[2]Rajkiya Engineering College, Churk, Sonbhadra,
Uttar Pradesh, INDIA
[3]Ex Faculty, Indian Institute of Technology, Jodhpur,
Rajasthan, INDIA
E-mail: [1]pranayphdk@gmail.com, [2]amodtiwari@gmail.com, [3] rajivs18@gmail.com

ABSTRACT

Demand for higher bandwidth is increasing as the day progresses. Current electronic switching used in data centers are now facing bottleneck to match data rates. To alleviate such problem, fiber optic technology is used in backbone network, and switching is still performed using electronic circuitry after performing O/E conversion. However, till recent past, data transfer technology heavily relies on electronics. Therefore, researchers propose the used of both electrical and optical technology for high speed data transfer. Data center technology is emerging over the period of time. However, due to the commercial un-availability of tunable wavelength converters all-optical data centers is not feasible till date. In the recent past many data centers core switch designs are proposed, which uses both electronic packet switching (EPS) and optical circuit switching (OCS) depending on network requirements, and rate at which data arrives. This paper presents the overview of the recently proposed data centers design along with their pros and cons. This paper reviews both well established and up-coming data centers design which can be considered as near future designs.

Keywords: *Data Centers, OPS, OCS and Bursty data*

1. INTRODUCTION

Data centers are physical or essential infrastructure used by enterprises to computer, server, network systems and components for the company's Information Technology (IT) needs, which usually involve storing, processing and serving huge amounts of critical data to clients in a Client/ Server environment [1]. In the present day environment most of the data center applications are provided free of charge, that's why datacenter operators are facing a big problem of meeting exponentially increasing demands for network bandwidth without excessive increase in power and infrastructure cost [2-4]. Currently, a datacenter typically contains tens of thousands of servers that form a massively parallel super-computing infrastructure. For improving the data center operations fiber optic technology can play a major role. The feasibility of wavelength routing in data centers relies primarily on wavelength tuning device that are ubiquitous in wavelength-routing network architecture. Tunable wavelength converters (TWCs) that enable dynamic routing of optical signals are the critical building blocks of a wavelength-routing interconnect due to their large numbers, as well as tuning range, speed, wavelength stability, and electronic control requirements. Although all-optical TWCs are not yet commercially available, recent demonstrations of highly-integrated optical switching infrastructure and high-performance, power-efficient wavelength conversion an filtering based on nano-scale silicon photonics [5]-[7] pave the way for building large-scale, cost-efficient WDM interconnects. Optical packet switching provides many advantages over their electronic counterparts, but lack of optical RAMs is major bottleneck. In electronic RAMs millions of packets can be stored for longer duration, but in fiber delay lines (for temporarily storage in optical switching) some hundreds of packets can be stored for very short durations. Thus, an efficient design of optical switch is necessary for loss-less system.

2. CHALLENGES IN DATA CENTERS

In this section we primarily focus on the current challenges associated with the data center architectures.

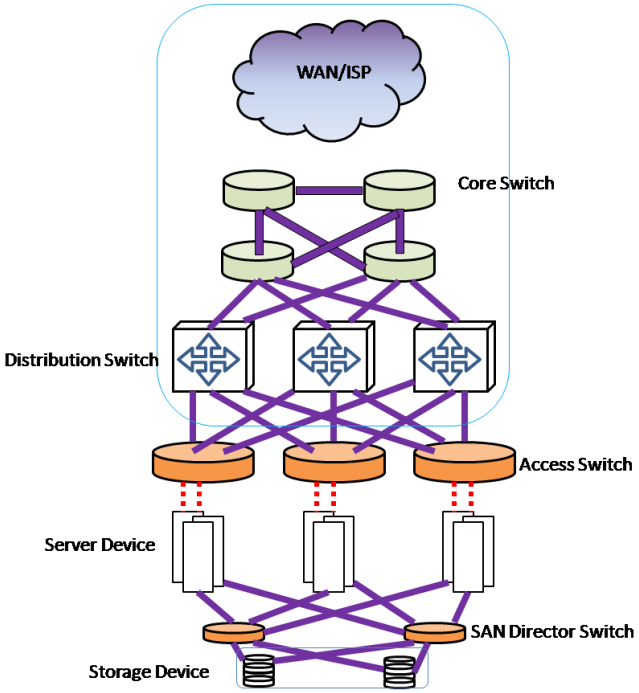

Figure.1: Hierarchal architecture of current data centers.

