# Constructions of Optical Priority Queues with Multiple Inputs and Multiple Outputs

Jay Cheng, *Senior Member, IEEE,* Hsien-Chen Chiu, Cheng-Shang Chang, *Fellow, IEEE,* and Duan-Shin Lee, *Senior Member, IEEE*

***Abstract*—In this paper, we consider the constructions of an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M} d_i$ by using a feedback system consisting of a single $(M+\max[N,K])\times(M+\max[N,K])$ (bufferless) optical crossbar switch, $\min[N,K]$ $1\times 2$ (bufferless) optical crossbar switches, and $M$ fiber delay lines with delays $d_1, d_2, \ldots, d_M$, where $N$ is the number of arrival links and $K$ is the number of departure links of the priority queue. We first obtain two sufficient conditions (the conditions (A1) and (A2) in Section I) for our constructions of $N$-to-$K$ optical priority queues. By establishing a *space-time advancement property* and a *monotonically decreasing/increasing property* for the packets stored in the fiber delay lines, we then use these sufficient conditions to show that with an appropriate choice for the delays $d_1, d_2, \ldots, d_M$, we can achieve a buffer size of $O(\frac{M^3}{N^2})$ for the case that $N = K$. For the special case that $N = K = 1$, our constructions achieve a buffer size of $O(M^3)$, which is much better than the $O(M^2)$ buffer size previously known in the literature for single-input single-output optical priority queues. Therefore, other than the extension from the constructions of optical priority queues with a single input and a single output to the constructions of optical priority queues with multiple inputs and multiple outputs, our constructions also achieve a larger buffer size than previous constructions of single-input single-output optical priority queues. Furthermore, we give another sufficient condition (the condition (A3) in Section I) for our constructions of $N$-to-$K$ optical priority queues and then use that condition to obtain choices for the delays $d_1, d_2, \ldots, d_M$ so that our constructions have the fault tolerant capability that can tolerate up to $F$ broken/malfunctioning fibers (e.g., fiber cut, fiber shorting out, etc), where $0 \le F \le M-1$.**

***Index Terms*—Fault tolerant capability, multiple inputs and multiple outputs (MIMO), optical buffers, optical queues, optical switches, priority queues, survivability, switched delay lines.**

## I. Introduction

One of the main problems in all-optical packet switching is the lack of optical buffers to resolve conflicts among packets competing for the same resources. Traditionally, such conflicts are resolved by first converting optical packets into electronic packets, storing them in electronic random-access memory (RAM), and then converting electronic packets back into optical packets when the resources can be accessed. However, such an optical-electrical-optical (O-E-O) approach incurs serious overheads so that it cannot keep up with the speed of the optical links and hence the tremendous bandwidth afforded by the optical links cannot be fully exploited. As the demand for the transmission speed/bandwidth is ever increasing, it has been well recognized that the design of optical buffers has become one of the most critically sought after optical technologies in all-optical packet-switched networks.

As optical packets, composed of a train of photons, cannot be easily stopped, stored, and forwarded, currently the only known way to "store" optical packets without converting them into other media is to direct them through a set of (bufferless) optical crossbar Switches and fiber Delay Lines (SDL) so that fiber delay lines are used as the storage devices to store optical packets. The main difference between such an optical buffer and an electronic RAM is that packets stored in such an optical buffer are constantly moving forward along the fiber delay lines instead of being stored at fixed positions as in an electronic RAM. Furthermore, an optical packet can only enter a fiber delay line in the optical buffer from one end of that optical fiber instead of from any position as in an electronic RAM, and it can only be accessed when it appears at the other end of that optical fiber instead of from any position as in an electronic RAM. As it takes time for an optical packet to travel through the optical fibers, such an optical buffer does not have the random-access capability.

Fortunately, one of the key observations in packet-switched networks is that we often do not need buffers with random-access capability for contention resolution. In packet-switched networks, many network elements have certain special arrival/departure patterns, and the key idea in the SDL constructions is to exploit such special patterns in the arrival/departure process of a network element to design a customized optical buffer for that network element. It has been shown in the SDL literature [1]–[45] that with appropriate choices for the delays of the fiber delay lines and appropriate designs for the connection patterns of the (bufferless) optical crossbar switches, optical packets can be routed to the right places at the right times and exact emulations of various types of network elements with certain special arrival/departure patterns can be achieved.

Note that in this paper, we adopt the following basic assumptions that are usually considered in the SDL literature: (i) Packets are of the same size; (ii) Time is slotted and

This work was supported in part by the National Science Council, Taiwan, R.O.C., under Contract NSC-95-2221-E-007-039, Contract NSC-95-2221-E-007-047-MY3, Contract NSC-96-2221-E-007-076, and the Program for Promoting Academic Excellence of Universities NSC-95-2752-E-007-002-PAE and NSC-96-2752-E-007-002-PAE. This paper was presented in part at the IEEE International Conference on Computer Communications (INFOCOM'07 Minisymposium), Anchorage, AK, USA, May 6–12, 2007.

Jay Cheng, Cheng-Shang Chang, and Duan-Shin Lee are with the Institute of Communications Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan, R.O.C. (e-mails: jcheng@ee.nthu.edu.tw; cschang@ee.nthu.edu.tw; lds@cs.nthu.edu.tw).

Hsien-Chen Chiu was with the Institute of Communications Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan, R.O.C. He is now with HTC Corporation, Taipei 23144, Taiwan, R.O.C. (e-mail: g935618@alumni.nthu.edu.tw).

synchronized so that a packet can be transmitted within a time slot; (iii) An $M \times M$ (bufferless) optical crossbar switch is a network element with $M$ input links and $M$ output links that realizes all of the $M!$ permutations between its inputs and its outputs; (iv) A fiber delay line with delay $d$ is an optical link that requires $d$ time slots for a packet to traverse through. As optical network elements are usually characterized by their arrival/departure processes, they can be viewed as special types of (discrete-time) optical queues. In the early works on the SDL constructions of optical queues [1]–[4], the focus is mainly on the feasibility of such an approach through numerical simulations rather than rigorous analytical studies. Recently, many interesting results on the theoretical SDL constructions of optical queues have appeared in the literature, including output-buffered switches in [5]–[10], FIFO multiplexers in [5] and [10]–[19], FIFO queues in [19]–[24], LIFO queues in [21]–[22], priority queues in [25]–[28], time slot interchanges in [19] and [29], and linear compressors, linear decompressors, non-overtaking delay lines, and flexible delay lines in [19] and [30]–[35], and FIFO, LIFO, and absolute contractors in [36]. Furthermore, results on the fundamental complexity of SDL constructions of optical queues can be found in [37] and performance analysis for optical queues has been addressed in [38]–[39]. For review articles on SDL constructions of optical queues, we refer to [40]–[45] and the references therein.

In this paper, we focus on the constructions of optical priority queues with multiple inputs and multiple outputs. In a priority queue, every packet is associated with a label, called *priority*. When the control input of the priority queue is enabled, the packets with the highest priorities are always the next ones to depart. When the buffer of the priority queue is full, the packets with the lowest priorities are always the next ones to be dumped. A formal definition of priority queues with multiple inputs and multiple outputs and its application to the implementation of optical output-buffered switches that support quality of service (QoS) will be given in Section II-A. Note that both FIFO queues and LIFO queues are special cases of priority queues as one can simply use the arrival time of a packet as its priority, i.e., the earliest arriving packet has the highest priority in a FIFO queue and the latest arriving packet has the highest priority in a LIFO queue. Therefore, the construction of an optical priority queue is considered to be much more difficult than that of an optical FIFO queue or LIFO queue.

The first construction of an optical priority queue with a single input and a single output was proposed by Sarwate and Anantharam [25]. In [25], they considered a feedback system (see Figure 1) consisting of an $(M+1)\times(M+1)$ (bufferless) optical crossbar switch, a $1 \times 2$ (bufferless) optical crossbar switch, and $M$ fiber delay lines with delays $d_1, d_2, \ldots, d_M$. If $M$ is an odd integer, say $M = 2k-1$ for some $k \geq 1$, $d_i = i$ for $i = 1, 2, \ldots, k$, and $d_i = 1$ for $i = k+1, k+2, \ldots, 2k-1$, then it was shown in [25] that such a feedback system can be used for exact emulation of a single-input single-output optical priority queue with buffer $\sum_{i=1}^{k} d_i = k(k+1)/2$. However, the proof in [25] is quite elaborate. A simpler proof was given in [26], and it was shown that if $M$ is an odd integer, say $M = 2k-1$ for some $k \geq 1$, then one can choose $d_i = i$ for $i = 1, 2, \ldots, k$ and $d_i = 2k-i$ for $i = k+1, k+2, \ldots, 2k-1$ for exact emulation of a single-input single-output optical priority queue with buffer $\sum_{i=1}^{M} d_i = k^2$. On the other hand, if $M$ is an even integer, say $M = 2k$ for some $k \geq 1$, then one can choose $d_i = i$ for $i = 1, 2, \ldots, k$ and $d_i = 2k+1-i$ for $i = k+1, k+2, \ldots, 2k$ for exact emulation of a single-input single-output optical priority queue with buffer $\sum_{i=1}^{M} d_i = k(k+1)$. We note that both constructions in [25] and [26] show that one can construct a single-input single-output optical priority queue with $O(M^2)$ buffer size by using the feedback system in Figure 1.

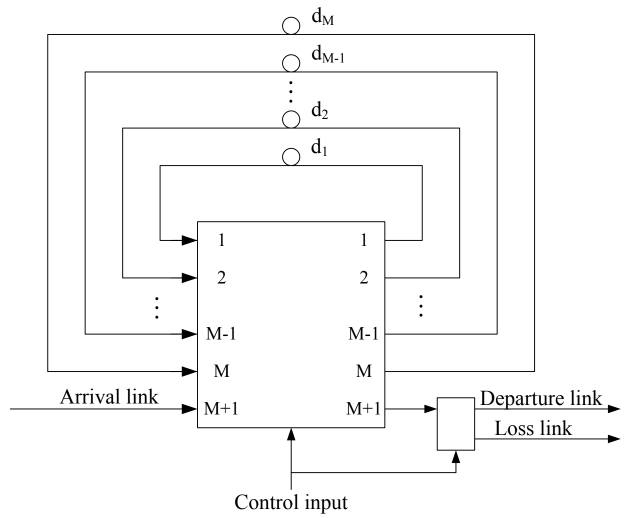


Fig. 1. A construction of an optical priority queue with a single input and a single output by using an $(M+1)\times(M+1)$ (bufferless) optical crossbar switch, a $1 \times 2$ (bufferless) optical crossbar switch, and $M$ fiber delay lines with delays $d_1, d_2, \ldots, d_M$.

Our main contribution in this paper is to show that by using a feedback system (see Figure 4(a) in Section II-B for the case that $N \geq K$ and Figure 5(a) in Section II-C for the case that $N < K$) consisting of a single [1] $(M+\max[N,K])\times(M+\max[N,K])$ (bufferless) optical crossbar switch, $\min[N,K]$ $1 \times 2$ (bufferless) optical crossbar switches, and $M$ fiber delay lines with appropriately chosen delays $d_1, d_2, \ldots, d_M$, we can construct an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M} d_i$, where $N$ is the number of arrival links and $K$ is the number of departure links of the priority queue. An in [26], the key idea in our constructions is to view empty time slots as the arrivals of *fictitious* packets, namely, if there is no packet arriving from an arrival link, then we regard that there is a fictitious packet arriving from that arrival link. Furthermore, we assume that fictitious packets have priorities lower than those of real packets, and the priorities among the fictitious packets are decreasing in the order of their arrival times and in the order of their arrival links (in the case of multiple fictitious packets with the same arrival time). Therefore, every packet in the queue has a *distinct* priority, meaning that we have a *total order* for the priorities of all of the packets (including

[1]Note that the sorter and the shifter in Figure 4(a) and Figure 5(a) can be combined together so that they can be implemented by using a single (bufferless) optical crossbar switch.

both the real packets and the fictitious packets) in the queue. We note that in the rest of this paper, a packet may refer to either a real packet or a fictitious packet, and it should be clear from the context whether a packet refers to a real packet or a fictitious packet.

We show in Theorem 2 (in Section II-B) and Theorem 3 (in Section II-C) that the construction in Figure 4(a) or Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M} d_i$ if the condition in (A1) below is satisfied at all times:

(A1) Any packet stored in the fiber delay lines in Figure 4(a) or Figure 5(a) cannot be either one of the $K$ *highest* priority packets or one of the $N$ *lowest* priority packets among all of the packets (including both the real packets and the fictitious packets) stored in the fiber delay lines until it appears at the first $M$ inputs of the sorter.

It is clear that if the condition in (A1) is satisfied at all times, then at the beginning of any time slot, the $K$ highest priority packets and the $N$ lowest priority packets in the queue will appear at the inputs of the sorter. The main idea in our constructions is to use the sorter to sort the packets at the inputs of the sorter according to their priorities so that the priorities of the packets at the outputs of the sorter are decreasing in the indices of the sorter's output links. Then the shifter and the $1 \times 2$ switches are used to route the $K$ highest priority packets to the $K$ departure links when the control input is enabled, and route the $N$ lowest priority packets to the $N$ loss links when the control input is disabled. By so doing, we show in the proofs of Theorem 2 and Theorem 3 that we achieve an exact emulation of an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M} d_i$.

In order to give appropriate choices for the fiber delays $d_1, d_2, \ldots, d_M$ in Figure 4(a) or Figure 5(a) so that the condition in (A1) is satisfied at all times, we show in Lemma 4 in Section III-A that the condition in (A2) below is a sufficient condition for the condition in (A1):

(A2) For $1 \le i \le M$, consider a packet that enters the $i^{\text{th}}$ fiber delay line at time $t$. Call this packet the tagged packet. There are at least $K(d_i - 1)$ packets stored in the $M$ fiber delay lines at time $t$ (at the end of the $t^{\text{th}}$ time slot) with priorities *higher* than that of the tagged packet, and there are at least $N(d_i - 1)$ packets in the $M$ fiber delay lines at time $t$ (at the end of the $t^{\text{th}}$ time slot) with priorities *lower* than that of the tagged packet.

Furthermore, we show that if the fiber delays $d_1, d_2, \ldots, d_M$ are given as in Theorem 5 in Section III-A, then the condition in (A2) is satisfied at all times, implying that the condition in (A1) is also satisfied at all times and hence the feedback system in Figure 4(a) or Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M} d_i$. Such choices of the fiber delays rely on establishing a *space-time advancement property* and a *monotonically decreasing/increasing property* for the packets stored in the fiber delay lines (see (35), (37), (57), and (58) in Appendix A) which are the key properties to showing that the condition in (A2) holds at all times if the fiber delays $d_1, d_2, \ldots, d_M$ are given as in Theorem 5.

In Section III-B, we find from numerical computations that in the case that $N = K$, the optimal choice of the value $m_1$ in Theorem 5 to maximize the buffer size is roughly $0.433M$ for large $M$, and the maximum buffer size is approximately $0.000929\frac{M^3}{N^2}$ for large $M$, i.e., the buffer size is $O(\frac{M^3}{N^2})$. These are further verified by approximating sums by integrals. Note that in the special case that $N = K = 1$, our constructions achieve a buffer size of $O(M^3)$, which is much better than the $O(M^2)$ buffer size previously obtained in [25] and [26]. As such, our constructions not only extend the previous constructions from a single input and a single output to multiple inputs and multiple outputs, but also improve (in the sense of achieving a larger buffer size) on the previous constructions of single-input single-output optical priority queues.

