Retransmission in OBS Networks with Fiber Delay Lines

Kyung Joon Kwak, E. G. Coffman, Jr.
Electrical Engineering Department
Columbia University, NY 10027
Email: {kjkwak, egc}@ee.columbia.edu

***Abstract*— While most transmission schemes in OBS networks relegate retransmission to higher protocol layers, the scheme proposed in this paper reduces retransmission delays by exploiting fiber delay lines at the optical layer. We present an analysis of the scheme that focuses not only on individual components but also on the end-to-end properties of the network. We also propose a facility location model suitable for optimizing the locations of sites with fiber delay lines when the number of such sites is limited. Evaluation of our proposed scheme on a common test network shows that major improvements in performance are possible.**

I. Introduction

Recently, *Optical Burst Switching* (OBS)[1]-[4] networks have received enormous attention as next generation DWDM (Dense Wavelength Division Multiplexing) core networks. Each ingress node assembles multiple IP packets into a data burst. Once a data burst is assembled, the ingress node sends a header (more precisely a *Burst Header Packet*) prior to data burst transmission in order to reserve available wavelengths at intermediate nodes. Once these reservations are set, the data burst will be transported along the designated path without any Optical-Electronic-Optical (OEO) conversion. According to a prominent signaling scheme, *Just-Enough-Time* (JET) [2], the control plane and data-burst plane are maintained separately. The header traverses the control plane to reserve wavelengths at the intermediate core nodes, and is followed by a data burst in the data-burst plane after a predetermined offset time. Most scheduling and wavelength assignment schemes use one way signaling and, in so doing, inevitably create data-burst contention and loss. In such cases, data burst retransmission is considered to be a higher-layer protocol such as an implementation of TCP in the application layer. Yu et al [5] examined the dynamics of TCP retransmission in OBS networks. They mention a retransmission penalty along with a correlation gain due to longer retransmission periods.

However, the higher application layer usually sets large time-out values so that a retransmission will be triggered long after the burst loss. In contrast, a retransmission in the OBS domain responds faster than that at higher layers, thus alleviating the problem of burst losses which waste bandwidth, increase packet delivery delays, and decrease throughput. Mach et al [6] were first to suggest a retransmission in the OBS domain. The ingress node keeps the original data burst and retransmits the saved burst when it receives a negative acknowledgement (NACK) from one of the intermediate nodes. If the ingress node does not receive a NACK during the source-to-destination round trip time, it discards the original burst from the electronic buffer. Also investigated in [6] is the effect of traffic shaping and congestion monitoring as an extension in the control plane. Zhang et al [7] proposed a retransmission scheme with electronic buffers. The ingress node keeps the original burst until it receives an acknowledgement (ACK) from the egress node. If the wavelength reservation fails, the intermediate core node sends an *Automatic Retransmission Request* (ARQ) back to the ingress node, then the ingress node retransmits the header. However, these two schemes require a significant electronic buffer capacity at every core node in order to store each data burst during the source-to-destination round trip delay. Owing to this limitation, as pointed out in [6], the retransmission schemes with electronic buffers can only be used for Metropolitan Area Networks (MANs) or Local Area Networks (LANs). In addition, these schemes produce unfairness in the sense that ingress nodes store only some fraction of a burst.

In this paper, we introduce a retransmission scheme in the OBS domain with fiber delay lines [8] and small electronic buffers, and develop an analytical model for computing end-to-end path setup delays and blocking probabilities. We also introduce solutions to versions of the *Facility Location Problem* [9] which provide optimal locations of nodes with fiber delay lines when only a limited number of such nodes can be made available. Our proposed retransmission scheme is similar to the schemes in [6] and [7] in that we also consider retransmission schemes in the OBS domain. However, our scheme is different from both of the earlier schemes in that our scheme uses fiber delay lines at a limited number of core nodes, whereas the two earlier schemes use fairly large electronic buffers at the source node without fiber delay lines.

Section II describes the retransmission scheme, and Section III analyzes an OBS network with the retransmission scheme. Section IV introduces the facility location problem with new formulations of the cost and objective function. Section V gives experimental results for the NSF network, and the last section concludes with a discussion of directions for further research.

II. Description of the Retransmission Scheme

For a limited number of core nodes, some number of fiber delay lines can be provided for optical buffering. Each fiber

