

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

SuKyoung Lee, *Member, IEEE*, Kotikalapudi Sriram, *Fellow, IEEE*, HyunSook Kim, *Member, IEEE*, and JooSeok Song, *Member, IEEE*

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 with 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 and 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 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.

Index Terms—Burst contention, burst loss mitigation, deflection routing, optical burst switching (OBS), optical Internet, performance.

I. INTRODUCTION

THE OPTICAL core networks have the capacity to carry terabytes of data per second through each node. The edge routers feed data into these networks. The data is typically carried over 10 Gb/s 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 Gb/s 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 ultrahigh capacity [terabytes per second (Tb/s)] core packet routers at several of the nodes in the core network. These high-capacity routers regroom the packet traffic arriving from various nodes and try to statistically multiplex and pack the 10 Gb/s wavelength channels efficiently. Another approach for efficient bandwidth usage in the core optical networks is via optical burst switching (OBS) [1]–[6], aspects of which are the focuses 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 onto a wavelength on an egress port. 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 wavelength-division-multiplexing/dense (WDM/DWDM) capability can be produced less expensively than equipment combining ultrahigh capacity core routers, optical switches, and WDM/DWDM. 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 [3], [6]. 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 (WXC or OXC) together with tunable lasers.

One of the challenging issues in the implementation of burst switching is the resolution of contentions that result 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 [7]–[15]. 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 often 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 reroutes (on the fly) the contending bursts to an output port other than the intended output port. Thus, deflection routing is a very practical approach to resolve contentions, and has been

Manuscript received February 14, 2004; revised November 12, 2004. This work was supported in part by the U.S. National Communications System and in part by the Korea Science and Engineering Foundation under the OIRC Project.

S. Lee is with the Graduate School of Information and Communications, Sejong University, Seoul 143-747, Korea (e-mail: sklee@sejong.ac.kr).

K. Sriram is with the National Institute of Standards and Technology, Gaithersburg, MD 20899-1070 USA.

H. Kim is with the Department of Electrical Engineering, Photonics and Networking Research Laboratory, Stanford University, Stanford, CA 94305 USA.

J. Song is with the Department of Computer Science, Yonsei University, Seoul 120-749, Korea.

Digital Object Identifier 10.1109/JSAC.2005.851742

examined through simulations, as well as analysis in [10]–[15]. Prior to the emergence of OBS networks, deflection routing was first used as a contention resolution method in optical networks with regular mesh topologies [16]. In [17] and [18], deflection routing is shown to provide much improved performance as compared with hot-potato routing in a network with high-connectivity topology, such as ShuffleNet. The authors of [19] have presented a heuristic that enhances unslotted deflection routing to provide similar performance level as slotted routing. In [20], the concept of priority is introduced and output ports are selected based on preassigned port priorities, while considering irregular mesh topologies.

With the emergence of OBS technology, a deflection routing protocol for OBS network was proposed in [12] and [13], demonstrating that deflection routing reduced the burst loss and the average delay as compared with the method of data retransmission from the source. Some recent work about deflection routing is reported in [10]–[14]. The authors of [10] investigate the performance of deflection routing in OBS networks with prioritized burst types and just-enough-time (JET) scheduling. In [11]–[13], it is demonstrated via simulation studies that the blocking probability improves when deflection routing is used as a means for contention resolution. The authors of [14] describe how deflection routing can be used in conjunction with the self-routing address scheme. However, these studies do not address the issue of how routing to an alternate path should be done, while considering some performance constraints.

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 with techniques currently known in the literature. While several variants of the basic deflection routing scheme have been proposed before [10]–[14], 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 and 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 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 recompute and store optical paths, with the aim of staying optimal while node and link congestion measures may be changing. 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 blocking rate 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 blocking rate. The analytical model results are compared with simulation results, and are used to help understand simulation results more intuitively. Additionally, the analytical modeling results are also used to generate estimates for 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 II and III describe the CLDR mechanism and other enhancements in detail. The analytic model for burst loss probability is presented in Section IV. In Section V, 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 VI.

II. RESERVATION PROTOCOL

In OBS networks, 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 acknowledgment for the connection establishment. The control packet enables reservation of time slots within available wavelengths on links along the burst path. Burst delay methods using an offset time or FDLs have 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 time slot reservation schemes. Two classes of such protocols are: offset-based schemes including JET [21] and just-in-time (JIT) [22], [23], and FDL-based schemes [24]. 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 control packet is being processed.

The proposed CLDR mechanism can be applied to both classes of burst scheduling/reservation schemes stated above, i.e., with offset time or with FDL. The emphasis in CLDR is on criteria for deciding on doing deflection routing and for selection of an alternative path. The need for deflection routing is somewhat less when FDLs are used, 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 implementation is not quite mature in practice, in our simulation study, which will be described later in detail, the CLDR is implemented and studied in conjunction with an offset-based scheme.

Fig. 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 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

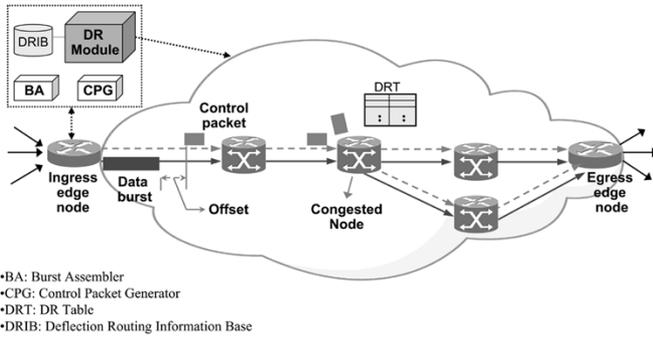


Fig. 1. Basic architecture of an OBS network with deflection routing.

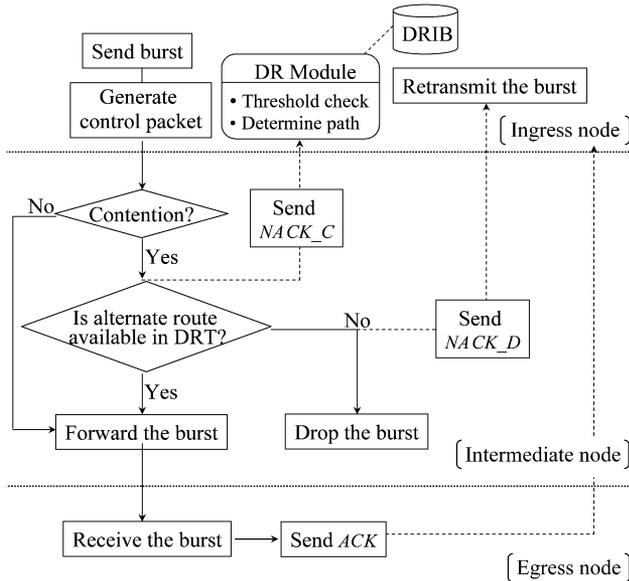


Fig. 2. Flowchart describing burst contention notification and measurement in the CLDR algorithm.

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, Fig. 2 shows the flowchart that describes the operation of the CLDR scheme in relation to the architecture of Fig. 1. For the implementation of the proposed CLDR method, as shown in Figs. 1 and 2, there exists a management database referred to 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 Internet protocol (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 the 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 uses 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 a 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 Fig. 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 Fig. 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 time slot 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).

III. CONTENTION-BASED LIMITED DEFLECTION ROUTING (CLDR) ALGORITHM

A. 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 Fig. 1. However, 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 periodic global reoptimization 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 Fig. 2. Even though this reoptimization of alternate routes is periodic, it is to be performed not too

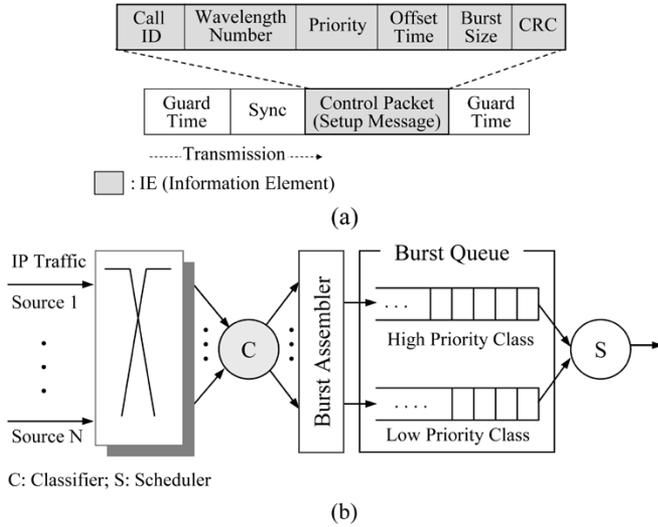


Fig. 3. (a) Control packet including priority field. (b) Classification of bursts into different priorities at ingress node.

frequently in order to stay within limits of the available computational power at a node. Based on experience, the network operator can decide the frequency of these computations. It is also possible for the reoptimization to be performed on demand.