Another contribution of our constructions is their fault tolerant capability. Such a survivability issue is of great concern to a practical system designer as a system consisting of hundreds or thousands of components may be in a total breakdown even if only a single component fails to function properly. For this, we show in Lemma 7 in Section IV that if the fiber delays $d_1, d_2, \ldots, d_M$ satisfy the condition in (A3) below, then the condition in (A2) is satisfied at all times and hence the feedback system in Figure 4(a) or Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M} d_i$:

(A3) $1 \le d_i \le \min\left[\lfloor \frac{i-1}{K} \rfloor, \lfloor \frac{M-i}{N} \rfloor\right] + 1$, for all $i = 1, 2, \ldots, M$.

We use the condition in (A3) to show that if the fiber delays $d_1, d_2, \ldots, d_M$ are given as in Theorem 8 in Section IV, then the feedback system in Figure 4(a) or Figure 5(a) can still be operated as an $N$-to-$K$ optical priority queue even when up to $F$ of the $M$ fibers are broken/malfunctioning (e.g., fiber cut, fiber shorting out, etc), where $0 \le F \le M - 1$.

This paper is organized as follows. In Section II, we first give the definition of $N$-to-$K$ priority queues and show that optical priority queues with multiple inputs and multiple outputs can be used to implement optical output-buffered switches that support QoS. Then we show that the condition in (A1) is a sufficient condition for our constructions of $N$-to-$K$ optical priority queues in Figure 4(a) or Figure 5(a). In Section III, we first show that the condition in (A2) is a sufficient condition for the condition in (A1) and give choices for the fiber delays $d_1, d_2, \ldots, d_M$ in Figure 4(a) or Figure 5(a) so that the condition in (A2) holds at all times. Then we perform an approximation analysis for the maximum buffer size that could be achieved by our constructions. In Section IV, we present our constructions of fault tolerant $N$-to-$K$ optical priority queues. Finally, the paper is concluded in Section V, where we summarize our results.

In the following, we provide a list of notations used in the paper for easy reference.

$N$: the number of arrival links of an $N$-to-$K$ priority queue

$K$: the number of departure links of an $N$-to-$K$ priority queue

$B$: the buffer size of an $N$-to-$K$ priority queue in Definition 1

$M$: the number of fiber delay lines in our constructions of an $N$-to-$K$ priority queue in Figure 4(a) or Figure 5(a)

$d_i$: the delay of the $i^{\text{th}}$ fiber delay line in Figure 4(a) or Figure 5(a), where $1 \le i \le M$

$F$: the fault tolerant capability of a fault tolerant $N$-to-$K$ priority queue, where $0 \le F \le M-1$

$B^*(N,M)$: the maximum buffer size of an $N$-to-$K$ priority queue that can be achieved by using the constructions in Theorem 5 in the case that $N = K$

$B^*(N,K,F,M)$: the maximum buffer size of an $N$-to-$K$ priority queue with fault tolerant capability $F$ by using the constructions in Theorem 8(i)

$a(t)$: the set of the "real" packets arriving from the $N$ arrival links at time $t$

$d(t)$: the set of the "real" packets departing through the $K$ departure links at time $t$

$\ell(t)$: the set of the lost "real" packets dumped through the $N$ loss links at time $t$

$q(t)$: the set of the "real" packets stored in the buffer at time $t$ (at the end of the $t^{\text{th}}$ time slot)

$\tilde{a}(t)$: the set of both the "real" packets and the "fictitious" packets arriving from the $N$ arrival links at time $t$

$\tilde{d}(t)$: the set of both the "real" packets and the "fictitious" packets departing through the $K$ departure links at time $t$

$\tilde{\ell}(t)$: the set of both the lost "real" packets and the lost "fictitious" packets dumped through the $N$ loss links at time $t$

$\tilde{q}(t)$: the set of both the "real" packets and the "fictitious" packets stored in the buffer at time $t$ (at the end of the $t^{\text{th}}$ time slot)

## II. Constructions of Optical Priority Queues with Multiple Inputs and Multiple Outputs

We first give the definition of $N$-to-$K$ priority queues and show that optical priority queues with multiple inputs and multiple outputs can be used to implement optical output-buffered switches that support QoS in Section II-A. Then in Section II-B (resp., Section II-C), we show that the condition in (A1) is a sufficient condition for our constructions of $N$-to-$K$ optical priority queues in Figure 4(a) (resp., Figure 5(a)) for the case that $N \ge K$ (resp., $N < K$).

### *A. Definition of $N$-to-$K$ Priority Queues*

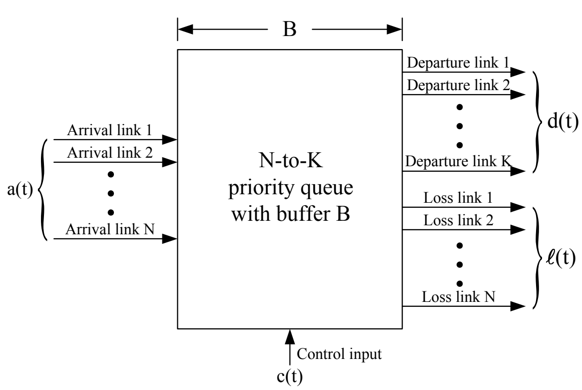


Fig. 2. An $N$-to-$K$ priority queue with buffer $B$.

In the following, we give a formal definition of $N$-to-$K$ priority queues. We note that all of the packets in this subsection refer to real packets.

**Definition 1** **($N$-to-$K$ Priority Queues)** *An $N$-to-$K$ priority queue with buffer $B$ is a network element with $N$ input links, one control input, and $N+K$ output links (see Figure 2). The $N$ input links are for arriving packets. Among the $N+K$ output links, $K$ of them are for departing packets and the other $N$ output links are for lost packets. When a packet arrives at the queue, it is associated with a label, called* priority. *We assume that there is a* total order *for the priorities of all of the packets in the queue (including the packets already stored in the buffer and the arriving packets from the arrival links), i.e., every packet in the queue has a* distinct *priority. As shown in Figure 2, let $c(t)$ be the state of the control input at time $t$. We say that the priority queue is enabled at time $t$ if $c(t) = 1$; otherwise, we say that the priority queue is disabled at time $t$ if $c(t) = 0$. Also, let $a(t)$ be the set of the packets arriving from the $N$ arrival links at time $t$* [2]*, $d(t)$ be the set of the packets departing through the $K$ departure links at time $t$, $\ell(t)$ be the set of the lost packets that are dumped through the $N$ loss links at time $t$, and $q(t)$ be the set of the packets stored in the buffer at time $t$ (at the end of the $t^{th}$ time slot). Then an $N$-to-$K$ priority queue with buffer $B$ satisfies the following five properties.*

(P1) *Flow conservation: Arriving packets from the $N$ arrival links are either stored in the buffer or transmitted through the $N+K$ output links, i.e.,*

$$q(t) = (q(t-1) \cup a(t)) \backslash (d(t) \cup \ell(t)). \tag{1}$$

(P2) *Nonidling: If the control input is enabled at time $t$, then there are always packets departing at time $t$ whenever there are packets in the queue at time $t$, i.e., whenever there are packets stored in the buffer at time $t-1$ or there are packets arriving from the $N$ arrival links at time $t$, under the constraint that there are at most $K$ departing packets at time $t$ as there are only $K$ departure links. Specifically, if $c(t) = 1$, then there are $\min[|q(t-1) \cup a(t)|, K]$ departing packets at time $t$; otherwise, if $c(t) = 0$, then there are no departing packets at time $t$. Thus, we have*

$$|d(t)| = \begin{cases} \min[|q(t-1) \cup a(t)|, K], & \textit{if } c(t) = 1, \\ 0, & \textit{if } c(t) = 0. \end{cases} \tag{2}$$

(P3) *Maximum buffer usage: There are lost packets only when the buffer is full. Specifically, if $|q(t-1) \cup a(t)| - |d(t)| > B$, then there are $|q(t-1) \cup a(t)| - |d(t)| - B$ lost packets at time $t$; otherwise, if $|q(t-1) \cup a(t)| - |d(t)| \le B$, then there are no lost packets at time $t$. Thus, we have*

$$|\ell(t)| = \max[|q(t-1) \cup a(t)| - |d(t)| - B, 0]. \tag{3}$$

[2]This means that $a(t)$ is an empty set when there are no packets arriving from the $N$ arrival links at time $t$, and $a(t)$ is a set consisting of $N$ arriving packets when there is a packet arriving from each of the $N$ arrival links at time $t$.

(P4) *Priority departure: If there are packets departing at time $t$, i.e., $|d(t)| > 0$, then the departing packets are the $|d(t)|$* highest *priority packets in $q(t-1) \cup a(t)$, and they depart from departure links $1, 2, \ldots, |d(t)|$ in the order of decreasing priorities.*

(P5) *Priority loss: If there are lost packets at time $t$, i.e., $|\ell(t)| > 0$, then the lost packets are the $|\ell(t)|$* lowest *priority packets in $q(t-1) \cup a(t)$, and they are dumped through loss links $1, 2, \ldots, |\ell(t)|$ in the order of decreasing priorities.*

We note that there is a control input in Figure 2 and its purpose is for enabling/disabling the $N$-to-$K$ priority queue. To see why we need a control input in Figure 2, consider the scenario that the $N$-to-$K$ priority queue shares the same resources, i.e., the usage of the $K$ departure links in Figure 2, with other network elements. If the resources are allocated to the $N$-to-$K$ priority queue, then the control input of the $N$-to-$K$ priority queue is enabled and the priority queue can send up to $K$ highest priority packets in the queue to the departure links. On the other hand, if the resources are allocated to other network elements, then the control input of the $N$-to-$K$ priority queue is disabled and no packets in the queue can be sent to the departure links. When the control input of the $N$-to-$K$ priority queue is disabled and there are packets arriving from the $N$ arrival links, the queue size of the priority queue will build up and we need buffer to store the packets in the queue no matter the number of arrival links $N$ is greater than, equal to, or less than the number of departure links $K$. In the case that the buffer of the priority queue is full, up to $N$ lowest priority packets in the queue are dumped through the loss links. As to when the $N$-to-$K$ priority queue will be allocated the resources for its use is a resources management issue that is more involved and is not the focus of the study in this paper. Furthermore, we note that the $K$ departure links in Figure 2 are either allocated to the $N$-to-$K$ priority queue altogether or not at all, so that they are not independently controlled and we only need one control input in Figure 2 to enable/disable the $K$ departure links.

One of the key applications of optical priority queues with multiple inputs and multiple outputs is to implement optical output-buffered switches that support QoS. For example, in Figure 3 we show a construction of a $4 \times 4$ optical output-buffered switch that can be used for implementing the packetized version of the generalized processor sharing (PGPS) policy in [46]. The PGPS policy is one of the most popular QoS schemes in the literature. In such a policy, every packet is assigned a virtual finishing time when it arrives, and packets are then scheduled according to their virtual finishing times.

In the construction in Figure 3, there are four 1-to-4 optical demultiplexers in the first stage and four 4-to-1 optical priority queues in the second stage. When a packet arrives at an input of the $4 \times 4$ optical output-buffered switch, it is routed via the corresponding 1-to-4 optical demultiplexer to one of the four optical priority queues according to its destination. In addition, we also compute the virtual finishing time outlined in [46] for that packet, and use the computed virtual finishing time as the priority of that packet. By enabling the control input of every

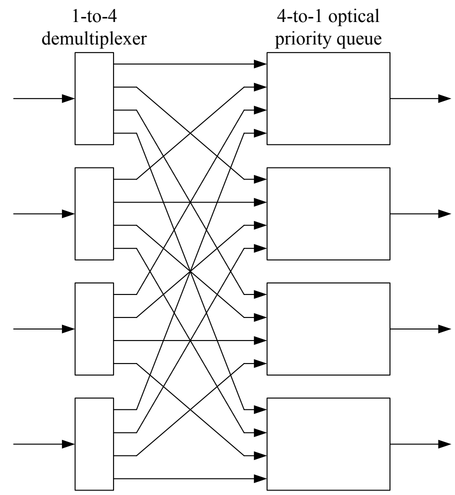


Fig. 3. An implementation of a $4 \times 4$ optical output-buffered switch that can support QoS by using four 1-to-4 optical demultiplexers in the first stage and four 4-to-1 optical priority queues in the second stage.

optical priority queue at all times, the packet with the highest priority, i.e., the packet with the earliest virtual finishing time, in each priority queue will depart from the priority queue. As such, we achieve exact emulation of an optical output-buffered switch that supports the PGPS policy. Note that if the virtual finishing times of the packets at the optical priority queue are increasing in the order of their arrival times and in the order of their arrival links (in the case of multiple packets with the same arrival time), then the construction in Figure 3 is the conventional optical FIFO output-buffered switches.

### *B. Constructions of an $N$-to-$K$ Optical Priority Queue With $N \geq K$*

As mentioned earlier in Section I that the key idea in our constructions of an $N$-to-$K$ optical priority queue is to view empty time slots as the arrivals of fictitious packets, and the priorities of the fictitious packets are assigned in such a way that there is a total order for the priorities of all of the packets (including both the real packets and the fictitious packets) in the queue. For ease of presentation in the rest of the paper, we let $\tilde{a}(t)$ be the set of the packets (including both the real packets and the fictitious packets) arriving from the $N$ arrival links at time $t$, $\tilde{d}(t)$ be the set of the packets (including both the real packets and the fictitious packets) departing through the $K$ departure links at time $t$, $\tilde{\ell}(t)$ be the set of the lost packets (including both the real packets and the fictitious packets) dumped through the $N$ loss links at time $t$, and $\tilde{q}(t)$ be the set of the packets (including both the real packets and the fictitious packets) stored in the buffer at time $t$ (at the end of the $t^{\text{th}}$ time slot). It is clear that the set $a(t)$ (resp., $d(t)$, $\ell(t)$, and $q(t)$) defined in Definition 1 is the subset of the real packets in $\tilde{a}(t)$ (resp., $\tilde{d}(t)$, $\tilde{\ell}(t)$, and $\tilde{q}(t)$).

In this subsection, we suppose that $N \geq K$. We will show in Theorem 2 below that the construction in Figure 4(a) can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M} d_i$ if the condition in (A1) given in Section I is satisfied at all times. (In Theorem 5 in Section III-A, we will give

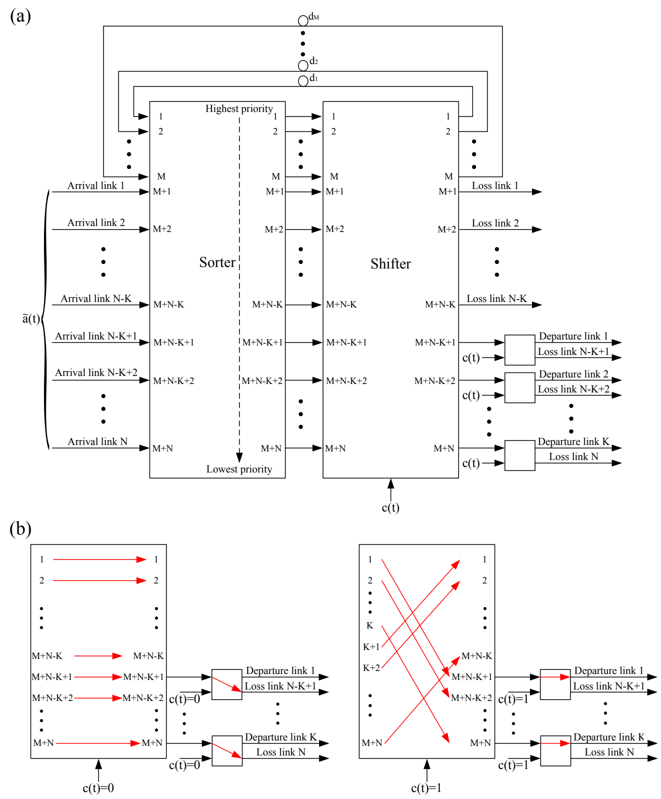


Fig. 4. (a) A construction of an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M} d_i$, where $N \geq K$. (b) The two possible connection patterns of the shifter and the $K$ $1\times 2$ switches in (a).

choices for the fiber delays $d_1, d_2, \ldots, d_M$ in Figure 4(a) so that the condition in (A1) is satisfied at all times.)