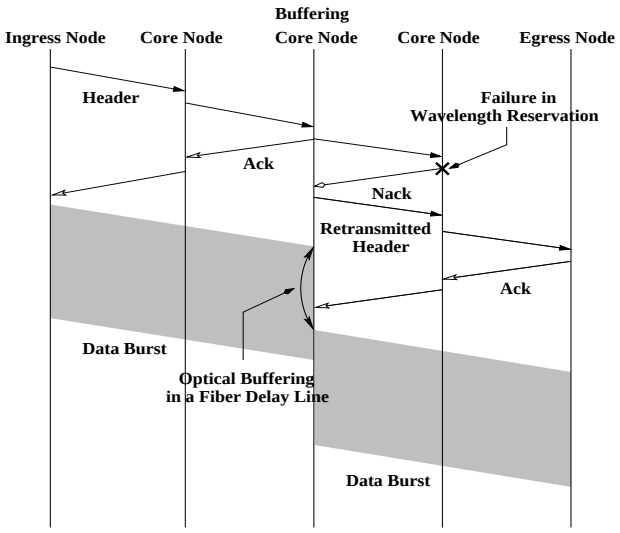

Fig. 1. Retransmission Scheme

delay line can generate either a fixed or variable delay ranging up to a maximum denoted by D_{max} expressed as a given multiple of the delay that a single fiber delay line can provide. Such core nodes are called Buffering Core Nodes (BCNs). In the proposed retransmission scheme, the ingress node must know for each burst the locations of the BCNs, if any, in the source-to-destination path. As shown below, this is because the signaling scheme differs depending on whether BCNs exist in the path. A detailed description follows.

Initially, the ingress node checks the routing table to see whether any BCNs are located on the source-to-destination path. BCNs are limited in number, so in general not all source-to-destination paths will encounter BCNs. When the path has no BCN, signaling and burst transmissions follow the usual two-way signaling scheme. The ingress node sends a header to (1) reserve wavelengths at each link on the path toward the egress node, and (2) reserve space in its electronic buffer until it receives an ACK from the egress node. Each intermediate core node forwards the header when the wavelength reservation is successful; otherwise, it sends a NACK back to the ingress node. If the ingress node receives such a signal, it resets the offset time and sends another header to reserve wavelengths. After a predetermined offset time following the receipt of an ACK from the egress node, the ingress node transmits the data burst.

Now suppose there is at least one BCN located along the path, as illustrated in Figure 1. The ingress node accumulates data packets until the accumulated data-burst size reaches its maximum. At this time, the ingress node starts by setting an offset time and burst time-out value; it then sends the header to the first BCN along the path, at the same time reserving space in its electronic buffer for the data burst while the ingress node waits for the appropriate ACK. The offset time must now be assigned to account for the two-way signaling so as to prevent the data burst from arriving at the BCN before wavelength reservation is completed. For guaranteed burst delivery, the offset time has to be set by the rule discussed in the next section. As soon as an intermediate BCN receives the header, the BCN simultaneously forwards it and sends an ACK back to the ingress node. Upon receiving the ACK, the ingress node transmits the data burst when the offset time has expired. Any nodes located between the source-to-destination path can send a NACK back to the ingress node or BCN if wavelength reservation is not successful. If the ingress node receives a NACK, it follows the same process as explained above. When a BCN receives a NACK, it checks its available fiber delay lines. If any fiber delay line is available, the BCN reserves an fiber delay line for the incoming data burst and forwards the updated header to the next BCN or egress node. If no fiber delay lines are available, the incoming data burst will be dropped and the NACK will be sent to the ingress node.

This scheme lies between one-way signaling and two-way signaling schemes. Wavelength reservation will take relatively more time in two-way signaling, but the data-burst drop rate will be much smaller. If a sufficient number of core nodes are BCNs, a moderate size electronic buffer is enough to store data bursts at the ingress node up to the time it receives an ACK from the first BCN. This is because the round trip delay to the first BCN will usually be significantly smaller than that of the entire source-to-destination path.

III. Retransmission Analysis

A. Modeling Issues

The blocking probability p for core nodes is obtained from the single-server $M/M/W$ queueing model as in [10], where W denotes the number of wavelengths. Lu et al [11] and Fayoumi et al [12] verified that a useful model of the blocking probability p' for buffering core nodes with fiber delay lines is that obtained from the $M/M/W/W+B$ queueing model, where B is the product of the numbers of fiber delay lines and wavelengths. This BCN queueing model is also used here. Because of the relatively high cost of BCNs, our study works out analytical details only for networks with at most two BCNs in every source-to-destination path. Our methods extend to the more general case, but the details become very elaborate and beyond our space constraints.

B. Path Setup Delay

Since a data burst always follows the header, characterizing the behavior of the header makes it possible to calculate the total delay of data-burst transport. The total delay to reserve wavelengths from source to destination is called the *path setup delay*. This delay, denoted by $T(n_1, n_2, \ldots, n_\ell)$, can be computed as the sum of the partial delays in the n_1 hop, n_2 hop, ..., n_ℓ hop intermediate sub-paths separated by BCNs.

The average processing delay δ includes OEO conversion; the average propagation delay D is assumed fixed, with $D \ll \delta$ in most cases. Let N_B be the number of BCNs along the source-to-destination path, and define the one-way transmission time $L_i := n_i \cdot (D+\delta)$ along the i^{th} intermediate

path. The expected setup delay, $\mathbb{E}T_i$, of the i^{th} intermediate path can be computed as

$$\mathbb{E}T_i = 2(D+\delta)\cdot\frac{1-(1-p)^{n_i-1}[1+p(n_i-1)]}{(1-p)^{n_i-1}(1-p')p} + \frac{2n_i(D+\delta)}{1-p'} \quad (1)$$

for $0 \le p, p' \le 1$. It is easy to verify that $1-(1-p)^{n_i-1}[1+p(n_i-1)] \approx \frac{n_i(n_i-1)p^2}{2} + O(p^3)$, so $T_i \to 2n_i(D+\delta)$ as $p \to 0$ with p' small.