The applications on the network may be classified into: 1) real-time and high-priority traffic and 2) nonreal-time and low-priority traffic. A burst belonging to the real-time class is allocated higher priority than a burst belonging to the nonreal-time class. An example of high-priority burst traffic is a virtual private network (VPN) with stringent service level agreement (SLA), e.g., voice-over-IP (VOIP) aggregate flows. The low-priority burst classification would be typically used for traffic streams that are delay tolerant and have somewhat stringent loss tolerance. The priority of each burst can be discerned by including a “priority” field in the control packet. Fig. 3(a) shows an example control packet containing a *Setup* message using the format proposed in [5]. Each field in the control packet can be either a hardware information element (IE) or a software IE, depending on the network architecture. Fig. 3(b) shows an example architecture for the packet classification, burst assembly, and priority burst queueing at the output port of an ingress node [6]. At the egress node, each burst is disassembled back into packets, which are forwarded to their destinations or the next hops through output links.

The low-priority bursts are more likely to be deflected, while the high-priority bursts, which normally have a much higher chance of wavelength and time-slot allocation, are far less likely to be deflected. In principle, the low-priority bursts may be further classified into multiple priority classes. In that case, different weighting factors should be applied to burst loss and delay for different classes in the optimization when computing alternate routes. In our formulation of the optimization problem for route computation, we assume for now that this finer classification is not done. However, the formulation can be easily generalized to the case of finer granularity within the low-priority class.

In this section, we describe how alternate routes are computed for updating the routing information in the DRTs, and these can-

didate routes are used when deflection routing is performed. In Section III-B, we describe an efficient algorithm for making a decision regarding burst deflection onto an alternate route from a congested node versus burst dropping followed by retransmission from the source node.

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 limits on those resources. 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 each with the same capacity of C Gb/s. At each node n , ($n = 1, \dots, N$), the numbers 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 topology.

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 flow 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.

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

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

where $i, j = \{1, 2, \dots, N\}$ 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 no more than the node’s out-degree and in-degree, respectively. Thus, only one lightpath per port can be setup at each node

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

There are some constraints related to the traffic flow on a 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 is not distributed fractionally

onto different links (except when it occurs naturally 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

$$\lambda_{ij}^{s_k d_k} \in \{0, \lambda_{s_k d_k}\} \quad \forall i, j \in N. \quad (3)$$

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 flow associated with the k th traffic demand

$$\lambda_{ij} = \sum_{s,d} \lambda_{ij}^{sd} + \lambda_{s_k d_k} \quad \forall i, j \in N. \quad (4)$$

For the traffic flow on each link, we define a constraint to assure that traffic flowing through a link cannot exceed the total link capacity

$$\lambda_{ij} \leq W_{ij}C \quad \forall i, j \in N \quad (5)$$

where W_{ij} and C are number of wavelengths and capacity per wavelength for link ij , respectively. If the link between node i and node j is not part of the alternate path, no burst associated with the k th traffic flow can exist on that link. This constraint can be expressed as

$$\lambda_{ij}^{s_k d_k} \leq x_{ij} \lambda_{s_k d_k} \quad \forall i, j \in N \quad (6)$$

where λ_{sd} ($\forall s, d \in N$) includes $\lambda_{s_k d_k}$. The constraint in (6) ensures that the bursts from the k th traffic flow can only flow through an existing alternate path. Finally, the constraint for flow conservation at each node becomes

$$\sum_j x_{ij} - \sum_j x_{ji} = \begin{cases} 1, & i = s_k \\ -1, & i = d_k \\ 0, & \text{otherwise} \end{cases} \quad \forall s_k, d_k, i \in N. \quad (7)$$

Equation (7) 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 .

Given the above constraints and the k th traffic flow of loss-sensitive service class, we can now specify an objective function to find an alternate path from the congested node to the destination. 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. The proposed objective function is a weighted sum of the end-to-end burst blocking rate and the distance for the route. Assuming that blocking events occur independently from link to link, the objective function is stated as follows:

$$\begin{aligned} \text{Minimize } g_d \sum_{i,j} x_{ij} D_{ij} \\ + g_b \left[\log_{10} \left[1 - \prod_{i,j} (1 - x_{ij} b_{ij}) \right] \right] \end{aligned} \quad (8)$$

where g_d , and g_b denote the weights for delay and blocking, respectively. To decrease the computational complexity, we can consider another similar objective function as follows:

$$\text{Minimize } g_d \sum_{i,j} x_{ij} D_{ij} + g_b \sum_{i,j} x_{ij} \log_{10} b_{ij}. \quad (9)$$

Equations (8) and (9) differ only in the second term. The second term in each is an increasing function of the burst-blocking rate (as would be required). However, the latter equation involves fewer computations due to much fewer multiplication operations. Also, for any given route it can be shown that the second term in (9) is equal to or somewhat higher than that in (8) for the range of values of parameters of interest (see Appendix A). This essentially tends to put a higher emphasis on burst losses over hop-count in (9) relative to that in (8). However, this can be balanced if desired by appropriately choosing the values of weights g_d and g_b .

The final outcome of each optimization run are the values of the x_{ij} that define the alternate path for deflection routing. Based on the above comments, it appears that the choice of (9) over (8) is implicitly justified.

As for values of the burst contention rate, b_{ij} , one can use the measured data that has been collected into the DRIB (see Figs. 1 and 2). The weights, g_d and g_b are usually supplied by the network manager or carrier responsible for maintenance of the network. As the final outcome, the alternate routes would be determined and loaded into the DRTs according to the values of the x_{ij} determined from the above integer linear programming (ILP) formulation.

The objective function in (8) or (9) has more practical importance than one involving distance (hop count) alone. It includes quality-of-service (QoS) requirements regarding burst loss, as well as distance (equivalently delay). 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 ((2)–(7)), for offset-based deflection routing schemes, we additionally need to consider the following. For bursts to arrive successfully at their 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 when 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 a constraint for the offset time is defined as

$$\sum_{i,j} x_{ij} D_{ij} \leq t_{o,c} \quad \forall i, j. \quad (10)$$

Alternatively, in (10), we can define a constraint for buffered delay simply by replacing $t_{o,c}$ in (10) 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. When a periodic update of the DRTs is done, a set of projected traffic demands for deflection routing is considered in the CLDR algorithm instead of a k th burst traffic flow emanating from deflection routing. 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 precomputed 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.

We will now discuss a way to enable scalability of the proposed CLDR algorithm for a realistic OBS network. To make the periodic alternate route computation for CLDR computationally scalable, the DR module in Figs. 1 and 2 can be implemented in a distributed manner at multiple hub routers. Here, we adopt the concept of hub routers as proposed in [25] and [26], where the idea is that some nodes in a network have an out degree sufficiently higher than other nodes to serve as major deflection routing nodes. A typical OBS network would have several such hub routers geographically distributed over the network topology. In this distributed architecture, each of the hub routers is responsible for periodic updating (or on-demand computation) of the candidate lists of alternate routes for a subset of the network nodes, i.e., those that are in its neighborhood. All routers (or most of the routers) in the network implement the deflection routing but only the hub routers perform the alternate route computations. Those routers that are not the hub routers communicate with their nearest hub-router to obtain the necessary DRT information. This distributed approach makes the CLDR implementation scalable to any OBS network size and topology.

All routers in an OBS network need to communicate with each other to share required CLDR information about traffic load, burst contention rate, etc. Such information could be obtained through periodic exchange of local network status among the hub router and its neighboring routers. The hub routers can further exchange such information with each other. For this information exchange, the generalized multiprotocol label switching (GMPLS) [27] can be employed as the control signaling protocol. Extended interior gateway protocols (IGP) such as OSPF-TE/IS-IS TE and the link management protocol (LMP) of GMPLS can distribute network status information between the OBS routers. Specifically for OBS networks, labeled optical burst switching (LOBS) has also been proposed to augment OBS nodes with IP/MPLS controllers [28], and there are also proposals on GMPLS-based photonic burst switching architectures [29].