In Figure 4(a), there are two $(M+N)\times(M+N)$ (bufferless) optical crossbar switches (a sorter on the left-hand side and a shifter on the right-hand side) and $K$ $1\times 2$ (bufferless) optical crossbar switches (at the last $K$ outputs of the shifter). As mentioned in Section I that the main idea of our construction is to use the sorter to sort the packets (including both the real packets and the fictitious packets) at the inputs of the sorter according to their priorities so that the priorities of the packets at the outputs of the sorter are decreasing in the indices of the sorter's output links. Then the shifter and the $K$ $1\times 2$ switches are used to route the $K$ highest priority packets to the $K$ departure links when the control input is enabled, and route the $N$ lowest priority packets to the $N$ loss links when the control input is disabled.

Suppose that the condition in (A1) holds at all times. The details of the operation rules in our constructions are described as follows.

(i) As the condition in (A1) holds at all times, the set of the packets appearing at the first $M$ inputs of the sorter at time $t$ contains the $K$ highest priority packets and the $N$ lowest priority packets in $\tilde{q}(t-1)$. It follows that the set of the packets appearing at the $M+N$ inputs of the sorter at time $t$ contains the $K$ highest priority packets and the $N$ lowest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$. The sorter on the left-hand side in Figure 4(a) then sorts the packets at its $M+N$ inputs according to their priorities. Thus, the $K$ highest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ appear at output links $1, 2, \ldots, K$ of the sorter in the order of decreasing priorities and the $N$ lowest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ appear at output links $M+1, M+2, \ldots, M+N$ of the sorter in the order of decreasing priorities.

(ii) There are only two connection patterns (see Figure 4(b)) for the shifter on the right-hand side of Figure 4(a) and the $K$ $1\times 2$ switches at the last $K$ outputs of the shifter. If $c(t) = 0$, then the connection pattern of the shifter is realized by the $(M+N)\times(M+N)$ identity matrix, i.e., the matrix $I = (I_{i,j})$ with $I_{i,j} = 1$ for $i = j$ and $I_{i,j} = 0$ otherwise, and the input of the $i^{\text{th}}$ $1\times 2$ switch is connected to loss link $N-K+i$ of the $N$-to-$K$ optical priority queue for $i = 1, 2, \ldots, K$. It follows that there are no packets departing from the departure links and the $N$ lowest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ are dumped through loss links $1, 2, \ldots, N$ in the order of decreasing priorities. Therefore, we have $|\tilde{d}(t)| = 0$ and $|\tilde{\ell}(t)| = N$ when $c(t) = 0$. On the other hand, if $c(t) = 1$, then the connection pattern of the shifter is realized by the $(M+N)\times(M+N)$ $K$-circular-shift matrix, i.e., the matrix $P = (P_{i,j})$ with $P_{i,j} = 1$ for $j = ((i+M+N-K-1) \bmod (M+N))+1$ and $P_{i,j} = 0$ otherwise, and the input of the $i^{\text{th}}$ $1\times 2$ switch is connected to departure link $i$ of the $N$-to-$K$ priority queue for $i = 1, 2, \ldots, K$. It follows that the $K$ highest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ depart from departure links $1, 2, \ldots, K$ in the order of decreasing priorities, and the $N-K$ lowest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ are dumped through loss links $1, 2, \ldots, N-K$ in the order of decreasing priorities. Therefore, we have $|\tilde{d}(t)| = K$ and $|\tilde{\ell}(t)| = N-K$ when $c(t) = 1$. Furthermore, we note that in both the case $c(t) = 0$ and the case $c(t) = 1$, the priorities of the packets at the first $M$ outputs of the shifter are decreasing in their link indices.

**Theorem 2** *Suppose that $N \geq K$. If the feedback system in Figure 4(a) is started from an empty system at time 0 and the condition in (A1) holds at all times, then it can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M} d_i$ under the operation rules described before the statement of the theorem.*

**Proof.** Suppose that $N \geq K$ and the condition in (A1) holds at all times. Note that to emulate an empty system at time 0, we can store $\sum_{i=1}^{M} d_i$ fictitious packets in the fiber delay lines, i.e., $\tilde{q}(0) = \sum_{i=1}^{M} d_i$. By viewing empty time slots as the arrivals of fictitious packets, there are always $N$ packets arriving from the $N$ arrival links, i.e., $|\tilde{a}(t)| = N$ for all $t$. From the operation rules, we also see that there are always $N$ packets leaving the feedback system in Figure 4(a), i.e., $|\tilde{d}(t)\cup\tilde{\ell}(t)| = N$ for all $t$ (note that we have $|\tilde{d}(t)| = 0$ and $|\tilde{\ell}(t)| = N$ when $c(t) = 0$, and we have $|\tilde{d}(t)| = K$ and $|\tilde{\ell}(t)| = N-K$ when $c(t) = 1$). Therefore, there are always $\sum_{i=1}^{M} d_i$ packets stored in the fiber delay lines, i.e., $|\tilde{q}(t)| = \sum_{i=1}^{M} d_i$ for all $t$. Furthermore, it is clear that (P1) in Definition 1 is satisfied at all times.

In the rest of the proof, we consider the two cases $c(t) = 0$ and $c(t) = 1$ separately.

*Case 1:* $c(t) = 0$. As the condition in (A1) holds at all times, we see from the operation rules that there are no packets departing from the departure links and the $N$ lowest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ are dumped through loss links $1,2,\dots,N$ in the order of decreasing priorities, and we have $|\tilde{d}(t)|=0$ and $|\tilde{\ell}(t)|=N$. As $d(t)$ is the subset of the real packets in $\tilde{d}(t)$, we also have $|d(t)|=0$. It follows that (P2) and (P4) in Definition 1 are satisfied in this case. As $q(t-1)\cup a(t)$ is the subset of the real packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ and there are $\sum_{i=1}^M d_i+N$ packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ for all $t$, we see that there are $|q(t-1)\cup a(t)|$ real packets and $\sum_{i=1}^M d_i+N-|q(t-1)\cup a(t)|$ fictitious packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$. As $0\le|q(t-1)\cup a(t)|\le\sum_{i=1}^M d_i+N$, we then consider the following two subcases.

*Subcase (1a):* $0\le|q(t-1)\cup a(t)|\le\sum_{i=1}^M d_i$. In this subcase, we have $\sum_{i=1}^M d_i+N-|q(t-1)\cup a(t)|\ge N$, i.e., there are at least $N$ fictitious packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$, implying that the $N$ lowest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ must be fictitious packets. As $\tilde{\ell}(t)$ consists of the $N$ lowest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$, we see that all of the packets in $\tilde{\ell}(t)$ are fictitious packets. As $\ell(t)$ is the subset of the real packets in $\tilde{\ell}(t)$, it then follows that $\ell(t)$ is an empty set, i.e., $|\ell(t)|=0$, and hence we have from $|d(t)|=0$ and $|q(t-1)\cup a(t)|\le\sum_{i=1}^M d_i$ that

$$|\ell(t)|=0=\max\left[|q(t-1)\cup a(t)|-|d(t)|-\sum_{i=1}^M d_i,0\right].$$

Thus, (P3) and (P5) in Definition 1 are satisfied in this subcase.

*Subcase (1b):* $\sum_{i=1}^M d_i<|q(t-1)\cup a(t)|\le\sum_{i=1}^M d_i+N$. In this subcase, we have $\sum_{i=1}^M d_i+N-|q(t-1)\cup a(t)|<N$, i.e., there are less than $N$ fictitious packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$. As $\tilde{\ell}(t)$ consists of the $N$ lowest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$, we see that $\tilde{\ell}(t)$ consists of the $\sum_{i=1}^M d_i+N-|q(t-1)\cup a(t)|$ fictitious packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ and the $|q(t-1)\cup a(t)|-\sum_{i=1}^M d_i$ lowest priority real packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$. As $\ell(t)$ is the subset of the real packets in $\tilde{\ell}(t)$, it then follows that $\ell(t)$ consists of the $|q(t-1)\cup a(t)|-\sum_{i=1}^M d_i$ lowest priority packets in $q(t-1)\cup a(t)$, i.e., $|\ell(t)|=|q(t-1)\cup a(t)|-\sum_{i=1}^M d_i$, and the packets in $\ell(t)$ are dumped through loss links $1,2,\dots,|\ell(t)|$ in the order of decreasing priorities. From $|d(t)|=0$ and $|q(t-1)\cup a(t)|>\sum_{i=1}^M d_i$, we immediately see that

$$\begin{aligned}|\ell(t)|&=|q(t-1)\cup a(t)|-\sum_{i=1}^M d_i\\&=\max\left[|q(t-1)\cup a(t)|-|d(t)|-\sum_{i=1}^M d_i,0\right].\end{aligned}$$

Thus, (P3) and (P5) in Definition 1 are also satisfied in this subcase.

*Case 2:* $c(t)=1$. As the condition in (A1) holds at all times, we see from the operation rules that the $K$ highest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ depart from departure links $1,2,\dots,K$ in the order of decreasing priorities and the $N-K$ lowest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ are dumped through loss links $1,2,\dots,N-K$ in the order of decreasing priorities, and we have $|\tilde{d}(t)|=K$ and $|\tilde{\ell}(t)|=N-K$. As $0\le|q(t-1)\cup a(t)|\le\sum_{i=1}^M d_i+N$, we then consider the following three subcases.

*Subcase (2a):* $0\le|q(t-1)\cup a(t)|\le K$. In this subcase, there are no more than $K$ real packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$. As $\tilde{d}(t)$ consists of the $K$ highest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$, we see that $\tilde{d}(t)$ consists of the $|q(t-1)\cup a(t)|$ real packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ and the $K-|q(t-1)\cup a(t)|$ highest priority fictitious packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$. As $d(t)$ is the subset of the real packets in $\tilde{d}(t)$, it then follows that $d(t)$ consists of the $|q(t-1)\cup a(t)|$ packets in $q(t-1)\cup a(t)$, i.e., $|d(t)|=|q(t-1)\cup a(t)|$, and the packets in $d(t)$ depart from departure links $1,2,\dots,|d(t)|$ in the order of decreasing priorities. From $|q(t-1)\cup a(t)|\le K$, we immediately see that

$$|d(t)|=|q(t-1)\cup a(t)|=\min[|q(t-1)\cup a(t)|,K].$$

Thus, (P2) and (P4) in Definition 1 are satisfied in this subcase. Furthermore, as there are $\sum_{i=1}^M d_i+N-|q(t-1)\cup a(t)|\ge\sum_{i=1}^M d_i+N-K$ fictitious packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ and $\tilde{\ell}(t)$ consists of the $N-K$ lowest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$, we see that $\tilde{\ell}(t)$ consists of only fictitious packets. As such, $\ell(t)$ is an empty set, i.e., $|\ell(t)|=0$, and hence we have from $|d(t)|=|q(t-1)\cup a(t)|$ that

$$|\ell(t)|=0=\max\left[|q(t-1)\cup a(t)|-|d(t)|-\sum_{i=1}^M d_i,0\right].$$

Therefore, (P3) and (P5) in Definition 1 are satisfied in this subcase.

*Subcase (2b):* $K<|q(t-1)\cup a(t)|\le\sum_{i=1}^M d_i+K$. In this subcase, there are more than $K$ real packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$. As $\tilde{d}(t)$ consists of the $K$ highest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$, we see that $\tilde{d}(t)$ consists of the $K$ highest priority real packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$. It follows that $d(t)$ consists of the $K$ highest priority packets in $q(t-1)\cup a(t)$, i.e., $|d(t)|=K$, and the packets in $d(t)$ depart from departure links $1,2,\dots,K$ in the order of decreasing priorities. From $|q(t-1)\cup a(t)|>K$, we immediately see that

$$|d(t)|=K=\min[|q(t-1)\cup a(t)|,K].$$

Thus, (P2) and (P4) in Definition 1 are satisfied in this subcase. Furthermore, as there are $\sum_{i=1}^M d_i+N-|q(t-1)\cup a(t)|\ge N-K$ fictitious packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$, we also see that $\tilde{\ell}(t)$ consists of only fictitious packets as in Subcase (2a) above. As such, $\ell(t)$ is an empty set, i.e., $|\ell(t)|=0$, and hence we have from $|d(t)|=K$ and $|q(t-1)\cup a(t)|\le\sum_{i=1}^M d_i+K$ that

$$|\ell(t)|=0=\max\left[|q(t-1)\cup a(t)|-|d(t)|-\sum_{i=1}^M d_i,0\right].$$

Therefore, (P3) and (P5) in Definition 1 are also satisfied in this subcase.

*Subcase (2c):* $\sum_{i=1}^M d_i+K<|q(t-1)\cup a(t)|\le\sum_{i=1}^M d_i+N$. As in Subcase (2b) above, we also see that $d(t)$ consists of the $K$ highest priority packets in $q(t-1)\cup a(t)$, and the packets in $d(t)$ depart from departure links $1,2,\dots,K$ in the order of decreasing priorities. Thus, (P2) and (P4) in Definition 1 are also satisfied in this subcase. Furthermore, as there are $\sum_{i=1}^M d_i+N-|q(t-1)\cup a(t)|<N-K$ fictitious packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$, we see that $\tilde{\ell}(t)$ consists of the $\sum_{i=1}^M d_i+N-|q(t-1)\cup a(t)|$ fictitious packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$

and the $|q(t-1) \cup a(t)| - K - \sum_{i=1}^{M} d_i$ lowest priority real packets in $\tilde{q}(t-1) \cup \tilde{a}(t)$. It follows that $\ell(t)$ consists of the $|q(t-1) \cup a(t)| - K - \sum_{i=1}^{M} d_i$ lowest priority packets in $q(t-1) \cup a(t)$, i.e., $|\ell(t)| = |q(t-1) \cup a(t)| - K - \sum_{i=1}^{M} d_i$, and the packets in $\ell(t)$ are dumped through loss links $1, 2, \ldots, |\ell(t)|$ in the order of decreasing priorities. From $|d(t)| = K$ and $|q(t-1) \cup a(t)| > \sum_{i=1}^{M} d_i + K$, we see that

$$\begin{aligned} |\ell(t)| &= |q(t-1) \cup a(t)| - K - \sum_{i=1}^{M} d_i \\ &= \max\left[|q(t-1) \cup a(t)| - |d(t)| - \sum_{i=1}^{M} d_i, 0\right]. \end{aligned}$$

Thus, (P3) and (P5) in Definition 1 are also satisfied in this subcase. ■

### *C. Constructions of an $N$-to-$K$ Optical Priority Queue With $N < K$*

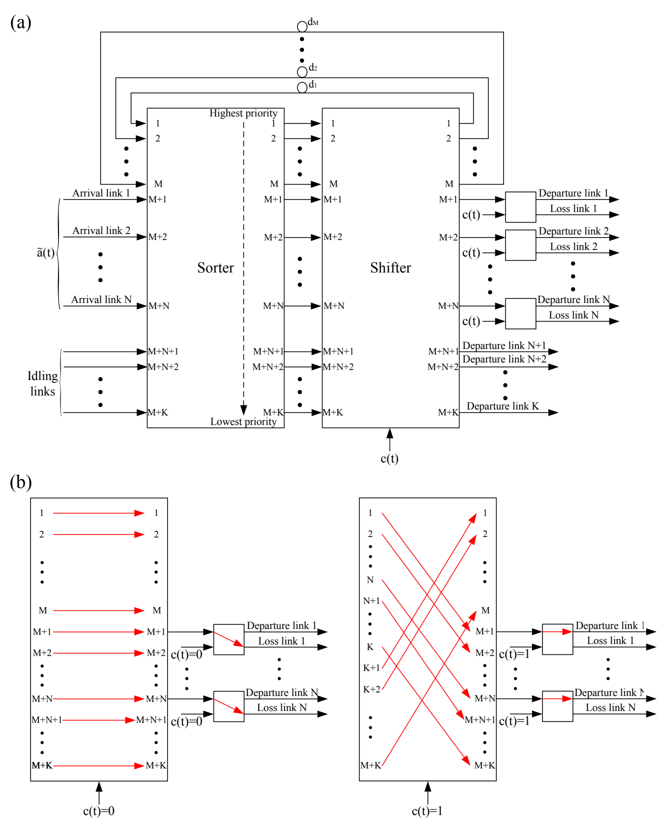


Fig. 5. (a) A construction of an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M} d_i$, where $N < K$. (b) The two possible connection patterns of the shifter and the $N$ $1 \times 2$ switches in (a).

