Burst Cloning: A Proactive Scheme to Reduce Data Loss in Optical Burst-Switched Networks

Xiaodong Huang[†], Vinod M. Vokkarane[‡], and Jason P. Jue[†]
Department of Computer Science, The University of Texas at Dallas, TX 75083-0688[†]
Computer and Information Science Department, The University of Massachusetts Dartmouth, MA 02747-2300[‡]
Email: {xxh020100@utdallas.edu, vvokkarane@umassd.edu, jjue@utdallas.edu}

Abstract— **In this paper, we propose a novel proactive scheme, *burst cloning*, to reduce data loss due to burst contention in optical burst-switched (OBS) networks. The idea is to replicate a burst and send duplicated copies of the burst through the network simultaneously. If the original burst is lost, the cloned burst may still be able to reach the destination. Primary design issues in burst cloning are to select the optimal nodes at which to do cloning and to prevent cloned bursts from contending for resources with original bursts. An analytical model is developed to evaluate the proposed scheme. The model is verified through extensive simulations. We observe that burst cloning could significantly reduce data loss in OBS networks.**

I. Introduction

Optical burst switching (OBS) is believed to be an effective paradigm to efficiently utilize the huge bandwidth of wavelength division multiplexing networks for bursty IP traffic [1]. At each ingress node, packets to the same egress node are packed together as a data burst, which will then be routed through the network all-optically. The control information for a data burst, contained in a burst head packet (BHP), is separated from the data and is transmitted on a dedicated control channel. BHPs are processed electronically at each intermediate node to reserve network resources before the data burst arrives.

A main concern in burst scheduling is the data loss due to burst contention. A contention occurs if and only if multiple bursts contend for the same outgoing channel on the same wavelength at the same time interval. Three well known contention resolution schemes are wavelength conversion [2], fiber delay line buffering [3], and deflection routing [4]. When a contention cannot be resolved, we can either drop an entire burst or just drop the contending part of a burst. The latter scheme, *burst segmentation* [5], can further reduce packet loss.

A common feature of the above solutions is that, if a burst (or a packet in a burst) is dropped, the burst (or the packet) can only be recovered by retransmission at a higher layer. In this paper, we propose a new proactive scheme, *burst cloning*, to reduce data loss in OBS networks. The idea is to replicate a burst and send duplicated copies of the burst through the network simultaneously. If the original burst is lost, the cloned burst may still be able to reach the destination. The destination egress nodes, with additional intelligence, will select one of the bursts, disassemble the burst, and forward the packets on to the corresponding destination hosts.

Primary design issues in burst cloning are to select the optimal nodes at which to do cloning and to prevent cloned bursts from contending for resources with original bursts. We investigate and provide solutions to both problems. An analytical model is developed to evaluate the proposed scheme. The model is verified through extensive simulations. We observe that burst cloning could significantly reduce data loss in OBS networks. It is worth noting that burst cloning is not a replacement of, but a complement to, existing contention resolution schemes. Burst cloning may be used along with any existing contention resolution scheme, which makes the proposed scheme quite universal and practical.

The rest of this paper is organized as follows. Section II explains the aspects of the proposed scheme. In Section III, we investigate node architectures to support burst cloning. An analytical model is developed in Section IV and numerical results are shown in Section V. We conclude the paper in Section VI.

II. Burst Cloning

In this section, we describe the details of burst cloning. We refer the original copy of a burst as the *original burst*, and the duplicated copy of a burst as the *cloned burst*. The traffic consisting of original bursts and cloned bursts is referred to as *original traffic* and *cloned traffic*, respectively. The node at which the cloning is done is referred to as the *cloning node*. In burst cloning, there are several aspects to be considered:

- the number of cloned bursts for each original burst,
- the selection of the cloning node, and
- the routing for the original burst and the cloned burst.

