

# Differentiated Contention Resolution for QoS in Photonic Packet-Switched Networks

Tao Zhang, Kejie Lu, and Jason P. Jue  
Department of Computer Science  
The University of Texas at Dallas, Richardson, TX 75083

**Abstract**—In this paper, we propose a framework for providing differentiated contention resolution in photonic packet-switched networks by exploiting recirculation buffering and deflection routing. We develop an analytical model to evaluate the packet loss probability and the end-to-end delay for different buffering and deflection routing schemes, and we investigate the effectiveness of the control schemes in providing differentiated loss and delay. The accuracy of the analytical model is confirmed by simulation.

## I. INTRODUCTION

With the exponential growth of Internet traffic, next-generation packet-switched networks must provide extremely fast data transmission rates as well as support for diverse quality of service (QoS). *Dense Wavelength Division Multiplexing* (DWDM), as a key transmission technology [1], can support transmission rates of more than 1 Tbit/s on a single fiber. To fully utilize such high data rates, photonic packet-switched networks have been proposed, in which data packets can be transmitted transparently in the optical domain.

In addition to bandwidth requirements, emerging packet-switched networks are also expected to satisfy loss and delay requirements for different applications. In most packet switching networks, packet loss and end-to-end delay result from packet contention, which occurs when two or more incoming packets need to be forwarded to the same output at the same time. In optical networks, approaches to resolving contention include wavelength conversion, buffering, and deflection routing.

In wavelength conversion, a contending packet can be converted from one wavelength to another in order to avoid conflict. Although this method has been studied in many works, it may have limited practicality, since all-optical wavelength conversion technology is immature and wavelength converters are expected to remain expensive in the foreseeable future [2].

In buffering, contending packets are temporarily stored and forwarded at a later time. In optical networks, optical fiber delay lines [3] can be utilized to delay packets for a fixed amount of time. A limitation of this scheme is that the fiber delay lines may be bulky and may increase end-to-end delay.

In deflection routing, a packet is forwarded to an alternative output port if its primary output port is occupied by another packet. Thus, the links in the network are acting as buffers [4],[5]. Consequently, packet loss probability in the network

can be decreased without increasing the cost of the photonic switch. Nevertheless, deflection routing may lead to increased end-to-end packet delay.

In this paper, we propose a framework for differentiated contention resolution which exploits recirculation buffering and deflection routing schemes. The proposed differentiated contention resolution scheme includes different buffering and different deflection schemes for different classes of traffic. Specifically, for buffering schemes, we can limit the buffer usage and the recirculation times of packets; for example, traffic with relaxed loss requirements may be restricted to use only a subset of the total available buffers. In this case, some buffers will be reserved only for higher priority traffic. The deflection policy can also be varied depending on the class of traffic. For example, traffic which is loss sensitive could be deflected, while traffic which is delay sensitive would not be deflected. Within the framework, we first define the buffering and deflection parameters that can be varied to provide differentiation, and then we specify the buffering and deflection policies applied to different classes of traffic. An analytical model is then developed to evaluate the performance of the proposed schemes, and the accuracy of the analysis is verified through simulation.

The rest of this paper is organized as follows. In Section II, we elaborate on differentiated contention resolution schemes for supporting different QoS requirements. An analytical model is proposed in Section III to evaluate the loss and delay performance of different buffering and deflection routing schemes. Section IV provides numerical results and discussions. Finally, Section V concludes the paper.

## II. DIFFERENTIATED CONTENTION RESOLUTION

In this section, we first provide the switch architecture with recirculation buffers, and we present a label-based forwarding scheme for supporting differentiated service. We then discuss various buffering and deflection routing schemes for providing differentiated contention resolution.

### A. Switch Architecture

One objective in network design is to minimize implementation costs. Thus, we assume that there are no wavelength converters in the network, and we also assume that only a small number of optical buffers, i.e., fiber delay lines, are available in a switch. Based on the assumptions above, we consider the generic switch architecture shown in Fig. 1.