In this subsection, we suppose that $N < K$. We will show in Theorem 3 below that the feedback system in Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M} d_i$ if the condition in (A1) given in Section I is satisfied at all times. (In Theorem 5 in Section III-A, we will give choices for the fiber delays $d_1, d_2, \ldots, d_M$ in Figure 5(a) so that the condition in (A1) is satisfied at all times.)

In Figure 5(a), there are two $(M+K) \times (M+K)$ (bufferless) optical crossbar switches (a sorter on the left-hand side and a shifter on the right-hand side) and $N$ $1 \times 2$ (bufferless) optical crossbar switches (at outputs $M+1, M+2, \ldots, M+N$ of the shifter). Note that there are $K-N$ idling links at the inputs of the sorter and we view these idling links at time $t$ as the arrivals of $K-N$ fictitious packets with priorities lower than those of the packets in $\tilde{q}(t-1) \cup \tilde{a}(t)$.

Suppose that the condition in (A1) holds at all times. The details of the operation rules in our constructions are described as follows.

(i) As the condition in (A1) holds at all times, the set of the packets appearing at the first $M+N$ inputs of the sorter at time $t$ contains the $K$ highest priority packets and the $N$ lowest priority packets in $\tilde{q}(t-1) \cup \tilde{a}(t)$ (note that the $K-N$ fictitious packets from the idling links are not contained in $\tilde{q}(t-1) \cup \tilde{a}(t)$). As in Section II-B, the sorter on the left-hand side of Figure 5(a) then sorts the packets at its $M+K$ inputs according to their priorities. Therefore, the $K$ highest priority packets in $\tilde{q}(t-1) \cup \tilde{a}(t)$ appear at output links $1, 2, \ldots, K$ of the sorter in the order of decreasing priorities, and the $N$ lowest priority packets in $\tilde{q}(t-1) \cup \tilde{a}(t)$ appear at output links $M+1, M+2, \ldots, M+N$ of the sorter in the order of decreasing priorities (note that the $K-N$ fictitious packets from the idling links appear at output links $M+N+1, M+N+2, \ldots, M+K$ of the sorter as they have priorities lower than those of the packets in $\tilde{q}(t-1) \cup \tilde{a}(t)$).

(ii) There are only two connection patterns (see Figure 5(b)) for the shifter on the right-hand side of Figure 5(a) and the $N$ $1 \times 2$ switches at outputs $M+1, M+2, \ldots, M+N$ of the shifter. If $c(t) = 0$, then the connection pattern of the shifter is realized by the $(M+K) \times (M+K)$ identity matrix, i.e., the matrix $I = (I_{i,j})$ with $I_{i,j} = 1$ for $i = j$ and $I_{i,j} = 0$ otherwise, and the input of the $i^{\text{th}}$ $1 \times 2$ switch is connected to loss link $i$ of the $N$-to-$K$ optical priority queue for $i = 1, 2, \ldots, N$. It follows that the $N$ lowest priority packets in $\tilde{q}(t-1) \cup \tilde{a}(t)$ are dumped through loss links $1, 2, \ldots, N$ in the order of decreasing priorities and the $K-N$ fictitious packets from the idling links depart from departure links $N+1, N+2, \ldots, K$ in the order of decreasing priorities. Therefore, we have $|\tilde{d}(t)| = K - N$ and $|\tilde{\ell}(t)| = N$ when $c(t) = 0$. On the other hand, if $c(t) = 1$, then the connection pattern of the shifter is realized by the $(M+K) \times (M+K)$ $K$-circular-shift matrix, i.e., the matrix $P = (P_{i,j})$ with $P_{i,j} = 1$ for $j = ((i + M - 1) \bmod (M+K)) + 1$ and $P_{i,j} = 0$ otherwise, and the input of the $i^{\text{th}}$ $1 \times 2$ switch is connected to departure link $i$ of the $N$-to-$K$ optical priority queue for $i = 1, 2, \ldots, N$. It follows that the $K$ highest priority packets in $\tilde{q}(t-1) \cup \tilde{a}(t)$ depart from departure links $1, 2, \ldots, K$ in the order of decreasing priorities and there are no packets dumped through the loss links. Therefore, we have $|\tilde{d}(t)| = K$ and $|\tilde{\ell}(t)| = 0$ when $c(t) = 1$. Furthermore, we note that in both the case $c(t) = 0$ and the case $c(t) = 1$, the priorities of the packets at the first $M$ outputs of the shifter are decreasing in their link indices.

**Theorem 3** *Suppose that $N < K$. If the feedback system in Figure 5(a) is started from an empty system at time 0 and the condition in (A1) holds at all times, then it can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M} d_i$*

*under the operation rules described before the statement of the theorem.*

**Proof.** Suppose that $N < K$ and the condition in (A1) holds at all times. As in the proof of Theorem 2, we store $\sum_{i=1}^M d_i$ fictitious packets in the fiber delay lines to emulate an empty system at time 0. Note that there are always $K$ packets arriving from the $N$ arrival links and the $K-N$ idling links. Also, we see from the operation rules that there are always $K$ packets leaving the feedback system in Figure 5(a), i.e., $|\tilde{d}(t)\cup\tilde{\ell}(t)| = K$ for all $t$ (note that we have $|\tilde{d}(t)| = K-N$ and $|\tilde{\ell}(t)| = N$ when $c(t) = 0$, and we have $|\tilde{d}(t)| = K$ and $|\tilde{\ell}(t)| = 0$ when $c(t) = 1$). Therefore, there are always $\sum_{i=1}^M d_i$ packets stored in the fiber delay lines, i.e., $|\tilde{q}(t)| = \sum_{i=1}^M d_i$ for all $t$. Furthermore, it is clear that (P1) in Definition 1 is satisfied at all times.

As in the proof of Theorem 2, we consider the two cases $c(t) = 0$ and $c(t) = 1$ separately.

*Case 1:* $c(t) = 0$. As the condition in (A1) holds at all times, we see from the operation rules that the $N$ lowest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ are dumped through loss links $1, 2, \ldots, N$ in the order of decreasing priorities and the $K-N$ fictitious packets from the idling links depart from departure links $N+1, N+2, \ldots, K$ in the order of decreasing priorities, and we have $|\tilde{d}(t)| = K-N$ and $|\tilde{\ell}(t)| = N$. It follows that $\tilde{d}(t)$ consists of only fictitious packets and hence $d(t)$ is an empty set, i.e., $|d(t)| = 0$. Thus, (P2) and (P4) in Definition 1 are satisfied in this case. The proof that (P3) and (P5) in Definition 1 are satisfied in this case is exactly the same as that of Subcase (1a) and Subcase (1b) in the proof of Theorem 2.

*Case 2:* $c(t) = 1$. As the condition in (A1) holds at all times, we see from the operation rules that the $K$ highest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ depart from departure links $1, 2, \ldots, K$ in the order of decreasing priorities and there are no packets dumped through the loss links, and we have $|\tilde{d}(t)| = K$ and $|\tilde{\ell}(t)| = 0$. Note that as $|\tilde{q}(t-1)\cup\tilde{a}(t)| = \sum_{i=1}^M d_i + N$ for all $t$ and $q(t-1)\cup a(t)$ is the subset of the real packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$, we have $0 \le |q(t-1)\cup a(t)| \le \sum_{i=1}^M d_i + N$. We then consider the following two subcases.

*Subcase (2a):* $0 \le |q(t-1)\cup a(t)| \le K$. As in Subcase 2(a) in the proof of Theorem 2, in this subcase we also see that $d(t)$ consists of the $|q(t-1)\cup a(t)|$ packets in $q(t-1)\cup a(t)$, i.e., $|d(t)| = |q(t-1)\cup a(t)|$, and the packets in $d(t)$ depart from departure links $1, 2, \ldots, |d(t)|$ in the order of decreasing priorities. Thus, (P2) and (P4) in Definition 1 are satisfied. Furthermore, $|\tilde{\ell}(t)| = 0$ implies that $|\ell(t)| = 0$, and hence it follows from $|d(t)| = |q(t-1)\cup a(t)|$ that

$$|\ell(t)| = 0 = \max\left[|q(t-1)\cup a(t)| - |d(t)| - \sum_{i=1}^M d_i, 0\right].$$

Therefore, (P3) and (P5) in Definition 1 are satisfied in this subcase.

*Subcase (2b):* $K < |q(t-1)\cup a(t)| \le \sum_{i=1}^M d_i + N$. As in Subcase 2(b) in the proof of Theorem 2, in this subcase we also have that $d(t)$ consists of the $K$ highest priority packets in $q(t-1)\cup a(t)$, i.e., $|d(t)| = K$, and the packets in $d(t)$ depart from departure links $1, 2, \ldots, K$ in the order of decreasing priorities. Thus, (P2) and (P4) in Definition 1 are satisfied. Furthermore, $|\tilde{\ell}(t)| = 0$ implies that $|\ell(t)| = 0$, and hence it follows from $|d(t)| = K$, $|q(t-1)\cup a(t)| \le \sum_{i=1}^M d_i + N$, and $N < K$ that

$$|\ell(t)| = 0 = \max\left[|q(t-1)\cup a(t)| - |d(t)| - \sum_{i=1}^M d_i, 0\right].$$

Therefore, (P3) and (P5) in Definition 1 are satisfied in this subcase. ■

Before we present the choices for the fiber delays $d_1, d_2, \ldots, d_M$ in Figure 4(a) and Figure 5(a) so that the condition in (A1) is satisfied at all times in Section III-A, we end this section by mentioning that each of the sorter and the shifter in Figure 4(a) or Figure 5(a) can be implemented by a single optical crossbar switch. The purpose of having two $(M+N)\times(M+N)$ optical crossbar switches in Figure 4(a) or two $(M+K)\times(M+K)$ optical crossbar switches in Figure 5(a) is for the ease of presentation. In practice, one can combine these two switches into one to reduce the hardware cost. Furthermore, if one would like to drop an arriving packet from an input link when the buffer is full, one can implement such an admission control scheme by simply adding a $1\times 2$ switch before that input link as in [20].

## III. Choices for the Fiber Delays and Maximum Buffer Size

In Section II, we have shown that if the condition in (A1) holds at all times, then the feedback system in Figure 4(a) or Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^M d_i$ under our operation rules. The remaining problem is how to choose the fiber delays $d_1, d_2, \ldots, d_M$ in Figure 4(a) or Figure 5(a) so that the condition in (A1) holds at all times. We give a solution to this problem in Theorem 5 in Section III-A. Then in Section III-B we present an approximation analysis for the maximum buffer size that could be achieved by using the constructions in Theorem 5.

### A. Choices for the Fiber Delays

We first show in the following lemma that the condition in (A2) given in Section I implies the condition in (A1), and hence it follows that the feedback system in Figure 4(a) or Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^M d_i$ under our operation rules if the condition in (A2) holds at all times.

**Lemma 4** *Suppose that the feedback system in Figure 4(a) or Figure 5(a) is started from an empty system at time 0. If the condition in (A2) holds at all times, then the condition in (A1) also holds at all times and hence the feedback system in Figure 4(a) or Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^M d_i$ under our operation rules.*

**Proof.** Suppose that the condition in (A2) holds at all times. As we emulate an empty system at time 0 by storing $\sum_{i=1}^M d_i$

fictitious packets in the fiber delay lines, the condition in (A1) holds at time $0$.

Assume as the induction hypothesis that the condition in (A1) holds up to time $t-1$ for some $t \geq 1$. From the operation rules described in Section II-B and Section II-C, it is clear that there are at most $K$ highest priority packets and at most $N$ lowest priority packets in $\tilde{q}(t-1) \cup \tilde{a}(t)$ leaving the feedback system in Figure 4(a) or Figure 5(a) at time $t$. Consider a packet that enters the $i^{\text{th}}$ fiber delay line at time $t$. Call this packet the tagged packet. Since (A2) holds at time $t$, the tagged packet cannot be either one of the $K$ highest priority packets or one of the $N$ lowest priority packets among all of the packets stored in the fiber delay lines for the next $d_i - 1$ time slots, implying that the tagged packet cannot be either one of the $K$ highest priority packets or one of the $N$ lowest priority packets among all of the packets stored in the fiber delay lines until it appears at the $i^{\text{th}}$ input of the sorter. It follows from the induction hypothesis that any packet stored in the fiber delay lines in Figure 4(a) or Figure 5(a) at time $t$ cannot be either one of the $K$ highest priority packets or one of the $N$ lowest priority packets among all of the packets stored in the fiber delay lines until it appears at the first $M$ inputs of the sorter. As such, the condition in (A1) holds at time $t$ and the induction is completed. ■

In the following theorem, we give choices for the fiber delays $d_1, d_2, \ldots, d_M$ so that the condition in (A2) holds at all times, and it then follows from Lemma 4 that the feedback system in Figure 4(a) or Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M} d_i$.