In burst cloning, one or more cloned bursts can be made for each original burst. On one hand, if more copies are made for a burst, the possibility of data loss for the burst is lower. On the other hand, if more copies are made, then more cloned traffic is added to the network. Cloned bursts may contend for network resources with original bursts, which may result in increasing loss for original bursts, which in turn may increase data loss instead of reducing it.

To prevent cloned traffic from interfering with original traffic, we introduce a *traffic isolation mechanism* by using priority-based preemptive burst scheduling. Original bursts are assigned high priority while cloned bursts are assigned low priority. When scheduling bursts, the high priority burst will always be scheduled if there is a contention between

0-7803-8938-7/05/$20.00 (C) 2005 IEEE

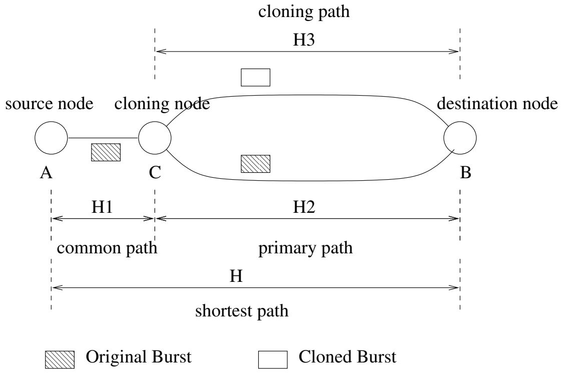

Fig. 1. General path structure for burst cloning

a high priority burst and a low priority burst, even if the low priority burst has already been scheduled. From the high priority traffic's point of view, there is *no* low priority traffic present in the network. The traffic isolation guarantees that the performance is at least as good with burst cloning as without cloning.

Another vital problem in burst cloning is the selection of the cloning node for a burst. In principle, the cloning node for each source destination pair could be different. A tightly related problem is the routing of the original burst and the cloned burst. The general path structure for burst cloning is shown in Fig. 1. An original burst is first sent along the *common path*. After the cloned copy is made at the cloning node, the original burst will then continue along the *primary path* while the cloned burst will be routed through the *cloning path*. The common path would be null if the cloning node is the source node. The primary path and cloning path would be null if there is no cloning for the burst. (We can also view this situation as the case in which the cloning node is the destination).

Since a cloned burst is a backup in case the original burst is lost, it is reasonable to keep the loss of original bursts as low as possible. Hence, we choose the common path and the primary path to be on the shortest path from the source to the destination. Accordingly, the cloning node is on the shortest path between the node pair. So we have:

$$H_1 + H_2 = H \tag{1}$$

where H_1, H_2, and H are the number of physical hops of the common path, the primary path, and the shortest path from the source to the destination, respectively.

After choosing the primary path, we must make a decision on whether the cloning path should be link-disjoint or even node-disjoint from the primary path. In this paper we are aiming to minimize data loss, and are not considering the protection issues. Therefore, the cloning path is allowed to be partially overlapped with the primary path, except on the first hop of both paths. The cloning path, as the second shortest path, will not be shorter than the primary path. Then we have:

$$H_3 - H_2 \geq 0. \tag{2}$$

Even with Equations (1) and (2), it is still not clear which node along the shortest path should be the cloning node. The scheduling of original bursts, due to the traffic isolation mechanism, is independent of the selection of the cloning nodes. Hence, we focus on the cloned bursts. On one hand, if the original burst is lost on the common path, the burst cannot be cloned. Thus, the common path should be shorter. On the other hand, the shorter the common path, the longer the primary path by equation (1), and the longer the cloning path by equation (2). This may in turn result in higher loss for cloned bursts. Thus a trade-off must exist between these two factors to achieve the best performance. Through our analytical model and simulations, we find that the former factor has a much greater effect on performance.

One brute force approach to find the optimal cloning node is to enumerate all possible cloning node configurations. For an OBS network with N nodes, there $N(N-1)$ source-destination pairs. It may be impractical to try all possible cloning configurations. As a compromise, in this paper we first classify source-destination pairs into D categories, where D is the diameter of a network. We set one cloning configuration for each category with d hops, where $d \in \{1, 2, ..., D\}$.