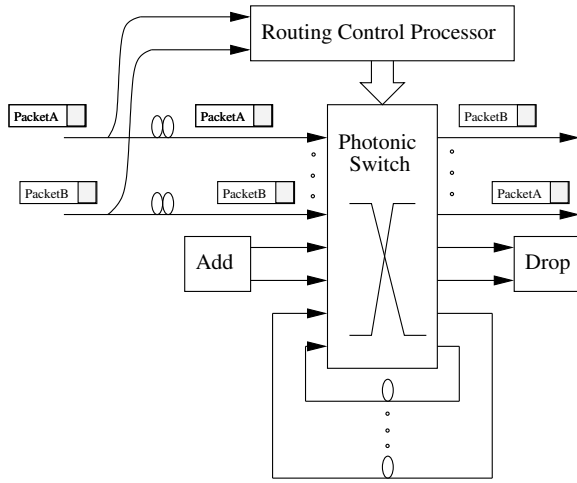


Fig. 1. A generic architecture for photonic packet switch.

In this switch, there is a central routing control processor which has the capability to perform control functions. Each input is connected to the switching element via a fixed-length fiber delay line, such that the propagation delay in the fiber is sufficient for the controller to process the header and configure the switch. Some ports of the switching element are connected to add/drop units, which can be used to add and drop local traffic. We assume that packets arrive asynchronously and the length of each packet is  $L$  time units.  $K$  recirculation buffers, which connect the switch outputs to the switch inputs via fiber delay lines, are used to resolve packet contention. In this manner, one fiber delay line can be shared by packets from different input ports at different times. In this paper, we assume that each fiber delay line in the recirculation buffers can delay a packet for  $L$  time units.

### B. Label-Based Forwarding Scheme

To support differentiated service in photonic packet-switched networks, we apply a label-based control scheme for forwarding packets. This method is similar to the method used in MPLS in that a label-path that satisfies certain QoS requirements must be set up before a packet can be sent into the network. The main difference between two methods is that, in the proposed scheme, the label may be distributed to all nodes in the network, and multiple alternative outputs may be specified for each label in the forwarding table so that deflection routing can be supported. Since each node forwards packets according to the label, the complexity in forwarding can be minimized. Furthermore, by establishing different label-paths for different services, the proposed scheme can support existing QoS schemes such as flow-based *Integrated Services (IntServ)* and class-based *Differentiated Services (DiffServ)*.

Label-paths can either be set up statically for different classes of traffic, or can be set up dynamically upon requests. In this paper, we assume that label-paths are established offline in a static manner.

### C. Differentiated Contention Resolution

Based on different loss and delay requirements, differentiated service can be implemented by different buffering and different deflection routing policies within the label-based control scheme. We define a control vector,  $V_l = [K^l, M^l, D^l, A^l]$ , corresponding to a label path, in which  $K^l$  is the subset of the total number of buffers that packets of a given label  $l$  can use.  $M^l$  is the maximum number of times that a packet is allowed to recirculate in the buffers.  $D^l$  represents the deflection routing policy, and  $A^l$  is the number of deflection alternatives.

In this paper, we consider a buffering-first policy. That is, a packet will be forwarded to buffers if the primary output port is occupied by another packet, then the packet will be forwarded to one of its predefined deflection output ports if the buffers are not available. The details of the control scheme are as follows.

◇  $K^l$ : limit of buffer usage

We can limit the number of buffers that a flow or class can use. Specifically, from a packet loss perspective, all of the buffers may be used by high priority traffic, while only a subset of the buffers may be used for low priority traffic. In this manner, some buffers are dedicated to higher priority applications in order to guarantee lower packet loss.

◇  $M^l$ : limit of recirculation time

In general, we must limit the overall time that a packet can stay in the switch, otherwise, one packet may recirculate indefinitely in the system. The maximum number of times that a packet may traverse a recirculation buffer may also be utilized for service differentiation purpose. Intuitively, a larger limit may lead to lower packet loss while increasing the buffering delay.