**Theorem 5** *Let $m_1 = nK$ and $m_2 = nN$ such that $m_1 + m_2 \leq M$, where $n \geq 1$, and let*

$$d_i = \left\lfloor \frac{i-1}{K} \right\rfloor + 1, \text{ for } i = 1, 2, \ldots, m_1, \tag{4}$$

$$d_i = \left\lfloor \frac{M-i}{N} \right\rfloor + 1, \quad \text{for } i = M - m_2 + 1, M - m_2 + 2, \ldots, M, \tag{5}$$

$$n \leq d_i \leq \min \left[ \left\lfloor \frac{\alpha_i - 1}{K} \right\rfloor, \left\lfloor \frac{\beta_i - 1}{N} \right\rfloor \right] + 1, \quad \text{for } i = m_1 + 1, m_1 + 2, \ldots, M - m_2, \tag{6}$$

*where*

$$\alpha_i = \begin{cases} i + \sum_{j=2}^{n} \left\lceil \left( \frac{i - M + 2m_1 - N - (4j-2)K}{2} \right)^{+} \right\rceil, & \text{if } N \geq K, \\ i + \sum_{j=2}^{n} \left\lceil \left( \frac{i - M + m_1 + m_2 - N - (4j-2)K}{2} \right)^{+} \right\rceil, & \text{if } N < K, \end{cases} \tag{7}$$

*and*

$$\beta_i = \begin{cases} M - i + 1 + \sum_{j=2}^{n} \left\lceil \left( \frac{m_1 + m_2 - i + 1 - K - (4j-2)N}{2} \right)^{+} \right\rceil, & \text{if } N \geq K, \\ M - i + 1 + \sum_{j=2}^{n} \left\lceil \left( \frac{2m_2 - i + 1 - K - (4j-2)N}{2} \right)^{+} \right\rceil, & \text{if } N < K. \end{cases} \tag{8}$$

*Suppose that the feedback system in Figure 4(a) or Figure 5(a) is started from an empty system at time 0. Then the feedback system in Figure 4(a) or Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M} d_i$ under our operation rules.*

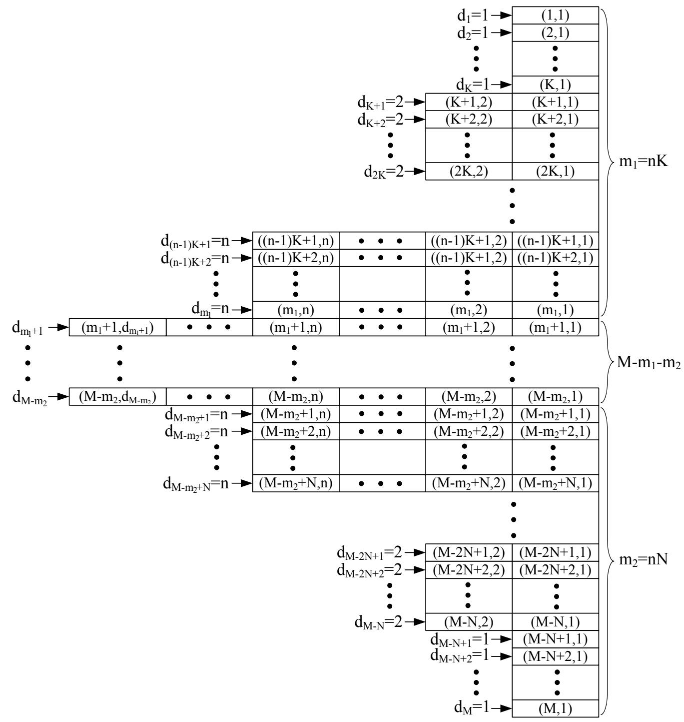


Fig. 6. The fiber delays given by (4)–(6) and the cells of the fiber delay lines.

In Figure 6, we show the fiber delays given by (4)–(6). The proof of Theorem 5 relies on establishing a *space-time advancement property* and a *monotonically decreasing/increasing property* for the packets stored in the fiber delay lines. For this, we view a fiber delay line with delay $d$ as a "sequential" buffer that consists of $d$ cells with each cell capable of holding one packet. The cells of the fiber delay lines are also shown in Figure 6. Note that we index the cells from the *input* of a fiber delay line in Figure 6. Specifically, the $(i,j)^{\text{th}}$ cell is the $j^{\text{th}}$ cell from the input of the $i^{\text{th}}$ fiber delay line for $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, d_i$. Note that since we view a fiber delay line as a sequential buffer, a packet entering the $i^{\text{th}}$ delay line at time $t$ will be stored in the $(i,j)^{\text{th}}$ cell at time $t + j - 1$ for $j = 1, 2, \ldots, d_i$, and can be accessed only when it appears at the *output* of the $i^{\text{th}}$ delay line at time $t + d_i$.

The reason why we divide the fiber delay lines into three sections, the section consisting of the first $m_1$ fiber delay lines (the first section), the section in between (the second section), and the section consisting of the last $m_2$ fiber delay lines (the third section), is as follows. By using the regular delay assignment for the fibers in the first section and the third section (there are $K$ fibers with delays equal to $i$ in the first section for $i = 1, 2, \ldots, n$ and there are $N$ fibers with delays equal to $i$ in the third section for $i = 1, 2, \ldots, n$) together with the space-time advancement property and the monotonically decreasing/increasing property developed in Appendix A for the packets stored in the fiber delay lines, we can obtain a lower bound on the total number of the packets stored in the cells of the first $n$ columns in Figure 6 with priorities higher/lower than that of a tagged packet when the tagged packet enters a fiber delay line (see (56) and (69) in Appendix A) if the delays of the fibers in the "in-between" section are greater than or equal to $n$. Such a lower bound allows the choices for the delays of the fibers in the in-between section to go beyond the regular delay assignment for the fibers in the first section and the third section, and it guarantees that the condition in (A2) is satisfied if the delays of the fibers in the "in-between" section are less than or equal to the upper bound in (6). It will be seen shortly in Section III-B that in the case that $N = K$, the increase of the delays of the fibers in the in-between section beyond the regular delay assignment as that in the first section and the third section makes it possible to increase the buffer size from $O(\frac{M^2}{N^2})$ for regular delay assignment to $O(\frac{M^3}{N^2})$ for the optimal delay assignment given by Theorem 5.

**Proof.** **(Proof of Theorem 5)** It is clear from Theorem 2, Theorem 3, and Lemma 4 that it suffices to show that the condition in (A2) holds at all times. For this, consider a packet from the $i_0^{\text{th}}$ output of the shifter that enters the $i_0^{\text{th}}$ fiber delay line at time $t_0$, where $1 \le i_0 \le M$. Call this packet the tagged packet. From the operation rules described in Section II-B and Section II-C, we know that the priorities of the packets at the first $M$ outputs of the shifter are decreasing in their link indices. Therefore, among the packets that enter the $M$ fiber delay lines at time $t_0$, there are $i_0 - 1$ packets with priorities higher than that of the tagged packet and there are $M - i_0$ packets with priorities lower than that of the tagged packet. We then consider the following three cases:

*Case 1.* $1 \le i_0 \le m_1$*:* In this case, we have from $i_0 \le m_1$, $M \ge m_1 + m_2$, $m_1 = nK$, and $m_2 = nN$ that

$$\begin{aligned}
&(i_0 - 1)N - (M - i_0)K \\
&= i_0(N + K) - MK - N \\
&\le m_1(N + K) - (m_1 + m_2)K - N \\
&= -N \le 0.
\end{aligned}$$

In other words, $\frac{i_0-1}{K} \le \frac{M-i_0}{N}$. It follows that

$$K(d_{i_0} - 1) = K \left\lfloor \frac{i_0 - 1}{K} \right\rfloor \le i_0 - 1, \tag{9}$$

$$N(d_{i_0} - 1) = N \left\lfloor \frac{i_0 - 1}{K} \right\rfloor \le N \left\lfloor \frac{M - i_0}{N} \right\rfloor \le M - i_0. \tag{10}$$

As we know that among the packets that enter the $M$ fiber delay lines at time $t_0$, there are $i_0 - 1$ (resp., $M - i_0$) packets with priorities higher (resp., lower) than that of the tagged packet, it follows from (9)–(10) that there are at least $K(d_{i_0} - 1)$ (resp., $N(d_{i_0} - 1)$) packets stored in the $M$ fiber delay lines at time $t_0$ with priorities higher (resp., lower) than that of the tagged packet. Therefore, the condition in (A2) holds at time $t_0$ in this case.

*Case 2.* $M - m_2 + 1 \le i_0 \le M$*:* In this case, we have from $i_0 \ge M - m_2 + 1$, $M \ge m_1 + m_2$, $m_1 = nK$, and $m_2 = nN$ that

$$\begin{aligned}
&(i_0 - 1)N - (M - i_0)K \\
&= i_0(N + K) - MK - N \\
&\ge (M - m_2 + 1)(N + K) - MK - N \\
&= MN - m_2(N + K) + K \\
&\ge (m_1 + m_2)N - m_2(N + K) + K \\
&= K \ge 0.
\end{aligned}$$

In other words, $\frac{i_0-1}{K} \ge \frac{M-i_0}{N}$. It follows that

$$K(d_{i_0} - 1) = K \left\lfloor \frac{M - i_0}{N} \right\rfloor \le K \left\lfloor \frac{i_0 - 1}{K} \right\rfloor \le i_0 - 1, \tag{11}$$

$$N(d_{i_0} - 1) = N \left\lfloor \frac{M - i_0}{N} \right\rfloor \le M - i_0. \tag{12}$$

Thus, we see from (11)–(12) that the condition in (A2) also holds at time $t_0$ in this case.

*Case 3.* $m_1 + 1 \le i_0 \le M - m_2$*:* In this case, the knowledge that among the packets that enter the $M$ fiber delay lines at time $t_0$, there are $i_0 - 1$ (resp., $M - i_0$) packets with priorities higher (resp., lower) than that of the tagged packet is not enough to prove that the condition in (A2) holds at time $t_0$ as we did in Case 1 and Case 2 above. Indeed, the proof for this case relies on establishing a space-time advancement property and a monotonically decreasing/increasing property for the packets stored in the fiber delay lines. By using the space-time advancement property and the monotonically decreasing/increasing property, we can show (see Appendix A for a proof) that the total number of packets stored in the cells of the first $n$ columns in Figure 6 at time $t_0$ with priorities higher/lower than that of the tagged packet is at least $\alpha_{i_0} - 1$ (resp., $\beta_{i_0} - 1$). This implies that the total number of packets stored in the fiber delay lines at time $t_0$ with priorities higher (resp., lower) than that of the tagged packet is also at least $\alpha_{i_0} - 1$ (resp., $\beta_{i_0} - 1$). Since it is clear that

$$\begin{aligned}
K(d_{i_0} - 1) &\le K \cdot \min\left[\left\lfloor \frac{\alpha_{i_0} - 1}{K} \right\rfloor, \left\lfloor \frac{\beta_{i_0} - 1}{N} \right\rfloor\right] \\
&\le K \left\lfloor \frac{\alpha_{i_0} - 1}{K} \right\rfloor \le \alpha_{i_0} - 1
\end{aligned}$$

and

$$\begin{aligned}
N(d_{i_0} - 1) &\le N \cdot \min\left[\left\lfloor \frac{\alpha_{i_0} - 1}{K} \right\rfloor, \left\lfloor \frac{\beta_{i_0} - 1}{N} \right\rfloor\right] \\
&\le N \left\lfloor \frac{\beta_{i_0} - 1}{N} \right\rfloor \le \beta_{i_0} - 1.
\end{aligned}$$

It then follows that the condition in (A2) also holds at time $t_0$ in this case, and the induction is completed. ■

### B. An Approximation Analysis for the Maximum Buffer Size

In this subsection, we present an approximation analysis for the maximum buffer size that could be achieved by using the constructions in Theorem 5. For the purpose of illustration and ease of presentation, we assume that $N = K$ in the following analysis. Therefore, the condition in Theorem 5 becomes $m_1 = m_2 = nN$, $n \geq 1$, $M \geq 2m_1$, and

$$d_i = d_{M-i+1} = \left\lfloor \frac{i-1}{N} \right\rfloor + 1, \text{ for } i = 1, 2, \ldots, m_1, \quad (13)$$

$$n \leq d_i = d_{M-i+1} \leq \left\lfloor \frac{\alpha_i - 1}{N} \right\rfloor + 1, \text{ for } i = m_1+1, m_1+2, \ldots, \left\lceil \frac{M}{2} \right\rceil, \quad (14)$$

where

$$\alpha_i = i + \sum_{j=2}^{n} \left\lceil \left( \frac{i - M + 2m_1 - (4j-1)N}{2} \right)^+ \right\rceil. \quad (15)$$

Among all of the choices of the fiber delays $d_1, d_2, \ldots, d_M$ that satisfy (13)–(14), the buffer size $\sum_{i=1}^{M} d_i$ achieves its maximum value when we choose the delays of the fiber in the in-between section to be given by the upper bound in (14), i.e.,

$$d_i = d_{M-i+1} = \left\lfloor \frac{i-1}{N} \right\rfloor + 1, \text{ for } i = 1, 2, \ldots, m_1, \quad (16)$$

$$d_i = d_{M-i+1} = \left\lfloor \frac{\alpha_i - 1}{N} \right\rfloor + 1, \text{ for } i = m_1+1, m_1+2, \ldots, \left\lceil \frac{M}{2} \right\rceil, \quad (17)$$

where $\alpha_i$ is given by (15) for $i = m_1+1, m_1+2, \ldots, \lceil \frac{M}{2} \rceil$. It follows that the maximum buffer size $B^*(N,M)$ that could be achieved by using the constructions in Theorem 5 in the case that $N = K$ is given by

$$B^*(N,M) = \sum_{i=1}^{M} d_i, \quad (18)$$

where $d_i$ is given by (16)–(17) for $i = 1, 2, \ldots, M$.

If we choose $n = \lceil \frac{M}{3N} \rceil$ so that $m_1 = N\lceil \frac{M}{3N} \rceil$, then we see from (17) and (15) that $B^*(N,M)$ can be lower bounded by

$$\begin{aligned}
&B^*(N,M) \\
&\geq \sum_{i=m_1+1}^{\lceil M/2 \rceil} \left( \left\lfloor \frac{\alpha_i - 1}{N} \right\rfloor + 1 \right) \\
&\geq \sum_{i=m_1+1}^{\lceil M/2 \rceil} \frac{\alpha_i - 1}{N} \\
&\geq \frac{1}{N} \sum_{i=N\lceil M/(3N) \rceil + 1}^{\lceil M/2 \rceil} \sum_{j=2}^{\lceil M/(3N) \rceil} \left\lceil \left( \frac{i - M + 2N\lceil \frac{M}{3N} \rceil - (4j-1)N}{2} \right)^+ \right\rceil \\
&\geq \frac{1}{N} \sum_{i=N\lceil 3M/(8N) \rceil + 1}^{\lceil M/2 \rceil} \sum_{j=2}^{\lfloor M/(192N) \rfloor} \left( \frac{i - M + 2N\lceil \frac{M}{3N} \rceil - (4j-1)N}{2} \right)^+ \\
&\geq \frac{1}{N} \sum_{i=N\lceil 3M/(8N) \rceil + 1}^{\lceil M/2 \rceil} \sum_{j=2}^{\lfloor M/(192N) \rfloor} \left( \frac{N\lceil \frac{3M}{8N} \rceil + 1 - M + 2N\lceil \frac{M}{3N} \rceil - (4\lfloor \frac{M}{192N} \rfloor - 1)N}{2} \right)^+ \\
&\geq \frac{1}{N} \sum_{i=N\lceil 3M/(8N) \rceil + 1}^{\lceil M/2 \rceil} \sum_{j=2}^{\lceil M/(192N) \rceil} \frac{M}{96} \\
&\approx \frac{M^3}{147456N^2}.
\end{aligned}$$

This shows that the buffer size that could be achieved by choosing $m_1 = N\lceil \frac{M}{3N} \rceil \approx 0.333M$ is at least $O(M^3/N^2)$ for large $M$. From the numerical results and approximation analysis below, it can be seen that this lower bound is of the same order as the buffer size achieved by choosing the optimal value of $m_1$.