III. Network Architecture

In OBS networks, there are two types of nodes: electronic edge nodes and optical core nodes. Edge nodes are the gateways between OBS networks and traditional networks, such as IP networks and ATM networks. Core nodes route bursts hop-by-hop all-optically through the OBS network. A general architecture of edge nodes and core nodes in OBS networks was proposed in [6].

Edge nodes are responsible for burst assembly and de-assembly. In burst cloning, if the cloning node is the source node (referred to as *source cloning*), ingress nodes are also responsible for duplicating bursts. In source cloning, cloning can be done in the electronic domain. With burst cloning, egress nodes also have more work to do than just de-assembling bursts. The designated egress node may receive both the original burst and the cloned burst. To avoid sending duplicated packets to higher layers, the identity number of packets may be buffered at the egress node until a time out occurs or the other copy is received.

Core nodes provide all-optical routing, enabling data bursts to bypass the node. If the cloning node is an intermediate node (referred to as *intermediate cloning*), the cloning core node should have optical splitting capability, which duplicates an incoming data burst into two or more copies. Multicast capable optical crossconnect [7], MC-OXC, can be considered for intermediate cloning. In general, MC-OXCs are much more expensive than regular OXCs. However, as we will show with our analytical model and simulations in the following sections, source cloning has the best loss performance. Hence, in our proposed scheme, we need only regular OXCs instead of expensive MC-OXCs.

Another architectural issue comes from the combination of burst segmentation and burst cloning. One of the side-effects of burst segmentation is that the length of the burst can decrease due to possible contentions as it travels towards the destination.

Hence, some packets in the data burst can be lost before they reach the destination. Specifically, when a strict tail-dropping technique is adopted, the probability of packets toward the tail being dropped is much higher than the packets toward the head of the burst. When combined with burst cloning, it is very likely that both bursts reach the destination but that each burst has lost a part of its tail.

In order to counter this effect, we can reverse the order of packets in the cloned burst at the cloning node while sending out the original burst as it is. This reversal of packets ensures that packets in the tail of the original burst will be in the head of the cloned burst. At the destination, if both burst copies are received, even though the tail of each burst may be lost due to segmentation, the entire burst may be recovered from these two burst copies. If the cloning node is not a source node, it may be too complex to implement the burst reversal operation in the optical domain. For source cloning, we investigate the performance of a *complete reversal* policy, under which the packet order in the cloned burst is a complete reversal of that in the original burst.

At destination nodes, after both copies are received or time out, a post-process procedure will begin. Let L_0, L_1 and L_2 be the number of packets in the original burst, in the received primary burst, and in the received cloned burst, respectively. It is easy to obtain the number of packets received for complete reversal and without burst reversal as $min\{L_0, L_1 + L_2\}$ and $min\{L_1, L_2\}$, respectively. If any copy of the burst is lost, we can think the corresponding L_1 or L_2 to be 0. It can be seen that complete reversal can further reduce data loss compared to no reversal.

IV. Analytical Model

In this section, we extend the analytical model in [5] to calculate the average packet loss probability in burst cloning with burst segmentation. Interested readers are referred to [5] for the details. First, let us define the following notation:

- $1/\mu$: average burst length;
- K^{sd}: total number of hops from the source s to the destination d;
- k_l^{sd}: total number of hops from the source s until to the link l, along the path sd;
- λ^{sd}: burst arrival rate from source s to destination d;
- λ_l^{sd}: arrival rate of original bursts to the link l on the path from source s to destination d;
- γ_l^{sd}: arrival rate of cloned bursts to the link l on the path from source s to destination d;
- $\lambda_l = \sum_{sd} \lambda_l^{sd}$: arrival rate of original bursts to link l, due to all source-destination pairs sd;
- $\gamma_l = \sum_{sd} \gamma_l^{sd}$: arrival rate of cloned bursts to link l, due to all source-destination pairs sd;
- $\lambda_{l_i^{sd}}$: arrival rate of all original bursts on the ith hop link of the path between source s and destination d. Let $l' = l_i^{sd}$, then $\lambda_{l_i^{sd}} = \lambda_{l'}$;
- r_{sd}: primary route from source s to destination d;
- r'_{sd}: cloning route from source s to destination d.