For the case of one BCN in an n hop, end-to-end path (i.e., $n = \sum_{i=1}^{\ell} n_i$), assume it is divided into two intermediate paths at the n_1^{st} hop from the ingress node. (i.e., $\ell = 2$ and the BCN divides the full path into an n_1 hop path and an $n - n_1 = n_2$ hop path). The expected total setup delay, $\mathbb{E}T(n_1, n_2)$, with two intermediate paths is

$$\mathbb{E}T(n_1,n_2) = \begin{cases} \mathbb{E}T_1 + \mathbb{E}T_2 - L_1 & for\ n_1 < \lceil \frac{2n}{3} \rceil \\ \mathbb{E}T_1 + \mathbb{E}(T_2 - L_1)^+ & for\ n_1 \ge \lceil \frac{2n}{3} \rceil \end{cases} \quad (2)$$

with

$$\mathbb{E}(T_2 - L_1)^+ = \sum_{a=1}^{\infty}[2\cdot(\lfloor\frac{n_1}{2}\rfloor + a) - n_1](D+\delta)\cdot\mathbb{P}(k,m)$$

where $k = n_2$, $m = \lfloor \frac{n_1}{2} \rfloor + a$, and

$$\mathbb{P}(k,m) = \sum_{b=0}^{m-k-\lceil\frac{m-k}{k}\rceil} C_{(b,m-k-b)}^{k-1}\cdot p^{m-k-b}(1-p)^{k+b}$$

where $C_{(x,z)}^{y}$ is a combinatorial quantity giving the number of ways to distribute x balls into z distinguishable urns with the constraint that y is the maximum number of balls which can be put into an urn. A discussion of the origin of these last two quantities is deferred to the appendix.

Now, assume an n hop, end-to-end path is divided into three intermediate paths by two BCNs, one at the n_1^{th} hop and one at the $(n_1+n_2)^{th}$ hop from the ingress node (i.e., divide the overall path into an n_1 hop path, a n_2 hop path, and an n_3 hop path). To compact notation define $\xi := T_2 - L_1$, $\phi := T_3 - L_2$, $\psi := (T_2 - L_1) + (T_3 - L_2)$, $\alpha := \lceil \frac{2(n_1+n_2)}{3} \rceil$ and $\beta := \lceil \frac{2(n-n_1)}{3} \rceil$. Then define the events $\Xi := \{\xi > 0\}$, $\Phi := \{\phi > 0\}$ and $\Psi := \{\psi > 0\}$. Then, in analogy with the two-intermediate-path case, the expected total setup delay is (details are given in the appendix)

$$\mathbb{E}T(n_1,n_2,n_3) = \begin{cases} \mathbb{E}T_1 + \mathbb{E}T_2 + \mathbb{E}T_3 - L_1 - L_2, \\ \mathbb{E}T_1 + \mathbb{E}T_2 + \mathbb{E}(\phi|\Phi)\cdot Pr(\Phi) - L_1, \\ \mathbb{E}T_1 + \mathbb{E}(\psi|\Psi)\cdot Pr(\Psi), \\ \mathbb{E}T_1 + \mathbb{E}(\xi|\Xi)\cdot Pr(\Xi) \\ \quad + \mathbb{E}(\phi|\Phi,\Xi)\cdot Pr(\Phi)\cdot Pr(\Xi) \\ \quad + \mathbb{E}(\psi|\Psi,\Xi^c)\cdot Pr(\Psi)\cdot Pr(\Xi^c). \end{cases} \quad (3)$$

for $n_1 < \alpha,\ n_2 < \beta$ and $n_1 < \alpha,\ n_2 \ge \beta$ and $n_1 \ge \alpha,\ n_2 < \beta$ and $n_1 \ge \alpha,\ n_2 \ge \beta$ respectively. For the $N_B \ge 3$ cases, the expected value of total setup delay is not provided here but its calculation is a straightforward extension of the $N_B = 1, 2$ cases.

C. Offset Time

To implement the suggested scheme, the i^{th} BCN compares the two-way transmission time on the $i+1^{st}$ intermediate path with the one-way transmission time plus propagation time of the data burst on the i^{th} intermediate path. If the former is smaller than the latter, which means that the ACK can arrive at the i^{th} BCN from the $i+1^{st}$ BCN or egress node before the data burst reaches the i^{th} BCN, then the i^{th} BCN reduces the offset time by the OEO processing delay, δ, in order that the data burst will continuously pass through the i^{th} BCN without optical buffering. Otherwise, the i^{th} BCN reserves one fiber delay line for the duration of the time difference between the arrival time of the data burst at the i^{th} BCN and the arrival time of an ACK at i^{th} BCN from the $i+1^{st}$ BCN. Each core node simply reduces the offset time by δ and forwards the header, if the wavelength reservation is successful. The ingress node sets the offset time as the two-way transmission time between it and the first BCN (or egress node if there are no BCNs along the source-to-destination path).