Note that for sufficiently large $m_1$ and $M$, we can approximate (17) as follows (note that the expression is exact when $N = 1$):

$$\begin{aligned}
d_i &= d_{M-i+1} \\
&\approx \left\lfloor \frac{i-1}{N} \right\rfloor + 1 \\
&\quad + \frac{1}{N} \sum_{j=2}^{n} \left\lceil \left( \frac{i - M + 2m_1 - (4j-1)N}{2} \right)^+ \right\rceil. \quad (19)
\end{aligned}$$

Suppose that $M$ is even, then $B^*(N,M)$ can be approximated by

$$\begin{aligned}
&B^*(N,M) \\
&\approx \sum_{i=1}^{M/2} 2 \left( \left\lfloor \frac{i-1}{N} \right\rfloor + 1 \right) \\
&\quad + \sum_{i=m_1+1}^{M/2} \sum_{j=2}^{n} \frac{2}{N} \left\lceil \left( \frac{i - M + 2m_1 - (4j-1)N}{2} \right)^+ \right\rceil \quad (20)
\end{aligned}$$

Denote the first term and the second term on the right-hand side of (20) as $B_1^*(N,M)$ and $B_2^*(N,M)$, respectively. It is clear that $B_1^*(N,M)$ is independent of the choice of $m_1$, but $B_2^*(N,M)$ depends on the choice of $m_1$.

If $M = q \cdot 2N + r$, where $q \geq 0$ and $0 \leq r \leq 2N - 1$ are the quotient and the remainder, respectively, of $M$ divided by $2N$. Then we can see that

$$B_1^*(N,M) = q(q+1)N + r(q+1) \approx \frac{M^2}{4N} \quad (21)$$

for large $M$. In Table I, Table II, and Table III, we numerically compute $B_1^*(N,M)$ (in the second column) and the maximum value of $B_2^*(N,M)$ (in the third column) that is obtained by using the optimal choice $m_1^*$ (in the first column) of the value $m_1$ for $N = 1$, $N = 2$, and $N = 4$, respectively. It is interesting to see from these tables that

$$m_1^* \approx 0.433M \text{ and } B_2^*(N,M) \approx 0.000929 \frac{M^3}{N^2} \quad (22)$$

for large $M$. In other words, the optimal choice $m_1^*$ of the value $m_1$ is roughly $0.433M$ and the maximum value of $B_2^*(N,M)$ is approximately $0.000929 \frac{M^3}{N^2}$ for large $M$.

| $M$ | $m_1^*$ | $B_1^*(N,M)$ | $B_2^*(N,M)$ | $B^*(N,M)/B_1^*(N,M)$ |
|---|---|---|---|---|
| 4 | $2 = 0.5M$ | $6 = M^2/4N + M/2$ | 0 | 1 |
| 8 | $4 = 0.5M$ | $20 = M^2/4N + M/2$ | 0 | 1 |
| 16 | $8 = 0.5M$ | $72 = M^2/4N + M/2$ | 0 | 1 |
| 32 | $15 \approx 0.469M$ | $272 = M^2/4N + M/2$ | $12 \approx 0.000366M^3/N^2$ | 1.044 |
| 64 | $28 \approx 0.438M$ | $1056 = M^2/4N + M/2$ | $162 \approx 0.000618M^3/N^2$ | 1.153 |
| 128 | $56 \approx 0.438M$ | $4160 = M^2/4N + M/2$ | $1604 \approx 0.000765M^3/N^2$ | 1.386 |
| 256 | $111 \approx 0.434M$ | $16512 = M^2/4N + M/2$ | $14172 \approx 0.000845M^3/N^2$ | 1.858 |
| 512 | $222 \approx 0.434M$ | $65792 = M^2/4N + M/2$ | $118932 \approx 0.000886M^3/N^2$ | 2.808 |
| 1024 | $444 \approx 0.434M$ | $262656 = M^2/4N + M/2$ | $974338 \approx 0.000907M^3/N^2$ | 4.710 |
| 2048 | $887 \approx 0.433M$ | $1049600 = M^2/4N + M/2$ | $7887320 \approx 0.000918M^3/N^2$ | 8.515 |
| 4096 | $1773 \approx 0.433M$ | $4196352 = M^2/4N + M/2$ | $63469912 \approx 0.000924M^3/N^2$ | 16.125 |
| 8192 | $3547 \approx 0.433M$ | $16781312 = M^2/4N + M/2$ | $509252408 \approx 0.000926M^3/N^2$ | 31.346 |
| 16384 | $7093 \approx 0.433M$ | $67117056 = M^2/4N + M/2$ | $4079995388 \approx 0.000928M^3/N^2$ | 61.789 |

TABLE I
THE MAXIMUM BUFFER SIZE BY THE OPTIMAL CHOICE $m_1^*$ FOR $N = 1$.

| $M$ | $m_1^*$ | $B_1^*(N,M)$ | $B_2^*(N,M)$ | $B^*(N,M)/B_1^*(N,M)$ |
|---|---|---|---|---|
| 8 | $4 = 0.5M$ | $12 = M^2/4N + M/2$ | 0 | 1 |
| 16 | $8 = 0.5M$ | $40 = M^2/4N + M/2$ | 0 | 1 |
| 32 | $16 = 0.5M$ | $144 = M^2/4N + M/2$ | 0 | 1 |
| 64 | $30 \approx 0.469M$ | $544 = M^2/4N + M/2$ | $20 \approx 0.000305M^3/N^2$ | 1.037 |
| 128 | $56 \approx 0.438M$ | $2112 = M^2/4N + M/2$ | $306 \approx 0.000583M^3/N^2$ | 1.145 |
| 256 | $112 \approx 0.438M$ | $8320 = M^2/4N + M/2$ | $3128 \approx 0.000746M^3/N^2$ | 1.376 |
| 512 | $222 \approx 0.434M$ | $33024 = M^2/4N + M/2$ | $27980 \approx 0.000834M^3/N^2$ | 1.847 |
| 1024 | $444 \approx 0.434M$ | $131584 = M^2/4N + M/2$ | $236444 \approx 0.000881M^3/N^2$ | 2.797 |
| 2048 | $888 \approx 0.434M$ | $525312 = M^2/4N + M/2$ | $1942930 \approx 0.000905M^3/N^2$ | 4.699 |
| 4096 | $1774 \approx 0.433M$ | $2099200 = M^2/4N + M/2$ | $15751284 \approx 0.000917M^3/N^2$ | 8.504 |
| 8192 | $3548 \approx 0.433M$ | $8392704 = M^2/4N + M/2$ | $126846724 \approx 0.000923M^3/N^2$ | 16.114 |
| 16384 | $7094 \approx 0.433M$ | $33562624 = M^2/4N + M/2$ | $1018130954 \approx 0.000926M^3/N^2$ | 31.335 |

TABLE II
THE MAXIMUM BUFFER SIZE BY THE OPTIMAL CHOICE $m_1^*$ FOR $N = 2$.

| $M$ | $m_1^*$ | $B_1^*(N,M)$ | $B_2^*(N,M)$ | $B^*(N,M)/B_1^*(N,M)$ |
|---|---|---|---|---|
| 16 | $8 = 0.5M$ | $24 = M^2/4N + M/2$ | 0 | 1 |
| 32 | $16 = 0.5M$ | $80 = M^2/4N + M/2$ | 0 | 1 |
| 64 | $32 = 0.5M$ | $288 = M^2/4N + M/2$ | 0 | 1 |
| 128 | $60 \approx 0.469M$ | $1088 = M^2/4N + M/2$ | $38 \approx 0.000290M^3/N^2$ | 1.035 |
| 256 | $112 \approx 0.438M$ | $4224 = M^2/4N + M/2$ | $595 \approx 0.000567M^3/N^2$ | 1.141 |
| 512 | $224 \approx 0.438M$ | $16640 = M^2/4N + M/2$ | $6178 \approx 0.000736M^3/N^2$ | 1.371 |
| 1024 | $444 \approx 0.434M$ | $66048 = M^2/4N + M/2$ | $55618 \approx 0.000829M^3/N^2$ | 1.842 |
| 2048 | $888 \approx 0.434M$ | $263168 = M^2/4N + M/2$ | $471477 \approx 0.000878M^3/N^2$ | 2.792 |
| 4096 | $1776 \approx 0.434M$ | $1050624 = M^2/4N + M/2$ | $3880131 \approx 0.000903M^3/N^2$ | 4.693 |
| 8192 | $3548 \approx 0.433M$ | $4198400 = M^2/4N + M/2$ | $31479398 \approx 0.000916M^3/N^2$ | 8.498 |
| 16384 | $7096 \approx 0.433M$ | $16785408 = M^2/4N + M/2$ | $253600425 \approx 0.000923M^3/N^2$ | 16.108 |

TABLE III
THE MAXIMUM BUFFER SIZE BY THE OPTIMAL CHOICE $m_1^*$ FOR $N = 4$.

To see this, we replace $m_1$ by $\alpha M$ with $\frac{1}{3} \le \alpha \le \frac{1}{2}$, $j\frac{N}{M}$ by $x$, and $\frac{i}{M}$ by $y$, then the double sum in $B_2^*(N,M)$ can be approximated by the following double integral

$$\begin{aligned}
&B_2^*(N,M) \\
&\approx \frac{M^3}{N^2}\int_\alpha^{\frac{1}{2}}\int_0^\alpha (y-1+2\alpha-4x)^+ dxdy \\
&= \frac{M^3}{N^2}\int_\alpha^{\frac{1}{2}}\int_0^{(y-1+2\alpha)/4} (y-1+2\alpha-4x)\ dxdy \\
&= \frac{M^3}{8N^2}\int_{3\alpha-1}^{2\alpha-\frac{1}{2}} z^2 dz \\
&= \frac{M^3}{24N^2}\left(\left(2\alpha-\frac{1}{2}\right)^3-(3\alpha-1)^3\right), \qquad (23)
\end{aligned}$$

where the first equality holds due to the restriction that $\alpha \ge \frac{1}{3}$ and in the second equality we have used the change of variable $z = y-1+2\alpha$. The optimal value $\alpha^*$ that maximizes (23) for $\alpha$ in $[\frac{1}{3}, \frac{1}{2}]$ can be obtained by solving the following quadratic equation:

$$2\left(2\alpha^*-\frac{1}{2}\right)^2 - 3(3\alpha^*-1)^2 = 0.$$

The result is

$$\alpha^* = \frac{14+\sqrt{6}}{38} \approx 0.433.$$

By using the optimal value $\alpha^* \approx 0.433$ for $B_2^*(N,M)$ in (23), we see that

$$B_2^*(N,M) \approx 0.000929 \frac{M^3}{N^2} \tag{24}$$

for large $M$. Therefore, we see from (21) and (24) that the maximum buffer size that could be achieved by using the constructions in Theorem 5 in the case that $N = K$ is

$$B^*(N,M) = B_1^*(N,M) + B_2^*(N,M) \approx 0.000929 \frac{M^3}{N^2} \tag{25}$$

for large $M$.

Note that $B_1^*(N,M) = O(\frac{M^2}{N^2})$ is the buffer size achieved by the regular delay assignment $d_i = d_{M-i+1} = \lfloor \frac{i-1}{N} \rfloor + 1$ for $i = 1, 2, \ldots, \lceil \frac{M}{2} \rceil$, and $B^*(N,M) = O(\frac{M^3}{N^2})$ is the buffer size achieved by the optimal delay assignment that could be given by (16)–(17). The improvement $B^*(N,M)/B_1^*(N,M)$ of the buffer size $B^*(N,M)$ over the buffer size $B_1^*(N,M)$ is shown in the last columns of Table I–Table III. In Table I–Table III, $M$ is a multiple of $2N$ and hence we have $B_1^*(N,M) = \frac{M^2}{4N} + \frac{M}{2}$. We make the following observations: (i) The improvement $B^*(N,M)/B_1^*(N,M)$ is small when $M$ is small (note that we have $B^*(N,M)/B_1^*(N,M) < 2$ for $\frac{M}{N} \leq 256$); (ii) The improvement $B^*(N,M)/B_1^*(N,M)$ increases linearly with $M$ when $\frac{M}{N}$ is sufficiently large.

When $N = K = 1$, the maximum buffer size achieved in [26] is $B_1^*(1,M) = M^2/4 + M/2$ and the maximum buffer size achieved by using the constructions in Theorem 5 is $B^*(1,M)$. The results in Table I show that the buffer size $B^*(1,M)$ is larger than $B_1^*(1,M)$ for $M \geq 32$, and the improvement $B^*(1,M)/B_1^*(1,M)$ increases linearly with $M$ when $M$ is large (as $B^*(1,M) = O(M^3)$ and $B_1^*(1,M) = O(M^2)$). Therefore, our constructions not only extend the previous constructions from a single input and a single output to multiple inputs and multiple outputs, but also improve on the previous constructions [25]–[26] of single-input single-output optical priority queues.

In the following theorem, we show an upper bound on the maximum buffer size of an $N$-to-$K$ priority queue by using fiber delay lines as the storage devices.

**Theorem 6** *Suppose that an $N$-to-$K$ priority queue with buffer $B$ is constructed by using SDL elements that contain $M$ fiber delay lines as the storage devices. Then we have*

$$B \leq (K^2 + 2K + N) 2^{(M-N-K)\log_2(1+\frac{1}{K})}. \tag{26}$$

*In particular, if $N = K$, then we have*

$$B \leq (N^2 + 3N) 2^{(M-2N)\log_2(1+\frac{1}{N})}. \tag{27}$$

**Proof.** See Appendix B. ■

Note that for the special case that $N = K = 1$, the exponential upper bound given by (27) is $O(2^M)$, which is the same as that obtained in [25]. Furthermore, there is a gap between the $O(\frac{M^3}{N^2})$ buffer size achieved by our constructions and the exponential upper bound $O(2^{(M-2N)\log_2(1+\frac{1}{N})})$ given in (27). Whether it is possible to achieve such an exponential bound, and if possible, how to do that, remains an open research problem.

## IV. Constructions of Fault Tolerant Optical Priority Queues with Multiple Inputs and Multiple Outputs

In Section III-A, we have shown that if the fiber delays $d_1, d_2, \ldots, d_M$ are given as in Theorem 5, then the feedback system in Figure 4(a) or Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^M d_i$ under our operation rules. The problem with such a construction is its fault tolerant capability. When some of the fiber delays lines are broken/malfunctioning (e.g., fiber cut, fiber shorting out, etc), the delays of the remaining unbroken/functioning fibers may not satisfy the condition in Theorem 5, and hence the construction in Theorem 5 no longer works. An interesting and challenging question is then: when some of the fiber delays lines are broken/malfunctioning, could the feedback system in Figure 4(a) or Figure 5(a) still be operated as an $N$-to-$K$ optical priority queue with a "smaller" buffer size by using the remaining unbroken/functioning fibers? The answer is affirmative if the fiber delays are carefully chosen as will be seen in Theorem 8 below.

Before we state Theorem 8, we first show in the following lemma that if the fiber delays $d_1, d_2, \ldots, d_M$ satisfy the condition in (A3) in Section I, then the condition in (A2) is also satisfied, and hence it follows from Theorem 2, Theorem 3, and Lemma 4 that the feedback system in Figure 4(a) or Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^M d_i$.