B. 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 in the preceding section and 2) the rules of limited deflection routing that we describe in this section. The authors of [13] pointed out the limitations of normal deflection

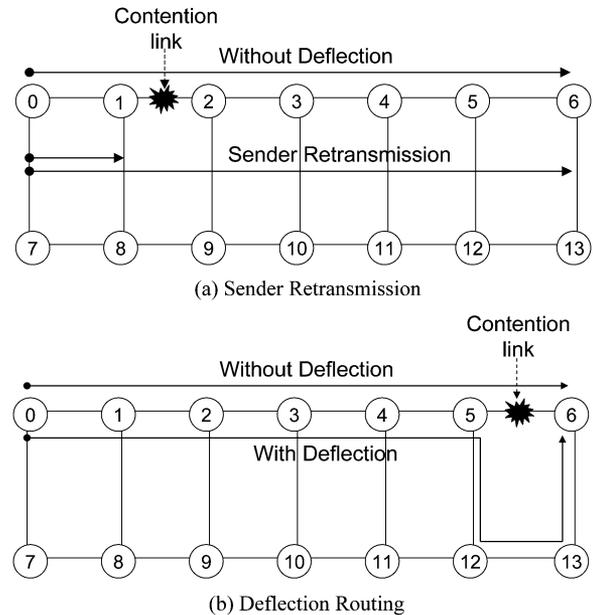


Fig. 4. Effect of limited deflection routing.

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 [13] does not take the current network performance into consideration. Limited deflection routing decision 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 Fig. 4(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 Section III-B1. Fig. 4(a) and (b) shows 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).

- 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 versus sender retransmission in each of the above four exemplary cases. The threshold check performed on a performance measure (described in Section III-B1) introduces intelligence in the decision to alternate route versus 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, Cases 2) and 3) would be executed in the event of congestion on link 1–2 and link 5–6, respectively. Thus, if a 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 [13], 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 Section III-B1. The steps of our mechanism are illustrated in Fig. 5 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 a 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 Section III-B1 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.

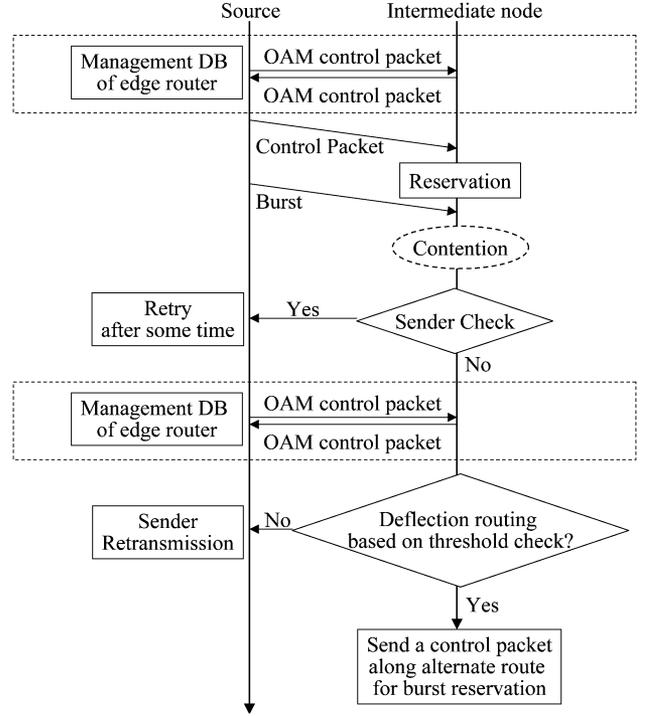


Fig. 5. Limited deflection routing procedure.

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 the 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 by the primary path, respectively. As in (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

$$C_h = \sum_{\forall i, i+1 \in N_c} x_{i,i+1} - \sum_{\forall j, j+1 \in N_d} x_{j,j+1}$$

$$\text{Decision: } \begin{cases} \text{if } (C_h \geq 0), & \text{deflection route the burst} \\ \text{otherwise,} & \text{drop the burst} \end{cases} \quad (11)$$

If the hop count for the primary path from source s to the congested node c is more than that for the primary path from the congested node c to destination d , that is, C_h is more than zero, we can keep deflection routing in mind to resolve the contention. Otherwise, the burst that is experiencing contention is dropped. When (11) 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 and overhead (i.e., processing time and resources reserved by control packets).

Let b^* denote a tolerable end-to-end blocking rate for a route. We now define a threshold check function to satisfy b^* :

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

$$\text{Decision: } \begin{cases} \text{if } C_b \geq 0, & \text{deflection route the burst} \\ \text{otherwise,} & \text{drop the burst} \end{cases} \quad (12)$$

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

Now, we generalize the threshold check function to include the path hop-count (or alternatively, the link distance), as well as the 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

$$Q_h = \begin{cases} 1, & \text{if } (C_h \geq 0) \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

and

$$Q_b = \begin{cases} 1, & \text{if } b_1^* \leq C_b \leq b_2^* \\ M, & \text{if } C_b < b_1^* \\ -M, & \text{if } C_b > b_2^* \end{cases} \quad (14)$$

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

$$Q_t = w_h Q_h + Q_b \quad (15)$$

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

$$C_t = \begin{cases} 1, & \text{if } Q_t \geq w_h + 1 \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

Deflection routing is done if $C_t = 1$. Fig. 6 further illustrates how the proposed combined threshold check function works when w_h is set to 1.

The desired actions for different ranges of blocking (or, alternatively, load) and hop-count are exemplified in Fig. 7. Deflection routing (DR) is desirable whenever the burst-blocking ratio is very small (low loads). 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 as defined above, the proposed CLDR is capable of operating in the most suitable and efficient way under different traffic and topological scenarios.

It may be noted here once again that the alternate route selection part of the CLDR scheme (i.e., the first stage of CLDR described in Section III-A) does yield candidate alternate paths (placed in the DRT) that are optimized for hop-count (or distance), as well as burst blocking ratios on the available alternate paths. It is just in the decision process (described

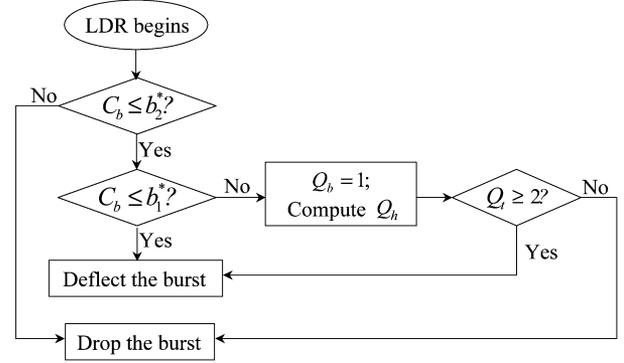


Fig. 6. Implementation example of the threshold check function.

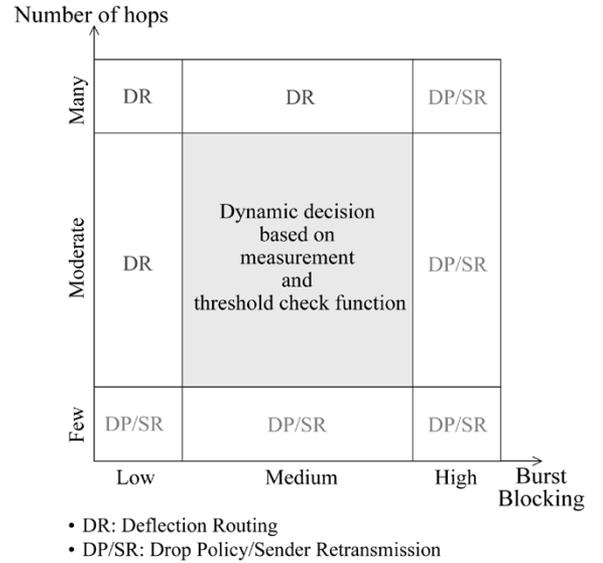


Fig. 7. Desired actions for different ranges of burst blocking and hop-count.

in this section) of whether or not to alternate route, where the primary path plays a role. In determining validity of the threshold check, we base the computation on the primary path instead of the deflection route because of the following reasons. First, there is expected to be a reasonable degree of correlation in terms of the average burst blocking between the primary path and the deflection route. This is because typically they would both be approximately in the same parts of the network topology. Second, deflecting nodes have information on the next hop for deflection routing purpose but not the entire deflection route. The nodes only keep information in DRTs about which input port is mapped to which output port for deflection routing purpose. It would involve significant control and processing overhead at hub routers to exchange the burst blocking and route information for all the deflection route candidates between various source-destination possibilities. Considering these complexities, the threshold-check function was designed to reflect the primary path, which implicitly conveys an adequate measure from the point of view of making a decision on whether to deflection route or to drop followed by source retransmit. Of course, once the decision is made

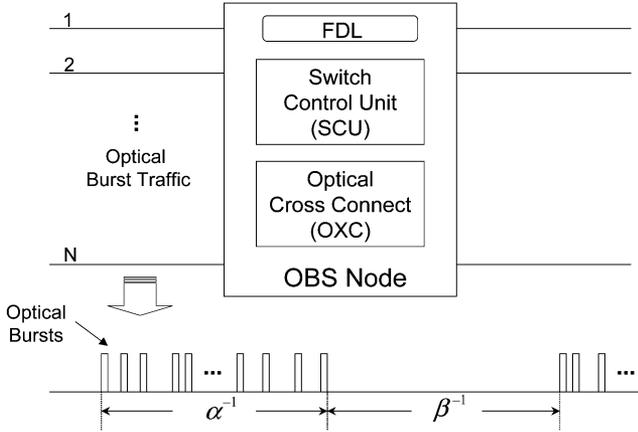


Fig. 8. Illustration of an OBS node and input traffic from a source node.

to deflection route, the nodes along the way would direct the bursts along the most desirable alternate route.