The offered load on link l by traffic from source s to destination d depends on whether link l is on the path from s to d. With segmentation, burst length may decrease along the path from s to d. However, there is no reduction of the arrival rate of bursts. Thus,

$$\lambda_l^{sd} = \begin{cases} \lambda^{sd}, & \text{if } l \in r_{sd} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Following [5], we obtain the packet loss probability of original bursts (with high priority) as

$$P_{loss0}^{sd} = 1 - \frac{\mu}{\sum\limits_{i=1}^{K^{sd}} \lambda_{l_i^{sd}} + \mu} \tag{4}$$

and the utilization due to original bursts on link l as

$$\rho_l = \sum_{s,d} \frac{\lambda_l^{sd}}{\sum\limits_{i=1}^{k_l^{sd}} \lambda_{l_i^{sd}} + \mu}. \tag{5}$$

Cloning traffic is treated as the same as the low priority traffic in [5]. Thus,

$$\gamma_l^{sd} = \begin{cases} \lambda^{sd}, & \text{if } l \in r'_{sd}, l = l_0'^{sd} \\ \gamma_h^{sd}(1-\rho_h), & \text{if } l, h \in r'_{sd}, h = l_{i-1}'^{sd}, i \geq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Following [5], we obtain the packet loss probability of cloned bursts (with low priority) as

$$P_{loss1}^{sd} = 1 - \frac{\prod\limits_{i=1}^{K^{sd}} (1-\rho_i) \cdot \mu}{\sum\limits_{j=1}^{K^{sd}} (\lambda_{l_j^{sd}} + \gamma_{l_j^{sd}}) + \mu}. \tag{7}$$

After we obtain the packet loss probabilities for both original bursts and cloned bursts, we can calculate the end-to-end packet loss probability P_{loss}^{sd} by

$$P_{loss}^{sd} = P_{loss0}^{sc} + (1 - P_{loss0}^{sc}) \cdot P_{loss0}^{cd} \cdot P_{loss1}^{cd} \tag{8}$$

where c is the cloning node of source-destination pair sd.

Taking the traffic weighted average of end-to-end packet loss probabilities, we obtain the average packet loss probability for the network as

$$P_{loss} = \sum_s \sum_d \frac{\lambda^{sd}}{\lambda} P_{loss}^{sd}. \tag{9}$$

V. Numerical Results

In this section, we present the numerical results from our analytical model and simulations. We evaluate the performance of our proposed scheme in the 14-node NSFNET as shown in Fig. 2, in which the number on a link is the distance in kilometers between two adjacent nodes. Bursts arrive to the network according to a Poisson process. Incoming traffic is evenly distributed among all source-destination pairs. Packets in a burst have a fixed length of 1250 bytes. The length of a burst is exponentially distributed. The link transmission rate is 10 Gb/s, and the speed of light in optical fibers is assumed to be 250 km/ms. There are no wavelength converters or optical buffers in the network. TAG signalling protocol is assumed, with each node equipped with fixed FDLs to buffer data bursts while BHPs are being processed. To avoid the effect of wavelength assignment algorithms, we run the simulation on one wavelength.

The diameter of the 14-node NSFNET is $D = 3$. According to the proposed scheme, we divide source-destination pairs into 3 categories: 1-hop pairs, 2-hop pairs, and 3-hop pairs. We number the nodes along the k-hop ($k = 1, 2, 3$) path as $0, 1, 2, \ldots, k$. Then, we use a vector $\mathcal{C} = [c_1, c_2, c_3]$ to denote cloning configurations, where c_k denotes the cloning node for all k-hop source-destination pairs. $c_k = 0$ means source cloning; $c_k = k$ means no cloning; $c_k \in \{1, 2, \ldots, k-1\}$ means intermediate cloning. For example, if we do cloning for all pairs at the source node, we set $\mathcal{C} = [0, 0, 0]$; if we do not do cloning for any node pair, we set $\mathcal{C} = [1, 2, 3]$. All possible cloning configurations with *cloning configuration index* (denoted by $I_{\mathcal{C}}$) are listed in Table I.