**Lemma 7** *Suppose that the feedback system in Figure 4(a) or Figure 5(a) is started from an empty system at time 0. If the fiber delays $d_1, d_2, \ldots, d_M$ satisfy the condition in (A3), then the feedback system in Figure 4(a) or Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^M d_i$ under our operation rules.*

**Proof.** As commented above, it suffices to show that the condition in (A2) holds at all times. Consider a tagged packet that enters the $i^{\text{th}}$ fiber delay line at time $t$. Under our operation rules, we know that among the packets that enter the $M$ fiber delay lines at time $t$, there are $i-1$ packets with priorities higher than that of the tagged packet and there are $M-i$ packets with priorities lower than that of the tagged packet. If the fiber delays $d_1, d_2, \ldots, d_M$ satisfy the condition in (A3), then we can see that

$$\begin{aligned} K(d_i - 1) &\leq K \min\left[\left\lfloor \frac{i-1}{K} \right\rfloor, \left\lfloor \frac{M-i}{N} \right\rfloor\right] \\ &\leq K \left\lfloor \frac{i-1}{K} \right\rfloor \leq i-1 \end{aligned}$$

and

$$\begin{aligned} N(d_i - 1) &\leq N \min\left[\left\lfloor \frac{i-1}{K} \right\rfloor, \left\lfloor \frac{M-i}{N} \right\rfloor\right] \\ &\leq N \left\lfloor \frac{M-i}{N} \right\rfloor \leq M-i. \end{aligned}$$

Therefore, the condition in (A2) holds at time $t$. ■

It is clear that the maximum buffer size that could be achieved under the condition in (A3) is by choosing $d_i = \min\left[\lfloor\frac{i-1}{K}\rfloor, \lfloor\frac{M-i}{N}\rfloor\right] + 1$ for $i = 1, 2, \ldots, M$. Suppose that $M = q(N+K) + r$, where $q \geq 0$ and $0 \leq r \leq N+K-1$ are the quotient and the remainder, respectively, of $M$ divided by $N+K$. Then the fiber delays are given by

$$d_i = \begin{cases} j, \text{ if } (j-1)K+1 \leq i \leq jK \text{ and } 1 \leq j \leq q, \\ q+1, \text{ if } qK+1 \leq i \leq qK+r, \\ q-j+1, \\ \quad \text{if } qK+r+(j-1)N+1 \leq i \leq qK+r+jN \\ \quad \text{and } 1 \leq j \leq q. \end{cases}$$

As such, the maximum buffer size that could be achieved under the condition in (A3) is equal to $\frac{q(q+1)}{2}(N+K) + r(q+1)$. In the special case that $N = K$, the maximum buffer size that could be achieved under the condition in (A3) is $q(q+1)N + r(q+1)$, i.e., $O(\frac{M^2}{N})$, which is smaller than the $O(\frac{M^3}{N^2})$ buffer size that could be achieved by using the constructions in Theorem 5. However, we will see shortly that the condition in (A3) can be used to design the fiber delays $d_1, d_2, \ldots, d_M$ as given in Theorem 8 below such that the feedback system in Figure 4(a) or Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue that can tolerate up to $F$ broken/malfunctioning fibers, where $0 \leq F \leq M-1$.

In our constructions of fault tolerant $N$-to-$K$ optical priority queue, we assume that the $(M + \max[N,K]) \times (M + \max[N,K])$ switch in Figure 4(a) or Figure 5(a) is accompanied with a detection circuitry that can detect whether a fiber is broken/malfunctioning (e.g., fiber cut, fiber shorting out, etc). This can be done by simply sending pilot signals through the fiber delay lines. If the pilot signals are not received at a pair of input/output ports, then a fiber cut may have occurred on the fiber connecting that pair of input/output ports, and if the pilot signals are received immediately at a pair of input/output ports, then the fiber connecting that pair of input/output ports may be shorting out. As such, if some of the fibers are broken/malfunctioning, then we can just disregard the input/output ports corresponding to the broken/malfunctioning fibers, and view the remaining input/output ports as the input/output ports of a smaller switch.

Note that even with the fault-detection mechanism as described above, the condition in (A3) still does not guarantee that the construction in Figure 4(a) or Figure 5(a) is fault tolerant. In order to make sure that the construction in Figure 4(a) or Figure 5(a) can tolerate up to $F$ broken/malfunctioning fibers, where $0 \leq F \leq M-1$, what we need is a more restrictive condition (we will give such a condition in Theorem 8 below) on the fiber delays $d_1, d_2, \ldots, d_M$ such that after up to $F$ of the fibers are broken/malfunctioning, the delays of the remaining unbroken/functioning fibers still satisfy the condition in (A3) (with $M$ replaced by the number of the remaining unbroken/functioning fibers in (A3)) after a proper re-indexing of the remaining unbroken/functioning fibers. For example, consider the case that $M = 12$, $N = 3$, and $K = 2$. If we choose $(d_1, d_2, \ldots, d_{12}) = (1,1,2,2,3,3,2,2,2,1,1,1)$, then the condition in (A3) is satisfied and hence the feedback system in Figure 4(a) or Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M} d_i = 21$. However, when the third fiber with delay $d_3 = 2$ is broken/malfunctioning, the delays of the remaining unbroken/functioning fibers do not satisfy the condition in (A3) (with $M = 11$ in (A3)) no matter how we re-index the remaining unbroken/functioning fibers.

On the other hand, if we choose $(d_1, d_2, \ldots, d_{12}) = (1,1,2,2,3,2,2,2,1,1,1,1)$ so that the delays $(d_1, d_2, \ldots, d_{11}) = (1,1,2,2,3,2,2,2,1,1,1)$ of the first eleven fibers satisfy the condition in (A3) (with $M = 11$ in (A3)) and the delay of the last fiber is equal to 1. Then it can be seen that when the third fiber with delay $d_3 = 2$ is broken/malfunctioning, we can re-index the the twelfth fiber as the third fiber. As such, after the re-indexing the delays of the remaining unbroken/functioning fibers satisfy the condition in (A3) (with $M = 11$ in (A3)). Similarly, it can be seen that when any one of the fibers is broken/malfunctioning, we can re-index the remaining unbroken/functioning fibers so that the condition in (A3) (with $M = 11$ in (A3)) is satisfied and hence the feedback system in Figure 4(a) or Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue with a buffer size at least $\sum_{i=1}^{M} d_i - d_5 = 16$ (note that $d_5$ is the largest fiber delay among the twelve fiber delays).

The idea of our choice of the fiber delays $d_1, d_2, \ldots, d_M$ in Theorem 8 below is to choose the delays $d_1, d_2, \ldots, d_{M-F}$ of the first $M-F$ fibers such that they satisfy the condition in (A3) (with $M$ replaced by $M-F$ in (A3)) and choose the delays $d_{M-F+1}, d_{M-F+2}, \ldots, d_M$ of the last $F$ fibers to be equal to 1. The last $F$ fibers with delays equal to 1 can be regarded as the "backup" fibers for the first $M-F$ fibers. When some of the first $M-F$ fibers are broken/malfunctioning, we can then replace them by using the unbroken/functioning fibers among the last $F$ fibers. As the delays of the last $F$ fibers are chosen to be equal to 1, it can be shown that the remaining unbroken/functioning fibers still satisfy the condition in (A3) (with $M$ replaced by the number of the remaining unbroken/functioning fibers in (A3)) and hence the feedback system in Figure 4(a) or Figure 5(a) can be operated as an $N$-to-$K$ optical priority queue with a smaller buffer size.

Suppose that $0 \leq F \leq M-1$. Let

$$d_i^* = \begin{cases} \min\left[\lfloor\frac{i-1}{K}\rfloor, \lfloor\frac{M-F-i}{N}\rfloor\right] + 1, \\ \quad \text{for } i = 1, 2, \ldots, M-F, \\ 1, \text{ for } i = M-F+1, M-F+2, \ldots, M. \end{cases} \tag{28}$$

Let

$$B^*(N,K,F,M) = \sum_{i=1}^{M-F} d_{(i)}^*, \tag{29}$$

where $d_{(i)}^*$ is the $i^{\text{th}}$ smallest delay in $\{d_1^*, d_2^*, \ldots, d_M^*\}$ for $i = 1, 2, \ldots, M$, i.e., the delays $d_{(1)}^* \leq d_{(2)}^* \leq \cdots \leq d_{(M)}^*$ are the delays $d_1^*, d_2^*, \ldots, d_M^*$ ordered from the smallest to the largest.

**Theorem 8** *Suppose that* $0 \leq F \leq M-1$.

*(i) Let*

$$1 \leq d_i \leq d_i^*, \text{ for } i = 1, 2, \ldots, M, \tag{30}$$

*where $d_i^*$ is given by (28) for $i = 1, 2, \ldots, M$. In other words, the delays $d_1, d_2, \ldots, d_{M-F}$ of the first $M-F$ fibers satisfy the condition in (A3) (with $M$ replaced by $M-F$ in (A3)) and the delays $d_{M-F+1}, d_{M-F+2}, \ldots, d_M$ of the last $F$ fibers are all equal to 1. Then after up to $F$ of the $M$ fibers in the feedback system in Figure 4(a) or Figure 5(a) are broken/malfunctioning, the remaining unbroken/functioning fibers can be re-indexed so that their delays after re-indexing still satisfy the condition in (A3) (with $M$ replaced by the number of the remaining unbroken/functioning fibers in (A3)) and hence the feedback system could be operated as an $N$-to-$K$ optical priority queue with a buffer size at least $\sum_{i=1}^{M-F} d_{(i)}$, where $d_{(i)}$ is the $i^{\text{th}}$ smallest delay in $\{d_1, d_2, \ldots, d_M\}$ for $i = 1, 2, \ldots M$. In particular, if we choose $d_i = d_i^*$ for $i = 1, 2, \ldots, M$, then the buffer size is at least $B^*(N, K, F, M)$, where $B^*(N, K, F, M)$ is given by (29).*

*(ii) Conversely, assume that when at most $F$ of the $M$ fibers in the feedback system in Figure 4(a) or Figure 5(a) are broken/malfunctioning, the remaining unbroken/functioning fibers can be re-indexed so that their delays after re-indexing still satisfy the condition in (A3) (with $M$ replaced by the number of the remaining unbroken/functioning fibers in (A3)) and hence the feedback system can be operated as an $N$-to-$K$ optical priority queue with a buffer size at least $B'$. Then we must have $B' \leq B^*(N, K, F, M)$.*

**Proof.** (i) Index the $M$ fibers with delays $d_1, d_2, \ldots, d_M$ from 1 to $M$. Assume that $F'$ fibers are broken/malfunctioning, where $0 \leq F' \leq F$. Let $m$ and $n$ be the numbers of broken/malfunctioning fibers with indices in the sets $I = \{1, 2, \ldots, M-F\}$ and $J = \{M-F+1, M-F+2, \ldots, M\}$, respectively. Note that $m + n = F'$. Let $I' = \{i_1, i_2, \ldots, i_m\} \subseteq I$ be the index set of the "broken/malfunctioning" fibers with indices in $I$ and let $J' = \{j_1, j_2, \ldots, j_{F-n}\} \subseteq J$ be the index set of the "unbroken/functioning" fibers with indices in $J$.

Now we re-index the remaining $M - F'$ unbroken/functioning fibers with indices in $(I \backslash I') \cup J'$. We re-index the $i^{\text{th}}$ fiber as the $i^{\text{th}}$ fiber for $i \in I \backslash I'$, re-index the $j_k^{\text{th}}$ fiber as the $i_k^{\text{th}}$ fiber for $k = 1, 2, \ldots, m$ (note that this is feasible as we have $m = F' - n \leq F - n$), and re-index the $j_k^{\text{th}}$ fiber as the $(M - F + k - m)^{\text{th}}$ fiber for $k = m+1, m+2, \ldots, F-n$ (see Figure 7 for an illustration of the re-indexing).

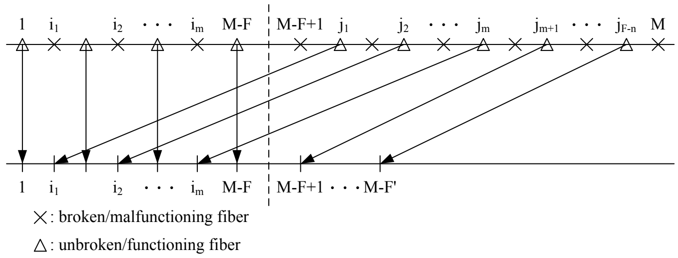


Fig. 7. An illustration of the re-indexing in the proof of Theorem 8.

Let $d_1', d_2', \ldots, d_{M-F'}'$ be the delays of the remaining $M - F'$ unbroken/functioning fibers after re-indexing. As $J' \subseteq J$ and $d_i = d_i^* = 1$ for all $i \in J$, we see that $d_i = 1$ for all $i \in J'$, and it then follows that

$$d_i' = \begin{cases} d_i, & \text{for } i \in I \backslash I', \\ 1, & \text{for } i \in I' \cup \{M-F+1, \ldots, M-F'\}. \end{cases} \tag{31}$$

If $i \in I \backslash I'$, then we see from (31), (30), (28), and $F' \leq F$ that

$$\begin{aligned} d_i' = d_i \leq d_i^* &= \min\left[\left\lfloor \frac{i-1}{K} \right\rfloor, \left\lfloor \frac{M-F-i}{N} \right\rfloor\right] + 1 \\ &\leq \min\left[\left\lfloor \frac{i-1}{K} \right\rfloor, \left\lfloor \frac{M-F'-i}{N} \right\rfloor\right] + 1. \end{aligned}$$

On the other hand, if $i \in I_1' \cup \{M-F+1, \ldots, M-F'\}$, then it is clear from (31) that

$$d_i' = 1 \leq \min\left[\left\lfloor \frac{i-1}{K} \right\rfloor, \left\lfloor \frac{M-F'-i}{N} \right\rfloor\right] + 1.$$

As such, the delays $d_1', d_2', \ldots, d_{M-F'}'$ of the remaining $M - F'$ unbroken/functioning fibers satisfy the condition in (A3) (with $M$ replaced by $M - F'$ in (A3)), and hence the feedback system in Figure 4(a) or Figure 5(a) could be operated as an $N$-to-$K$ optical priority queue with buffer $\sum_{i=1}^{M-F'} d_i'$. As $F' \leq F$ and $d_{(1)}, d_{(2)}, \ldots, d_{(M-F)}$ are the smallest $M - F$ delays in $\{d_1, d_2, \ldots, d_M\}$, we have

$$\sum_{i=1}^{M-F'} d_i' \geq \sum_{i=1}^{M-F} d_i' \geq \sum_{i=1}^{M-F} d_{(i)}.$$

(ii) First note that there are at least $F+1$ fibers with delays equal to 1, say $d_i = 1$ for $i = M-F, M-F+1, \ldots, M$. Otherwise, if all of the fibers with delays equal to 1 are broken/malfunctioning, then the delays of the remaining unbroken/functioning fibers cannot satisfy the condition in (A3) as the condition in (A3) requires that at least one of the remaining unbroken/functioning fibers has a delay equal to 1.

Let $d_{(i)}$ and $d_{(i)}^*$ be the $i^{\text{th}}$ smallest delay in $\{d_1, d_2, \ldots, d_M\}$ and $\{d_1^*, d_2^*, \ldots, d_M^*\}$, respectively, for $i = 1, 2, \ldots, M$. As $d_i = d_i^* = 1$ for $i = M-F+1, M-F+2, \ldots, M$, it follows that $d_{(i)} = d_{(i)}^* = 1$ for $i = 1, 2, \ldots, F$, and $d_{(F+i)}$ and and $d_{(F+i)}^*$ are the $i^{\text{th}}$ smallest delay in $\{d_1, d_2, \ldots, d_{M-F}\}$ and $\{d_1^*, d_2^*, \ldots, d_{M-F}^*\}$, respectively, for $i = 1, 2, \ldots, M-F$.