◇  $D^l$ : deflection routing algorithm

Deflection schemes can be classified based on whether or not loops are allowed. Loops occur when a packet is forwarded to a node which has already been visited on the path to the destination. An example of deflection routing which allows loops is *shortest-path deflection (SPD)*, in which all possible outputs are sorted by distance to the destination. If the primary output of an incoming packet is occupied, the switch will try to forward it to the output that is available and has the shortest distance to the destination. By allowing loops in the network, SPD provides lower packet loss under lower traffic load but incurs higher end-to-end delay. An example of loopless deflection is *destination-based loopless deflection (D-LLD)* [5], in which forwarding tables are set in a manner that avoids loops. Since D-LLD does not allow loops in the network, the maximum end-to-end propagation delay is limited.

◇  $A^l$ : number of deflection alternatives

The number of deflection alternatives is the maximum number of alternative output ports on which a packet can be forwarded. In non-deflection schemes, the number of deflection alternatives is equal to one. Therefore, each node has

to maintain only the primary output port in forwarding table for a given label path. In this case, if the primary output is occupied, then the packet will be dropped. For example, when the *shortest path (SP)* scheme is used, the primary output to a given destination can be determined by running Dijkstra's shortest path algorithm to find the output. In this case, the control vector for the label path can be expressed as:  $V_i = [K^l, M^l, \text{SPD}, 1]$  or  $V_i = [K^l, M^l, \text{SP}, 1]$ . If the number of deflection alternatives is greater than one, then each node must maintain multiple alternative output ports for a given label path. In this case, the packet loss ratio may be smaller than non-deflection schemes. However, deflection may also increase the end-to-end propagation delay.

We now use an example to illustrate the differentiation scheme. Consider a network where each node has 3 recirculation buffers and suppose three classes of traffic, loss sensitive, delay sensitive, and best effort traffic, have to be supported. To support loss sensitive traffic, we can define control vector  $V_0 = [3, 3, \text{SPD}, 2]$ , in which three buffers are used, recirculating times are limited to three, SPD is used, and the number of deflection alternatives is equal to two. To support delay sensitive traffic, a control vector  $V_1 = [3, 1, \text{D-LLD}, 2]$  can be defined, in which three buffers are used, recirculating times are limited to one, D-LLD is used, and the number of deflection alternatives is equal to two. To support best effort traffic, a control vector  $V_2 = [0, 0, \text{SP}, 1]$  can be defined, in which no buffering is allowed, and no deflection is allowed. We will show the performance of this scheme in Section IV.

### III. ANALYTICAL MODEL

In this section, we present an analytical model to evaluate the packet loss and delay performance of different buffering and deflection routing schemes. The analysis may be applied to any arbitrary mesh topology. The main idea in this analytical model is to calculate loss probability for each connection by using a reduced load approximation algorithm.

We first separate traffic in a network into different flows, where each flow consists of packets that have the same source and destination and that require the same quality of service. For any flow  $f$ , we assume that packets arrive according to a Poisson process with rate  $\lambda^f$ .

To simplify the analysis, we assume that there is no wavelength conversion; thus, we only consider a single wavelength plane. Therefore, we can model each link as an M/D/1/1 queuing system. Let  $\lambda_{ij}$  be the packet arrival rate on link  $ij$  and  $P_{ij}$  be the packet blocking probability on link  $ij$ . The packet blocking probability on link  $ij$  can be given as:

$$P_{ij} = \frac{\lambda_{ij}L}{1 + \lambda_{ij}L}. \quad (1)$$

For a given flow  $f$ , we denote  $\lambda_{ij}^f$  as the packet arrival rate on link  $ij$ . Then we have:

