Software-Defined Optical Burst Switching for High Performance Computing

Ashok Kumar Turuk *

Department of Computer Science & Engineering
National Institute of Technology Rourkela, India-769008

September 9, 2017

Abstract: we assess the performance of techniques for optical burst switching (OBS) designed for high performance computing (HPC) and cloud computing data center networks (DCNs) by using network-level simulation. We consider short-duration bursts using the faster switching technologies that are now available. The modeled switch architecture features fast optical switches in a single-hop topology with a centralized, software-defined optical control plane. Instead of using OBS with traditional methods (i.e., one-way reservation), we consider OBS with a two-way reservation protocol that results in zero burst loss. We model different workloads with various data rates by considering different edge-to-core network over-subscription ratios to investigate the performance of such designs across usage patterns. Our results reveal that the proposed technique shows considerable improvement in terms of throughput and packet loss ratio and comparable performance in terms of delay when compared to traditional methods of OBS.
Index Terms: Data center networking; Optical burst switching; Optical interconnects; Software-defined optical networks.

1 Introduction

OBS [1] paradigm offers a standardized backbone to transmit IP traffic in a successful yet realistic manner. OBS isolates the control plane activities from the data plane in order to exploit their distinct advantages in electronic and optical domains respectively. Control messages (or Burst Header Packets (BHPs)) are processed electronically at every node en routed, while the data bursts are transmitted optically from end to end [2]. In brief, OBS paradigm maintains a trade off between Optical Circuit Switching (OCS) and Optical Packet Switching (OPS).
Our simulation model consists of 40 ToR switches. Each ToR switch has 40 servers connected to it. The controller and ToR switches are connected to the management network via an electrical switch. We use one fast switch that is interfaced to the management network. The two different cases of network oversubscription ratios, are considered to investigate the impact of traffic aggregation on the performance of the system. In a fully subscribed network, all servers in a rack send their traffic to the servers in other racks, i.e., 100% of the traffic is inter-rack, while in a 21 oversubscribed network, 50% of the traffic is inter-rack, and 50% of it is intra-rack.

*akturuk@gmail.com

In this article, we address the general problem of route optimization in a given OBS backbone network. The routing problem is modeled as a bi-objective ILP problem where the objectives are based on minimizing both the number of wavelengths used, and the number of hops traversed to fulfill the burst transmission requests for a given set of node pairs. The ILP is solved by using a novel approach based on a DE algorithm. The DE algorithm results in a route configuration that has the ability to produce a better network-wide BLP by using fewer network resources.
To assess the performance of OBS for a DCN, we developed simulation models in the OMNeT++ simulation framework. This required the control plane algorithms to be implemented in C++ within the OMNeT++ models. The term software-defined refers to our own implementation of the control plane in OMNeT++. Porting this code to interact with real switch hardware using, for example, appropriate extensions to the OpenFlow protocol, rather than with the simulation model will allow the control plane to be deployed on generic hardware. The problem of packaging our solution for real-world deployment in a manner compatible with existing SDN frameworks will be addressed in our future work.

2 Routing in OBS Networks

The ToR switch also has a burst disassembler and packet extractor module to disassemble the bursts received through the receivers. The packets are extracted from the burst and are sent to the electronic switch fabric and finally to the destination servers using electronic switching.
When the control packet arrives back at the ToR switch, the scheduler module of the ToR switch generates a burst according to the timeslot assigned by the controller. The timeslot refers to the duration of time assigned for a burst in an optical switch path. The generated burst is then sent to the queue of the allocated port. The scheduler module also initiates a new timer if the VOQ is not empty after the burst generation because new packets might have arrived during the RTT of the control packet.
The above analysis considers only data plane issues. Evaluating the performance of the control plane is an open issue as it is dependent on implementation and can only be addressed after deployment of the proposed technique in a real-world scenario. This may be the bottleneck in scaling up the architecture and will be considered in future studies of the architecture. In addition, the feasibility of deployment of multiple controllers could also be investigated.

The control packet is 440 bits long and contains two main fields: routing and reservation. The routing field contains the IP address of the source ToR switch, the IP address of the controller, and the IDs of the source and destination ToR switches. We consider 128 bits for IPv6 addresses; however, this length could be reduced to 32 bits using IPv4 addresses, making overall control packet length 31 bytes. The reservation field is 96 bits long and is divided into three sub-fields: 1) burst length, 2) start time, and 3) port number. The burst length field is filled by the ToR switch to request a timeslot from the controller. The controller fills the other two fields after processing the control packet. All of these three fields are 4 bytes long. The burst length field contains the burst length expressed in bytes; the start time contains the time when the burst will be sent; and the port number is the port of the ToR switch in which the burst is to be sent. A field is reserved for a cyclic redundancy check (CRC), and a couple of optional fields are reserved for flags.
Our design features separate control and data planes. The control plane comprises a centralized controller that performs routing, scheduling, and switch configuration functions. It receives connection setup requests from all ToR switches, finds routes, assigns timeslots to the connection requests, and configures optical switches with respect to the timeslots allocated. In order to perform these tasks, the controller keeps a record of the connection states of all optical switches. The data plane comprises optical switches that perform data forwarding on pre-configured lightpaths set up by the controller.