D. Path Blocking Probability

For a given source and destination, as the number of BCN's increases, the *path blocking probability*, P, from source to destination will decrease, since a BCN has a much smaller burst-drop probability due to its fiber delay lines. If the graph of core nodes is strongly connected and the data-burst losses on each link are independent, then the path blocking probability is

$$P = 1 - (1-p)^{n_c}\cdot(1-p')^{n_b} \quad (4)$$

where n_c is the number of core nodes and n_b is the number of BCN's along the source-to-destination path.

E. Number of Control Packets

A shortcoming of the proposed scheme is that it produces more traffic in the control plane, since both ACK and NACK packets are used. The average number of control packets, C, can be modeled by (5), as was the average path setup delay.

$$C_{\text{ACK/NACK}} = \sum_{i=1}^{\ell}\{2\cdot\frac{1-(1-p)^{n_i-1}}{(1-p)^{n_i-1}(1-p')p} + \frac{2}{1-p'}\} \quad (5)$$

for $0 \le p,\ p' \le 1$. If, to reduce the traffic in the control plane, only the NACK packet is used, then the average number of control packets can be modeled as in (6) and decreases by $n = \sum_{i=1}^{\ell} n_i$ those counted in (5).

$$C_{\text{NACK}} = \sum_{i=1}^{\ell}\{2\cdot\frac{1-(1-p)^{n_i-1}}{(1-p)^{n_i-1}(1-p')p} + \frac{2-(1-p')\cdot n_i}{1-p'}\} \quad (6)$$

for $0 \le p,\ p' \le 1$.

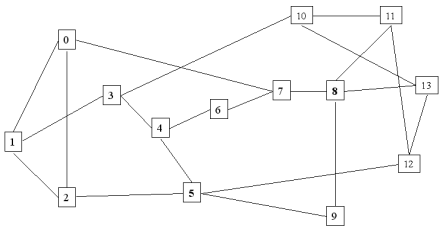

Fig. 2. NSF 14 Core Node Network

IV. Optimal Location Analysis

The NSF 14 core node network topology in Figure 2 is our case study. Each core node has 2 edge nodes and each edge node (as an ingress node) transmits data bursts to every other edge node in the network. The JET signaling scheme and Latest Available Unscheduled Channel (LAUC) [14] scheduling scheme are used. The maximum data-burst size is 40000 bytes. Each source-destination pair generates 1000 bytes of TCP packets at 16Mbps rate. The capacity of the links is 10Gbps, 2 wavelengths are used for the control channel, and 8 wavelengths are used for the data channel. The numbers of dropped bursts in the NS2 [13] simulation are given by

$$304, 269, 265, 245, 238, 116, 89, 79, 68, 56, 18, 15, 2 \qquad (7)$$

These data imply that some core nodes experience heavier congestion than others. If we can model the network properly to predict the nodes/links with specific properties (e.g., experiencing heavier congestion) and provide extra functionality such as fiber delay lines, then the overall performance of the network can be dramatically improved. Two promising methods for finding the desired locations are solutions to the *k-median* and *k-center facility location problems* for given $k \in \mathbb{N}$. The main objective and detailed descriptions are given below.

A. New Cost Definition

Usually, client costs are defined as distances to a facility. However, this definition is unsuitable for OBS networks. The distances between adjacent core nodes do not have a significant impact on cost, since the propagation delay in fiber is truly small compared to processing delays and OEO conversion delays. In addition, some source-destination pairs might not have any BCNs along their paths, whereas other pairs might have two or more, so simple distances do not reflect real costs. For these reasons, one needs to introduce a new cost function applicable to OBS networks with the proposed scheme.

The proposed retransmission scheme in OBS networks induces longer path setup delays and more control packets, as do similar two-way signaling schemes, but at the same time, the proposed scheme reduces path blocking probabilities. In this design problem, path setup delay (see Section III, part B) is defined as the new cost function, since data-burst delivery is governed by time-out and offset times.

B. The k-median Problem

A solution to the *k-median problem* places k new facilities serving n demand nodes so that the total distance (or cost) minimizes the cost incurred by the demand nodes. Here, the problem is formulated with the assumption that if the ingress node finds any BCNs along the path, the ingress node uses the proposed retransmission scheme; otherwise, it uses the two-way signaling scheme. Note that the average setup delay, as a cost function, increases almost linearly as the number of hops between the consecutive BCNs increases, for small blocking probabilities. Let $cost(\overrightarrow{n}(p))$ denote the cost of the path p with intermediate hop count vector $\overrightarrow{n}(p) = (n_1, \ldots, n_\ell)$, which in our earlier notation is $\mathbb{E}T(n_1, n_2, \ldots, n_\ell)$; and let the cost of communication between the ingress node v_i and egress node v_e be

$$cost(v_i, v_e) := cost(\overrightarrow{n}(w(v_i, v_e)))$$

where $w(v_i, v_e)$ is the shortest path from the ingress node v_i to the egress node v_e. Let $\mathbb{I}$ be the set of nodes in a given graph and let $\mathbb{S}_k$ be the set of all k-marked graphs in $\mathbb{I}$. Define the cost of $G \in \mathbb{S}_k$ to be

$$cost(G) = \sum_{(v_i, v_e) \in G} cost(v_i, v_e)$$

Then the objective is to find

$$\arg\min_{G \in \mathbb{S}_k} cost(G)$$

The complexity of the k-median problem with k part of the instance is NP-hard. However, for fixed values of k, the problem can be solved in $O(n^{k+2})$ time. To see this, note that Dijkstra's shortest-path algorithm runs in $O(n^2)$ time and that there are $\binom{n}{k} \leq n^k$ k-element subsets of the nodes of the given graph.

