Contention-Based Limited Deflection Routing Protocol in Optical Burst-Switched Networks

SuKyoung Lee*, K. Sriram, H.S. Kim, and J.S. Song

*Corresponding author: National Institute of Standards and Technology

100 Bureau Drive, Stop 8920, Gaithersburg, MD, 20899, USA.

Email: sukyoung@antd.nist.gov, ksriram@nist.gov

Abstract

Optical Burst Switching (OBS) is a very promising switching technology for realization of an economical optical Internet. In OBS networks, when contention occurs at an intermediate switch, two or more bursts that are in contention can be lost because a forwarding path reservation is not made for a burst until a control message for the burst arrives. That is the reason why one of the critical design issues in OBS is finding ways to minimize burst dropping resulting from resource contention. In this paper, we propose and analyze a novel deflection routing protocol, which mitigates and resolves contention with significantly better performance as compared to techniques currently known in the literature. While several variants of the basic deflection routing scheme have been proposed before, they all lacked the ability to determine the alternate route based on clear performance objectives. In this paper, we present an on-demand deflection routing scheme, which sequentially performs the following: (1) based on certain performance criteria, dynamically determines if the burst should be deflection routed or retransmitted from source, (2) if the decision is to deflection route, then the same is done using a path that is based on minimization of a performance measure that combines distance and blocking due to contention. The proposed Contention-Based Limited Deflection Routing (CLDR) scheme prevents injudicious deflection routing. Our simulation results show that the scheme proposed here has much superior performance both in terms of burst loss probability and increased network throughput. Through analytical and simulation modeling, a number of useful insights into the OBS network protocols and performance are provided.

Keywords: Optical burst switching, Deflection routing, Performance, Burst loss mitigation, Burst contention, Optical Internet

1 Introduction

The optical core networks have the capacity to carry terra bytes of data per second through each node. The edge routers feed data into these networks. The data is typically carried over 10 Gbps wavelength channels. Once a wavelength channel is setup between any two end-points, it can only carry packet traffic between those end-points. If the edge-routers feed the traffic sparsely, then the 10 Gbps channel is highly underutilized. One way to overcome this bandwidth inefficiency in the core networks is to setup the wavelength channels with short hops and use ultra-high capacity (Tbps) core packet routers at several of the nodes in the core network. These high-capacity routers re-groom the packet traffic arriving from various nodes and try to statistically multiplex and pack the 10 Gbps wavelength channels efficiently. Another approach for efficient bandwidth usage as well as resource usage in the core optical networks is potentially achievable via Optical Burst Switching (OBS)[1][2][3], aspects of which are the focus of this study.

The OBS switches can potentially perform traffic grooming in the optical domain using tunable lasers and wavelength cross-connect (all optical) switches. The OBS switches would statistically multiplex traffic from different incoming ports and wavelengths to an egress port and wavelength. The statistical multiplexing occurs at the burst level, each burst consisting of numerous packets. There is a possibility that the OBS switches together with the WDM/DWDM capability can be produced less expensively than a combination of ultra-high capacity core routers, optical switches, and WDM/DWDM equipment. Also, the switching delay for OBS is dropping down to the range of tens or hundreds of nanoseconds which makes a good case for feasibility of OBS implementation [2]. Although promising, OBS still has implementation challenges, which need to be overcome. These challenges include limited optical buffering and optical power and distortion management. The OBS implementation strategy includes both an electronic control processing mechanism for optical burst scheduling and an optical transmission technology utilizing wavelength cross-connects (WXCs or OXCs) together with tunable lasers.

One of the challenging issues in the implementation of burst switching is the resolution of contentions that results from multiple incoming bursts that are directed to the same output port. In an optical burst switch, various techniques designed to resolve contentions include optical buffering, wavelength conversion, and deflection routing[4]-[11]. In comparison to other techniques, deflection routing has an advantage in that it can work with limited Fiber Delay-Line (FDL) buffer capacity. Fiber buffer capacity is indeed very limited, and a larger amount of it is needed in pure buffering schemes for contention resolution. However, deflection routing can work with limited optical buffering (or even no buffering) because it deflects or re-routes (on the fly) the contending bursts to an output port other than the intended output port.

A brief review of the work related to deflection routing in OBS networks is in order. Some recent work about deflection routing is reported in [7]-[10]. The authors of [7] investigate the performance of deflection routing in prioritized Just Enough Time (JET)-based OBS networks. In [8] and [9], it is demonstrated via simulation tests that the blocking probability improves when deflection routing is used as a means for contention resolution. The authors of [10] describe how deflection routing can be used in conjunction with the self-routing address scheme. However, they do not address the issue of how routing to an alternate path should be done, given that some constraints may apply to the selection of an alternate path.

In this paper, we propose and analyze a novel Contention-based Limited Deflection Routing (CLDR) protocol, which mitigates and resolves contention with significantly better performance as compared to techniques currently known in the literature. While several variants of the basic deflection routing scheme have been proposed before [7]-[10], they all lacked the ability to determine the alternate route based on clear performance objectives. In this paper, we present an on-demand deflection routing scheme, which sequentially performs the following: (1) based on certain performance criteria, dynamically determines if the burst should be deflection routed or retransmitted from source, (2) if the decision is to deflection route, then the same is done using a path that is based minimization of a performance measure that combines distance and blocking due to contention. The proposed CLDR scheme prevents injudicious deflection routing. Our simulation results show that the scheme proposed here has much superior performance both in terms of burst loss probability and increased network throughput. In this paper, we have also proposed that the network nodes should periodically re-compute and store optical paths, with the aim of staying optimal in the face of changing node and link congestion measures. This allows for deflection routed bursts to traverse the alternate optical paths that are not necessarily shortest path but are optimized for best performance (i.e., blocking and delay). This technique calls for monitoring the link and node congestion and updating the same in a periodic manner so that the path computation can be as optimal as possible (albeit with some minor lag).

Further, we have presented here an analytical model for computation of burst loss ratio due to contention on congested links in the network. Typically, the traffic originating from the edge nodes of the network would be correlated and such correlations would have a significant impact on the burst contentions at the edge as well on internal links in the network. Our analytical model accounts for these correlations (including various parameters that help quantify the correlations) in the prediction of burst loss ratio or probability. The analytical model results are compared with simulation results. Additionally, the analytical modeling results also used to create some relevant inputs in the design of the simulation experiments for studying CLDR and comparing it with other known schemes.

The rest of this paper is organized as follows. Sections 2 and 3 describe CLDR mechanism and other enhancements in detail. The analytic model for burst loss probabilty is presented in Section 4. In Section 5, we present the simulation model used, and performance of the proposed CLDR is examined via numerical results obtained using analytical and simulation models. Finally, we summarize and state our conclusions in Section 6.

2 Reservation Protocol

In OBS, a control packet is sent first to set up a connection by reserving an appropriate amount of bandwidth and configuring the switches along a path, followed by a data burst without waiting for an acknowledgement for

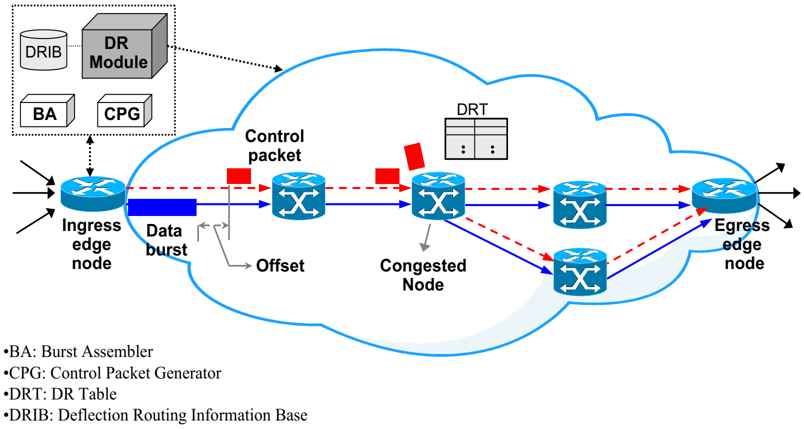


Figure 1: Basic architecture of an OBS network with deflection routing

the connection establishment. The control packet enables reservation of time slots within available wavelengths on links along the burst path. A delay referred to as offset time or FDL has been proposed to bring this form of reservation to fruition. Recently, several reservation protocols have been proposed to implement burst switching with different wavelength and timeslot reservation schemes. Two examples of such protocols are: offset-based scheme including JET [12] and Just-In Time (JIT) [13][14], FDL-based scheme [15]. The offset time allows for adequate time for the control packet to be processed at each node while the burst is buffered electronically at the source; thus, no FDLs are necessary at the intermediate nodes to delay the burst while the packet is being processed.

The proposed Contention-based Limited Deflection Routing (CLDR) mechanism can be applied to both styles of burst scheduling/reservation stated above, i.e., with offset time or with FDL. This is so because the CLDR is not a mechanism for how to reserve the wavelength and timeslot but it is about a criterion for deciding on doing deflection routing and further about a methodology for selection of the alternative path. With FDL, the need for deflection routing is somewhat less but can still be invoked when the FDL by itself does not resolve contention (i.e., FDL buffer overflow occurs). However, in consideration of the fact that FDL implementations is not quite mature in practice, in our simulation study, which will be described later in detail, the CLDR is implemented and studies in conjunction with the offset-based scheme.