The proposed topology has two layers, i.e., the edge and core. The edge contains the electronic top of the rack (ToR) switches, while the core comprises a group of fast optical switches. Servers in each rack are connected to the ToR switches using bidirectional optical fibers. The ToR switches are linked to the optical switches using unidirectional optical fibers.

3 Framework to Reduce Burst Loss in OBS Network

In this section, first we explain the assumptions and notations used. Then, we specify the burst loss model adopted.

3.1 Assumptions and Notations

1. We consider an OBS network without FDLs. The OBS network is modeled as a graph $G(V, E)$ where $V = \{v_1, v_2, v_3, ..., v_N\}$ is the set of nodes that are capable of wavelength conversion and $E = \{e_1, e_2, ..., e_M\}$ is the set of directed links. If a link e connects an output port of v_i to an input port of v_j, then we refer to v_i and v_j as the tail and head of e respectively. Accordingly, we define the following notations:

$$\begin{aligned} \omega^+(v) &= \{e | v \, is \, a \, tail \, of \, e\} \\ \omega^-(v) &= \{e | v \, is \, a \, head \, of \, e\} \end{aligned} \tag{1}$$

2. We consider source-based routing [3]. Let $Z = \{(s_i, d_j, \rho_{ij})\}$ denotes the burst transmission requests for a set of given node pairs where s_i is the source node, d_j is the destination node, and ρ_{ij} is the requested burst traffic intensity between the node pair (s_i, d_j).

3. Let $P = \underset{(s_i, d_{j\neq i}) \in Z}{\cup} p_{ij}$ denotes the set of selected paths for all the node pairs $(s_i, d_{j\neq i}) : 1 \leq i, j \leq N$ in Z where p_{ij} denotes the selected path between the node pair $(s_i, d_{j\neq i})$ for burst transmission. Let $P_e \subseteq P$ be the set of all paths routed through the link $e \in E$. To elevate the problem of out-of-order burst arrival at the destination, we assume that the network operates with single-path routing approach [4].

4. The request for burst transmission among the node pairs in the OBS network is assumed to be uniformly distributed.

5. Dynamic routing algorithm is used to select an efficient path for every node pair $(s_i, d_{j\neq i})$ in order to improve the network state information in terms of the number of wavelengths used, and total number hops traversed.

3.2 Burst Loss Model

We assume the non-reduced load calculation technique [5] to estimate the burst loss in the OBS network. We denote by, ρ_e, the offered load of all the paths passing through the link e and is stated below:

$$\rho_e = \sum_{p \in P_e} \rho_p; \; \exists (s_i, d_j) \in Z : \rho_{ij} \vdash \rho_p \tag{2}$$

The BLP, B_e, in network link e can be calculated by using Erlang-B formula [6]:

$$B_e(\rho_e, W) = \frac{(\rho_e)^W / W!}{\sum_{c=1}^{W} (\rho_e)^c / c!} \tag{3}$$

where W is the number of allocated wavelengths.
The BLP, B_p, along the path p can be calculated as:

$$B_p = 1 - \prod_{e \in p} (1 - B_e) \tag{4}$$

Finally, the network-wide BLP, B_N, is calculated below:

$$B_N = \frac{\sum_{p \in P} \rho_p B_p}{\sum_{p \in P} \rho_p} \tag{5}$$

4 The RWA Problem Formulation

Burst losses can be reduced by judiciously selecting the paths for all the node pairs (s_i, d_j) in Z. The key form of path optimization is load balancing [7]. Given the network topology $G(V, E)$ and the demand set, Z, the load balancing strategy minimizes the maximum flow traversing through any link in $E(G)$.
In an OBS network, the network state information is maintained centrally and is easily controlled by a dedicated control channel [8]. The RWA approach for burst switching generally uses this information to optimize its objective function(s). In this article, the RWA problem is formulated as a bi-objective optimization problem [9]. The first objective function ($Y1$) minimizes the congestion of the most congested link in $E(G)$ and the second objective function ($Y2$) minimizes the total number of hops traversed by all the paths $p_{ij} \in P$. A convenient way to produce a single unique solution for a bi-objective optimization problem is the weighted sum method where the weights reflect, a priori, the user's preferences [10]. The RWA problem is modeled by the following ILP:
ILP Model:
Input:
The OBS network $G(V, E)$: where $|V(G)| = N$ and $|E(G)| = M$
Variables:

$$x_e^{z,w} = \begin{cases} 1;\ if\ path\ selected\ for\ \boldsymbol{z} \in Z\ uses\ wavelength\ \boldsymbol{w}\ to\ send\ burst\ traffic \\ along\ link\ \boldsymbol{e} \\ 0;\ otherwise \end{cases} \tag{6}$$