$$\lambda_{ij} = \sum_f \lambda_{ij}^f. \quad (2)$$

$\lambda_{ij}^f$  depends on whether link  $ij$  is on the primary path or on the deflection path. To analyze  $\lambda_{ij}^f$ , we first define  $\lambda_i^f$  as the total arrival rate of  $f$  from all incoming links of node  $i$ , define  $\lambda_i^{fb}$  as the arrival rate of packets that pass through buffers and eventually depart from the primary output of  $f$  on node  $i$ . Let  $S_f$  and  $D_f$  be the source and destination node of flow  $f$ . If link  $ij$  is on the primary path to  $D_f$ , then the applied load will be both the load offered by all previous-hop links toward node  $i$  and the load from local buffers. If link  $ij$  is the  $l$ -th ( $l > 1$ ) deflection link at node  $i$ , and link  $ik$  is the  $(l-1)$ -th deflection link to destination  $D_f$ , then the applied load on link  $ij$  will be the load that contends and is blocked on link  $ik$ . To simplify the notation, we let  $w_i^f(l)$  be the next node corresponding to the  $l$ -th alternate deflection port at node  $i$ . For example,  $w_i^f(1)$  is the primary output port at node  $i$ . The parameter  $\lambda_{ij}^f$  can be calculated as

$$\lambda_{ij}^f = \begin{cases} \lambda_i^f + \lambda_i^{fb}, & j = w_i^f(1) \\ \lambda_i^f \cdot P_{iw_i^f(1)} - \lambda_i^{fb}, & j = w_i^f(2) \\ \lambda_{iw_i^f(l-1)}^f \cdot P_{iw_i^f(l-1)}, & j = w_i^f(l), (l > 2). \end{cases} \quad (3)$$

The parameter  $\lambda_i^f$  can be calculated as

$$\lambda_i^f = \begin{cases} \lambda^f, & i = S_f \\ \sum_{k, i=w_k^f(1)} [\lambda_k^f \cdot (1 - P_{ki}) + \lambda_k^{fb}] \\ \quad + \sum_{k, i \neq w_k^f(1)} [\lambda_{ki}^f \cdot (1 - P_{ki})], & i \neq S_f. \end{cases} \quad (4)$$

From Eq. (3) and Eq. (4) we can observe that, for a given flow  $f$ , the attempted arrival rate to buffers at node  $i$  is

$$\lambda_i^f \cdot P_{iw_i^f(1)}.$$

We let  $P_i^{fb}$  be the conditional probability that packets can not depart from the primary output  $w_i^f(1)$  given that these packets have entered buffers on node  $i$ . Therefore,  $\lambda_i^{fb}$  can be calculated as

$$\lambda_i^{fb} = (\lambda_i^f \cdot P_{iw_i^f(1)}) \cdot (1 - P_i^{fb}). \quad (5)$$

To calculate the conditional probability  $P_i^{fb}$ , we make the following assumptions:

- each packet can enter the fiber delay line at a given node up to  $M^f$  times;
- the probability that no fiber delay line is available on node  $i$  is  $P_i^b$ .

Since the length of any fiber delay line is  $L$  units, one packet that leaves a buffer and enters the switching element can use the same buffer if there is no limitation to enter again. Therefore, we have

$$P_i^{fb} = 1 - (1 - P_i^b) \times \left(1 - (r_i^f)^{M^f}\right) \quad (6)$$

where we define  $r_i^f$  be the probability that the primary output  $j$  ( $j = w_i^f(1)$ ) is not available when the packet leaves the buffer.  $r_i^f$  can be estimated by using the following approximation

$$r_i^f = 1 - e^{(-\lambda_{ij} \times L/2)}. \quad (7)$$

We now discuss how to calculate  $P_i^b$  in Eq. (6). Let  $d_i^f$  be the average duration that a packet in flow  $f$  stays in a buffer. Since a packet in flow  $f$  can stay in a buffer for up to  $M^f$  times, we have

$$d_i^f = \sum_{m=1}^{M^f-1} \left[ m \cdot L \cdot (r_i^f)^{m-1} \cdot (1 - r_i^f) \right] + M^f \cdot L \cdot (r_i^f)^{M^f-1}. \quad (8)$$

Let  $d_i$  be the average duration that a packet of any flow stays in a buffer on node  $i$ . Notice that the attempted arrival rate of flow  $f$  to buffers on node  $i$  is  $\lambda_i^f \cdot P_{iw_i^f(1)}$ , we have

$$d_i = \frac{\sum_f d_i^f \cdot (\lambda_i^f P_{iw_i^f(1)})}{\sum_f \lambda_i^f P_{iw_i^f(1)}}. \quad (9)$$