Figure 1 shows a basic OBS architecture where deflection routing algorithm operates. While processing a control packet for sending a burst on a primary path, if it is determined that the burst is experiencing contention, then the another control packet is originated from the congested intermediate node and the burst is sent via an alternate path from that intermediate node. However, in the proposed CLDR method, there is an added element to the decision process as follows: It is first determined whether to alternate route a burst or to drop and do retransmission from the ingress node. This determination is based on a performance criterion. Further,

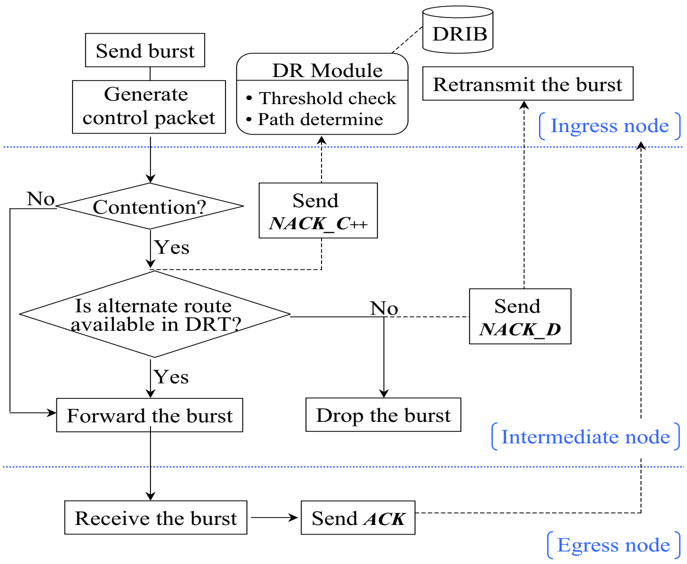


Figure 2: A flowchart describing burst contention notification and measurement in the CLDR algorithm

Figure 2 shows the flowchart that describes the operation of the CLDR scheme in relation to the architecture of Figure 1. For the implementation of the proposed CLDR method, as shown in Figures 1 and 2, there exists a management database referred to in Figure 2 as the Deflection Routing Information Base (DRIB) at the OBS edge node. The DRIB stores the management information for the optical burst layer together with the traditional DWDM transport and IP layers of network.

The edge node sends special control packets to carry the control information necessary for the OBS network to perform Operation, Administration, and Maintenance (OAM) functions. These functions include updating the DRIB to assist in deflection routing. These control packets are not associated individually with data bursts. When network status changes and management DB should be updated, these OAM control packets are generated and sent on a separate control channel. The separate control channel could be what is commonly known as the Optical Supervisory Channel (OSC). The OSC is made of a separate wavelength that is reserved for it on all fiber links. Thus by the use of these OAM control packets, each core switch could be informed of network status including burst loss rate due to contention, egress OBS node, and hop counts for each burst-mode connection.

The usual control packets are those that are associated individually with each burst. These control packets carry information regarding the number of hops traversed by a burst and the burst length. The control packet for burst is processed in order to schedule the burst through a node. If it is determined that the burst is experiencing a contention with another burst, the proposed CLDR protocol is invoked and it makes use of information in the associated control packet and the available information from the DRIB at the congested node. The congested node already has the associated attributes about its ports including contention status and hop counts from the OAM control packets. Additionally, a core node can also request an OAM control packet from the edge node when necessary.

An updated measurement about burst contentions is needed at all the nodes in the network for the CLDR algorithm to perform well. The flow chart of Figure 2 illustrates the mechanism for signaling contention occurrences and updating the burst contention measurement. An ingress node is a node from where a burst-mode connection originates and the egress node for that connection is the node where it terminates. Each ingress node receives updates about the burst congestion status along the primary and alternates routes for the bursts that have originated from it. These updates come in the form of two kinds of $NACK$ messages, $NACK\_C$ and $NACK\_D$, which are defined for primary and alternate paths, respectively. These messages help update the DRIB at the ingress nodes of each burst-mode connection. As illustrated in Figure 2, the $NACK\_C$ message is sent with an incremented value by an intermediate congested node to the ingress node when contention occurs on the primary path due to the lack of a timeslot in a wavelength for the burst in consideration. $NACK\_D$ is also sent by the intermediate congested node when there is no available alternate route in the Deflection Routing Table (DRT).

3 Contention-based Limited Deflection Routing Algorithm

3.1 Computation of Alternate Routes

In an OBS network, the deflection routing functions implemented in each switch automate the selection of alternate path setups when a control packet encounters a congested node over the primary path as illustrated in Figure 1. However, the each switch has current information only for the status of its own resources (wavelength availability, link congestion status, etc.). Similar information regarding other nodes and links may be stale. Thus, a local routing decision for the alternate route made at a node may result in a degraded global network performance in the long run. However, this is mitigated in the proposed CLDR algorithm by performing periodical global re-optimization of alternate routes based on updates received from other nodes regarding their most recent contention status. The messaging needed for the updating process was illustrated in Figure 2. Even though this re-optimization of alternate routes is periodical, it is to be performed not too frequently in order to stay within limits of the available computational power at a node; it can be performed perhaps once a day. It is also possible for the re-optimization to be performed on demand.

We now formulate the deflection routing problem by means of the following components: the network topology, node configuration, a set of attributes pertaining to node and link resources, and constraints pertaining to resource limits. The demands that are to be routed through alternate paths in the network are described by a set of attributes as well. Then, the problem is to find an optimal alternate path minimizing a cost function, which explicitly accounts for the contention rate as well as the burst hop distance. The aforementioned deflection routing problem can be formulated as follows: Consider a physical network represented by a graph $G(N, L)$, where N is the set of nodes and L the set of links (*i.e.*, fibers) connecting the nodes. It is assumed that each link between nodes i and j, has W_{ij} wavelengths with the same capacity of C. At each node n, $(n = 1, \ldots, N)$, the number of transmitters and receivers are defined as $P_n^{(t)}$ and $P_n^{(r)}$, respectively. If a node n has the number, P_n of ports, clearly, at most $\sum_n P_n$ wavelengths are needed to realize any possible virtual topology.

The applications on the network may be classified into (1) reserved and contention free, (2) contention oriented and loss-sensitive. Here we need to focus only on the loss sensitive traffic. For simplicity, we assume the first class of traffic to be absent. Let Λ be the set of traffic demands belonging to the loss-sensitive service class between a pair of edge nodes, where $\lambda_{ij}^{sd} \in \Lambda$ represents the arrival rate of bursts from source s to destination d that flows over a virtual link between node i and node j. Further, let $\lambda_{s_k d_k}$ denote the average flow of bursts associated with the k^{th} traffic demand requesting service. Let $D = \{D_{ij}\}$ be the distance matrix from node i to node j representing a propagation delay from node i to node j $(i \neq j)$. As the cost of contention from node i to node j $(i \neq j)$, let b_{ij} denote the burst blocking rate, which is collected periodically from the network.

In the deflection routing problem formulation, the variable, x_{ij} is defined as

$$x_{ij} = \begin{cases} 1 & \text{if alternate route includes a link } (i,j) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $i, j = \{1, 2, \ldots, N\}$ and $i \neq j$. This decision variable x_{ij} pertains to the specific k^{th} traffic demand at hand which is characterized by the $\lambda_{s_k d_k}$ average flow of bursts. Here, for the purpose of routing decisions, we are treating each burst-oriented Variable Bit-Rate (VBR) connection request as a Constant Bit-Rate (CBR) connection with an effective bandwidth of $\lambda_{s_k d_k}$. It should be noted that a specific burst requires one whole wavelength momentarily for a certain short duration needed for that burst to complete transmission onto a link. Then a subsequent burst from possibly a different demand might go over the same wavelength.

The constraint conditions are defined as follows. The number of lightpaths originating from and terminating at a node is not more than the node's out-degree and in-degree, respectively. Thus, only one lightpath per port can be setup at each node:

$$\sum_{\forall j \in N} x_{ij} \leq P_i^{(t)}, \sum_{\forall i \in N} x_{ij} \leq P_j^{(r)} \tag{2}$$

There are some constraints related to the traffic flow on virtual topology for all i and j. First, since we are setting up an alternate path for the optical bursts coming from a specific traffic flow the bursts of the demand $\lambda_{s_k d_k}$ are not segmented at any congested node in the network. Further, the flow of bursts belonging to a specific demand are not distributed fractionally onto different links except when they are switched to an alternate path as a result of deflection routing. Thus, we can state that the traffic demand $\lambda_{s_k d_k}$ is routed from node i to j on a single deflected path:

$$\lambda_{ij}^{s_k d_k} \in \{0, \lambda_{s_k d_k}\} \qquad \forall i, j \in N \tag{3}$$

Second, the total flow on the simplex link from node i to node j is expressed as the superposition of the existing traffic (*i.e.*, bursts) and the new burst associated with the link.

$$\lambda_{ij} = x_{ij} \sum_{s,d} \lambda_{ij}^{sd} + x_{ij} \lambda_{s_k d_k} \qquad \forall i, j \in N \tag{4}$$

For flow conservation at each node, the third constraint becomes

$$\sum_j x_{ij} - \sum_j x_{ji} = \begin{cases} 1, & i = s_k \\ -1, & i = d_k \\ 0, & \text{otherwise} \end{cases} \qquad \forall s_k, d_k, i \in N \tag{5}$$

Eq. 5 captures the fact that traffic flowing into a node should be equal to that flowing out of that node for any node other than the source and destination for each traffic flow k.
With regard to traffic, the final constraint assures that traffic flowing through a link can not exceed the total link capacity:

$$\lambda_{ij} \leq W_{ij} C \qquad \forall i, j \in N \tag{6}$$

where W_{ij} and C are number of wavelengths and capacity per wavelength for link ij.