Objective Function:

$$\underset{\mathbf{x}}{Minimize}\ Y = w_1\, Y1(\mathbf{x}) + w_2\, Y2(\mathbf{x}) \tag{7}$$

where

$$Y1 = Minimize[\underset{e \in E}{Max} \sum_{z=1}^{|Z|} \sum_{w=1}^{W} x_e^{z,w}] \tag{8}$$

$$Y2 = Minimize[\sum_{z=1}^{|Z|} \sum_{e=1}^{M} \sum_{w=1}^{W} x_e^{z,w}] \tag{9}$$

$$w_i \geq 0;\ \sum_i w_i = 1 \tag{10}$$

Subject to:
Flow Reservation Constraint:

$$\sum_{e \in \omega^+(v|v \in V)} \sum_{w=1}^{W} x_e^{z,w} - \sum_{e \in \omega^-(v|v \in V)} \sum_{w=1}^{W} x_e^{z,w} = \begin{cases} 1;\ if\ v = s_i \\ 0;\ if\ v \neq s_i,\ d_j \\ -1;\ if\ v = d_j \end{cases} ;\forall z \in Z \tag{11}$$

Loop-less Constraint:

$$\sum_{e \in \omega^{+}(v|v \in V)} \sum_{w=1}^{W} x_e^{z,w} \leq 1;\ \forall z \in Z \tag{12}$$

$$\sum_{e \in \omega^{-}(v|v \in V)} \sum_{w=1}^{W} x_e^{z,w} \leq 1;\ \forall z \in Z \tag{13}$$

Integer Constraint:

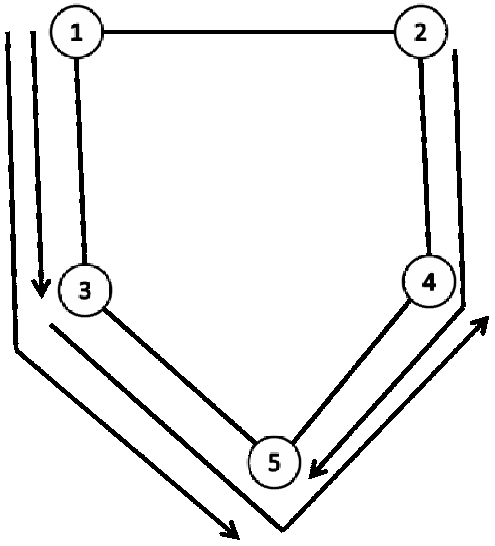
$$x_e^{z,w} \in \{0, 1\} \tag{14}$$

Equation 11 is analogous to classical flow conservation constraint in a multi-commodity flow problem [11]. Equation 12 and Equation 13 together eliminate the possibility of the formation of potential loops along a path. Equation 14 requires the variable $x_e^{z,w}$ to possess integer values. We employ OBS in the proposed data center network architecture. We aggregate packet traffic to create a burst of short duration. A control packet is created to request the allocation of resources needed to transmit the burst from the controller by using a two-way reservation process similar to that proposed for OBS networks [27]. Although such two-way reservation is not feasible in a long-haul backbone network, it is suitable in data centers for the reasons presented earlier. The controller assigns resources and sends the control packet back to the originating node as an acknowledgment. The burst is then transmitted on the preestablished path.

4.1 An Example of RWA Problem

An exemplary network has been taken into consideration to understand the above formulated RWA problem as shown in Fig. 1. We have considered the following set of node pairs and their selected paths for burst transmission:

$Z = \{(1,3), (2,5), (3,4), (1,5)\}$



Demand Number	Selected Path
1. (1,3)	1-3
2. (2,5)	2-4-5
3. (3,4)	3-5-4
4. (1,5)	1-3-5

Figure 1: An Exemplary Network

As we can see in Fig. 1, the maximum link congestion is $Y1 = 2$ and the total number of hops traversed is $Y2 = Y_{13} + Y_{25} + Y_{34} + Y_{15} = 1 + 2 + 2 + 2 = 7$. The solution presented in Fig. 1 may not be the optimal one; but helps to understand the RWA problem formulation and the stated objective functions.

4.2 The DE Algorithm

The burst is sent at a particular time after sending the control packet, which is called the offset time. During the offset time, these bursts are temporarily stored at edge node before transmission. During this time, the switch controller at the core node processes the control information and sets up the switching matrix for the incoming burst. Burst loss due to output port contention is the major limitation of the OBS network. Several techniques exist in the literature to avoid contention, such as FDLs, deflection routing, wavelength conversion, and segmentation- based dropping, but none of them can guarantee zero burst loss. OBS with two-way reservation ensures zero burst loss in which a control packet reserves resources in all nodes from the source to the destination and is sent back to the source as an acknowledgment. The control packet has a high roundtrip time (RTT) for a large wide area optical network due to high propagation and switching delay.

4.2.1 Individual Encoding

Individual encoding justifies how the problem is structured using DE algorithm. For every node pair (s_i, d_j) in the demand set Z, we employ Dijkstra's algorithm to find its shortest path. For each link (v_y, v_{y+1}) in the shortest path, one link is disabled at a time and a new shortest path, if available, is calculated as shown in Table 1. Fig. 2 depicts an individual coding for the path configuration shown in Fig. 1. As we can observe, the individual is a vector of size $|Z|$ and every z^{th} position of the individual stores a pointer to the path selected for the node pair (s_i^z, d_j^z) given that $1 \leq z \leq |Z|$.