The Figure 1 shows an hierarchal system of current data center system. In this figure the servers present at the lowest

Volume 7, Issue 2
April 2018

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

©2012-18International Journal of Information Technology and Electrical Engineering

level of hierarchy, are organized in to the server racks (usually consist of 40-80 blades). Each of the blade server is connected with the Top of the Rack (TOR) switch with a 1 Gbps Ethernet link. For redundancy purpose, each of the TOR switch is connected with more than one aggregation switch and the traffic through from these switches aggregates further through routers. At the top level the core routers are used to connect data center with the internet (multi routed tree topology). Here all link uses Ethernet as the physical layer protocol, with a mix of copper and multi-mode fiber cabling. All switches below each pair of access routers form a single layer-2 domain, typically connecting thousands of servers. This hierarchal system has some disadvantages which makes it an inappropriate choice in the case of large data center system.

A. Limited Scalability

As the traffic flowing within the data centre network continues to grow, so this is very difficult to manage the traffic through the switches based on Ethernet links. According to the recent survey report given by Cisco, the global datacenter IP traffic will nearly triple during 2012-17, this measure the compound annual growth rate of about 25% (up from 220 exabytes per month in 2013 to 648 exabytes per month in year 2017). In next several years we can expect the doubling of transistor density on chips in every 18-24 months. This stems from the possibility of integrating multiple cores on a single die, the multicore processor technology. The hierarchal design cannot support the growing traffic heavy data center systems that will be equipped with more numbers of servers and also more microprocessor cores on each server.

B. Power Consumption

The one of the most challenging issue in the deployment of data center systems is the power consumption. According to the recent Greenpeace survey report, the global demand of electricity in data center system is expected to triple from 340 billion kWh in 2006 to 1000 billion kWh in 2020. Studies also suggest that, in data center system around 40%, 37% and 23 % of the IT power is consume in servers, storage and in networking equipment's respectively. That's why the savings in the power consumption of network elements makes an positive impact on the overall power consumption of data center sites. Electronic links and switches in the data center design do not leave much space for improvement in this area as they are power hungry devices especially as the bit rate increases. The other shortcoming of current designs in terms of energy efficiency arises from the fact that the power consumption of servers is not necessarily low when they are idle. They could consume as much as 60 percent of their peak power when idle [12]. Measurements on current deployments have recorded average server utilization as low as 30% [13], indicating a huge energy wastage due to idling hardware starved for data.

3. OPTICAL CIRCUIT SWITCHING

There is a committed communication path between the sending and receiving devices in circuit switching. The dedicated or committed path is a connected sequence of links switching nodes. Communication through circuit switching includes three stages. These are circuit establishment; data transfer; and circuit termination. Circuit switching is primarily used for voice telephone network, but it is not that much fruitful for data communication network, as channel capacities are not completely utilized, as data communication equipments do not produce data persistently

Native optical packet switching (OPS) has for quite some time been an objective of the optics group. Various basic difficulties leave this vision a leap forward far from broad commercial adoption. While anticipating such an achievement, OCS guarantees to drastically change the substance of the data center. OCS holds a number of benefits relative to electronic packet switching (EPS). OCS holds various advantages with respect to EPS. OCS is (to a great extent) information rate agnostic and to a great degree energy effective. MEMS-based OCS essentially reflects light starting with one port then onto the next port; so as the information rate enhances and in addition the quantity of per-port wavelengths expands, an OCS can scale without substitution. In the same way, due to the fact that there is no per- packet processing, there is no additional latency, and per-bit consumption of energy can be requests of magnitude not more than EPS partners.

Data center economic, scale and performance challenges impose a number of requirements for OCS hardware:

• *Lower Cost*: The integrated MEMS-based OCS expense is at present an obstacle to get in the data center. In the meantime, the fundamental chip innovation is inherently economical.