We first study the performance of different clone configurations without the burst reversal operation. Fig. 3 and Fig. 4 show the packet loss probability with different cloning configurations by simulation and analysis, respectively. In both figures, each curve denotes the loss performance under one specific network load, which varies from 0.1 to 64.

We observe that the results for different cloning configurations are quite consistent under different network loads for simulation and for the analytical model. Source cloning ($I_{\mathcal{C}} = 0$ with $\mathcal{C} = [0, 0, 0]$) always has the best loss performance, followed by configuration $I_{\mathcal{C}} = 12$ (i.e., $\mathcal{C} = [1, 0, 0]$, no cloning for 1-hop node pairs and source cloning for all other node pairs).

From Fig. 3 and Fig. 4, we find that any cloning configuration (with $I_{\mathcal{C}}$ between [0, 22]) has better loss performance than without cloning (i.e., $I_{\mathcal{C}} = 23$ with $\mathcal{C} = [1, 2, 3]$). This performance is due to the traffic isolation mechanism and preemptive scheduling in our proposed scheme. Thus, cloned bursts do not interfere with original bursts. Cloned bursts just try to utilize network resources which are not occupied by original bursts.

It is quite interesting to notice in Fig. 3, Fig. 4, and Fig. 5 that there are consistently 6 increasing segments in each curve. We find that these segments have cloning configuration index $I_{\mathcal{C}}$ as follows: $\{0, 1, 2, 3\}$, $\{4, 5, 6, 7\}$, $\{8, 9, 10, 11\}$, $\{12, 13, 14, 15\}$, $\{16, 17, 18, 19\}$ and $\{20, 21, 22, 23\}$. In each segment, the four configurations have the same setting for 1-hop and 2-hop node pairs, while the cloning node for 3-hop pairs moves further and further from the source node. The further the cloning node is from the source, the less chance that the burst is cloned.

Fig. 5 and Fig. 6 compare the loss performance of the analytical model with that of the simulations. We can see that the analytical model is quite accurate. Fig. 5 emphasizes the loss performance with different cloning configurations under a fixed network load (0.1). Under other network loads, there is similar relative performance. Fig. 6 gives a global view of

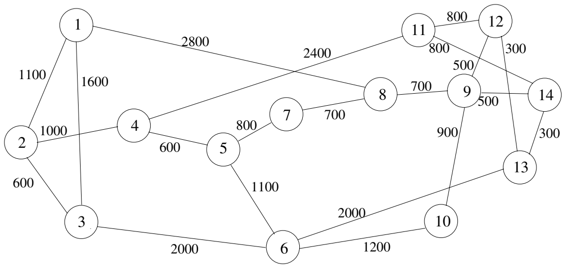

Fig. 2. 14-node NSFNET with distance in kilometers

TABLE I

NUMBERING THE CLONE CONFIGURATIONS

$I_{\mathcal{C}}$	$\mathcal{C}$	$I_{\mathcal{C}}$	$\mathcal{C}$	$I_{\mathcal{C}}$	$\mathcal{C}$
0	[0, 0, 0]	8	[0, 2, 0]	16	[1, 1, 0]
1	[0, 0, 1]	9	[0, 2, 1]	17	[1, 1, 1]
2	[0, 0, 2]	10	[0, 2, 2]	18	[1, 1, 2]
3	[0, 0, 3]	11	[0, 2, 3]	19	[1, 1, 3]
4	[0, 1, 0]	12	[1, 0, 0]	20	[1, 2, 0]
5	[0, 1, 1]	13	[1, 0, 1]	21	[1, 2, 1]
6	[0, 1, 2]	14	[1, 0, 2]	22	[1, 2, 2]
7	[0, 1, 3]	15	[1, 0, 3]	23	[1, 2, 3]