Table 1: List of Possible Paths

Reference Number	(1,3)	(2,5)	(3,4)	(1,5)
#1	1-3	2-4-5	3-5-4	1-3-5
#2	1-2-4-5-3	2-1-3-5	3-1-2-4	1-2-4-5

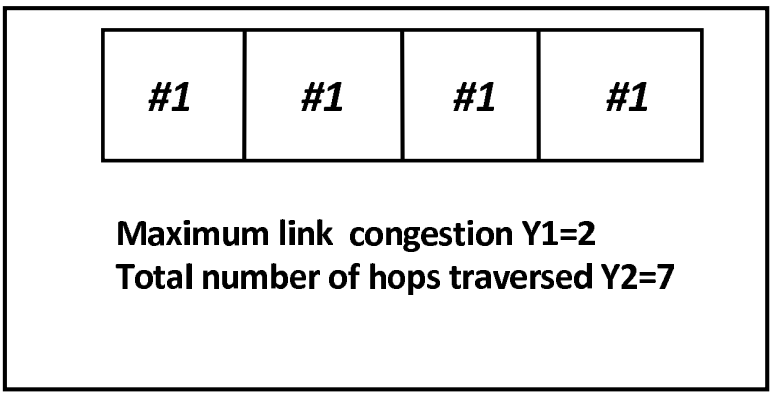


Figure 2: Individual Encoding

4.2.2 Working Principle of DE Algorithm

During the initialization step, we generate randomly *popsize* number of individuals by using the procedure explained in section 4.2.1. At each generation, all *popsize* number of individuals are evaluated and the individual under evaluation is termed as the *target vector*. The fitness of an individual *ind* is a target function to be minimized and is stated as below:

$$Fitness(ind) = w_1 \ \frac{con(ind)}{|Z|} + w_2 \ \frac{tot\_hop\_count(ind)}{(V(G)-1)|Z|} \tag{15}$$

where

$con(ind)$ denotes the congestion of the individual and is measured by the most appeared link among all the paths in the individual

$tot\_hop\_count(ind)$ denotes the total number of hops traversed by all the paths in the individual

For every target vector $b_{i,g} : i = 1, 2, ..., popsize$, a mutant individual m_i is created as stated below:

$$m_{i,g+1} = b_{r1,g} + MF(b_{r2,g} - b_{r3,g}) \ \ \forall b_{i,g} : i = 1, 2, ..., popsize \tag{16}$$

where $b_{r1,g} \neq b_{r2,g} \neq b_{r3,g} \neq b_{i,g}$ and g is the current generation. The MF operator maintains a trade-off between global exploration and local exploitation in the search space.

The crossover phase is activated, once the mutation phase is completed. The CP operator is applied to increase diversity in the mutation process. The perturbed individual $m_{i,g+1}$ and the target vector $b_{i,g}$ are subject to the crossover operation to create the trial vector $t_{i,g+1}$ as stated below:

$$t^z_{i,g+1} = \begin{cases} m^z_{i,g+1} \, if \, rand_z \ < CP \\ b^z_{i,g} \, otherwise \end{cases} \quad \forall z \in Z \tag{17}$$

where $rand$ is a random number in the interval $[0, 1]$. After the creation of the trial vector $t_{i,g+1}$, it is necessary to verify the boundary constraints of every element in $t_{i,g+1}$ to avoid the creation of infeasible solutions. If any element of the trial vector goes off the search space, then its replacement is generated by using Eq. 18:

$$t^z_{i,g+1} = l_z + round((u_z - l_z)rand(1)) \tag{18}$$

where u_z and l_z are the lower and upper bound of the reference numbers for the paths assigned to z^{th} node pair respectively.

Finally, the selection operation is performed by comparing the fitness between the target vector and the trial vector. The selection scheme is as follows (for a minimization problem):

$$b_{i,g+1} = \begin{cases} b_{i,g} \, if \, f(b_{i,g}) < f(t_{i,g+1}) \\ t_{i,g+1} \, Otherwise \end{cases} \tag{19}$$

4.2.3 An illustration of DE Algorithm

Now, we exemplify the DE algorithm using a simple 6-nodes network as shown in Fig. 3. For the sake of illustration, we assume the following set of node pairs in Z:

$Z = \{(4, 1), (5, 1), (2, 3), (6, 2), (2, 5), (1, 5), (3, 4), (5, 6), (5, 3), (1, 3)\}$

Using the procedure mentioned in section 4.2.1, we calculate a set of admissible paths for every node pair in Z as shown in Table 2. Then, we execute the DE algorithm and the values of different parameters of DE algorithm are listed in Table 3. The best individual in the current generation and the global best individual found so far are recorded to keep track of the best solution. In Fig. 4, and Fig. 5, we made a comparison between the path configurations obtained by SP algorithm and DE algorithm respectively. As SP algorithm does not have the ability to