• *Larger scale*: The biggest OCS we know about at present backings couple of many duplex ports. With the purpose of integration into data centers even though they are of moderate level, OCS must be scaled to several hundred or may be tens of thousands of ports.

• *Faster switching time*: Commercial OCS time of switching is commonly between 10-20 ms. Such switching times are to a great extent driven by the prerequisites of the telecom industry, which just needs failover under 50 ms. While, speaking about the other side of the spectrum, per-packet switching would need switching times calculated in nanoseconds. It is concluded by us that in the data center, there exists a number of options for largescale OCS supporting under 100 µs switching time.

• *Lower insertion loss*: As of now, insertion loss differs relying upon the correct port combination and coupling method utilized as a part of a large-scale OCS, yet goes upto 5dB. While providing support to bigger scale optical circuits witches and coordinating economical optical transceivers with direct link power budget plan into the data center needs driving down the insertion loss with the help of the OCS, in a perfect situation of being under 2 dB.

4. NOTABLE DESIGNS

In this section notable switch designs are presented with their advantages and dis-advantages.

A. C-Through

Cut-through is a hybrid electro-optic switch for data-center application. It combines the advantages of traditional electrical packet switch and new developed optical circuit switch. This architecture was proposed by Wang. The configuration of the network, is shown in Figure 2, consists of a tree-structure of electrical network with ToR switches at the bottom of the tree. These ToR switches are accessed using aggregated switch and these aggregated switches are further connected to core switch sitting on top of hierarchy. ToR switches also connects to

Volume 7, Issue 2
April 2018

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

©2012-18International Journal of Information Technology and Electrical Engineering

optical network. Due to the higher cost of optical components, ToR switches are connected to re-configurable fiber links. This arrangement reduces cost of optical network. It is customary to note that, ToR to ToR connections are established using fiber links. Depending on traffic demands new connections can be established among different ToR switches. In c-through network, connections are established Edmonds' algorithms for the maximum weight perfect matching problem [14].

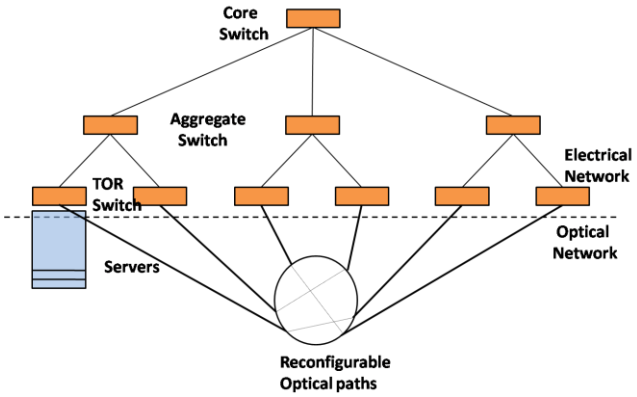

Figure 2: Design of C-Through switch design

For successful operation of the network, servers in control plane monitor the bandwidth requirements with other hosts and buffer limits of the sockets are set.

Advantages:

1. Depending on traffic flow and network requirements both electrical and optical networks can be used simultaneously.
2. This technique is effective in case of loose synchronization, bulk data transfer.
3. Very simple ad cost effective optical switching design.

Dis-advantages:

1. This design is not easily scalable.
2. It is not easy to have so many servers in data centers.
3. When two ToR switches try to connect using full bandwidth simultaneously, to third ToR switch then bottleneck will be faced.

B. Helios

Helios is one of the notable projects in data centers applications. This design is also hybrid in nature, means it uses both electrical and optical switches (Figure 3). The Helios was proposed by Farrington et al., and structural design is shown in figure this design is similar to c-through switch design as it s also a two level data center networks, but it is based on WDM links. This switch is a combination of core and ToR switches. Core switches can be optical or electronic as shown in figure. ToR switches are electronic packet switches [15].

The electrical packet switches in Helios are used for all-to-all communication of the ToR switches which is used to distribute the bursty traffic. While the optical circuit switches offer high bandwidth slowly changing traffic and long lived communication between the pod switches. Same as c-through, the Helios architecture tries to make full use of the optical and the electrical networks.