IV. QUEUEING MODEL FOR BURST LOSS PROBABILITY

We present here an analytical model for determination of burst loss ratio at a single OBS switch. This model provides a number of significant insights that aid in the design of experiments for simulations, as well as complement our simulation results. This type of analysis and the analytical insights are crucial to understand burst loss performance in OBS networks both with and without FDLs, and such insights have not been reported elsewhere in the OBS literature. In this paper, the numerical results from analytical queueing models have been generated for a variety of traffic assumptions and parameter values such as ON-OFF times, FDL buffer size, etc. These analytical results combined with our numerous simulation results (presented in Section V-C) for various deflection routing protocols (CLDR, SPDR) provide significant insights about the performance of OBS switches and networks. This section also clearly defines the various system and traffic parameters that are common to our network simulation model (see Sections V-B and C) as well.

A. Source Traffic Model

The ingress edge node creates bursts by assembling incoming packets from each traffic source. 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 (Mb), 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 Mb) creation; then the assembled burst up to that time is padded into a fixed size burst and placed in the queue.

Fig. 8 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 Fig. 8. In effect we assume that the each source behaves as an ON-OFF source. Bursts arrive with exponential random interarrival 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.

B. Queueing Model

We closely follow the analytical model presented in [31] 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.

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

- L burst size (Mb);
- C link capacity (Gb/s);
- $1/\alpha$ average ON period (ms);
- $1/\beta$ average OFF period (ms);
- λ burst generation rate during ON period (bursts/s);
- n number of sources simultaneously multiplexed on a link (offered load);
- B burst queue size or the number of FDLs per output port (specified in total milliseconds worth of buffering at link speed C Gb/s);
- 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 B ms;
- 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);
- N_s number of sources multiplexed at link saturation (100% or almost 100% offered load);

We can readily write the following equations for n_0 and N_s :

$$n_0 = \left\lfloor \frac{10^3 C}{\lambda L} \right\rfloor \quad (17)$$

$$N_s = \left\lfloor \frac{10^3 C(\alpha + \beta)}{\beta \lambda L} \right\rfloor. \quad (18)$$

When the system is in state i 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., $(n_0 + 1), (n_0 + 2), \dots, i - 1$ 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

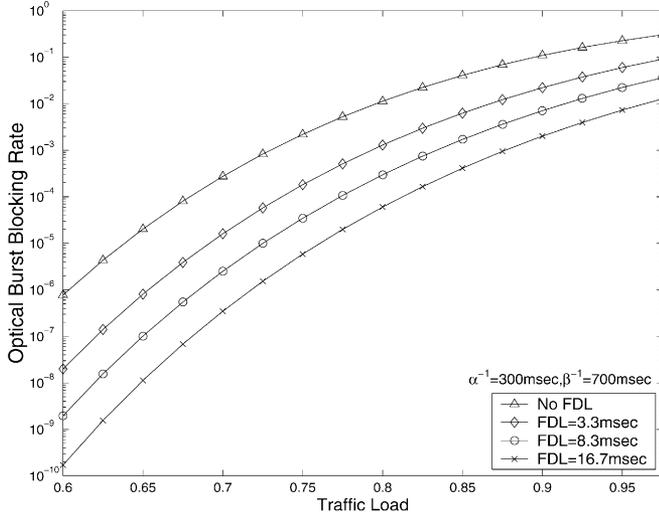


Fig. 9. Burst blocking rate under varying FDL sizes.

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

$$\tau_i = \frac{\frac{BC}{L\lambda}}{(i - n_0)(i - n_0 + 1)}, \quad i \geq (n_0 + 1). \quad (19)$$

The system-state probabilities p_i are binomial distributed and given by

$$p_i = \binom{n}{i} \left(\frac{\beta}{\alpha + \beta}\right)^i \left(\frac{\beta}{\alpha + \beta}\right)^{n-i}. \quad (20)$$

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

$$P_L = \sum_{i=n_0+1}^n p_i e^{-i\alpha\tau_i}. \quad (21)$$

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 of 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 [31], [32], which influence the burst loss ratio and make it fairly sensitive to numerous parameters. The model does well to capture the effect of these correlations on burst loss ratio.

V. PERFORMANCE RESULTS AND COMPARISONS

A. Analytical Model-Based Performance Results

In our numerical results, we have used transmission bandwidth of $C = 6$ Gb/s per link (six wavelengths each with 1 Gb/s capacity) and $L = 1$ Mb. The burst generation rate λ during ON period is assumed to be 100 per second. Fig. 9 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 to 16.7 ms. The values of FDL = 0, 3.3, 8.3, 16.7 ms correspond to buffer sizes that

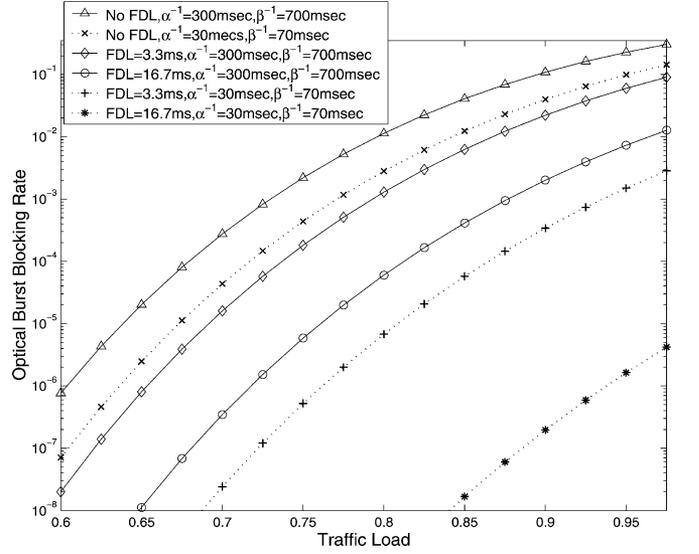


Fig. 10. Burst blocking rate for high and low ON-OFF periods when activity is 0.3.

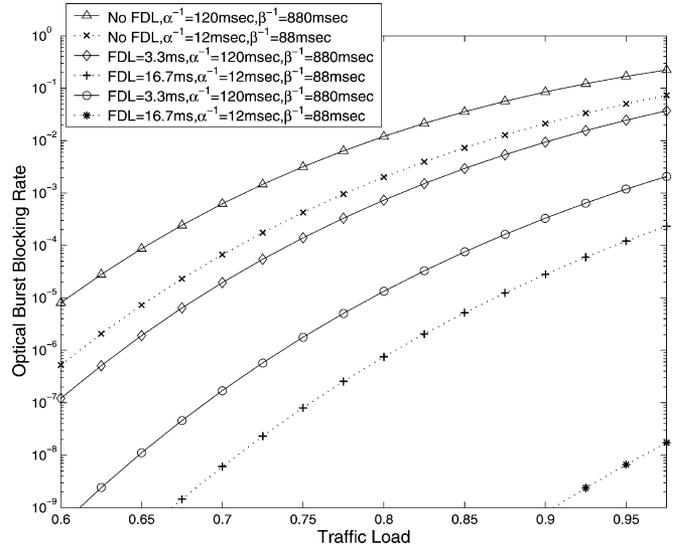


Fig. 11. Burst blocking rate for high and low ON-OFF periods when activity is 0.12.

can simultaneously delay (i.e., queue) 0, 20, 50, and 100 bursts, respectively.