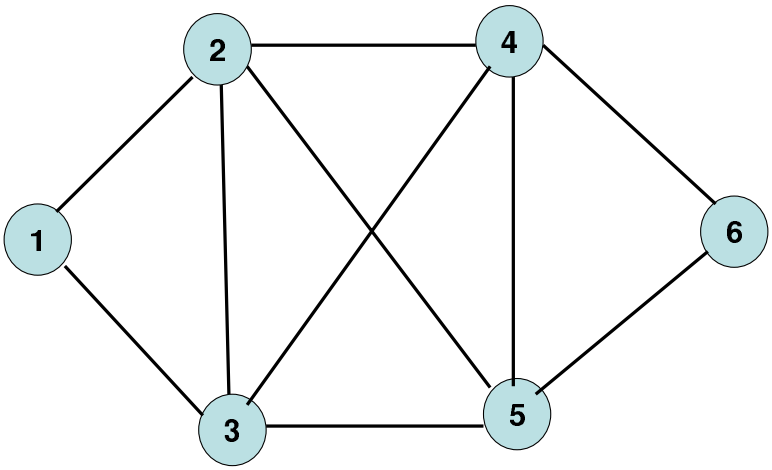


Figure 3: A 6-Nodes Network to Illustrate DE Algorithm

choose alternate routes, the only possible solution to it would be represented by the following individual vector:

$b = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1]$

On contrary, after applying DE algorithm, the optimal solution obtained is the following individual vector:

$b = [1, 2, 1, 1, 2, 1, 1, 1, 1, 2]$

Reference Number	(4,1)	(5,1)	(2,3)	(6,2)	(2,5)	(1,5)	(3,4)	(5,6)	(5,3)	(1,3)
#1	(4,3,1)	(5,3,1)	(2,3)	(6,5,2)	(2,5)	(1,3,5)	(3,4)	(5,6)	(5,3)	(1,3)
#2	(4,2,1)	(5,2,1)	(2,5,3)	(6,4,2)	(2,4,5)	(1,2,5)	(3,5,4)	(5,4,6)	(5,4,3)	(1,2,3)

Table 2: List of Admissible Paths for The Node Pairs in Z

Parameters of DE	Value
Number of Generations	100
Population Size (*popsize*)	100
Mutation Factor (*MF*)	0.2
Crossover Probability (*CP*)	0.5

Table 3: Parameters of DE Algorithm

Algorithm	Hop Count	Congestion
SP	14	4
DE	16	2

Table 4: DE Algorithm VS. SP Algorithm

4.2.4 Strategies to Enhance The Performance of DE Algorithm

DE algorithms are not originally designed to solve constrained optimization problems [12]. However, by adjusting the control parameters and incorporating effective constraint handling techniques, we can considerably improve the search ability of DE algorithm. In this section, we explain two modifications adopted to enhance the overall performance of DE algorithm in solving constraint optimization problems.

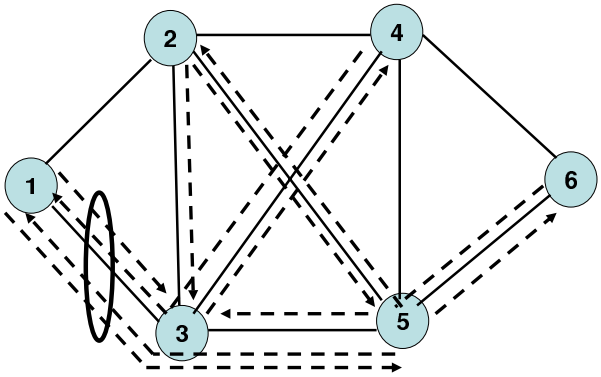


Figure 4: Path Establishment by using SP Algorithm

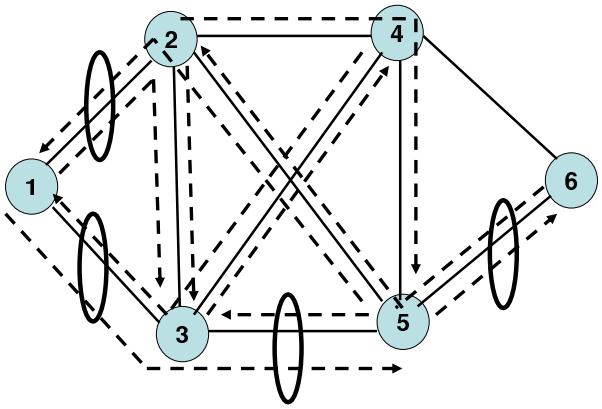


Figure 5: Path Establishment by using DE Algorithm

1. Modification of Mutation Operation: Population based search algorithms maintain a trade off between global exploration and local exploitation. Moreover, the mutation operation affects the convergence rate of DE algorithm. DE algorithm generally suffers from weak local exploitation ability. The mutation operation stated in Eq. 16 is able to maintain population diversity but slows down the convergence rate. To enhance the local exploitation ability, we adopted a directed mutation scheme proposed in [12] as stated below:

$$m_{i,g+1} = b_{r1,g} + MF(b_{best,g} - b_{worst,g}) \quad \forall b_{i,g} : i = 1, 2, ..., popsize \tag{20}$$

where $b_{r1,g}$ is a randomly chosen individual and $b_{best,g}$, and $b_{worst,g}$ are the best and the worst vector in the entire population of generation g respectively. This directed mutation operation favors local exploitation since the mutated vector follows the same direction of the best individual. We embed the new mutation operation in the DE algorithm and it is combined with the basic mutation operation as stated below:

$$\begin{gathered} if \ \ (rand < 0.5) \ \ then \\ m_{i,g+1} = b_{r1,g} + MF_1(b_{best,g} - b_{worst,g}) \\ else \ \ m_{i,g+1} = b_{r1,g} + MF_2(b_{r2,g} - b_{r3,g}) \end{gathered} \tag{21}$$

where both MF_1 and MF_2 take the value between $[0, 1]$.