Then, for the above constraints and the k^{th} traffic flow of loss-sensitive service class, we can design the objective function to find an alternate path from the congested node to the destination. This objective function is a weighted sum of the end-to-end burst blocking rate and the distance for the route, and is stated as follows:

$$Minimize\ g_d \sum_{i,j} x_{ij} D_{ij} + g_b[\log_{10}[1 - \prod\nolimits_{i,j}(1 - x_{ij} b_{ij})] \tag{7}$$

where g_d, and g_b denote the weights for delay and blocking, respectively. To decrease the computational complexity, we can express the above objective function, Eq. 7 as

$$Miminize \qquad g_d \sum_{i,j} x_{ij} D_{ij} + g_b \sum_{i,j} x_{ij} \log_{10} b_{ij}, \tag{8}$$

As for the burst contention rate, b_{ij}, the real data can be used that has been collected into the DRIB (see Figure 1). The weights, g_d and g_b are usually supplied by the network manager or carriers responsible for designing the network cost. In the end, the alternate route would be set up according to the values of the x_{ij} determined from the above integer linear programming formulation.

The objective function in Eq. 8 is more of a practical value than one involving distance alone. It includes Quality of Service (QoS) requirements regarding loss as well as distance. This objective function can be easily generalized to the case of multiple Classes of Service (CoS), where bursts of different CoS may have different QoS requirements regarding loss. The disparate CoS and their required QoS can be reflected into the routing decision by having different weights associated with each in our objective function.

In addition to the above constraints (Eq. 2- 6), for offset-based deflection routing schemes, we additionally need to consider the following. For bursts to arrive successfully at its destination over the alternate route computed by CLDR, an extra offset time or buffering delay needs to be allowed. When deflection routing is performed due to a contention at an intermediate node, the offset time on the alternate route is different from (usually, longer than) that on the primary path. One solution to this problem is to render sufficient extra offset time to each burst; another solution is to have the control packet reserve FDL buffer to delay the burst at intermediate nodes. Even though the above solutions are applied, it may happen that the significantly increased distance on an alternate route causes longer delay than the expected offset time or buffering time. Thus, if $t_{o,c}$ denotes a maximum limit on offset time for service class c, including the basic offset time and extra offset time, then another constraint for offset time is defined as

$$\sum_{i,j} x_{ij} D_{ij} \leq t_{o,c} \qquad \forall i,j. \tag{9}$$

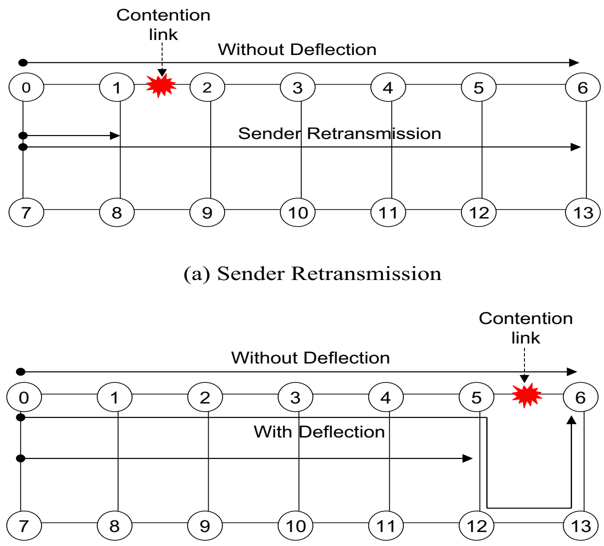
Alternatively, in Eq. 9, we can define a constraint for buffered delay simply by replacing $t_{o,c}$ in Eq. 9 with $t_{b,c}$, which denotes buffered delay limit for service class c.

As mentioned above, the optimization algorithm described in this section can be performed offline or online. In the former approach, multiple fixed alternate routes are considered when a contention occurs. Thus, each node in the network is required to maintain a DRT that contains an ordered list of a number of fixed alternate routes to each destination node. In the on-demand CLDR method, the alternate route from a congested node to a destination node is chosen dynamically, depending on the current network state. The on-demand CLDR method will require more computations and a longer response time than CLDR based on pre-computed alternate routes and lookup-table. But the on-demand CLDR approach is more flexible and would result in better resource utilization and performance than the latter approach.

3.2 Limited Deflection Routing Rules for CLDR

Our CLDR algorithm consists of (1) the optimal alternate routing methodology for Contention-based Deflection Routing that was described the preceding section and (2) the rules of Limited Deflection Routing that we describe in this section. The authors of [9] pointed out the limitations of normal deflection routing on WDM networks, and added two sender control functions in deflection routing protocol to reduce unnecessary deflection routing. One is sender check function and the other is sender retransmission function. In the optical switching node, if there are no available output links, it performs the sender check function before deflection routing, and selects the sender retransmission instead of deflection routing if the congested node is the sender itself. However, the minor modification proposed in [9] does not take the current network performance into consideration. Limited deflection routing decision tree should ideally include consideration of the current network performance. Therefore, we propose to add a threshold-check function which decides whether deflection routing is efficient or not at the congested node in light of the network performance.

The significance and an intuitive understanding of our enhancements to limited deflection routing can be stated using Figures 3(a) and (b). We propose that the decision whether to deflection route or drop and retransmit from source node should be performed at the congested node based on a performance measure, which is checked against a threshold. The threshold check is described in the next subsection. Figures 3(a) and (b) show some examples of burst transmissions in an OBS network, including the effects of deflection routing. We assume that a source or sender is node 0 and a destination is node 6. A burst transmitted from the sender would normally take the shortest path (0-1-2-3-4-5-6).



(a) Sender Retransmission

(b) Deflection Routing

Figure 3: The effect of limited deflection routing

- Case 1: A contention occurs on the link between node 5 and node 6, and the burst is dropped and retransmitted from the sender. In this case, the total number of hops is 11(=5+6) including the hops by sender retransmission assuming that the retransmission is successful.
- Case 2 : A contention occurs between node 1 and node 2, and the burst is dropped and retransmitted from the sender. The total number of hops becomes 7(=1+6) including the hops by sender retransmission assuming that the retransmission is successful.
- Case 3 : Deflection routing is used (rather than drop and retransmit) in case 1 at node 5, and the burst is sent over an alternate path. Thus the total number of hops is $5+\alpha$, where α denotes the number of hops in the deflection route.
- Case 4 : Deflection routing is used (rather than drop and retransmit) in case 2 at node 1, and the burst is sent over an alternate path. Thus the total number of hops is $1+\alpha$, where α denotes the number of hops in the deflection route.

We propose in our CLDR algorithm that a threshold check be executed before deciding on deflection routing vs. sender retransmission in each of the above four exemplary cases. The threshold check performed on a performance measure (described in the next subsection) introduces intelligence in the decision to alternate route vs. drop followed by sender retransmission. The threshold check function is designed to minimize resource consumption as well as provide higher network throughput. For a moment, just for simplicity, let us say that we use only the number of hops from the congested node to destination as the threshold check function. Then, case 2 and case 4 would be executed in the event of congestion on link 1-2 and link 5-6, respectively. Thus, if a

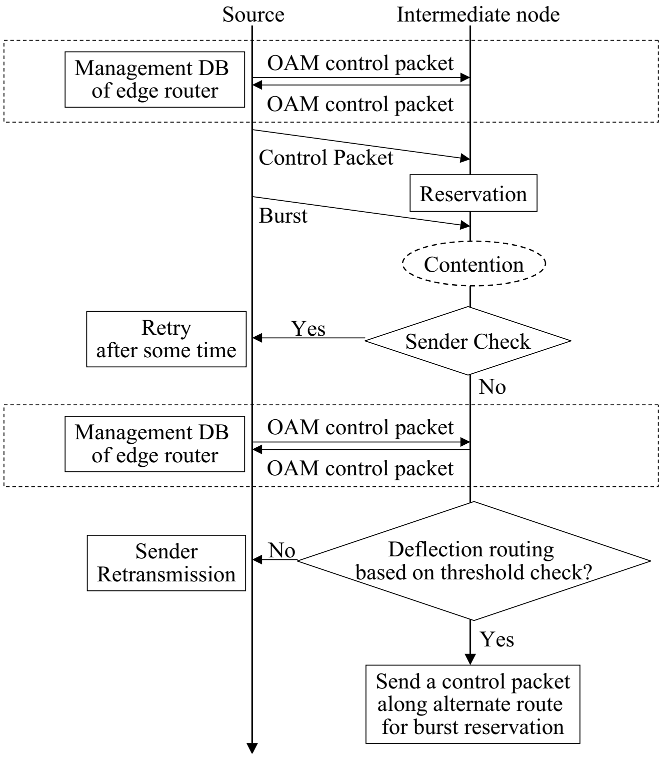


Figure 4: Limited deflection routing procedure

contention occurs on the link closer to the sender, such as 1-2 link, then drop followed by sender retransmission is performed instead of deflection routing. Further, if a contention occurs on the link closer to the destination, such as 5-6 link, then deflection routing is performed. As compared with the simple sender check function of [9], this approach reduces unnecessary deflection routing at the intermediate nodes as well as at the sender, and prevents contentions which are caused by inefficient deflection routing. More complex and useful performance measure and threshold check function are described in the next subsection. The steps of our mechanism are illustrated in Figure 4 and proceed as follows.