C. The k-center Problem

The *k-center problem*, first introduced by Hakimi, addresses the problem of minimizing the maximum distance between any node and its closest of k given facility locations. Here we take the distance as the number of hops between each core node and its closest BCN. This problem is formulated with the assumption that the distances between all pairs of adjacent nodes are equal. Define $\mathbb{I}$ and $\mathbb{S}_k$ as before, and let $\mathbb{J}$ be the set of all marked nodes in $\mathbb{S}_k$ such that $\mathbb{J} \subset \mathbb{S}_k$. Define the binary indicator x_j to tell whether a BCN is located at node j; x_j must satisfy $\sum_{j \in \mathbb{J}} x_j = k$ which stipulates that k BCNs are to be deployed. Define another binary indicator y_{ij} which indicates whether a core node at node i is assigned to a BCN at node j; y_{ij} must satisfy $\sum_{j \in \mathbb{J}} y_{ij} = 1$ for $\forall i \in \mathbb{I}$, which requires that each core node be assigned to exactly one BCN. Define $W(j)$ to be the maximum number of hops between core nodes and their closest BCNs for a given node $j \in \mathbb{J}$. Then the objective is to find

$$\arg\min_{j \in \mathbb{J}} W(j)$$

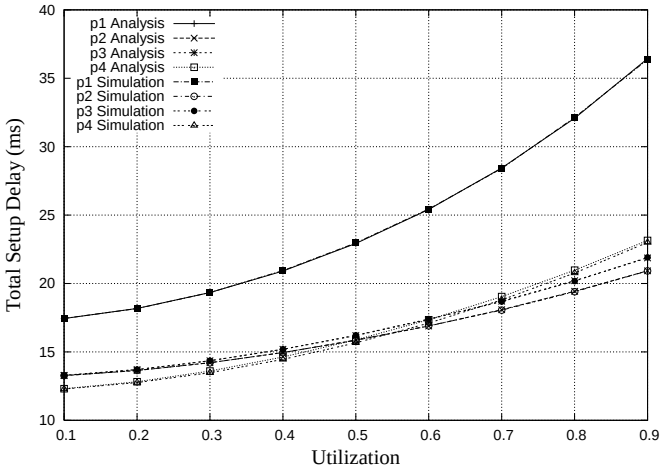

Fig. 3. Total Setup Delay

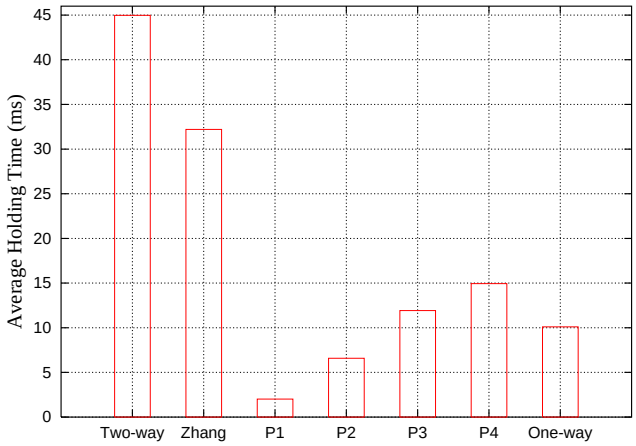

Fig. 4. Average Holding Time

The complexity of the k-center problem is NP-complete for $k \geq 3$. However, for fixed values of k, the k-center problem can be solved in time proportional to $n^{O(\sqrt{k})}$[15].

V. Evaluation of Networks with BCNs

To assess the benefit of BCNs, computation and simulation are applied to four different path vectors, each with the same path length 10 : $\overrightarrow{n}(p_1) = (1,2,7)$, $\overrightarrow{n}(p_2) = (3,5,2)$, $\overrightarrow{n}(p_3) = (5,2,3)$, $\overrightarrow{n}(p_4) = (6,3,1)$. The propagation delay is assumed fixed at $D = 10\ \mu s$ and the processing delay is assumed to be $\delta = 1ms$. The number of wavelengths is $W = 2$ and the number of fiber delay lines is 3. Regular core nodes and BCNs have fixed burst drop probabilities p and p' according to our queueing models, and wavelength reservation attempts are independent (Bernoulli) trials. A total of 10^6 simulations were performed and a constant total setup delay was assumed. Figure 3 shows the analytical computations and simulation results of average total setup delay. Note that the path vector which has decreasing(increasing) order has shorter(greater) average total setup delay due to the two-way signaling scheme at the last intermediate path. Path vector $\overrightarrow{n}(p_1)$ and $\overrightarrow{n}(p_4)$ have a similar but inverse order, and the last intermediate path in $\overrightarrow{n}(p_1)$, which is the longest, leads to a much greater total setup delay. Path vectors $\overrightarrow{n}(p_2)$ and $\overrightarrow{n}(p_3)$ have similar total setup delays when link utilization is low. However, the total setup delay of $\overrightarrow{n}(p_3)$ exceeds that of $\overrightarrow{n}(p_2)$ as link utilization increases. We can conclude that the last intermediate path should not be the longest one if we are to keep the total setup delay relatively small.