2. Modification of the Mutation Factor: The MF has a considerable influence on global exploration: A small value leads to premature convergence whereas a large value makes the search process unstable. In Eq. 20, we observe that it is a directed difference vector from the worst to the best vector. Hence, MF must be a positive value to push the trial vectors to follow the same direction of the best individual. Thus, MF_1 should possess a value between $(0, 1]$. However, in Eq. 16, it is a pure random difference as fitness values are not used. In this case, the best direction that can lead to good exploration is not known. Thus, in order to maintain uniformity in the search space, MF_2 is required to possess a value in the interval $[-1, 0) \cup (0, 1]$.

   In general, for both the above alternatives, MF value may vary to differ the mutant vectors by different direct weights.

5 Results & Discussion

In this section, we present the analytical results to showcase the superior performance of DE algorithm against SP algorithm in terms of its ability to achieve a desired network-wide BLP (B_N) requirement by using lesser number of wavelengths. Our analysis is verified in a Pentium(R) 4 CPU with a clock cycle of 3.2 GHz and a 2GB of RAM. The simulator used is MATLAB 7.0.1. The networks considered for analysis are 20-nodes ARPANET, 16-nodes NSFNET, 9-nodes TORUS network, and 12-nodes random network as shown in Figure 6, Figure 7, Figure8, and Figure 9 respectively.

Let $1/\mu$ denotes the mean burst holding time for all the node pairs (s_i, d_j) in Z. We use the notion $\rho_{ij} = \lambda_{ij}/\mu$ to denote the offered load of burst traffic from s_i to d_j. The traffic is characterized by a Poisson process [13]. We consider an independent and exponential distribution of burst arrivals together with an independent and identical distribution of burst durations.

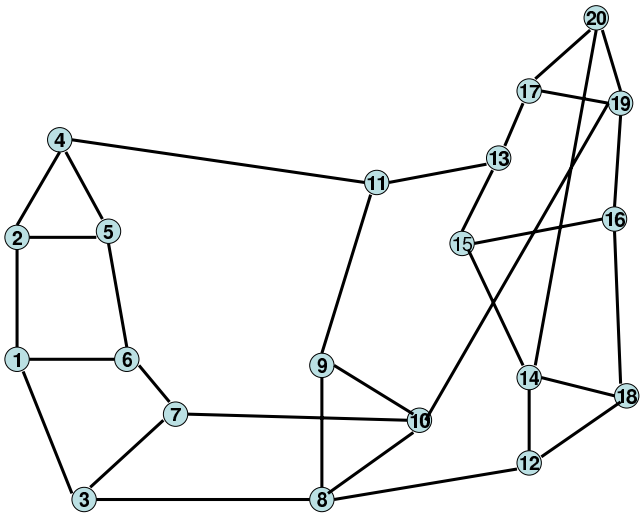


Figure 6: Advanced Research Project Agency Network (ARPANET)

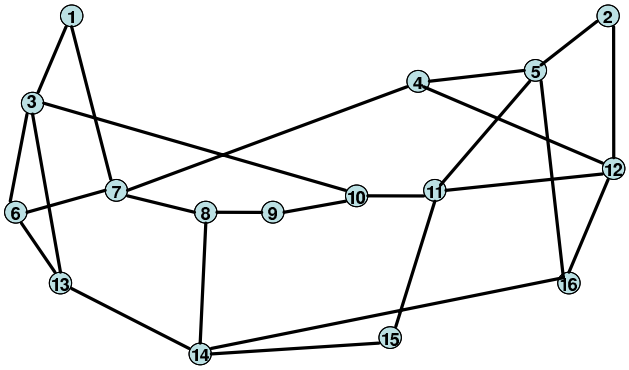


Figure 7: National Science Foundation Network (NSFNET)