- Step 1: Source node sends out a burst control packet.
- Step 2: Intermediate nodes process the control packet and attempt to reserve the channel in anticipation of the burst that would follow.
- Step 3: Source node sends out the burst after offset time.
- Step 4: If there is no available egress channel for the burst at a node, at first it is checked whether the current node is sender or not. If the current node is the sender, then deflection routing is not done. Instead, after some wait time, the sender retransmits a burst control packet and subsequently the burst is retransmitted. If the current node is an intermediate node, then go to Step 5.

- Step 5: The current node is identified as an intermediate node. So the current node computes a performance measure and does the threshold check on that performance measure. Accordingly, it decides whether to deflection route or drop and notify sender to retransmit (see the next subsection for details). If the decision is to deflection route, then the alternate route selection is chosen as per the DRT. However, if there are no available routes in the DRT, then the current node drops the burst and sends NACK packet to sender for retransmission from the source.

3.2.1 Threshold Check Function

In this section, we formulate some threshold check functions to assist in deciding whether dropping or deflection routing should be done. Let s, d, and c denote the source, destination, and current nodes, respectively. Let N be the set of nodes in core network. Further, let N_c and N_d be the set of nodes that have been passed from source to the current node (i.e., the congested node in consideration) and the set of nodes that would be passed from the current node to destination, respectively. As in Eq. 1, $x_{i,i+1}$ is a binary variable associated with link $(i, i+1)$ between a node i and the next node $i+1$. So $(x_{i,i+1} = 1)$ indicates that link $(i, i+1)$ is inclusive as part of the route from source to destination.

We first define a threshold check function, which is based on hop counts alone.

$$\begin{aligned} C_h &= \sum_{\forall i,i+1 \in N_c} x_{i,i+1} - \sum_{\forall j,j+1 \in N_d} x_{j,j+1} \\ &= H_{sc} - H_{cd} \end{aligned} \tag{10}$$

$$\begin{cases} \textit{if } C_h \geq 0, & \text{deflection route the burst} \\ \textit{otherwise}, & \text{drop the burst} \end{cases}$$

If the hop count for nodes passed from source s to the congested node c are more than those that would be passed over the deflection route from the congested node c to destination d, that is, C_h is more than zero, then deflection routing is done. Otherwise, the burst that is experiencing contention is dropped. When Eq.10 is used as a threshold check function, the goals that are accomplished are: (1) economize network resources and improve performance for bursts by deflection routing if the current node is closer to destination or dropping and retransmitting if the current node is closer to source, and (2) decrease the control processing load (i.e., processing time and resources reserved by control packets) on hops.

Let p and q denote the number of links between the source and the congested node and the number of links between the congested node and the destination, respectively. Let b^* denote a tolerable end-to-end blocking rate for a route. We now define a threshold check function to satisfy b^*:

$$C_b = \log_{10} b^* - \log_{10}[1 - \prod_{i=1}^{d-1}(1 - b_{i,i+1})] \qquad \forall i, i+1 \in N_d \tag{11}$$

$$\begin{cases} \textit{if } C_b \geq 0, & \text{deflection route the burst} \\ \textit{otherwise}, & \text{drop the burst} \end{cases}$$

where $b_{i,i+1}$ denotes contention blocking probability between node i and node $i+1$. It is expected that selecting the burst on a path with smaller mean contention probability results in decreasing the burst loss rate and the blocking rate in overall network.

Now we generalize the threshold check function to include the path hop-count (or alternatively, the end-to-end distance) as well as the end-to-end burst blocking probability. The two performance measures in this threshold check are given different relative weights to emphasize one or the other, as desired. With a relatively large value M and burst blocking decision parameters b_2^* and b_1^* ($b_2^* > b_1^*$), we introduce two decision variables:

$$Q_h = \begin{cases} 1 & \text{if } C_h \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

and

$$Q_b = \begin{cases} 1 & \text{if } b_1^* \leq C_b \leq b_2^* \\ M & \text{if } C_b < b_1^* \\ -M & \text{if } C_b > b_2^* \end{cases} \tag{13}$$

Using the above two variables, we can now express a decision variable as

$$Q_t = w_h Q_h + Q_b \tag{14}$$

where $w_h \ll M$ is weight for emphasizing/de-emphasizing the hop count relative to the burst loss ratio. Then, a combined threshold check function can be stated as

$$C_t = \begin{cases} 1 & \text{if } Q_t \geq w_h + 1 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Figure 5 further illustrates how the proposed combined threshold check function works when w_h is set to 1.

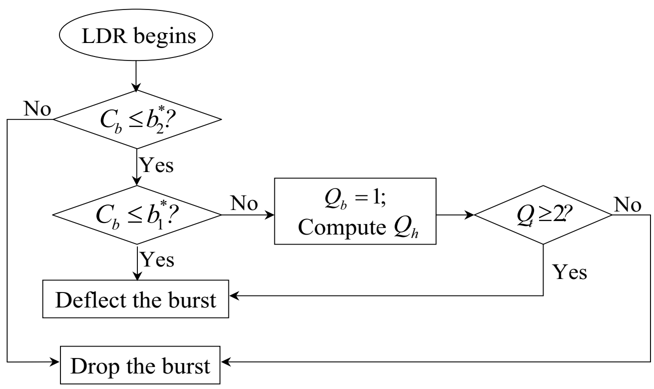


Figure 5: An example implementation of the threshold check function

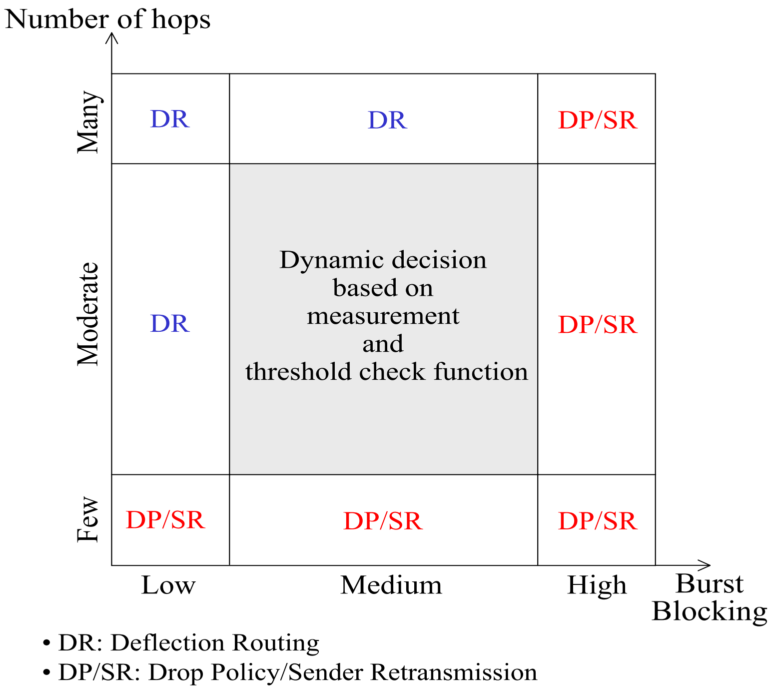


Figure 6: The desired actions for different ranges of burst blocking and hop-count

The desired actions for different ranges of blocking (or, alternatively, load) and hop-count are exemplified in Figure 6. Deflection Routing (DR) is desirable whenever the burst-blocking ratio is very small (low loads). And when the blocking ratio is very high (at high loads), Drop Policy with Sender Retransmission (DP/SR) is more suitable. In the middle range of blocking ratios, DR or DP/SR can be judiciously used depending on the outcome of the combined threshold check described above. Thus with the threshold check function defined above, the proposed CLDR is capable of operating in the most suitable and efficient way under different traffic and topological scenarios.

4 Queueing Model for Burst Loss Probability

The numerical results from analytical queueing models have been generated, in this paper, for a variety of traffic assumptions and parameter values such as on-off times, FDL buffer size, etc. These analytical results combined with numerous simulation results for various deflection routing protocols have provided significant insights about OBS networks. The analytical results are meant to baseline the performance behavior of statistical burst multiplexing, while the simulation results examine and compare the relative performance of the proposed CLDR method vis--vis other previously known methods.