We can then use the well-known Erlang-B formula to calculate  $P_i^b$ , since the Erlang-B formula is insensitive to the distribution of the service time.

$$P_i^b = \frac{\rho_i^K / K!}{\sum_{k=0}^K \rho_i^k / k!} \quad (10)$$

where

$$\rho_i = d_i \times \sum_f \lambda_i^f P_{iw_i^f(1)} = \sum_f d_i^f \cdot (\lambda_i^f P_{iw_i^f(1)}). \quad (11)$$

The packet loss probability of flow  $f$  can be calculated as

$$P^f = 1 - \frac{\lambda_{D_f}^f}{\lambda^f}, \quad (12)$$

and then the average packet loss probability  $P$  in the network can be calculated as:

$$P = \frac{\sum_f P^f \cdot \lambda^f}{\sum_f \lambda^f}. \quad (13)$$

To calculate packet loss probability recursively, we use the following reduced load approximation algorithm.

- 1) Initialize all  $P_{ij} = 0$  for any link  $ij$ .
- 2) Calculate all  $\lambda_{ij}$ ,  $\lambda_i^f$  and  $\lambda_{ij}^f$  by using Eq. (2)-(4).
- 3) Calculate all  $P_{ij}$  by using Eq. (1).
- 4) Calculate all  $\lambda_i^{f^b}$  by using Eq. (5)-(11).
- 5) Calculate  $P$  by using Eq. (12) and Eq. (13).
- 6) Stop if  $P$  converges; otherwise, go to Step (2).

For a given flow  $f$ , the average end-to-end delay for packets that successfully reach their destination  $D_f$  can be calculated as:

$$D^f = \frac{\sum_{pq} d_{pq} \hat{\lambda}_{pq}^f}{\lambda^f (1 - P^f)} + \frac{\sum_p d_p \hat{\lambda}_p^f}{\lambda^f (1 - P^f)} \quad (14)$$

where  $d_{pq}$  denotes the propagation delay on link  $pq$ ,  $\hat{\lambda}_{pq}^f$  denotes the  $f$  traffic that passes through link  $pq$  and successfully reaches the destination  $D_f$ , and  $\hat{\lambda}_p^f$  denotes the  $f$  traffic that passes through node  $p$  and successfully reaches the destination  $D_f$ .

To calculate  $\hat{\lambda}_{pq}^f$ , we first define  $\lambda_i^f(pq)$  and  $\lambda_{ij}^f(pq)$  as packet arrival rate on node  $i$  and link  $ij$ , provided that all packets have traversed link  $pq$ . By using the method we

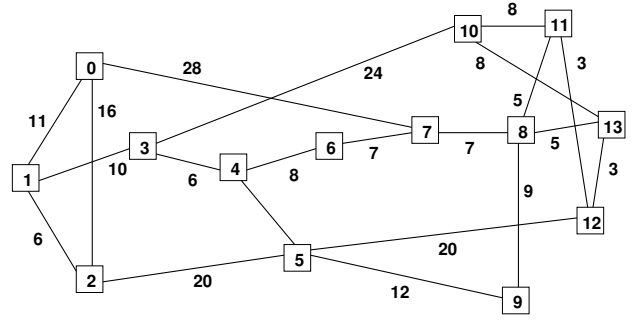


Fig. 2. A 14-node network topology.

developed for calculating  $\lambda_i^f$  and  $\lambda_{ij}^f$ , we can calculate  $\lambda_i^f(pq)$  and  $\lambda_{ij}^f(pq)$ . Therefore, we have:

$$\hat{\lambda}_{pq}^f = \lambda_{D_f}^f(pq). \quad (15)$$

We can calculate  $\hat{\lambda}_p^f$  in a similar manner.