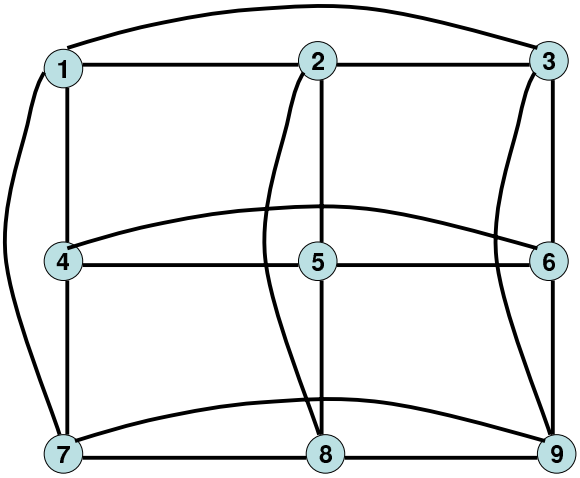


Figure 8: TORUS 9 Network

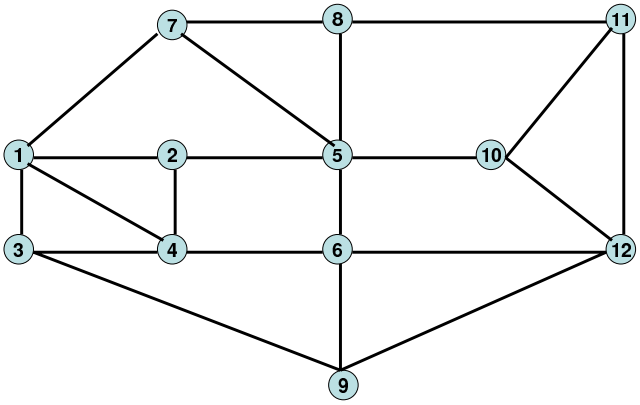


Figure 9: Random 12 Network

The average throughput is identical in all the networks because the burst loss is zero in OBS with traditional methods until a 95% offered load because bursts are generated in an order and

all the bursts are destined to only one destination network. Thus, no collision happens, which results in zero burst loss.
The average throughput in traditional methods of OBS is considerably low as compared to average throughput achieved in the proposed scheme because of burst losses in traditional methods of OBS, whereas in the proposed scheme, burst loss is zero. Due to zero burst loss, our scheme demonstrates comparable performance to the baseline electrical network until a very high load. Another important point is that the average throughput decreases with the increase of data rate at a very high load because bandwidth is wasted during assignment of the timeslot in a link. This wasted bandwidth is 4 times higher in 40 Gbps as compared to 10 Gbps data rates.

1. Optimality: In most of the cases (see Fig. 10, Fig. 11, Fig. 12, and Fig. 13), the primary optimization objective i.e. congestion produces optimal results.