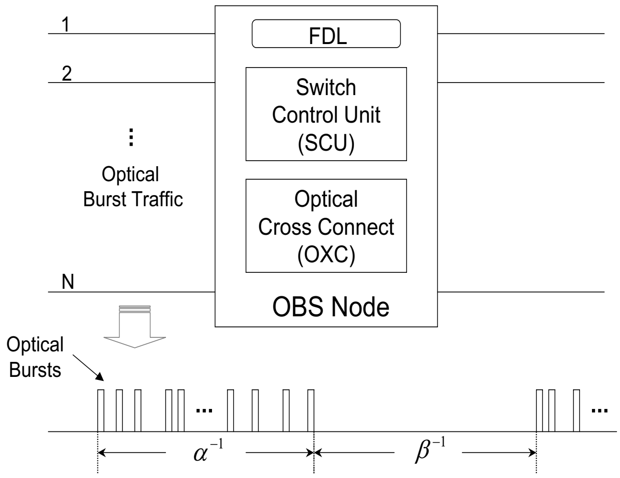


Figure 7: Illustration of an OBS node and input traffic from a source node

4.1 Source Traffic Model

We assume that a threshold-based burst assembly scheme is applied at the edge node, where once the length of a burst being created reaches a threshold value, L(Mbits), the burst is generated and placed in a burst queue. In practice, there would also be a burst assembly timer, which may sometimes expire sooner than full burst (L Mbits) creation; then the assembled burst up to that time is padded into a fixed size burst and placed in the queue.

Figure 7 illustrates an OBS node with multiple input and output fibers or ports. The incoming traffic on each wavelength of an incoming fiber is an aggregation of many individual sources of burst traffic. The bursts from each source are assumed to behave in an on-off manner as shown in Figure 7. In effect we assume that the each source behaves as an on-off source. Many bursts arrive with exponential random inter-arrival times or at fixed intervals during the on period, and the off period is typically much longer than the on period. As an example on average 12 bursts may arrive during the on period over 120 ms average duration, and the off period may have 880 ms average duration. We varied these parameters widely in our analytical and simulation results so as to capture the sensitivity to these parameters.

4.2 Queueing Model

We closely follow the analytical model presented in [17] for queueing analysis involving multiplexing of many on-off sources. We make suitable modifications to the model to make it suit the statistical burst multiplexing problem at hand. The numerical results from the model were generated for a variety of traffic assumptions and parameter values such as on-off times, FDL buffer size, etc. These analytical numerical results have provided significant insights about OBS networks together with the simulation results for various deflection routing protocols.

To describe the analytical model, let us define the following parameters:

- L: burst size (Mbits)
- C: link capacity (Gbps)
- $1/\alpha$: average On period (ms)
- $1/\beta$: average Off period (ms)
- λ: burst generation rate during On period
- n: number of sources simultaneously multiplexed on a link
- B: the burst queue size or the number of FDLs per output port (specified in total ms worth of buffering at link speed)
- i: system state in terms of number of sources simultaneously in On period($0 \leq i \leq n$)
- τ_i: effective time (ms) spent in system state i for burst delay to exceed buffer size Bms
- p_i: probability that system is in state i
- n_0: number of sources in On period simultaneously above which the system is considered to be in temporary overload (i.e., when $n_0 + 1$ or more sources are in On period, the instantaneous total burst arrival rate exceeds the burst service rate)

When the system is in state i for the rate for $i \geq (n_0 + 1)$, the rate at which bursts fill the queue is $(i - n_0)\lambda$ because bursts are assembled at rate $i\lambda$ and are served at rate $n_0\lambda$. When the system is in temporary overload state i, it has to have gone through other temporary overload states, i.e., $(n_0 + 1), (n_0 + 2), \cdots, i - 1$ before it reaches state i. When the system state $i = n_0 - 1$ then the system is not in overload. When the system state $i = n_0$ then the system is at the verge of overload. The quantity $(i - n_0 + 1)$ is a measure of the depth of excursion of the system into the possible set of overload states. The deeper the system has gone into the overload states, the less time it needs to spend there to cause large delays (or buffer overflow). Hence, we allow for an approximate adjustment factor $(i - n_0 + 1)$ in the denominator of the following equation for the duration of time τ_i that the system needs to be in state i for burst delay to exceed B ms

$$\tau_i = \frac{\frac{BC}{L\lambda}}{(i - n_0)(i - n_0 + 1)}. \tag{16}$$

In our numerical results, for example, we have used C = 6 Gbps (six wavelengths of 1 Gbps each) and $L = 1$ Mb. The system-state probabilities p_i are binomial distributed and given by

$$p_i = \binom{n}{i} (\frac{\beta}{\alpha + \beta})^i (\frac{\beta}{\alpha + \beta})^{n-i}. \tag{17}$$

Now, an approximate expression for the probability of burst loss due to buffer overflow P_L is given as follows:

$$P_L = \sum_{i=n_0+1}^{n} p_i \times \exp^{-i\alpha\tau_i} . \tag{18}$$

The above equation for burst blocking is very useful in that, unlike many other approximations available in the literature, it captures the effects of numerous traffic parameters. These parameters include the lengths on- and off- periods, burst arrival rate while the source is in on-period, the buffer size (FDL), the link load, and the link bandwidth. The superposition of on-off sources has significant temporal correlations [17][18] which influence the burst loss ratio and make it fairly sensitive to numerous parameters. The model and the result on predicted on burst loss do well to capture these correlations.

5 Performance Results and Comparisons

5.1 Analytical Model Based Performance Results

Figure 8 shows the burst loss probability as a function of link load for the case of 300 ms average on-period and 700 ms average off-period. The sensitivity to FDL size is shown. The burst loss ratio significantly decreases as the FDL size increases from 0 ms to 16.7 ms. The values of FDL = 0, 3.3, 8.3, 16.7 ms correspond to buffer sizes that can simultaneously delay (i.e., queue) 0, 20, 50, and 100 bursts, respectively.

Figures 9 and 10 show the sensitivity of the burst loss ratio to the average on- and off-periods. In Figure 9, the comparison is made between (α^{-1} = 300 ms, β^{-1} = 700 ms) vs. (α^{-1} = 30 ms, β^{-1} = 70 ms); both cases

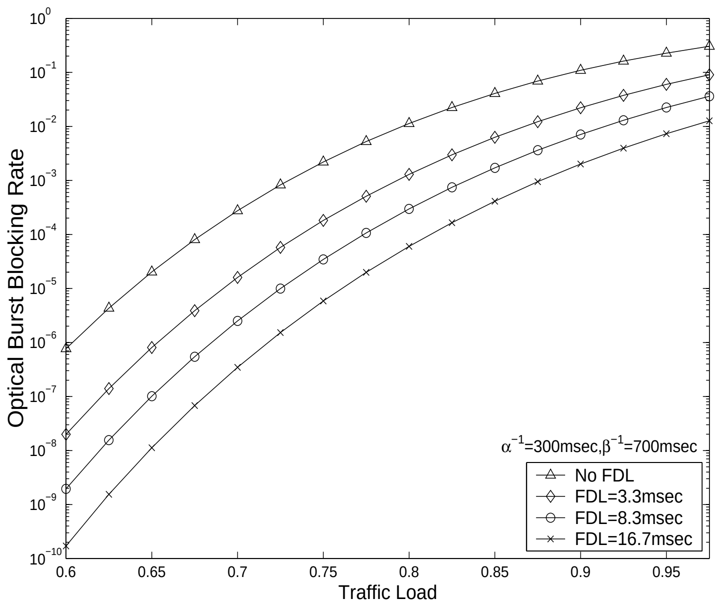


Figure 8: Burst blocking rate under varying FDL sizes

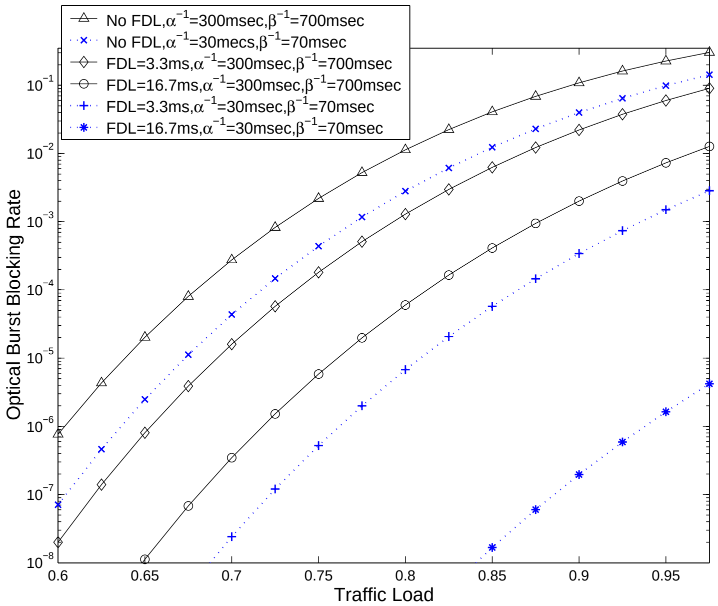


Figure 9: Burst blocking rate for high and low On/Off periods when activity is 0.3