As mentioned in Section I, using the electronic buffer can be a limitation of retransmission in the OBS domain, since the buffer size required is quite large. We investigated the average time that a single source node must keep a given data burst until it receives the desired acknowledgement, which can be from either an intermediate node or a destination node. We define this as an average holding time. Various schemes are considered: The two-way signaling scheme with acknowledgement from the destination node, Zhang's retransmission scheme [7]; our proposed scheme with 4 different path vectors as introduced above (i.e., p_1, p_2, p_3, p_4); and the JET scheme representing one-way signaling schemes without any acknowledgement. The offset time is used as the average holding time for the JET scheme. The calculation of average holding time of the two-way signaling scheme and Zhang's scheme is analogous to the calculation of $\mathbb{E}T_i$ in the Appendix. The number of hops from the source to the destination is taken to be 10 and our queueing models are used for the calculation of p and p' with a utilization of 0.5. The propagation delay and the processing delay are also set to be the same as above. The size of an electronic buffer needed to store a given data burst is directly proportional to the average holding time with fixed traffic rate, so we can estimate the required electronic buffer size by investigating the average holding time. As shown in Figure 4, the average holding time of the two-way signaling scheme with acknowledgements exceeds by a factor of almost four and a half the average holding time under the JET scheme. Zhang's scheme requires more than three times the average holding time required by the JET scheme. Our proposed scheme requires relatively small average holding time, since the source node needs to hold a data burst until it receives an acknowledgement from the first BCN. We can conclude that the first intermediate path is better be short if we are to keep the size of electronic buffer small at the source.

To investigate the results of location optimization, we adopt the well known NSF 14-core-node topology. This network is considered to reflect the dynamcis of real world networks and has become a standard test case; many researchers have based their performance validations on this network. (e.g., [5], [7]) We assume that one edge node is connected to every core node and every edge node (as an ingress node) sends data bursts to every other edge node. Every edge node chooses the shortest path, which has at most 2 BCNs, to the destination edge node. $\mathbb{E}T(n_1, n_2, \ldots, n_\ell)$ and $W(\cdot)$ (see Section IV, part C) are used as cost functions for the k-median and k-center problems, respectively; the propagation and processing delays are assumed to be the same as before. The computations are

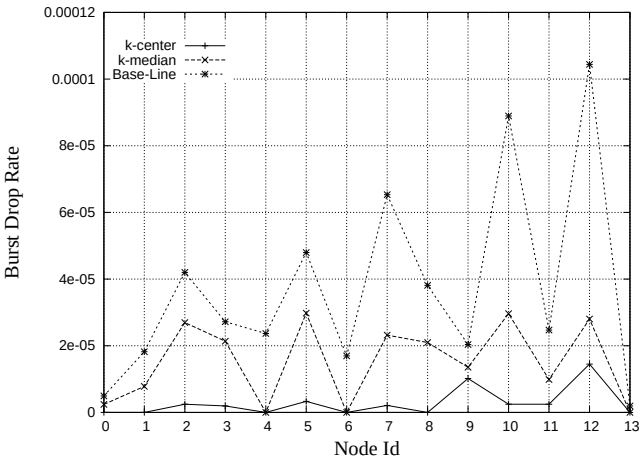

Fig. 5. Data-burst Drop Rate

given in Table I.

Figure 5 shows the NS2 simulation result with the parameter settings as follows. The NSF 14 core node network topology is used. Each core node has 2 edge nodes and each edge node (as an ingress node) transmits data bursts to every other edge node in the network. The processing delay is set to be 1ms and the propagation delay at each link to be 10 μs. The burst timeout is set to be 0.5 second and delay that one fiber delay line can produce is 10 μs. The maximum data-burst size is 40000 bytes. Each source-destination pair generates 1000 bytes of TCP packets at 16Mbps rate. The capacity of the links is 10Gbps, 2 wavelengths are used for the control channel, and 8 wavelengths are used for the data channel. We assumed all the core nodes have full wavelength conversion capability. The JET signaling and LAUC scheduling schemes are used as a base-line and our proposed signaling scheme with 4 BCNs located at optimal locations, as in Table I, are simulated for 30 minutes respectively. Figure 5 shows uneven burst drop rates which reflects the inhomogeneity of the nodes and the traffic they bear. Figure 5 reveals a key finding of our study: A small number of optimally located BCNs (i.e., 4 BCNs) can very significantly reduce the number of data-burst drops. As an added result, we found that the k-center problem can be expected to yield the better solutions for BCN placement, as it minimizes the longest intermediate paths, those paths most likely to have wavelength reservation failures. The k-center problem predicts relatively highly congested core nodes fairly well.