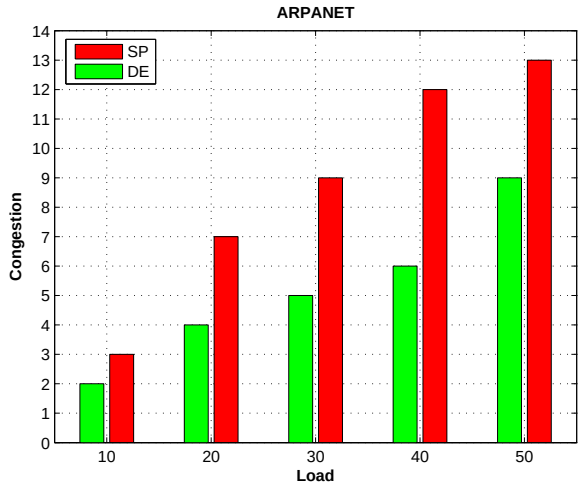


Figure 10: Network Congestion VS. Load

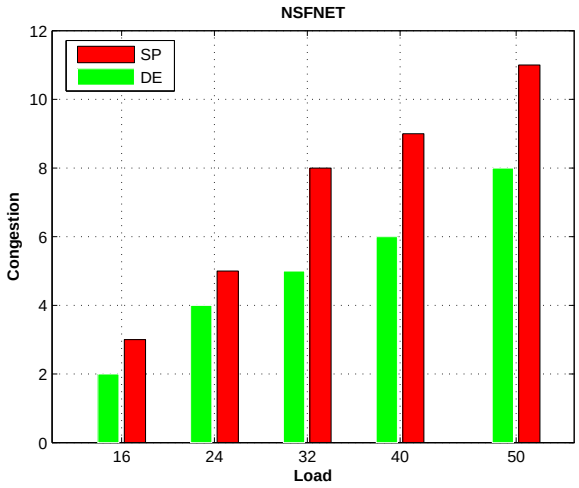


Figure 11: Network Congestion VS. Load

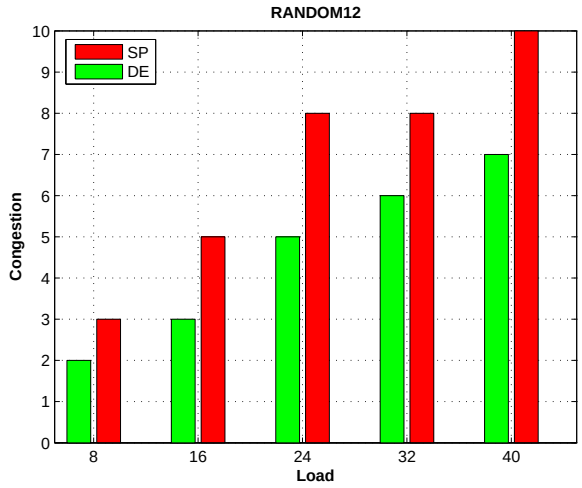


Figure 12: Network Congestion VS. Load

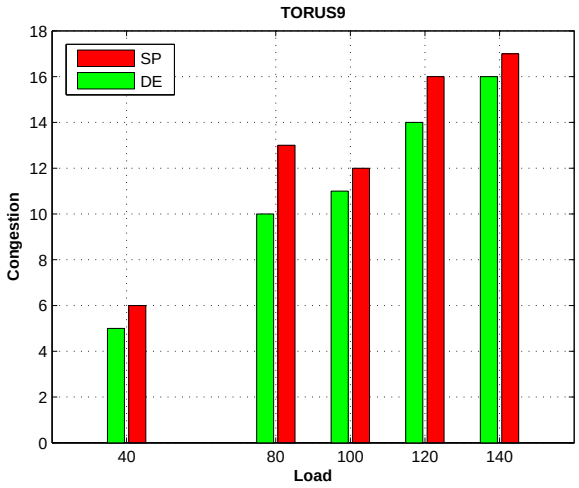


Figure 13: Network Congestion VS. Load

It can be noticed that the execution time of routing and scheduling operations is in a nanoseconds scale for all types of networks. Execution time is the lowest in the 41 oversubscribed network, but it increases slightly as we decrease network oversubsription. Similarly, the execution time of the switch configuration operations is at a minimum when P is minimum, and it increases slightly with an increase in the number of optical switches. The overall execution time of switch configuration operations is negligible (at most a few nanoseconds). We obtain total execution time of the control plane processing by adding up the execution times of routing/scheduling and switch configuration operations, which is in the nanoseconds range. Thus, our algorithms in the control plane demonstrate efficient performance for all types of network oversubscriptions.

6 Conclusion

We proposed a novel optical interconnect based on fast optical switches. The proposed design features fast optical switches in a single-hop topology with a centralized, software-defined optical control plane. The single-stage core topology can be easily scaled up (in capacity) and scaled out (in the number of racks) without requiring major recabling and network reconfiguration. We use OBS with two-way reservation to obtain zero burst loss. Two-way reservation is not appropriate for conventional backbone optical networks due to the high RTT of the control packet, but in a DCN, the RTT is not high. We use network-level simulation to model different workloads with various data rates by considering different edge-to-core network over-subscription and investigate the performance of such designs across usage patterns. Our results reveal that the proposed technique shows considerable improvement in terms of throughput and packet loss ratio as compared to conventional methods of OBS, while comparable performance in terms of delay with conventional methods of OBS is also achieved. The proposed technique also demonstrates performance comparable to that of electrical data center networks.

References

[1] Yang Chen, Chunming Qiao, and Xiang Yu. Optical Burst Switching: A New Area in Optical Networking Research. *IEEE Network*, 18(3):16–23, 2004.

[2] Yijun Xiong, Marc Vandenhoute, and Hakki C. Cankaya. Control Architecture in Optical Burst-Switched WDM Networks. *IEEE Journal on Selected Areas in Communications*, 18(10):1838–1851, 2000.

[3] Yufei Wang and Zheng Wang. Explicit Routing Algorithms for Internet Traffic Engineering. In *Eight International Conference on Computer Communications And Networks*, pages 582–588. IEEE, 1999.

[4] Sebastian Gunreben and Guoqiang Hu. A Multi-Layer Analysis of Reordering in Optical Burst Switched Networks. *IEEE Communications Letters,*, 11(12):1013–1015, 2007.

[5] Arie Koster. *Graphs And Algorithms in Communication Networks*. Springer, 2010.

[6] James Jewett, J. Shrago, and B. Yomtov. *Designing Optimal Voice Networks for Businesses, Government, And Telephone Companies*. Telephony Pub. Corp., 1980.

[7] Kee Chaing Chua, Mohan Gurusamy, Yong Liu, and Minh Hoang Phung. *Quality of Service in Optical Burst Switched Networks*. Springer, 2007.

[8] Miroslaw Klinkowski, Pedro Pedroso, Davide Careglio, Michal Pioro, and Josep Solé-Pareta. Joint Routing And Wavelength Allocation Subject to Absolute QoS Constraints in OBS Networks. *Journal of Lightwave Technology*, 29(22):3433–3444, 2011.

[9] Kalyanmoy Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. Wiley, 2005.

[10] R. Timothy Marler and Jasbir S. Arora. The Weighted Sum Method for Multi-objective Optimization: New Insights. *Structural And Multidisciplinary Optimization*, 41(6):853–862, 2010.

[11] Asuman E Ozdaglar and Dimitri P Bertsekas. Optimal Solution of Integer Multicommodity Flow Problems with Application in Optical Networks. *Nonconvex Optimization And Its Applications*, 74:411–436, 2003.

[12] Ali Wagdy Mohamed and Hegazy Zaher Sabry. Constrained Optimization based on Modified Differential Evolution Algorithm. *Information Sciences*, 194:171–208, 2012.

[13] M. Izal and J. Aracil. On The Influence of Self-Similarity on Optical Burst Switching Traffic. In *Global Telecommunications Conference, GLOBECOM'02*, pages 2308–2312. IEEE, 2002.