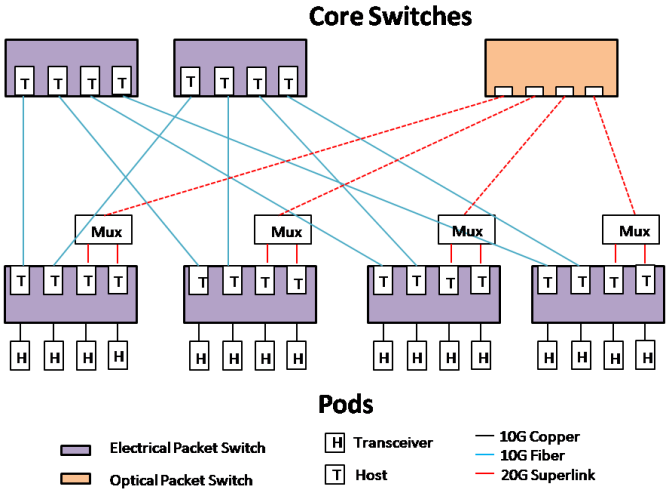

Figure 3: Structure design of Helios switch

Each of the ToR switches comprises of transceivers. Half of the transmitters connect electronic core switches. In the simple arrangement it is make sure that each of the ToR switch connects to all of the core switches. Other half optical transceivers on each ToR are used for connecting to optical core switches through a passive optical multiplexer in the form of super links for full flexible bisection bandwidth assignment.

In the optical circuit switches, Helios chooses MEMS technology, which is not only power constant in independence of bandwidth, but also consumes much less power compared to electronic packet switches. Also, in MEMS system, there is no optical-electronic signal conversion through the full crossbar mirrors switches, which leads to high performance and less delays.

Helios uses two algorithms for configuring the maximal traffic demand. One is from Hedera to allocate rack-to-rack bandwidth share. The other is Edmonds' Algorithm, which is also used in c through for solving the maximum weight match problem.

The software of Helios control scheme is based on three primary components: Pod Switch Manager, Circuit Switch Manager and Topology Manager. Every module has a distinct role to act coordinately when required and the relationship between all of them.

Pod Switch Manager provides statistical data about traffic sent out from its pods. It interfaces with the Topology Manager and configures the switch appropriately based on the input from traffic routing decision made. The Pod Switch Manger is set to rout traffic accordingly either through the WDM transceivers from the optical circuit switch or the colour-less transceivers.

Advantages:

1. Helios is deployable for commercially available optical modules and transceivers to use in optical communication networks.
2. There is no need for end-host or switch hardware modifications.

Dis-advantages:

1. The main drawback concerns the issue with the reconfiguration time of the MEMS switches.
2. The inherent limitation of electronics requires several milliseconds for the process, which isseen to be long.

Volume 7, Issue 2
April 2018

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

©2012-18International Journal of Information Technology and Electrical Engineering

C. Calient

Calient is a renounced company in the field of telecom industry. Recently Calient has commercialized high-level hybrid data centre, which is a blend of both the packet switching and optical circuit switching (OCS). [16] This switch is designed for both light and heavy data conditions. When information arrives in hurry is well known as bursty traffic arrival. The calient switch supports bursty traffic data using OCS. Using OCS a higher throughput can be obtained, with low latency. Under lower loading conditions, the system uses typical ToR switches. Calient uses a software-defined networking (SDN) to separate the control plane from the data plane.

Advantages:
1. It supports bursty traffic arrivals.
2. It provide unlimited bandwidth capacity.
3. Latency is small, and throughput is higher.
4. Scaling can be done very easily.

D. Mordia

Merida (Microsecond Optical Research Data-center Interconnect Architecture) is a 24-node hybrid switch architecture design, which uses optical circuit switching (OCS) and wavelength selective switch (WSS)[17] The main advantage of design is that it operates in microsecond-scale OCS technology which is fast enough to completely replace electronic packet switches. This switch design supports a wide range of services and applications.

Advantages:
1. Mordia switch has total time of 11.5μs of the optical circuit switch, including the signal acquisition by NIC. With such low time more than one application can be run simultaneously.
2. With multiple parallel rings, Mordia can possibly be scaled to build up large bisection bandwidth data centres required in the future.