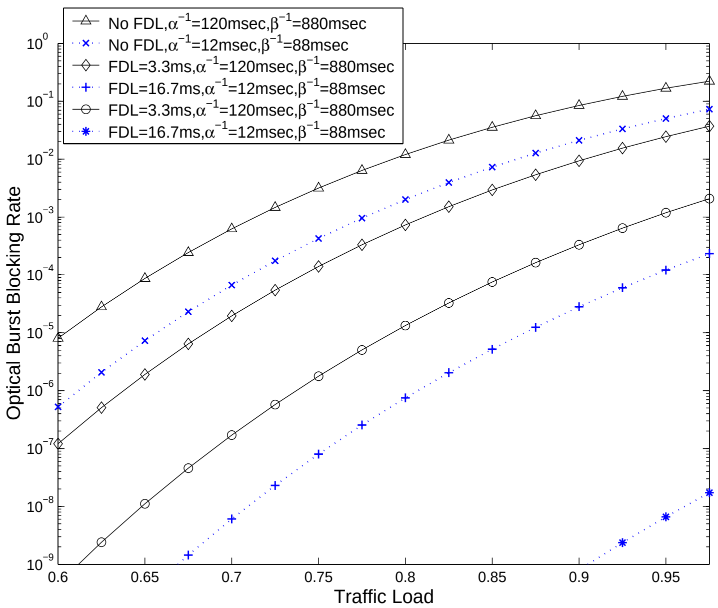


Figure 10: Burst blocking rate for high and low On/Off periods when activity is 0.12

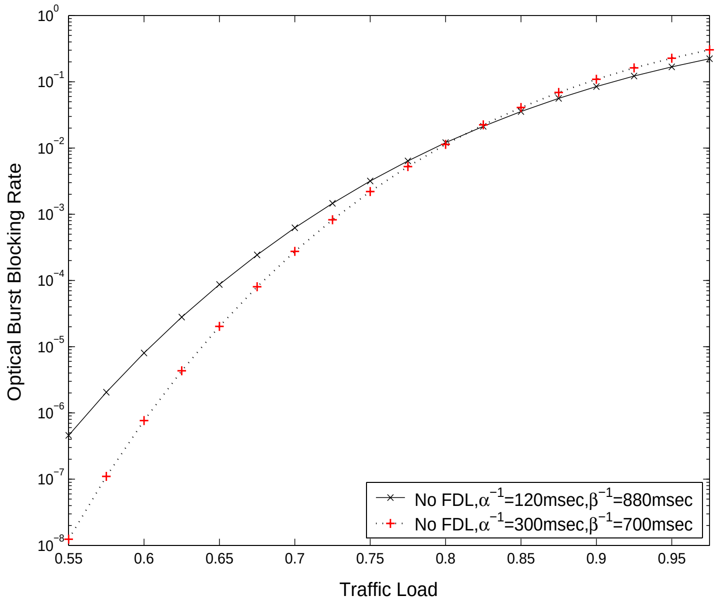


Figure 11: Burst blocking rate sensitivity to On/Off periods when there is no FDL

have an average activity factor of a=0.3. In Figure 9, the comparison is made between (α^{-1} = 120 ms, β^{-1} = 880 ms) vs. (α^{-1} = 12 ms, β^{-1} = 88 ms); both cases have an average activity factor of a=0.12. What we see is that while the activity factor is constant, the burst loss ratio is higher for the case of higher average on-period. The difference between 300 ms (or 120 ms) vs. 30 ms (or 12 ms) average on-periods is only marginally higher for the zero buffer case (FDL = 0), but it is about an order of magnitude higher for the cases when FDL is used. This is because buffering allows the temporal correlations in the combined burst traffic (superposition of many burst sources) to be manifest in a more influential manner. That is, when the different sources of burst traffic interact with each over a period of time in the buffer, then the temporal correlations get manifested. Figures 11, 12 and 13 highlight another very interesting phenomenon. Here we are comparing (α^{-1} = 300 ms, β^{-1} = 700 ms) vs. (α^{-1} = 120 ms, β^{-1} = 880 ms) for different buffer sizes. The activity factors are 0.3 and 0.12, respectively. The average on-periods are different (300 ms vs. 120ms), but not quite an order of magnitude different as was the case in Figures 9 and 10. The comparisons are shown in Figures 11, 12 and 13 for buffer sizes 0, 3.3, and 16.7 ms, respectively. What is very interesting is that for the no FDL case, the burst loss behavior is entirely the opposite of what was stated in the previous paragraph involving Figures 9 and 10. Burst loss ratio in Figure 11 is in fact lower for the case of the larger on-period. Only when the buffer size increases, do we see that the burst loss ratios go higher for the higher on-period in the higher load region (see Figures 12 and 13. This again can be explained by a combination of these observations: (1) at a given percentage link load level, the number of burst sources multiplexed are 2.5 times more for the case of a = 0.12 as compared to that for a = 0.3, and (2) the temporal correlations in the superposed burst traffic is not manifest at low loads and smaller FDL sizes, while it is quite influentially manifest at high loads with larger queues or buffer sizes.

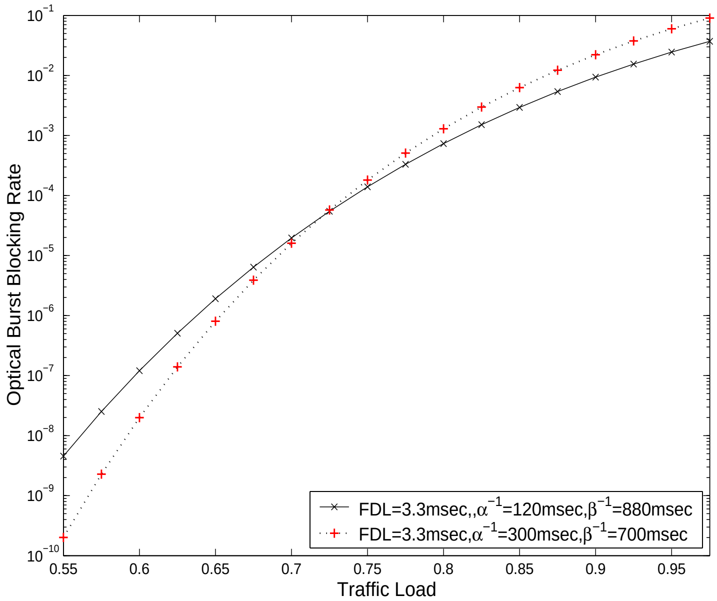


Figure 12: Burst blocking rate sensitivity to On/Off periods with moderate FDL size (FDL=3.3msec)

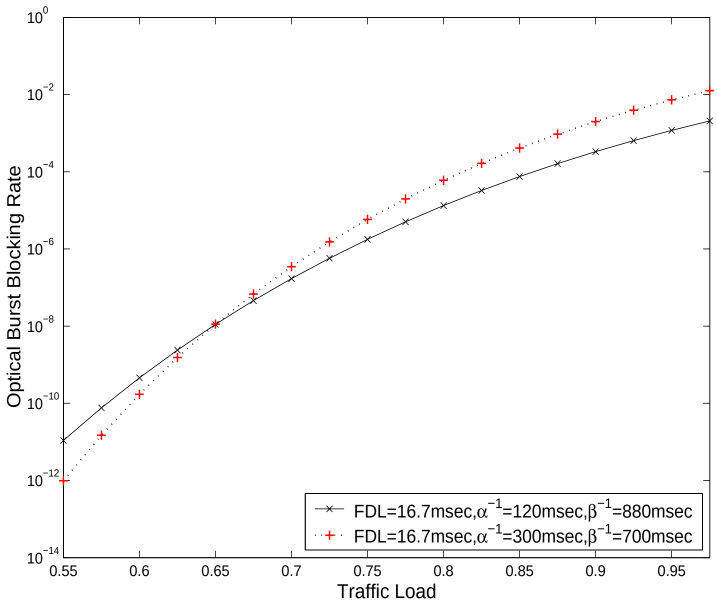


Figure 13: Burst blocking rate sensitivity to On/Off periods with large FDL size (FDL=16.7msec)

5.2 Simulation Environment

In order to evaluate the effectiveness of proposed CLDR technique, we did simulation tests for the CLDR algorithm and the well-known Shortest Path-based Deflection Routing (SPDR) algorithm [8]-[10]. As was mentioned in Section 2, we use the JET method of offset-based reservation in our simulation.

In our simulation, we assume that each fiber link is composed of same number of wavelengths. The burst sources were individually simulated with the on-off model as explained in the Queueing Model section. The burst source traffic parameters were varied similar those mentioned in the section above on Analytical Model Based Results. The tests were carried out using a 14-node NSFNET topology as shown in Figure 14. The transmission rate of each link is 6 Gbps, consisting of six wavelengths each operating at 1 Gbps. The simulations can be extended to more wavelengths and 10 Gbps, but we wanted to capture the key performance comparisons between CLDR and other defection routing schemes using a smaller capacity network in order to keep the simulation time manageable. FDL is assumed to be not used in the simulations.