# of BCNs	Optimal Locations	
	k-median	k-center
1	0	5
2	10, 13	4, 7
3	6, 10, 13	3, 4, 5
4	6, 9, 10, 13	3, 5, 7, 8
5	0, 6, 9, 10, 12	0, 3, 4, 5, 8
6	0, 7, 8, 9, 10, 12	0, 1, 3, 4, 5, 8

TABLE I
THE OPTIMAL LOCATIONS

VI. CONCLUSIONS

We presented a retransmission scheme based on sparse BCNs, and we developed an analytic model to evaluate it. In the proposed scheme, BCNs secure the wavelength reservations up to the BCN itself using ACK and NACK packets; the header is then forwarded to reserve wavelengths up to the next BCN (or egress node). The scheme produces longer end-to-end delay, but smaller data-burst drop rates than those of one-way signaling scheme. The scheme also requires smaller electronic buffer size than that of two-way signaling scheme. However, as the link utilization increases (i.e., the link becomes heavily congested), the one-way reservation scheme will experience more frequent data-burst drops that eventually cause the end-to-end delay to exceed that of our scheme, because of the protracted retransmission mechanism triggered by the higher layer.

We focused on the analysis of end-to-end properties rather than individual components of a network. We also introduced the Facility Location Model to find the optimal locations of a few BCNs operating under the proposed scheme; the computations were found to be quite time-consuming. Our simulation results indicated that (1) a few BCNs placed at proper locations can decrease remarkably the data-burst drop rates, and (2) the k-center problem can be expected to be more suitable than the k-median problem in finding efficient solutions to the BCN placement problem.

Motivated by our positive results so far, we intend to investigate, as future directions of research, the BCN location optimization problem with uneven offered loads, varying propagation delays, and different network topologies.

REFERENCES

[1] J. Turner *"Terabit Burst Switching"* Journal of High Speed Networks, Vol. 8, No.1, 1999, pp. 3-16

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

[3] J.Y. Wei and R.I. McFarland Jr. *"Just-in-time Signaling for WDM Optical Burst Switching Networks"* Journal of Lightwave Technology, vol. 18, no. 12, pp. 2019-2037, Dec. 2000.

[4] I. Baldine, G.N. Rouskas, H.G. Perros, and D. Stevenson *"JumpStart: A Just-in-Time Signaling Architecture for WDM Burst-Switched Network"* IEEE Communications, vol. 40, no. 2, pp. 82-89, Feb. 2002.

[5] X. Yu *et al "Traffic Statistics and Performance Evaluation in Optical Burst Switched Networks"*, Journal of Lightwave Technology, Dec 2004.

[6] A. Maach *et* al *"Robust optical burst switching"*, Proceedings of Networks 2004, Vienna, Austria, June 2004, pp. 447-452..

[7] Q. Zhang *et* al *"Evaluation of Burst Retransmission in Optical Burst-Switched Networks"*, BroadNet, Oct 2005.

[8] I. Chlamtac, A. Fumagalli *et al "CORD: Contention Resolution by Delay Lines"*, IEEE Journal on Selected Areas Communication, vol. 14, pp.1014-1029, Jun 1996.

[9] Mark S. Daskin *"Network and discrete location : models, algorithms, and applications"*, New York : Wiley, 1995.

[10] M.Yoo *et* al *"QoS performance of optical burst switching in IP-over-WDM networks"*, IEEE Journal on Selected Areas in Communications, vol.18 Oct 2000.

[11] X. Lu, B. L. Mark *"Performance Modeling of Optical Busrt Switching With Fiber Delay Lines"*, IEEE Transactions on Communications, vol. 52, No. 12, Dec 2004.

[12] A. G. Fayoumi, A. P. Jayasumana *"Performance Model of an Optical Switch using Fiber Delay Lines for Resolving Contention"*, Proceedings of the IEEE International Conference on Local Computer Networks, Page(s):178 - 186 Oct 2003.
[13] http://dawn.cs.umbc.edu/
[14] Y. Xiong, et al *"Control architecture in optical burst-switched WDM metworks"*, IEEE Journal on Selected Areas in Communications, vol. 18, Oct 2000
[15] R.Z. Hwang, R.C.T. Lee, and R.C. Chang *"The slab dividing approach to solve the euclidean k-center problem"*, Algorithms 9 (1993), 1-22.