Dis-advantages:
1. it is not compatible with Ethernet packet granularity.
2. Cost is very high, because of use of WSS.
3. If any point link cut, complete ring dies out.

E. REACToR

This switch design uses hybrid ToRs and combines of packet and circuit switching known as REACToR is proposed by California University, San Diego.[18] it is fast operated design, and in terms of reaction time it is superior to other compatible designs (Figure 4).. REACToR is an advanced version of Mordia. In REACToR, the optical circuit switching used to connect ToRs directly in the data centre, thus reduces cost and reduces the need of optoelectronic transceivers [16].
There are two important design features in REACToR. Firstly, in case of optical circuit switching, low cost buffering is provided. It support both electronic packet switched at the rate of 10Gb/s while optical circuit switching at the rate of 100 Gbps.

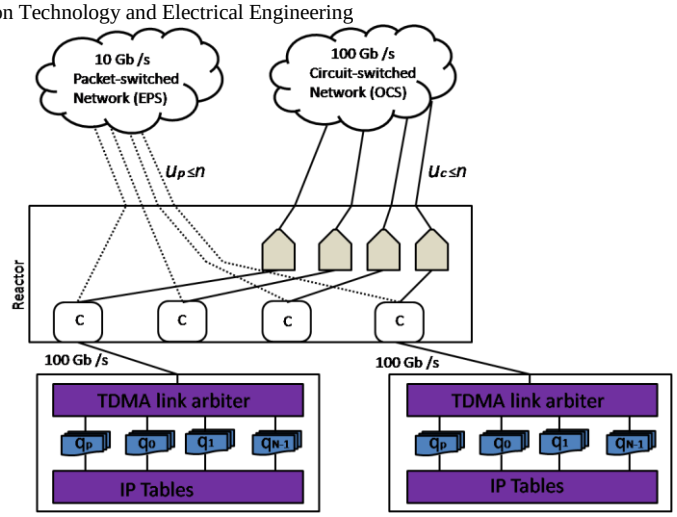

Figure 4: Structure design of REACToR switch

The designed protocol makes sure that both EPS and OCS can be used efficiently, using TDMA. This type of design also reduces latency.

Advantages:
1.This design of packets provides packet-switch-like execution with sufficient data transfer capacity usage.
2. This design is faster than MORDIA switch.
3. The model can possibly scale to serve data center requests with its hybrid leverage of joining the packet and circuit switches.

Dis-advantages:
1. It is difficult to configuration interconnect among numerous REACToRs and furthermore the synchronization is difficult.

F. NACK Based Optical Switch

In the above designs, sometimes information is blocked or dropped. In the similar context an AWGR based scheme is proposed in [19], here in case of blocking or dropping of packet a Negative Acknowledgement (NACK) is send to the senders to notify and senders re-send the packet again as shown in Figure 5. In this design a buffer-less AWGR is considered. In the figure, the path between circulators C_2 and C_3 is used for the label extractor where old label is extracted and after O/E conversion fed to the electronic controller, and new label is inserted and fed with payload at the output of the switch.
In Figure 5, the architecture proposed by Proietti et al. is shown. Here T_i represnets the i^{th} transmitter, and R_i represnets the i^{th} reciever. Here the generated information is first pases through the circulator C_1 and then passes thorugh FDL of length D, and circulator C_2. Over here, label is extracted and as per the destination wavelenth of the incoming paket is tuned to direct them to the appropiate output port of the AWG.

Volume 7, Issue 2
April 2018

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

©2012-18International Journal of Information Technology and Electrical Engineering