Over NSFNET topology, five source-destination node pairs were chosen randomly and optical bursts are generated from the source nodes. Just as an example, looking at Figure 14, let us say that some bursts whose source and destination are CA1 and NJ, respectively, experience burst contention at UT node on UT-MI link on the primary path (CA1-UT-MI-NJ). In our simulation, let us say that the DRT lists the (UT-CO-MN-IL-PA-NJ) and (UT-CO-TX-MD-NJ) as alternate candidate paths. Of these, (UT-CO-MN-IL-PA-NJ) is the shortest alternate path from UT to NJ. However, the CLDR scheme can very well select (UT-CO-TX-MD-NJ) as the preferred alternate path if that happens to be the only one that meets the requirement on performance objective (which includes distance and burst blocking measures).

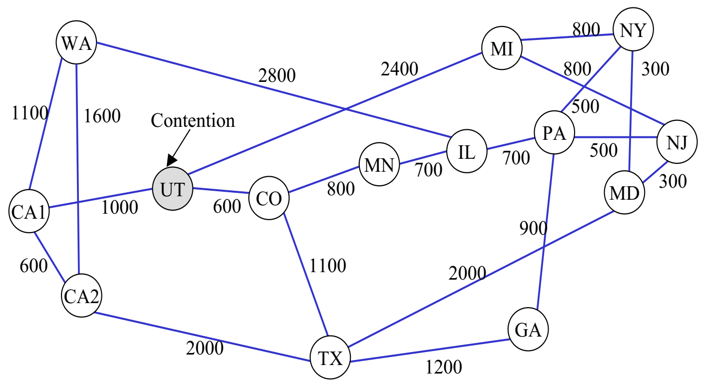


Figure 14: Simulation network topology

5.3 Performance of CLDR and Comparisons with Other Techniques

The focus of our performance evaluation is on burst (or data) loss rate caused by contention. A burst would be dropped if both primary and deflection paths are blocked. The data loss rate for the entire network is found by calculating the average of the burst drops over all source-destination pairs.

Figures 15 and 16 show simulation results comparing the burst blocking or loss rate for our CLDR method with the Shortest Path Deflection Routing (SPDR) method. The on-off durations are (α^{-1} = 120 ms, β^{-1} = 880 ms) and (α^{-1} = 300 ms, β^{-1} = 700 ms) in Figures 15 and 16, respectively. The SPDR algorithm simply picks the shortest path alternate route available from the DRT, whereas with the CLDR scheme the alternate path selection is based on minimizing a composite performance measure consisting of the alternate path distance as well as burst blocking along that path. For typical operating load values up to 0.75, the CLDR algorithm improves burst blocking by more than an order of magnitude as compared to SPDR in the test cases that we have studied through simulation runs.

Figure 17 shows a comparison of the burst blocking ratios for the CLDR scheme under two different scenarios, namely, (α^{-1} = 120 ms, β^{-1} = 880 ms) and (α^{-1} = 300 ms, β^{-1} = 700 ms). The reason we made this study of parameter sensitivity was to see if the burst loss results from simulations would be qualitatively in alignment with the analytical results of Figure 11, and it turns out they are. In both cases (simulation and analytical results), the case of longer on-period has lower burst loss than the case of shorter on-period up to fairly high values of load (when no FDL buffering is used). In the discussion of the analytical results, we discussed why this was counter-intuitive but could be explained.

Figure 18 shows a comparison of the CLDR scheme under two scenarios involving no FDL buffering vs. 3.3 ms of FDL buffering (i.e., capable of delaying/queueing up to 20 optical bursts) at each port in the network.

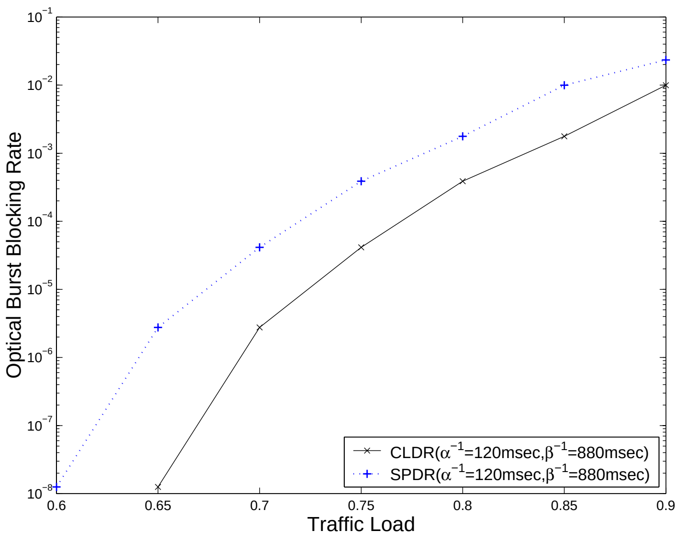


Figure 15: Burst blocking rate for CLDR and SPDR without FDL when activity is 0.12

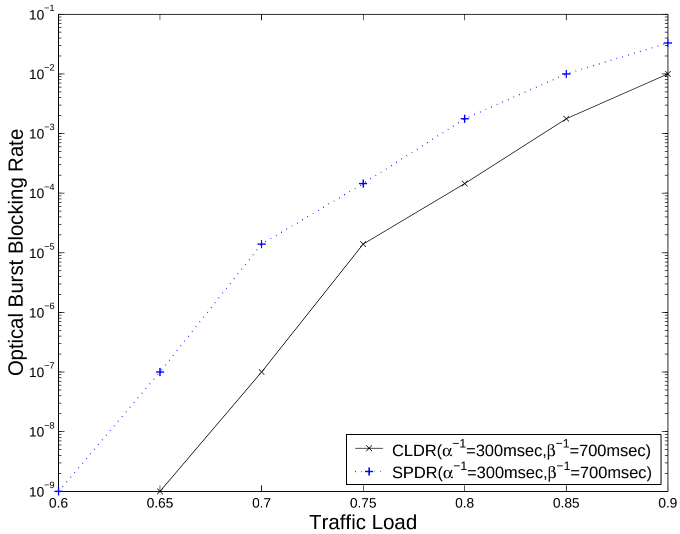


Figure 16: Burst blocking rate for CLDR and SPDR without FDL when activity is 0.3

The on-off parameters in this simulation study were set to ($\alpha^{-1} = 120$ ms, $\beta^{-1} = 880$ ms). The results are once again what we would have expected based on our prior analytical predictions; for this, we can compare Figure 18 with Figure 9. There is consistent improvement in burst blocking throughout the range of moderate to high loads. The burst blocking improvement is better than an order of magnitude in the typical operating range (say, about 0.65 to 0.75).

In the above simulation results, we showed how CLDR performs superior due to its consideration of a composite performance measure (path distance and burst blocking ratio) for pre-computation of alternate routes. There is another aspect of the CLDR scheme, which gives additional performance improvement (and consequently, throughput and delay improvement). As discussed before in the paper, this additional aspect is the limited deflection routing decision, based on a threshold check function, dynamically made at an intermediate node experiencing burst contention (see Section 4.2). The CLDR scheme examines whether the path performance degradation measures are better than certain threshold values for the alternate routes that are available from the DRT. If none of the available alternate routes pass this threshold check, then the burst is dropped at that node, and source node is sent a message to indicate retransmission is needed. If the path performance threshold check is valid, then deflection routing is done using a selected alternate route. Figure 19 highlights the advantage of this threshold check aspect of the CLDR by comparing the CLDR with a normal DR algorithm. The normal DR algorithm always deflection routes the burst experiencing contention, regardless of the quality of the alternate route available. From Figure 19, we see that the CLDR has better burst loss performance by about an order of magnitude as compared to the normal DR scheme, both under moderate and high link loads.