Figs. 10 and 11 show the sensitivity of the burst loss ratio to the average ON and OFF periods. In Fig. 10, the comparison is made between ($\alpha^{-1} = 300$ ms, $\beta^{-1} = 700$ ms) versus ($\alpha^{-1} = 30$ ms, $\beta^{-1} = 70$ ms); both cases have an average activity factor of $a = 0.3$. In Fig. 10, the comparison is made between ($\alpha^{-1} = 120$ ms, $\beta^{-1} = 880$ ms) versus ($\alpha^{-1} = 12$ ms, $\beta^{-1} = 88$ 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 burst loss rate for average ON period of 300 ms (or 120 ms) versus 30 ms (or 12 ms) is fractionally higher for the zero buffer case (FDL = 0), and it is about several orders of magnitude higher for the cases when FDL is used (see Figs. 10 and 11). This amplification of the difference in burst

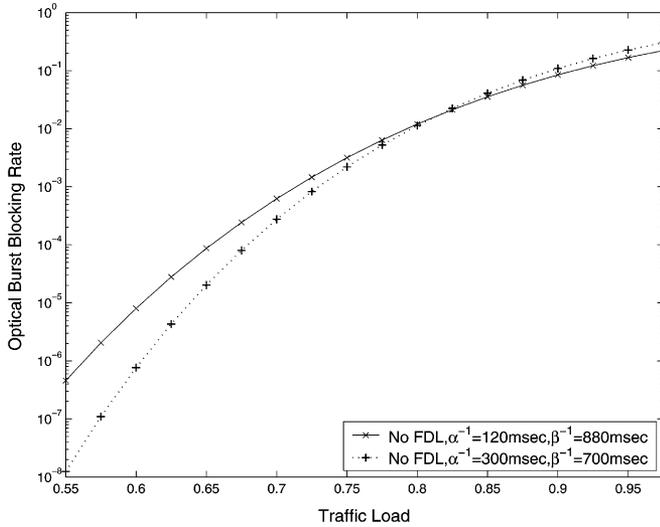


Fig. 12. Burst blocking rate sensitivity to ON-OFF periods when there is no FDL.

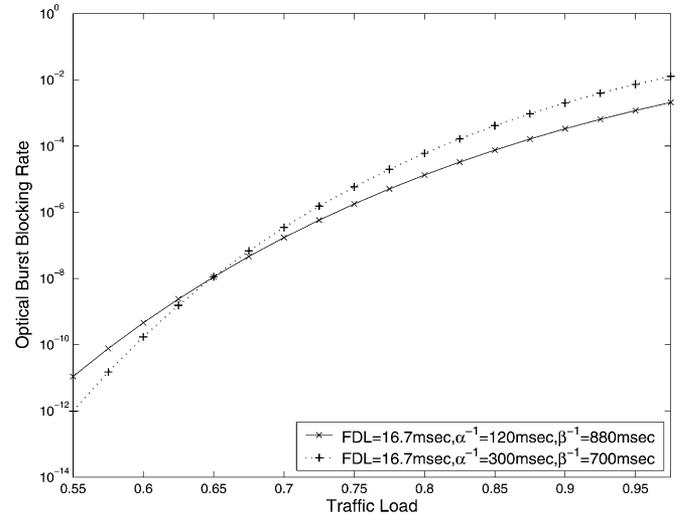


Fig. 14. Burst blocking rate sensitivity to ON-OFF periods with large FDL size (FDL = 16.7 ms).

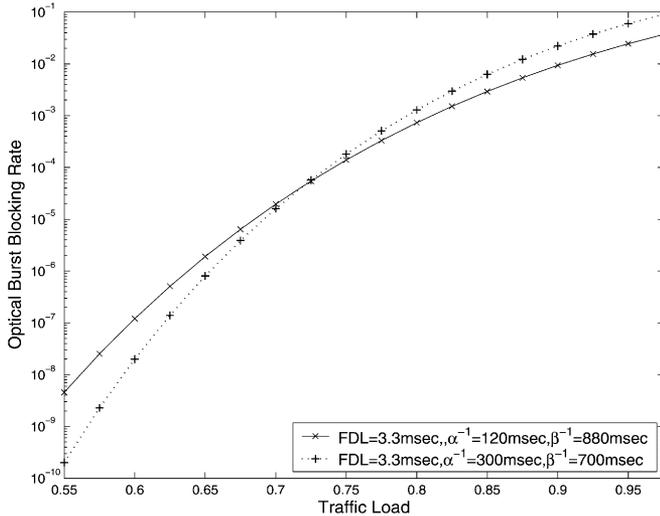


Fig. 13. Burst blocking rate sensitivity to ON-OFF periods with moderate FDL size (FDL = 3.3 ms).

loss rate when FDLs are used is because buffering allows the temporal correlations in the combined burst traffic (i.e., superposition of many sources of burst streams) to be manifest in a more influential manner. In other words, when the different sources of burst traffic interact with each other over a period of time in the buffer, then the temporal correlations get manifested more prominently. Figs. 12–14 highlight another very interesting phenomenon. Here, we are comparing ($\alpha^{-1} = 300$ ms, $\beta^{-1} = 700$ ms) versus ($\alpha^{-1} = 120$ ms, $\beta^{-1} = 880$ ms) for different buffer sizes. The activity factors are 0.3 and 0.12, respectively. The average ON periods are different (300 versus 120 ms), but not quite an order of magnitude different as was the case in Figs. 10 and 11. The comparisons are shown in Figs. 12–14 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 Figs. 10 and 11. Burst loss ratio in Fig. 12 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 Figs. 13 and 14). 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 is 2.5 times more for the case of $a = 0.12$ as compared with that for $a = 0.3$ and 2) the temporal correlation 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.

B. Description of the Simulation Model

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 [11]–[14].

In our simulation model, the simulations are run in two stages to test the effectiveness of the CLDR technique. The first stage involves only provisioning of alternate routes, and uses the CLDR algorithm to compute the alternate routes. The second stage covers contention resolution in the simulated network by using the limited deflection routing rules with the alternate routes yielded by the first stage. The simulation is performed on the basis of the periodic, offline CLDR approach, as discussed in Section III-A. The alternate routes computed in the first stage are stored in the DRTs. For the periodic computation of the alternate routes, a mixed integer linear programming (ILP) solver called *lp_solve* is used, which is based on the simplex method [33]. The computational complexity depends on the number of integer variables which are related to the number of paths to be computed, the network topology, and resource constraints. In our CLDR algorithm, because we do not aim to obtain all the paths for every source-destination pair but compute only the alternate paths relevant to resolving contentions, there is a significant reduction in the number of variables. Thus, here the ILP solver is dealing with a much smaller number of variables as compared with when it is solving a typical routing and wavelength assignment (RWA) problem. For on-demand computation, the alternate routes are computed only for the

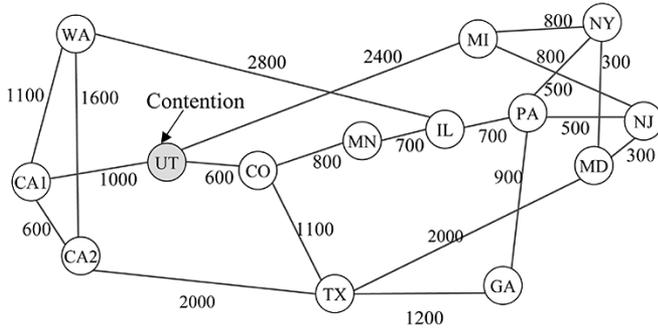


Fig. 15. Simulation network topology.

current congested node to the destination of a burst source. On the other hand, for periodic computations, all the nodes with DRTs are updated. In our simulation runs with 1050 variables, we observed that it took on average 4.7 iterations and a maximum of 10 iterations in the ILP to find an optimal alternate route. When we increased the number of variables to 2100 and ran the simulation to satisfy 100 deflection routing requests, the CPU execution time was around 0.08 s. It is to be noted that these CPU run times are not of concern for getting the alternate paths in a realistic OBS network, because they were not too large as such and were in fact obtained while using a noncommercial ILP solver on a simple desktop computer (1.7 GHz with 256 MB memory).

In our simulation tests, we use the JET method of offset-based reservation that was described in Section II. The burst sources are individually simulated with the ON-OFF model, as explained in Section IV-B. The system and traffic source parameter values for the simulation model are the same as those used in the section on results based on analytical modeling (see Section V-A). The burst source traffic parameters were also varied over a range here just as in Section V-A. The tests were carried out using a 14-node NSFNET topology as shown in Fig. 15. We assume that each fiber link is composed of the same number of wavelengths. The transmission rate of each link is 6 Gb/s, consisting of six wavelengths each operating at 1 Gb/s. The simulations can be extended to more wavelengths and 10 Gb/s, but we wanted to capture the key performance comparisons between CLDR and other deflection routing schemes using a smaller capacity network in order to keep the simulation time manageable. Our simulation results are well within the 95% confidence intervals because we have consistently run the simulations to generate 10–100 times more events (i.e., burst arrivals) than would be needed for 95% confidence [34] for the range of burst loss probabilities of interest.