Finally, the average end-to-end delay can then be calculated by using:

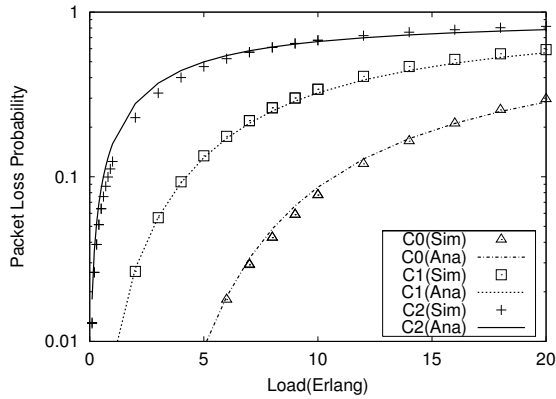
$$D = \frac{\sum_f D^f \cdot \lambda^f (1 - P^f)}{\sum_f \lambda^f (1 - P^f)}. \quad (16)$$

By using reduced load approximation, the above analysis can be applied to any loopless routing scheme. In order to analyze schemes which result in routing loops, the analysis must be modified slightly in order to avoid infinite path lengths. When calculating  $P_{ij}$  and  $\lambda_{ij}^f$ , the analysis will stop evaluating a path if the additional load on the next link in the path is less than some small value  $\epsilon$ . When evaluating the packet loss probabilities, the analysis will stop evaluating a path once it reaches a certain number of hops.

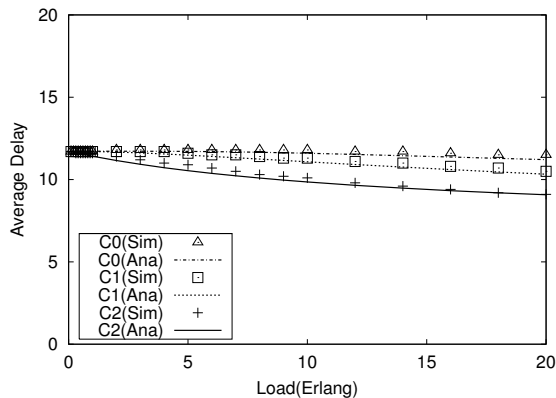
#### IV. NUMERICAL RESULTS

In this section, we evaluate the performance of different contention resolution schemes through simulation and analysis. A 14-node topology, illustrated in Fig. 2, is used to conduct the experiments. Each link represents a bi-directional channel with 10 Gb/s transmission rate, and the numbers on each link denote the physical length in units of 100 kilometers. We assume that the arrival of packets follows a Poisson process and that the length of each packet is 10 Kb. We further assume that the traffic is uniformly distributed over all source-destination pairs. For buffering we assume there are three recirculation buffers on each node, all with one unit of length.

Fig. 3 shows the simulation and analysis results for different buffering schemes, where we use SP as the routing algorithm. We assume that there are three classes of traffic in the network and that the total traffic load is distributed evenly among these three classes. We evaluate the three different buffering schemes for supporting these three classes of traffic. Specifically, we use control vector  $V_0 = [3, 3, \text{SP}, 1]$  to support Class 0 ( $C_0$ ) traffic; and use control vectors  $V_1 = [3, 1, \text{SP}, 1]$  and  $V_2 = [3, 0, \text{SP}, 1]$  for class 1 ( $C_1$ ) and class 2 ( $C_2$ ) traffic, respectively. Fig. 3 (a) illustrates packet loss versus load for the different schemes. We observe that, by limiting the number of times that a packet can



(a) loss



(b) average delay

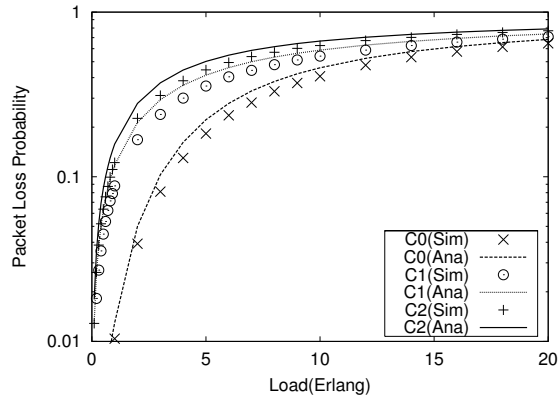
Fig. 3. Performance of different buffering schemes (SP routing).

enter a buffer, we can provide differentiated packet loss. As expected, a larger limit can lead to a smaller loss ratio. Fig. 3 (b) illustrates the delay performance of different schemes. We see that a larger buffer limit also leads to higher end-to-end delay. We observe that our buffering schemes have successfully provided differentiated service.