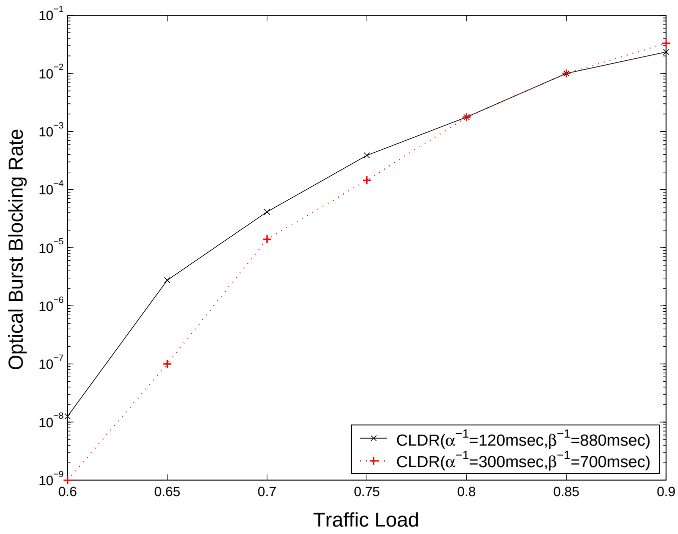


Figure 17: Burst blocking rate for CLDR under two different cases of On/Off periods

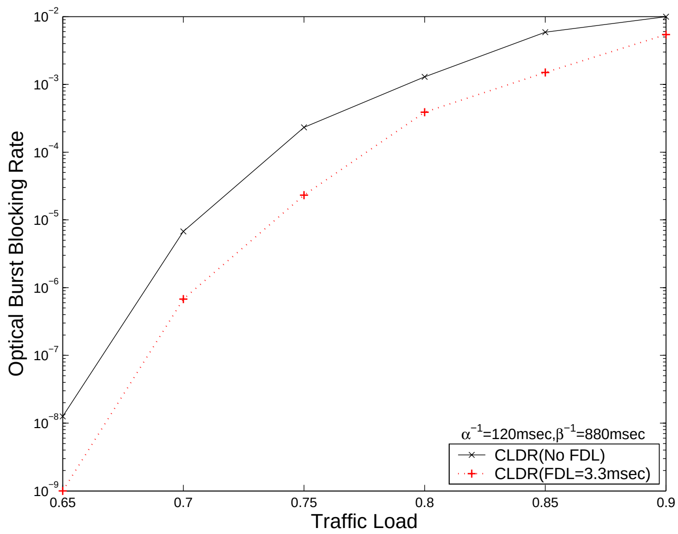


Figure 18: Burst blocking rate for CLDR without FDL and with FDL (of length 3.3 ms) when activity is 0.12

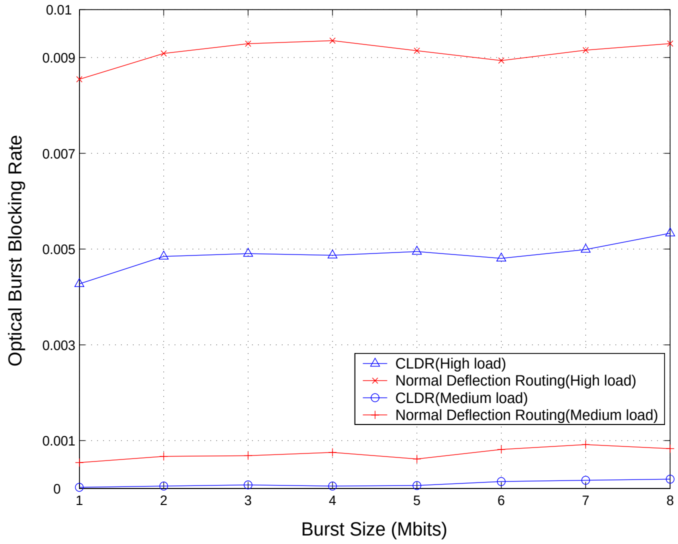


Figure 19: Burst loss rate comparison for normal deflection routing vs. CLDR under moderate and high load conditions

6 Conclusion

In this paper, we have shown that, in OBS networks, when deflection routing is used as a means for burst contention resolution, it is important to design alternate routes in an optimized fashion based on a composite performance measure that considers path distances as well as the expected burst loss probability along that alternate route. The proposed CLDR scheme was shown to perform significantly better than the Shortest Path Deflection Routing (SPDR) scheme that is known in the literature.

An additional salient feature of CLDR is that the limited deflection routing decision is based on a threshold-check function and is dynamically made at any intermediate node experiencing burst contention. The CLDR scheme examines whether current path performance degradation measures are better than certain threshold values for the alternate routes that are available. If none of the available alternate routes pass this threshold check, then the burst is dropped at that node, and source node is sent a message indicating that retransmission is needed. Otherwise, burst is deflection routed on the best available alternate route. A normal DR algorithm always deflection routes the burst experiencing contention, regardless of the quality of the alternate route available. Our simulation results have also shown that the CLDR scheme has better burst loss performance by at least an order of magnitude as compared to the normal DR scheme, both under moderate and high link loads.

We also presented numerical results based on analytical queueing models that have provided significant insights into the nature of statistical burst multiplexing at the edge as well as intermediate nodes. These models are useful in understanding the sensitivity of the burst loss to various traffic and system parameters. A number of simulation results were intuitively understandable due to the insights obtained from the analytical modeling.

We plan to carry out further simulations, as part of our future work, to study the effects of FDL buffering in more detail; for example, the effect of buffer partitioning principles (per port or per wavelength) on burst blocking ratio and network throughput improvement. The case of hybrid OBS multiplexing, involving circuit-switched (guaranteed bandwidth) connections and statistically multiplexed burst connections, is also of interest. In a separate technology overview paper [3], we have discussed numerous additional issues concerning OBS such as economic benefits, physical layer challenges for OBS implementation, protection and restoration, and enhancements to the control and signaling features.

Acknowledgement

This work was supported, in part, by the National Institute of Standards and Technology (NIST), the Advanced R&E Activity (ARDA), the Laboratory for Telecommunications Sciences (LTS) MENTER project.

References

[1] S. Verma, H. Chaskar, and R. Ravikanth, "Optical Burst Switching: A Viable Solution for Terabit IP Backbone", *IEEE Network Magazine* , vol. 14, no. 6 , pp. 48 -53, November/December 2000.

[2] "Optical Phased Array Technology for High-Speed Switching", *White Paper*, http://www.chiaro.com/pdf/ CHI100_OPA_1.pdf, Chiaro Networks, 2002.

[3] K. Sriram, D.W. Griffith, N.T. Golmie, S.K. Lee, D. Montgomery, and D.H. Su, "Optical Burst Switching: Benefits and Challenges," Technical Report (Draft), National Institute of Standards and Technology (NIST), 2003.

[4] C. Gauger, "Contention Resolution in Optical Burst Switching Networks", *In Advanced Infrastructures for Photonic Networks: WG 2 Intermediate Report*, http://www.ikr.uni-stuttgart.de/Content /Publications/Archive/Ga_COST266WG2_34734.ps.gz, 2002.

[5] M. Yoo, C. Qiao, and S. Dixit, "A Comparative Study of Contention Resolution Policies in Optical Burst Switched WDM Networks", *Internaional Conference on Terabit Optical Networking: Architecture, Control and Management Issues, Proceedings of SPIE*, vol. 4213, pp. 124-135, Boston, November, 2000.

[6] C. Gauger, "Performance of Converter Pools for Contention Resolution in Optical Burst Switching", *Proceedings of the SPIE Optical Networking and Communications Conference (OptiComm 2002)*, Boston, pp. 109-117, July 2002.

[7] C-F. Hsu, T-L Liu and F-F Huang, "Performance Analysis of Deflection Routing in Optical Burst-Switched Networks", *IEEE Proceedings INFOCOM'02*, vol. 1, pp. 66-73, NewYork, USA, June 2002.

[8] S. Kim, N. Kim, and M. Kang, "Contention Resolution for Optical Burst Switching Networks Using Alternative Routing", *IEEE Proceedings ICC'02*, vol. 5, pp. 2678-2681, NewYork, USA, Apr/May. 2002.

[9] X. Wang, H. Morikawa, and T. Aoyama, "Burst Optical Deflection Routing Protocol for Wavelength Routing WDM Networks", *Optical Networks Magazine*, November/December 2002, pp. 12-19.

[10] C.Y. Li et al, "Deflection Routing in Slotted self-routing Networks with Arbitrary Topology", *IEEE ICC'02*, vol. 5, pp. 2781-2785, NewYork, USA, April/May 2002.

[11] Vinod Vokkarane and Jason Jue, "Prioritized Routing and Burst Segmentation for QoS in Optical Burst-Switched Networks", *Proc. Optical Fiber Communication Conference (OFC) 2002*, pp. 221-222, March 2002.

[12] C. Qiao, and M. Yoo, "Optical Burst Switching (OBS) - A New Paradigm for an Optical Internet," *Journal of High Speed Networks*, vol. 8, no. 1, pp. 69-84, January 1999.

[13] J. Y. Wei, J. L. Pastor, R. S. Ramamurthy, and Y. Tsai, "Just-In-Time Optical Burst Switching for Multi-wavelength Networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 18. no. 12, pp. 2019-2037, December 2000.

[14] I. Baldine, George N. Rouskas, H.Perros, and D. Stevenson, "Jumpstart: a Just-in-Time Signaling Architecture for WDM Burst-Switched Networks", *IEEE Communications Magazine*, vol. 40, no. 2, pp. 82-89, February 2002.

[15] C. Gauger, K. Dolzer, and M. Scharf, "Reservation strategies for FDL buffers in OBS networks", *Report*, IND, University of Stuttgart, 2001

[16] R. Ramaswami and K. N. Sivarajan, "Optical Networks-A Practical Perspective" , Second edition, *Morgan Kaufmann Publishers*, pp.649, 2002.

[17] K. Sriram, T.G. Lyons, and Y.T. Wang, "Anomalies due to Delay and Loss in AAL2 Packet Voice Systems: Performance Models and Methods of Mitigation," *IEEE Journal on Selected Areas in Commun.*, vol. 17 , no. 1, pp. 4-17, January 1999.

[18] K. Sriram and W. Whitt, "Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data," *IEEE Journal on Selected Areas in Commun.*, vol. SAC-4, no. 6, pp. 833-846, September 1986.