Over the 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 Fig. 15, 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-distance alternate path from UT to NJ. However, the CLDR scheme can very

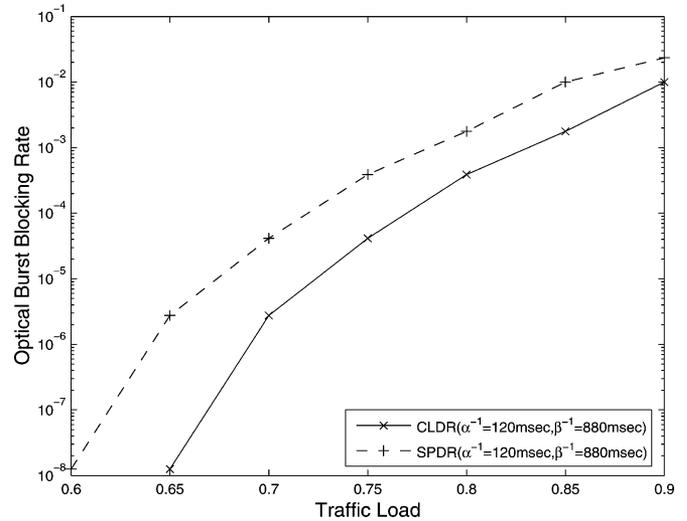


Fig. 16. Burst blocking rate for CLDR and SPDR without FDL when activity is 0.12.

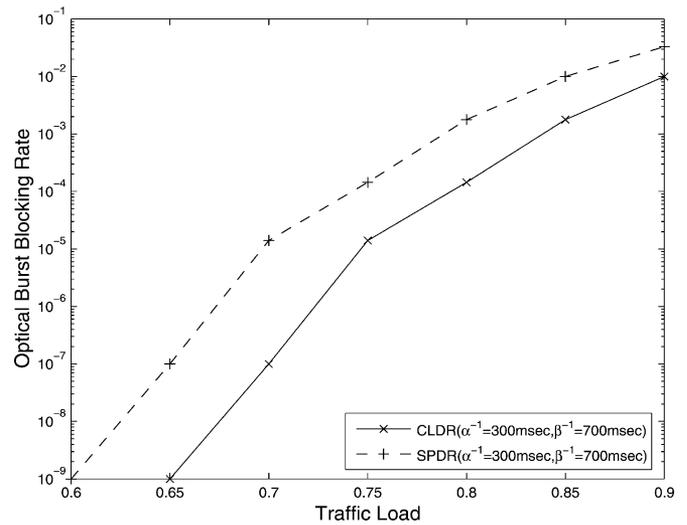


Fig. 17. Burst blocking rate for CLDR and SPDR without FDL when activity is 0.3.

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).

C. 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 throughput and data loss rate for the entire network are found by calculating the average of the burst blocking rates over all source-destination pairs.

Figs. 16 and 17 show simulation results comparing the burst blocking or loss rate for our CLDR method with that of the shortest path deflection routing (SPDR) method. The ON-OFF durations are ($\alpha^{-1} = 120$ ms, $\beta^{-1} = 880$ ms) and ($\alpha^{-1} = 300$ ms, $\beta^{-1} = 700$ ms) in Figs. 16 and 17, respectively. The SPDR algorithm simply picks the shortest path alternate route

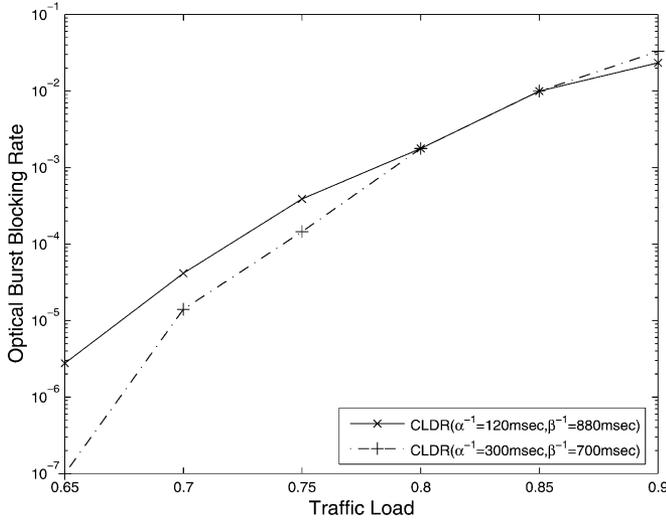


Fig. 18. Burst blocking rate for CLDR under two different cases of ON-OFF periods.

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 with the SPDR in the test cases that we have studied through simulation runs.

Fig. 18 shows a comparison of the burst blocking ratios for the CLDR scheme under two different scenarios, namely, ($\alpha^{-1} = 120$ ms, $\beta^{-1} = 880$ ms) and ($\alpha^{-1} = 300$ ms, $\beta^{-1} = 700$ ms). We talked about a similar study in Section V-A of the burst loss sensitivity to the source ON-OFF parameters using the analytical model for a single OBS link (see Fig. 12). We note that the burst loss results from the simulations in Fig. 18 are qualitatively in agreement with those from analytical results in Fig. 12. In both cases (i.e., 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 (Fig. 12), we discussed why this was counterintuitive but could be explained. The same explanation can also be extended to the network simulation results (compare Figs. 12 and 18).

Fig. 19 shows a comparison of the CLDR scheme under two scenarios involving no FDL buffering and FDL buffering of size 3.3 ms (i.e., capable of delaying/queueing up to 20 optical bursts) at each port in the network. The ON-OFF parameters in this simulation study were set to ($\alpha^{-1} = 120$ ms, $\beta^{-1} = 880$ ms). The results are once again qualitatively consistent with our prior analytical results; for this, we can compare Fig. 19 with Fig. 11. There is consistent improvement, attributable to the use of FDL, in the burst blocking rate throughout the range of moderate to high loads. The burst blocking improves due to use of FDL by about an order of magnitude in the typical operating range (say, about 0.65–0.75).

In the above simulation results, we showed how CLDR performs better than SPDR and the inherent reasons are: 1) the CLDR considers a composite performance measure (path distance and burst blocking ratio) for precomputation of alternate

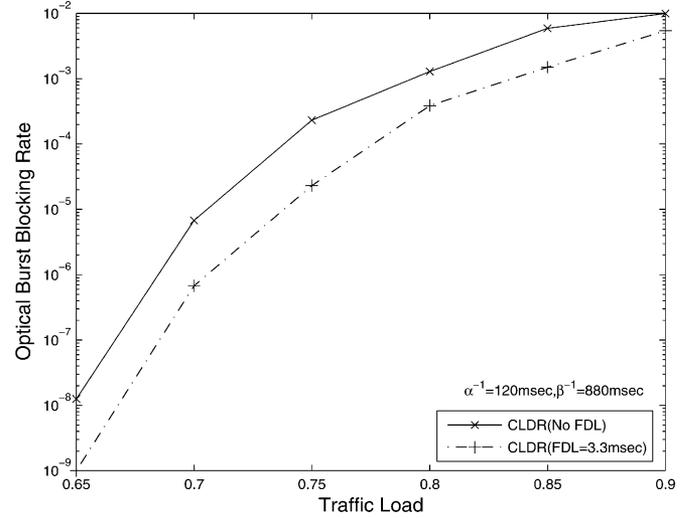


Fig. 19. Burst blocking rate for CLDR without FDL and with FDL (of length 3.3 ms) when activity is 0.12.