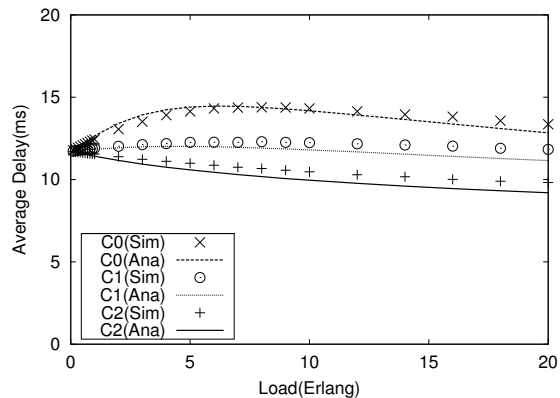
Fig. 4 compares the simulation and analysis results for different deflection routing schemes without buffers. We also assume that the total traffic load is distributed evenly among three classes of traffic. The corresponding control vectors for the label paths are  $V_0 = [0, 0, SPD, 2]$ ,  $V_1 = [0, 0, D-LLD, 2]$  and  $V_2 = [0, 0, SP, 1]$  respectively. Fig. 4 (a) shows the packet loss for the different schemes. It can be observed that  $C_0$  has the lowest packet loss rate; thus, it can be used to support loss-sensitive traffic that does not have strict delay requirements. Fig. 4 (b) shows the average delay versus load for the different schemes. We note that the average delay of  $C_1$  traffic remains stable over a large range of traffic loads; thus,  $C_1$  can be used to support traffic with guaranteed delay requirements. Finally,  $C_2$  can be used to support best effort traffic.

In addition to the experiments above, we also found that the performance could be further improved by combining the buffering and deflection routing schemes together. However, due to limited space, these results are not shown.

The above results also show that the packet loss and delay



(a) loss



(b) average delay

Fig. 4. Performance of different deflection routing schemes (no buffers).

analysis is highly accurate for all buffering and deflection routing schemes under different traffic loads.

## V. CONCLUSION

In this paper, we proposed a framework for differentiated contention resolution in photonic packet-switched networks. In this framework, different recirculation buffering and deflection routing schemes can be applied. An analytical model is also developed to evaluate the packet loss probability and the end-to-end delay for different schemes. Simulation and analysis results show that our scheme can provide sufficient options for supporting differentiated service requirements. The results also show that our analytical model is highly accurate.

## REFERENCES

- [1] R. Ramaswami, K. N. Sivarajan, *Optical Networks: A Practical Perspective*, 2nd Edition, Morgan Kaufmann Publishers, 2001.
- [2] J. M. H. Elmirghani and H. T. Mouftah, "All-Optical Wavelength Conversion: Technologies and Applications in DWDM Networks," *IEEE Comm. Mag.*, March 2000, pp. 86-92.
- [3] I. Chlamtac, A. Fumagalli, L. G. Kazovsky, et al., "CORD: Contention Resolution by Delay Lines," *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 5, pp. 1014-1029, June 1996.
- [4] A. S. Acampora and I. A. Shah, "Multihop Lightwave Networks: A Comparison of Store-and-Forward and Hot-Potato Routing," *IEEE Transactions on Communications*, vol. 40, no. 6, pp. 1082-1090, June 1992.
- [5] J. P. Jue, "An Algorithm for Loopless Deflection in Photonic Packet-Switched Networks," *Proceeding, IEEE ICC'02*, Vol. 5, pp. 2776-2780, New York, NY, USA, Apr. 2002.