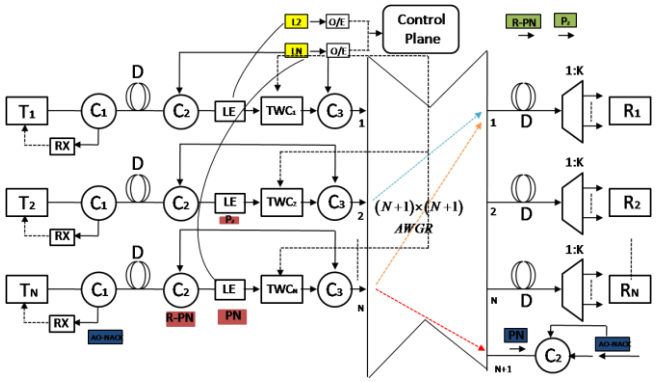

Figure 5: Structure design of NACK based optical switch

In case of blocking, appropiate branch TWC tunes the wavelength of the packet such that it is blocked by AWG, and the AWG acts as reflecting grating, and reflected information choosing other path it appears at circulator C_1 and it is transferred to appropiate transmitter where information is detected with assoicated Rx, and this negative acknowledgement mechanism allows the re-transmission of bocked packets.

Advantages:

1. In this design packet will not be dropped.
2. Due to absence of buffer, control unit design is simple.
3. WDM can easily be incorporated.

Dis-advantages:

1. Due to absence of buffer a large number of packets will be re-transmitted.
2. The re-transmission of large number of packets, will lead to the congestion of network.

5. CONCLUSIONS

The use of optical technology is not only limited to switching technology but also in data centers applications. But due to limited technological advances, optical switching cannot be applied. Therefore, both electrical and optical technologies are used simultaneously. This paper discusses the notable switch designs in the field of optical communications and data centers technologies. Full detail description of the switches is detailed, with their advantages and dis-advantages. After critical evaluation of the switch designs followings conclusions are made:

1. In the next a few years both electrical and optical technologies will co-exist.
2. For high speed data networks and for bursty traffic arrivals OCS will remain as next generation technology.
3. In the current scenario MEMS and WSS will be preferred in wavelength switching due to the unavailability of TWCs.
4. SDN will play a major role in next generation data transfer technology.
5. In the next generation both techniques OPS and OCS will be used depending on the data rates of arriving traffic.

REFERENCES

[1] M. Al-Fares, A. Loukissas, and A. Vahdat, "A Scalable, Commodity Data Center Network Architecture," in *Proc. SIGCOMM*, 2008.

[2] L. Aronson, B. Lemoff, L. Buckman, and D. Dolfi. Low-Cost Multimode WDM for Local Area Networks up to 10 Gb/s.*IEEE Photonics Technology Letters*, 10(10):1489–1491, 1998.

[3] K. Barker and etal., "On the feasibility of optical circuit switching for high performance computing systems," in *Proc. SC*, 2005.

[4] K. V. Vishwanath, A. Greenberg, and D. A. Reed. Modular Data Centers: How to Design Them? In *Proc. of the 1st ACM Workshop on Large-Scale System and Application Performance (LSAP)*, 2009.

[5] Priolo, F., Gregorkiewicz, T., Galli, M. and Krauss, T.F., 2014. Silicon nanostructures for photonics and photovoltaics. *Nature nanotechnology*, *9*(1), p.19.

[6] Hu, H., Ji, H., Galili, M., Pu, M., Peucheret, C., Mulvad, H.C.H., Yvind, K., Hvam, J.M., Jeppesen, P. and Oxenløwe, L.K., 2011. Ultra-high-speed wavelength conversion in a silicon photonic chip. *Optics Express*, *19*(21), pp.19886-19894.

[7] Daldosso, N. and Pavesi, L., 2009. Nanosilicon photonics. *Laser & Photonics Reviews*, *3*(6), pp.508-534.

[8] G. Wang, D. G. Andersen, M. Kaminsky, M. Kozuch, T. S. E. Ng, K. Papagiannaki, M. Glick, and L. Mummert. Your Data Center Is a Router: The Case for Reconfigurable Optical Circuit Switched Paths In*ACM HotNets '09*.

[9] T. Benson, A. Anand, A. Akella, and M. Zhang, "The Case for Fine-grained Traffic Engineering in Data-centers," in *Proc. USENIX INM/WREN*, 2010.

[10] T. Benson, A. Akella, and D. Maltz, "Network Traffic Characteristics of Data Centers in the Wild," in *Proc. IMC*, 2010.

[11] K. Chen, A. Singla, A. Singh, K. Ramachandran, L. Xu, Y. Zhang, X. Wen, and Y. Chen, "OSA: An Optical Switching Architecture for Data Center Networks with Unprecedented Flexibility," Northwestern University, Tech. Rep., 2012.