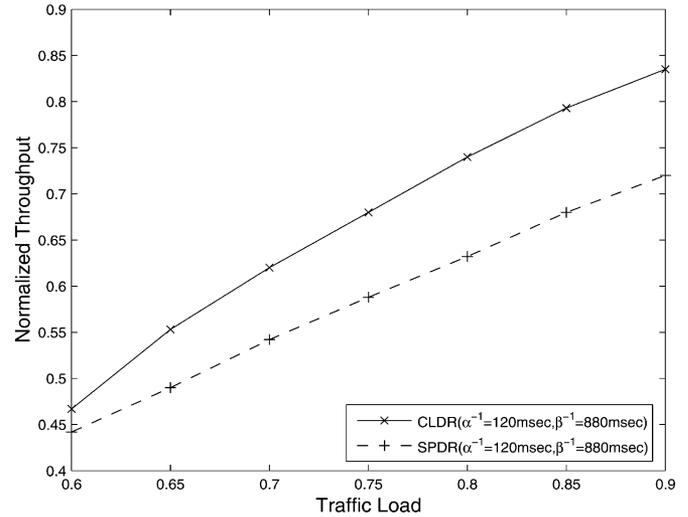


Fig. 20. Normalized throughput for SPDR versus CLDR under moderate and high load conditions when activity is 0.12.

routes and 2) the CLDR scheme runs an efficient threshold check algorithm for deciding if a burst encountering congestion should be deflection routed or dropped and retransmitted from source. Now, we turn our attention to comparisons of CLDR and SPDR based on their throughput performance. Throughput is defined as the amount of information successfully delivered per unit of time and normalized to the link capacity. The throughput improvement of CLDR over SPDR is shown in Figs. 20 and 21, where normalized throughput is plotted for varying traffic loads under two different cases of ($\alpha^{-1} = 120$ ms, $\beta^{-1} = 880$ ms) and ($\alpha^{-1} = 300$ ms, $\beta^{-1} = 700$ ms). In both cases, the CLDR scheme exhibits a higher throughput than the SPDR scheme for all traffic loads due to the aforementioned advantages of performance-metric-based alternate route computation and the decision algorithm with a threshold check function. Finally, in Fig. 22, we compare the throughput of the CLDR scheme for two cases of the ON-OFF durations. The somewhat counterintuitive nature of the results in Fig. 22 can be once again

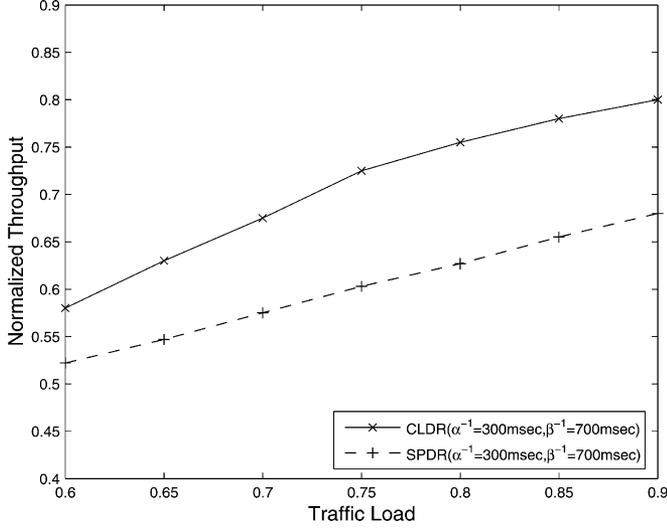


Fig. 21. Normalized throughput for SPDR versus CLDR under moderate and high load conditions when activity is 0.3.

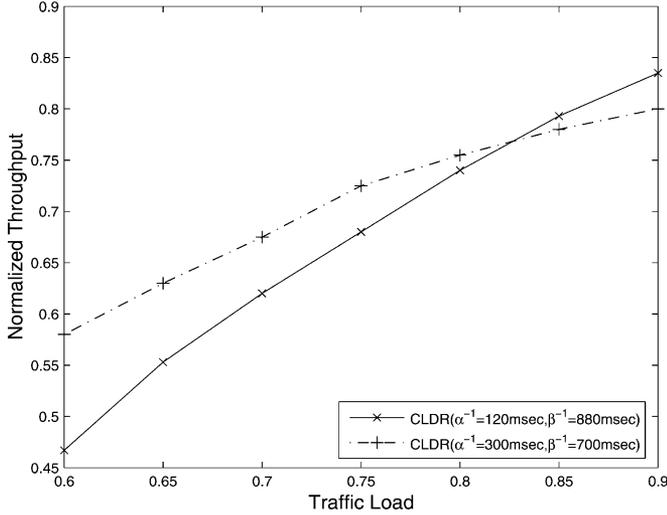


Fig. 22. Normalized throughput for CLDR under moderate and high load conditions with two different cases of ON-OFF periods.

understood by invoking the explanations given in the context of Figs. 12 and 18.

VI. 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 SPDR scheme that is known in the literature.

An additional salient feature of the proposed CLDR scheme is that the CLDR scheme runs a threshold-based dynamic decision algorithm to decide whether to deflection route or not. Based on the distance of the congested node from the source node and likely burst-blocking characteristics of available paths to the destination from the congested intermediate node, the

threshold-check algorithm decides whether to deflection route the burst or drop and retransmit it from the source.

We also presented numerical results based on analytical queueing models that provide 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.

APPENDIX

Due to the Taylor's theorem, we have

$$\ln(1-x) = -\sum_{k=1}^{\infty} \frac{x^k}{k} \text{ for } -1 < x < 1. \quad (22)$$

Further, it is known that

$$\prod_{i=1}^N a_i \leq \sum_{i=1}^N a_i \text{ for } 0 < a_i < 1. \quad (23)$$

Using (22), (23), and keeping in mind that the values of c_{ij} are between 0 and 1, we have the following:

$$\begin{aligned} & \text{Min} \log_{10} \left[1 - \prod_{i,j} (1 - b_{ij}) \right] \\ &= \frac{1}{\ln 10} \text{Min} \ln \left[1 - \prod_{i,j} c_{ij} \right] \text{ where } c_{ij} = 1 - b_{ij} \\ &= \frac{1}{\ln 10} \text{Min} - \sum_{k=1}^{\infty} \frac{1}{k} \prod_{i,j} c_{ij}^k \\ &= -\frac{1}{\ln 10} \text{Max} \sum_{k=1}^{\infty} \frac{1}{k} \prod_{i,j} c_{ij}^k \\ &\leq -\frac{1}{\ln 10} \text{Max} \sum_{k=1}^{\infty} \frac{1}{k} \sum_{i,j} c_{ij}^k \\ &= \frac{1}{\ln 10} \text{Min} - \sum_{k=1}^{\infty} \frac{1}{k} \sum_{i,j} c_{ij}^k \\ &= \frac{1}{\ln 10} \text{Min} \sum_{i,j} \left(-\sum_{k=1}^{\infty} \frac{c_{ij}^k}{k} \right) \\ &= \frac{1}{\ln 10} \text{Min} \sum_{i,j} \ln(1 - c_{ij}) \\ &= \text{Min} \sum_{i,j} \log_{10} b_{ij}. \end{aligned} \quad (24)$$

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for their constructive criticism and many useful suggestions to improve the paper.