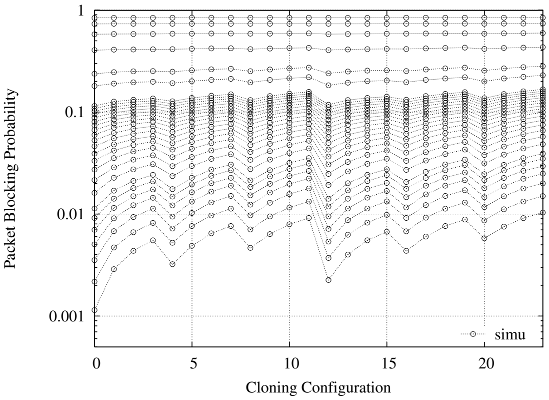

Fig. 3. Packet loss vs. clone configurations under load (0.1-64) (Simulation)

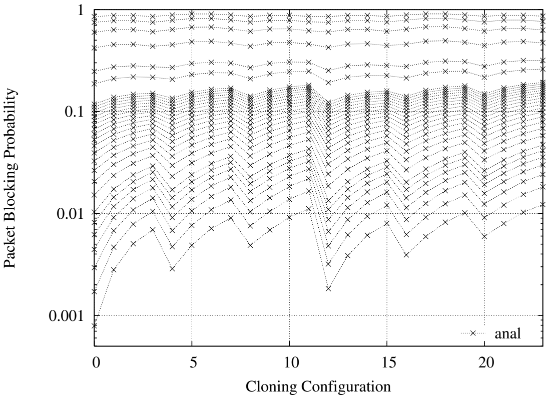

Fig. 4. Packet loss vs. clone configurations under load (0.1-64) (Analytical)

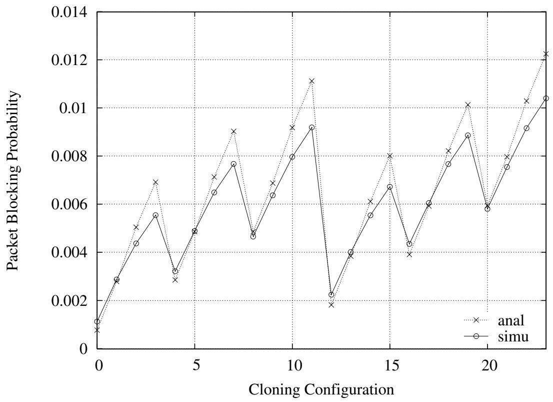

Fig. 5. Packet loss vs. clone configurations @ load (0.1)

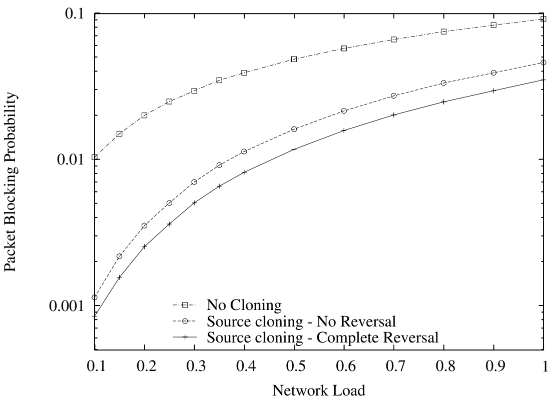

Fig. 7. Packet loss comparison of burst reversal scheme

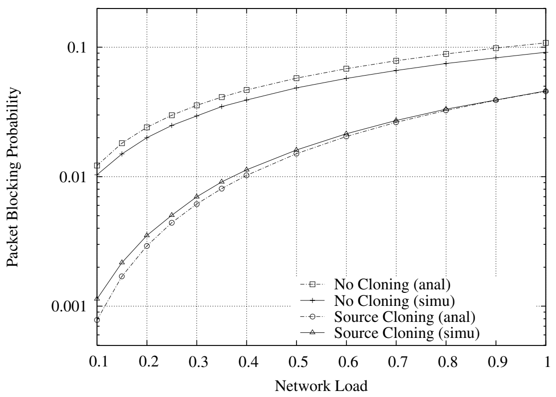

Fig. 6. Packet loss vs. network load (0.1-1)