[12] A. Greenberg, J. R. Hamilton, N. Jain, S. Kandula, C. Kim, P. Lahiri, D. A. Maltz, P. Patel, and S. Sengupta, "VL2: A Scalable and Flexible Data Center Network," in *Proc. ACM SIGCOMM*, 2009.

[13] H. Liu, C. F. Lam, and C. Johnson, "Scaling Optical Interconnects in Datacenter Networks Opportunities and Challenges for WDM," in *Proc. IEEE Symposium on High Performance Interconnects*, 2010.

[14] G. Wang, D. G. Andersen, M. Kaminsky, K. Papagiannaki, T. S. E. Ng, M. Kozuch, and M. Ryan, "c-Through: Part-time Optics in Data Centers," in *Proc. ACM SIGCOMM*, 2010.

[15] N. Farrington, G. Porter, S. Radhakrishnan, H. H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen, and A. Vahdat, "Helios: A Hybrid Electrical/Optical Switch Architecture for Modular Data Centers," in *Proc. ACM SIGCOMM*, 2010.

[16] Bowers, J., Raza, A., Tardent, D. and Miglani, J., 2014, July. Advantages and control of hybrid packet optical-

©2012-18International Journal of Information Technology and Electrical Engineering

circuit-switched data center networks. In *Photonics in Switching* (pp. PM2C-4). Optical Society of America.

[17] Farrington, N., Forencich, A., Sun, P.C., Fainman, S., Ford, J., Vahdat, A., Porter, G. and Papen, G.C., 2013, March. A 10 µs hybrid optical-circuit/electrical-packet network for datacenters. In *Optical Fiber Communication Conference* (pp. OW3H-3). Optical Society of America.

[18] Liu, H., Lu, F., Forencich, A., Kapoor, R., Tewari, M., Voelker, G.M., Papen, G., Snoeren, A.C. and Porter, G., 2014, April. Circuit Switching Under the Radar with REACToR. In *Nsdi* (Vol. 14, pp. 1-15).

[19] Yin, Y., Proietti, R., Ye, X., Nitta, C.J., Akella, V. and Yoo, S.J.B., 2013. LIONS: An AWGR-based low-latency optical switch for high-performance computing and data centers. *IEEE Journal of Selected Topics in Quantum Electronics*, *19*(2), pp.3600409-3600409.

AUTHOR PROFILES

Mr. Pronaya Bhattacharya is pursuing his PhD in Computer Science and Engineeringat Dr. A.P.J Abdul Kalam Technical University, Lucknow. Currently he is working as an Assistant Professor in Department of Information Technology at Nirma University, Ahmedabad, Gujarat. He has a teaching experience of more than 8 years as an Assistant Professor in department of Information Technology at PSIT, Kanpur. He has published papers in various reputed national and International Conferences like IEEE and ACM. His areas of interest include Computer Networks, Optical Communication and Design and Analysis of Algorithms.

Dr. Amod Kumar Tiwari is an Associate Professor in Department of Computer Science and Engineering in Rajkiya Engineering College, Sonbhadra, Uttar Pradesh. He has completed his PhD in Computer Science from Dr. A.P.J Abdul Kalam Technical University. He has a vast experience of more than 6 years of guiding many PhDs and M.Techs scholars. His areas of research include: Image Processing, Algorithms Design, Geometrical Modeling, Artificial, Intelligence, Computer Networks and Optical Communication.

Dr. Rajiv Srivastavais director at scholar-tech education Kanpur. He received M.Sc. from Kanpur University in 1997. He was awarded Gold medal in M.Sc. He received his M. Tech.: 2003 from IIT, Kanpur and PhD: 2009 (Optical Communication) from IIT, Kanpur. Currently he is also acting as Reviewer of IEEE Journal and Conference papers, also as a nominated reviewer for JESTEC Journal. He has published more than 40 research papers in highly reputed Journals like, Springer, Elsevier and IEEE. He also holds two patents. His research interests include: optical communication, wireless communication, image processing, wireless sensor networks etc.