APPENDIX

First, we compute the expected value of the delay, T_i, at the i^{th} intermediate, n_i hop path from the $i-1^{st}$ BCN to the i^{th} BCN, $1 \leq i \leq \ell$, where the 0^{th} and ℓ^{th} BCN denote the ingress and egress node, respectively. T_i consists of the times taken by the control header in its attempts to make it to the next BCN securing reservations at each node. The times taken by each attempt are i.i.d., so it is natural to express $\mathbb{E}T_i$ as a recurrence. The probability that a new attempt must be started because of a failure to get a reservation at the j^{th} node is $(1-p)^{j-1}p$, for $j < n_i$ and is $(1-p)^{n_i-1}p'$ for $j = n_i$; the time taken is $2j(D+\delta)$ for all $1 \leq j \leq n_i$. After each failure, the time remaining has the same distribution as T_i, and so

$$\begin{aligned}\mathbb{E}T_i &= \sum_{j=1}^{n_i-1} p(1-p)^{j-1} \cdot [2j(D+\delta) + \mathbb{E}T_i] \\ &+ p'(1-p)^{n_i-1} \cdot [2n_i(D+\delta) + \mathbb{E}T_i] \\ &+ (1-p')(1-p)^{n_i-1}[2n_i(D+\delta)]\end{aligned}$$

Solving for $\mathbb{E}T_i$, we find

$$\begin{aligned}&\mathbb{E}T_i \cdot \{[1 - p \cdot \sum_{j=1}^{n_i-1}(1-p)^{j-1}] - (1-p)^{n_i-1}p'\} \\ &= 2(D+\delta)p \cdot \sum_{j=1}^{n_i-1} j \cdot (1-p)^{j-1} + 2(D+\delta)n_i(1-p)^{n_i-1}\end{aligned}$$

and then evaluating the sums and simplifying gives (1).

For $N_B = 1$ case, define N_h^2 to be the total number of hops that the header traverses at the second intermediate path to first secure n_2 consecutive successful reservations, then it is easy to see that

$$Pr\{T_2 > L_1\} = \sum_{a=1}^{\infty} Pr\{N_h^2 = \lfloor \frac{n_1}{2} \rfloor + a\}$$

Let $C^y_{(x,z)}$ be a combinatorial quantity giving the number of ways to distribute x balls into z distinguishable urns with the constraint that y is the maximum number of balls which can be put into an urn. $C^y_{(x,z)}$ can be calculated by the recursion below.

$$C^y_{(0,j)} = 1 \quad j = 0, 1, 2, \ldots$$

$$C^y_{(j,1)} = \begin{cases} 1 & 0 \leq j \leq y \\ 0 & j > y \end{cases}$$

$$C^y_{(x,z)} = \begin{cases} \sum_{i=0}^{y} C^y_{(x-i,z-1)} & for\ x > y,\ z \geq 2 \\ \sum_{i=0}^{x} C^y_{(x-i,z-1)} & for\ x \leq y,\ z \geq 2 \end{cases}$$

Each reservation trial is an i.i.d. Bernoulli trial having an outcome of Fail (F) with probability p and Success (S) with probability $1-p$. The event that the first k consecutive successes occur at the m^{th} trial means that a sequence of length m ends with k consecutive S's. In the first $m-k$ trials in that sequence, we could have $m-k$ F's, $m-k-1$ F's, ... ,$\lceil \frac{m-k}{k} \rceil$ F's but no fewer than $\lceil \frac{m-k}{k} \rceil$ F's; otherwise, we could have k consecutive S's in the sequence before the $m-k^{th}$ trial. The probability that there are $m-k-i$ F's in a sequence of length $m-k$ is $C^{k-1}_{(i,m-k-i)} \cdot p^{m-k-i}(1-p)^{k+i}$. Define $\mathbb{P}(k,m)$ as the probability that the first k consecutive successes occur at the m^{th} trial, then

$$\mathbb{P}(k,m) = \sum_{b=0}^{m-k-\lceil \frac{m-k}{k} \rceil} C^{k-1}_{(b,m-k-b)} \cdot p^{m-k-b}(1-p)^{k+b}$$

for $k, m \geq 1$ and so

$$\mathbb{E}(T_2 - L_1)^+ = \sum_{a=1}^{\infty}[2 \cdot (\lfloor \frac{n_1}{2} \rfloor + a) - n_1](D+\delta) \cdot \mathbb{P}(k,m)$$

where $k = n_2$ and $m = \lfloor \frac{n_1}{2} \rfloor + a$, respectively. Note that $\mathbb{P}(0,m) = p^m$ and $\mathbb{P}(k,0) = 0$ for $k, m \geq 1$ and $\mathbb{P}(0,0) = 1$. Analogously, the following probabilities are defined for the $N_B = 2$ case.

$$Pr(\xi > 0) = \sum_{a=1}^{\infty} \mathbb{P}(n_2 - 1, \lfloor \frac{n_1}{2} \rfloor + a - 1) \cdot (1-p')$$

$$Pr(\chi > 0) = \sum_{a=1}^{\infty} \mathbb{P}(n_3, \lfloor \frac{n_1+n_2}{2} \rfloor + a)$$

$$\begin{aligned}Pr(\psi > 0) &= \sum_{a=1}^{\infty} Pr\{N_h^2 + N_h^3 = \theta + a\} \\ &= \sum_{a=1}^{\infty} \sum_{b=0}^{\theta+a+n_1-n} \mathbb{P}(n_2 - 1, n_2 + b - 1) \\ &\quad * \mathbb{P}(n_3, \theta + a - n_2 - b) \cdot (1-p')\end{aligned}$$

where $\theta = max(n_3, \lceil \frac{n_1+n_2}{2} \rceil)$.