REFERENCES

- [1] S. Verma, H. Chaskar, and R. Ravikanth, "Optical burst switching: A viable solution for terabit IP backbone," *IEEE Netw. Mag.*, vol. 14, no. 6, pp. 48–53, Nov./Dec. 2000.
- [2] "Optical phased array technology for high-speed switching," White Paper, http://www.chiaro.com/pdf/CHI100_OPA_1.pdf, 2002.
- [3] K. Sriram, D. Griffith, S. Lee, and N. Golmie, "Optical burst switching: Benefits and challenges," in *Proc. 1st Int. Workshop Optical Burst Switching (WOBS), Conjunction With OptiComm*, Dallas, TX, Oct. 16, 2003, pp. 1–12.
- [4] T. Zhang, K. Lu, and J. P. Jue, "Differentiated contention resolution for QoS in photonic packet-switched networks," in *Proc. IEEE ICC*, Paris, France, Jun. 2004, pp. 1599–1603.
- [5] "JumpStart JIT Signaling Definition, Revision 1.17," Tech. Rep., MC-NCRDI ANR and North Carolina State Univ, Raleigh, NC, [Online]. Available: http://jumpstart.anr.mcnc.org/Docs/03_05_12/SignalingDefinition.pdf, 2003.
- [6] Y. Chen, C. Qiao, and X. Yu, "Optical burst switching: A new area in optical networking research," *IEEE Netw. Mag.*, vol. 18, no. 3, pp. 16–23, May/June 2004.
- [7] C. Gauger, "Contention resolution in optical burst switching networks," In advanced infrastructures for photonic networks: WG 2 Intermediate Rep., [Online]. Available: http://www.ikr.uni-stuttgart.de/Content/Publications/Archive/Ga_COST266WG2_34734.ps.gz, 2002.
- [8] M. Yoo, C. Qiao, and S. Dixit, "A comparative study of contention resolution policies in optical burst switched WDM networks," in *Proc. SPIE Int. Conf. Terabit Optical Networking: Architecture, Control, Manage, Issues*, vol. 4213, Boston, MA, Nov. 2000, pp. 124–135.
- [9] C. Gauger, "Performance of converter pools for contention resolution in optical burst switching," in *Proc. SPIE Optical Netw. Commun. Conf. (OptiComm 2002)*, Boston, MA, Jul. 2002, pp. 109–117.
- [10] C.-F. Hsu, T.-L. Liu, and F.-F. Huang, "Performance analysis of deflection routing in optical burst-switched networks," in *Proc. IEEE INFOCOM*, vol. 1, New York, Jun. 2002, pp. 66–73.
- [11] S. Kim, N. Kim, and M. Kang, "Contention resolution for optical burst switching networks using alternative routing," in *Proc. IEEE ICC*, vol. 5, New York, Apr./May 2002, pp. 2678–2681.
- [12] X. Wang, H. Morikawa, and T. Aoyama, "Burst optical deflection routing protocol for wavelength routing WDM networks," in *Proc. SPIE/IEEE OptiComm*, Dallas, TX, USA, 2000, pp. 257–266.
- [13] —, "Burst optical deflection routing protocol for wavelength routing WDM networks," *Opt. Netw. Mag.*, pp. 12–19, Nov./Dec. 2002.
- [14] C. Y. Li *et al.*, "Deflection routing in slotted self-routing networks with arbitrary topology," in *Proc. IEEE ICC*, vol. 5, New York, Apr./May 2002, pp. 2781–2785.
- [15] A. Zalesky, H. L. Vu, Z. Rosberg, E. Wong, and M. Zukerman, "Modeling and performance evaluation of optical burst switched networks with deflection routing and wavelength reservation," in *Proc. IEEE INFOCOM*, vol. 3, Hong Kong, China, Mar. 2004, pp. 1864–1871.
- [16] F. Borgonovo, L. Fratta, and J. A. Bannister, "On the design of optical deflection-routing networks," in *PROC. IEEE INFOCOM*, vol. 1, Toronto, ON, Canada, Mar. 1994, pp. 120–129.
- [17] F. Forghieri, A. Bononi, and P. R. Prucnal, "Analysis and comparison of hot-potato and single-buffer deflection routing in very high bit rate optical mesh networks," *IEEE Trans. Commun.*, vol. 43, no. 1, pp. 88–98, Jan. 1995.
- [18] A. Bononi, G. A. Castanon, and O. K. Tonguz, "Analysis of hot-potato optical networks with wavelength conversion," *IEEE J. Lightw. Technol.*, vol. 17, no. 4, pp. 525–534, Apr. 1999.
- [19] T. Chich, J. Cohen, and P. Fraigniaud, "Unslotted deflection routing: A practical and efficient protocol for multihop optical networks," *IEEE/ACM Trans. Netw.*, vol. 9, pp. 47–59, Feb. 2001.
- [20] G. Castanon, L. Tancevski, and L. Tamil, "Routing in all-optical packet switched irregular mesh networks," in *Proc. IEEE GLOBECOM*, Rio de Janeiro, Brazil, Dec. 1999, pp. 1017–1022.
- [21] C. Qiao and M. Yoo, "Optical burst switching (OBS)—A new paradigm for an optical internet," *J. High-Speed Netw.*, vol. 8, no. 1, pp. 69–84, Jan. 1999.
- [22] J. Y. Wei, J. L. Pastor, R. S. Ramamurthy, and Y. Tsai, "Just-in-time optical burst switching for multiwavelength networks," *IEEE/OSA J. Lightw. Technol.*, vol. 18, no. 12, pp. 2019–2037, Dec. 2000.
- [23] I. Baldine, G. N. Rouskas, H. Perros, and D. Stevenson, "Jumpstart: A just-in-time signaling architecture for WDM burst-switched networks," *IEEE Commun. Mag.*, vol. 40, no. 2, pp. 82–89, Feb. 2002.
- [24] C. Gauger, K. Dolzer, and M. Scharf, *Reservation strategies for FDL buffers in OBS networks*. Stuttgart, Germany: Report, IND, Univ. Stuttgart, 2001.
- [25] S. Yao, B. Mukherjee, S. J. B. Yoo, and S. Dixit, "A unified study of contention-resolution schemes in optical packet-switched networks," *IEEE/OSA J. Lightw. Technol.*, vol. 21, no. 3, pp. 672–683, Mar. 2003.
- [26] S. K. Lee, L. Y. Kim, J. S. Song, D. Griffith, and K. Sriram, "Dynamic deflection routing with virtual wavelength assignment in optical-burst switched networks," *Photonic Netw. Commun.*, vol. 9, no. 3, pp. 347–356, May 2005.
- [27] L. Berger, Ed., "Generalized multiprotocol label switching (GMPLS) signaling functional description," RFC 3471, Jan. 2003.
- [28] C. Qiao, "Labeled optical burst switching for IP-over-WDM integration," *IEEE Commun. Mag.*, vol. 38, no. 9, pp. 104–114, Sep. 2000.
- [29] S. Ovadia, C. Maciocco, and M. Paniccia, "Photonic burst switching (PBS) architecture for hop and span constrained optical networks," *IEEE Commun. Mag.*, vol. 41, no. 11, pp. s24–s32, Nov. 2003.
- [30] R. Ramaswami and K. N. Sivarajan, *Optical Networks—a Practical Perspective*, 2nd ed. San Mateo, CA: Morgan Kaufmann, 2002, pp. 649–649.
- [31] 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 J. Sel. Areas Commun.*, vol. 17, no. 1, pp. 4–17, Jan. 1999.
- [32] K. Sriram and W. Whitt, "Characterizing superposition arrival processes in packet multiplexers for voice and data," *IEEE J. Sel. Areas Commun.*, vol. SAC-4, no. 6, pp. 833–846, Sep. 1986.
- [33] M. Weidenauer, *lp_solve: Linear programming solver*. [Online]. Available: ftp://ftp.es.ele.tue.nl/pub/lp_solve/
- [34] R. Jain, *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation and Modeling*. New York: Wiley, 1991.



SuKyoung Lee (M'04) received the B.S., M.S., and Ph.D. degrees in computer science from Yonsei University, Seoul, Korea, in 1992, 1995, and 2000, respectively.

From 2000 to 2003, she was with the Advanced Networking Technologies Division, National Institute of Standards and Technology (NIST), Gaithersburg, MD, where she investigated issues relating protection/restoration for optical data networks and contention resolution in optical burst-switched networks. She has been an Assistant Professor with

Sejong University, Seoul, Korea, since September 2003. Her current research interests include vertical handoff in heterogeneous wireless networks, mobile IP and secure routing in ad hoc networks, as well as optical data networking.



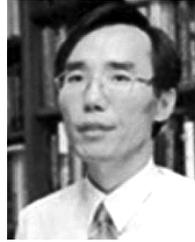
Kotikalapudi Sriram (SM'97–F'00) received the B.S. and M.S. degrees from the Indian Institute of Technology, Kanpur, and the Ph.D. degree from Syracuse University, Syracuse, NY, all in electrical engineering.

He is currently a Senior Researcher in the Advanced Networking Technologies Division, National Institute of Standards and Technology (NIST), Gaithersburg, MD. From 1983 to 2001, he was a Consulting and Distinguished Member of Technical Staff in the Performance Analysis Department, Bell Laboratories, Lucent Technologies. He is a contributing author and a coeditor of *Cable Modems: Current Technologies and Applications* (Piscataway, NJ: IEEE Press, 1999). His interests and responsibilities include performance modeling, network architecture, Internet routing protocol security, design of protocols and algorithms for multiservice broadband networks, ATM traffic controls, and hybrid fiber-coax networks. He holds 16 patents and is a co-inventor on ten other pending patents.



HyunSook Kim (M'05) received the B.S. degree in computer science from Duksung Women's University, Seoul, Korea, in 1997, and the M.S. and Ph.D. degrees in computer science from Yonsei University, Seoul, Korea, in 1999 and 2004, respectively.

She is currently a Postdoctoral Fellow in the Electrical Engineering Department, Stanford University, Stanford, CA, working on next-generation access networks in the Photonics and Networking Research Laboratory (PNRL).



JooSeok Song (S'77–M'88) received the B.S. degree in electrical engineering from Seoul National University, Seoul, Korea, in 1976, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, in 1979, and the Ph.D. degree in computer science from the University of California at Berkeley, in 1988.

He was an Assistant Professor of the Naval Postgraduate School from 1988 to 1989. He is currently a Professor of Computer Science at Yonsei University, Seoul, Korea. His research interests include optical network, wireless communication, cryptography, and information security.