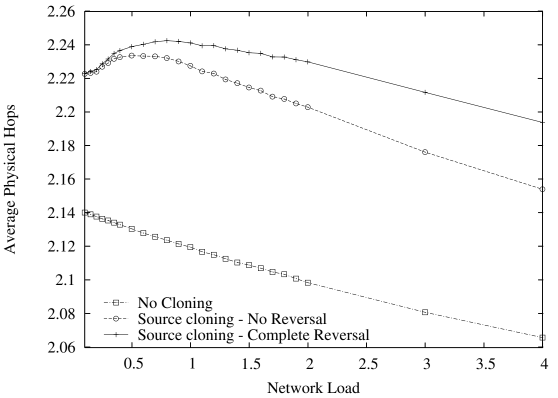

Fig. 8. Average physical hop comparison of burst reversal scheme

source cloning under different network loads (0.1-1). Fig. 6 clearly shows again, from another point of view, that source cloning can significantly improve the loss performance.

We also study the performance of burst reversal. Since we find that source cloning has the best loss performance among all possible cloning configurations, we will focus on source cloning with burst reversal. From Fig. 7, we observe that complete reversal can significantly reduce data loss. Fig. 8 shows that burst cloning results in a small increase in the average number of packet hops than without cloning. With burst cloning, some otherwise lost packets will arrive at the destination in the cloned burst. Between any node pair, cloned bursts undergo a greater number of hops than original bursts. Thus, bursts have a greater number of hops with burst cloning than without burst cloning. For the same reason, complete reversal results in more packets in the cloned burst reaching the destination compared to cloning without burst reversal. Thus, complete reversal also results in more hops. However, the increase in hops is not significant.

VI. Conclusions

This paper addresses the issue of data loss in OBS networks due to burst contention. A new proactive scheme, called *burst cloning*, was proposed, and an analytical model was developed to calculate the packet loss probability. Extensive simulations verified the analytical model and showed that burst cloning can significantly improve the loss performance without significant increase in packet delay.

References

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

[2] B. Ramamurthy and B. Mukherjee, “Wavelength Conversion in WDM Networking,” *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 7, pp. 1061-1073, Sept. 1998.

[3] C. Gauger, “Dimensioning of FDL Buffers for Optical Burst Switching Nodes,” *Proceedings, Optical Network Design and Modeling (ONDM 2002)*, Torino, Feb. 2002.

[4] C. Hsu, T. Liu, and N. Huang, “Performance Analysis of Deflection Routing in Optical Burst Switched Networks,” *Proceedings, IEEE INFOCOM 2002*, vol. 1, Jun. 2002.

[5] V. M. Vokkarane and J. P. Jue, “Prioritized Burst Segmentation and Composite Burst Assembly Techniques for QoS Support in Optical Burst-Switched Networks,” *IEEE Journal of Selected Areas of Communications*, vol. 21, no. 7, pp. 1198-1209, Sept. 2003.

[6] Y. Xiong, M. Vandenhoute, and H. Cankaya, “Control Architecture in Optical Burst-Switched WDM Networks,” *IEEE Journal of Selected Areas of Communications*, Vol. 18, No. 10, pp. 1838-1851, Oct. 2000.

[7] W.S. Hu and Q.J. Zeng, “Multicasting optical cross connects employing splitter-and-delivery switch,” *IEEE Photonics Technology Letters*, vol. 10, no. 7, pp. 970-972, Jul. 1998.