Consider the case that the $F$ fibers with delays $d_{M-F+1}, d_{M-F+2}, \ldots, d_M$ (the smallest $F$ delays in $\{d_1, d_2, \ldots, d_M\}$) are broken/malfunctioning. As we assume that the remaining $M - F$ unbroken/functioning fibers can be re-indexed so that their delays satisfy the condition in (A3) (with $M$ replaced by $M - F$ in (A3)), there exists a permutation $\sigma$ on $\{1, 2, \ldots, M-F\}$ such that the $i^{\text{th}}$ fiber is re-indexed as the $(\sigma(i))^{\text{th}}$ fiber and after re-indexing the delays $d_i'' = d_{\sigma^{-1}(i)}$, $i = 1, 2, \ldots, M-F$, satisfy

$$\begin{aligned} d_i'' &= d_{\sigma^{-1}(i)} \\ &\leq \min\left[\left\lfloor \frac{i-1}{K} \right\rfloor, \left\lfloor \frac{M-F-i}{N} \right\rfloor\right] + 1 = d_i^*, \\ &\qquad \text{for } i = 1, 2, \ldots, M-F. \end{aligned} \tag{32}$$

It then follows that

$$d_{(F+i)} \leq d_{(F+i)}^*, \ \text{for } i = 1, 2, \ldots, M-F. \tag{33}$$

To see this, note that $d^*_{(F+i)}$ is the $i^{\text{th}}$ smallest delay in $\{d^*_1, d^*_2, \ldots, d^*_{M-F}\}$, and hence there exist $k_1, k_2, \ldots, k_i \in \{1, 2, \ldots, M-F\}$ such that $d^*_{k_j} \le d^*_{(F+i)}$ for all $j = 1, 2, \ldots, i$. Thus, we see from (32) that $d_{\sigma^{-1}(k_j)} \le d^*_{k_j} \le d^*_{(F+i)}$ for all $j = 1, 2, \ldots, i$, i.e., there are at least $i$ delays in $\{d_1, d_2, \ldots, d_{M-F}\}$ that are less than or equal to $d^*_{(F+i)}$. This implies that the $i^{\text{th}}$ smallest delay in $\{d_1, d_2, \ldots, d_{M-F}\}$ is less than or equal to $d^*_{(F+i)}$, i.e., $d_{(F+i)} \le d^*_{(F+i)}$.

Now consider the case that the $F$ fibers with delays $d_{(M-F+1)}, d_{(M-F+2)}, \ldots, d_{(M)}$ (the largest $F$ delays in $\{d_1, d_2, \ldots, d_M\}$) are broken/malfunctioning. As the remaining $M-F$ unbroken/functioning fibers with delays $d_{(1)}, d_{(2)}, \ldots, d_{(M-F)}$ can only store a maximum of $\sum_{i=1}^{M-F} d_{(i)}$ optical packets, and the feedback system with the remaining $M-F$ unbroken/functioning fibers can still be operated as an $N$-to-$K$ optical priority queue with a buffer buffer at least $B'$, we must have

$$B' \le \sum_{i=1}^{M-F} d_{(i)}. \tag{34}$$

It then follows from (34), $d_{(i)} = d^*_{(i)} = 1$ for $1 = 1, 2, \ldots, F$, (33), and (29) that

$$B' \le \sum_{i=1}^{M-F} d_{(i)} \le \sum_{i=1}^{M-F} d^*_{(i)} = B^*(N, K, F, M).$$

The proof is completed. ■

## V. Conclusion

In this paper, we have considered the constructions of $N$-to-$K$ optical priority queues by using a single $(M + \max[N, K]) \times (M + \max[N, K])$ optical (bufferless) crossbar switch, $\min[N, K]$ $1 \times 2$ optical (bufferless) crossbar switches, and $M$ fiber delay lines with delays $d_1, d_2, \ldots, d_M$. By establishing a space-time advancement property and a monotonically decreasing/increasing property for the packets stored in the fiber delay lines, we showed that such constructions can be used for exact emulation of an optical priority queue with buffer $\sum_{i=1}^{M} d_i$. We also showed that our constructions achieve a buffer size of $O(\frac{M^3}{N^2})$ in the case that $N = K$. In particular, for the special case that $N = K = 1$, we can achieve a buffer size of $O(M^3)$, which is better than the $O(M^2)$ buffer size previously obtained in [25] and [26].

Furthermore, we showed that there is a gap between the $O(\frac{M^3}{N^2})$ buffer size achieved by our constructions and the exponential upper bound $O(2^{(M-2N)\log_2(1+\frac{1}{N})})$ given in (27) for the case that $N = K$. Whether it is possible to achieve such an exponential bound, and if possible, how to do that, remains an open research problem. For the special case that $N = K = 1$, it can be seen from a comment in [26] that if an optical priority queue has only $L$ priority classes of packets and one is willing to relax the requirement of exact emulation of an optical priority queue, then the exponential upper bound $O(2^M)$ can be achieved by using $L$ optical FIFO queues for these $L$ classes of packets. This is possible because an optical FIFO queue with buffer $B$ can be constructed with $O(\log B)$ $2 \times 2$ optical crossbar switches [20].

## Appendix A
## Proof of the Claim in Case 3 in the Proof of Theorem 5

Recall that in the proof of Theorem 5, the tagged packet under consideration enters the $i_0^{\text{th}}$ fiber delay line at time $t_0$. In this appendix, we prove the claim in Case 3 in the proof of Theorem 5 that the total number of packets stored in the fiber delay lines at time $t_0$ with priorities higher (resp., lower) than that of the tagged packet is at least $\alpha_{i_0} - 1$ (resp., $\beta_{i_0} - 1$), where $m_1 + 1 \le i_0 \le M - m_2$ and $\alpha_{i_0}$ (resp., $\beta_{i_0}$) is given by (7) (resp., (8)). To do so, we will first establish a space-time advancement property and a monotonically decreasing/increasing property for the packets stored in the fiber delay lines, and then show that there are at least $\alpha_{i_0}$ (resp., $\beta_{i_0}$) packets stored in the fiber delay lines at time $t_0$ with priorities higher (resp., lower) than or equal to that of the tagged packet. As there is a total order among all of the packets stored in the fiber delay lines, the only packet that has the same priority as that of the tagged packet is the tagged packet itself. Therefore, the total number of packets stored in the fiber delay lines at time $t_0$ with priorities higher (resp., lower) than that of the tagged packet is at least $\alpha_{i_0} - 1$ (resp., $\beta_{i_0} - 1$), and the claim is proved.

In the following, we divide our proof into two parts.

### *A-1. The first half of the proof*

In the first half of the proof, we show that there are at least $\alpha_{i_0}$ packets stored in the fiber delay lines at time $t_0$ with priorities higher than or equal to that of the tagged packet. For this, we let $p_{i,j}(t) = 1$ if the priority of the packet stored in the $(i, j)^{\text{th}}$ cell at time $t$ is higher than or equal to that of the tagged packet, and let $p_{i,j}(t) = 0$ otherwise.

Since a packet stored in the $(i, j)^{\text{th}}$ cell at time $t$ must be stored in the $(i, j-1)^{\text{th}}$ cell at time $t-1$, we have $p_{i,j}(t) = p_{i,j-1}(t-1)$. In general, we have the following *space-time advancement property* (see Figure 8):

$$\begin{aligned} p_{i,j}(t) &= p_{i,j-1}(t-1) = \cdots = p_{i,1}(t-(j-1)), \\ &\quad \text{for } j = 1, 2, \ldots, d_i \text{ and } i = 1, 2, \ldots, M. \end{aligned} \tag{35}$$

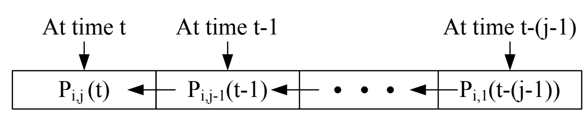


Fig. 8. The space-time advancement property in (35) for $j = 1, 2, \ldots, d_i$ and $i = 1, 2, \ldots, M$.

According to the operation rules of our constructions of an optical priority queue, the packets at the inputs of the $M$ fiber delay lines are sorted according to their priorities. Thus, we have

$$p_{1,1}(t) \ge p_{2,1}(t) \ge \cdots \ge p_{M,1}(t), \text{ for all } t. \tag{36}$$

As we assume that $d_i = \lfloor \frac{i-1}{K} \rfloor + 1$ for $i = 1, 2, \ldots, m_1$, $d_i = \lfloor \frac{M-i}{N} \rfloor + 1$ for $i = M - m_2 + 1, M - m_2 + 2, \ldots, M$, and $d_i \ge n$ for $i = m_1 + 1, m_1 + 2, \ldots, M - m_2$, where $m_1 = nK$, $m_2 = nN$, and $m_1 + m_2 \le M$, the definition of $p_{i,j}(t)$ for $j = 1, 2, \ldots, n$ and $i = (j-1)K + 1, (j-1)K +$

$2, \ldots, M-(j-1)N$ is *feasible*. It then follows from (36) and the space-time advancement property in (35) that we have the following *monotonically decreasing property* (see Figure 9):

$$p_{(j-1)K+1,j}(t) \geq p_{(j-1)K+2,j}(t) \geq \cdots \geq p_{M-(j-1)N,j}(t), \quad \text{for } j = 1, 2, \ldots, n. \quad (37)$$

We note that the space-time advancement property in (35) and the monotonically decreasing property in (37) are the key properties for our proof.

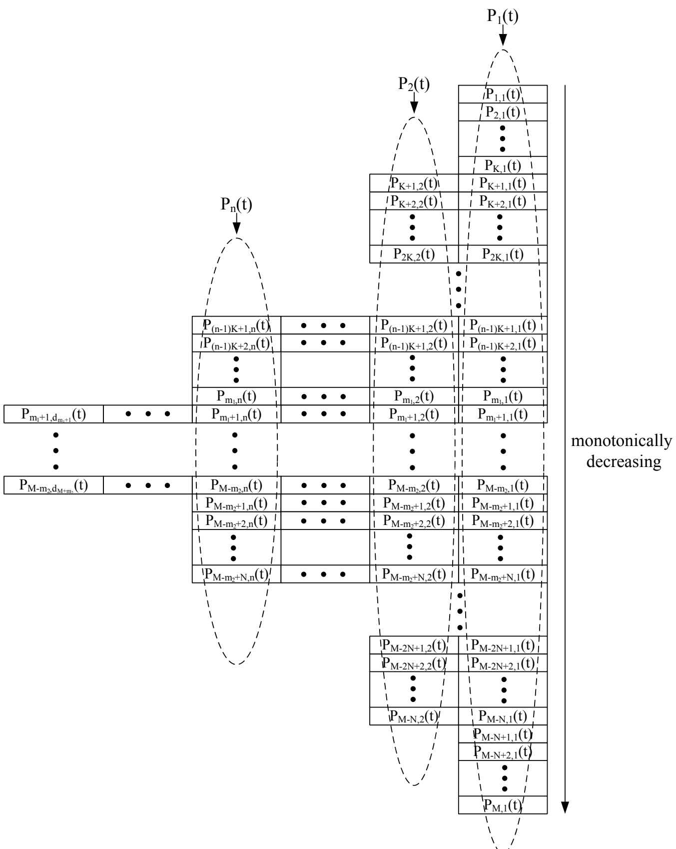


Fig. 9. The monotonically decreasing property in (37) and the definition of $p_j(t)$ in (38) for $j = 1, 2, \ldots, n$.

We need some more notations for our proof. Let $p_j(t)$ be the total number of packets stored in the cells of the "$j^{\text{th}}$ column" in Figure 6 at time $t$ with priorities higher than or equal to that of the tagged packet for $j = 1, 2, \ldots, n$ (see Figure 9), i.e.,

$$p_j(t) = \sum_{i=(j-1)K+1}^{M-(j-1)N} p_{i,j}(t), \text{ for } j = 1, 2, \ldots, n. \quad (38)$$

As $p_{i,j}(t)$ only assumes two values, i.e., $p_{i,j}(t) = 0$ or 1, we have from (38) that

$$0 \leq p_j(t) \leq M - (j-1)(N+K), \text{ for } j = 1, 2, \ldots, n. \quad (39)$$

Note that for $j = 1, 2, \ldots, n$ and $i = (j-1)K+1, (j-1)K+2, \ldots, jK$, we have from (4) that $d_i = j$ and hence the last cell of the $i^{\text{th}}$ fiber delay line is the $(i,j)^{\text{th}}$ cell. Let $x_j(t)$ be the total number of packets stored in the last cells of the $((j-1)K+1)^{\text{th}}$, the $((j-1)K+2)^{\text{th}}$, ..., and the $(jK)^{\text{th}}$ fiber delay lines at time $t$ with priorities higher than or equal to that of the tagged packet for $j = 1, 2, \ldots, n$, i.e.,

$$x_j(t) = \sum_{i=(j-1)K+1}^{jK} p_{i,d_i}(t) = \sum_{i=(j-1)K+1}^{jK} p_{i,j}(t), \quad \text{for } j = 1, 2, \ldots, n, \quad (40)$$

and let $x(t)$ be the total number of packets stored in the last cells of the first $m_1$ fiber delay lines at time $t$ with priorities higher than or equal to that of the tagged packet, i.e.,

$$x(t) = \sum_{i=1}^{m_1} p_{i,d_i}(t) = \sum_{j=1}^{n} \sum_{i=(j-1)K+1}^{jK} p_{i,d_i}(t) = \sum_{j=1}^{n} x_j(t). \quad (41)$$

Similarly, for $j = 1, 2, \ldots, n$ and $i = M-jN+1, M-jN+2, \ldots, M-(j-1)N$, we have $d_i = j$ and hence the last cell of the $i^{\text{th}}$ fiber delay line is the $(i,j)^{\text{th}}$ cell. Let $y_j(t)$ be the total number of packets stored in the last cells of the $(M-jN+1)^{\text{th}}$, the $(M-jN+2)^{\text{th}}$, ..., and the $(M-(j-1)N)^{\text{th}}$ fiber delay lines at time $t$ with priorities higher than or equal to that of the tagged packet for $j = 1, 2, \ldots, n$, i.e.,

$$y_j(t) = \sum_{i=M-jN+1}^{M-(j-1)N} p_{i,d_i}(t) = \sum_{i=M-jN+1}^{M-(j-1)N} p_{i,j}(t), \quad \text{for } j = 1, 2, \ldots, n, \quad (42)$$

and let $y(t)$ be the total number of packets stored in the last cells of the last $m_2$ fiber delay lines at time $t$ with priorities higher than or equal to that of the tagged packet, i.e.,

$$y(t) = \sum_{i=M-m_2+1}^{M} p_{i,d_i}(t) = \sum_{j=1}^{n} \sum_{i=M-jN+1}^{M-(j-1)N} p_{i,d_i}(t) = \sum_{j=1}^{n} y_j(t). \quad (43)$$

In Figure 10, we illustrate the definitions of $x_j(t)$ in (40) and $y_j(t)$ in (42) for $j = 1, 2, \ldots, n$, and $x(t)$ in (41) and $y(t)$ in (43).

Now we present three lemmas that will be used in our proof. In the following lemma, we show that if the inequality on the right-hand side of (39) holds with strict inequality for some $2 \leq j \leq n$, then we can obtain an upper bound on $p_{j-j'}(t-j')$ in terms of $p_j(t)$ for $j' = 1, 2, \ldots, j-1$.

**Lemma 9** *If $p_j(t) < M-(j-1)(N+K)$ for some $2 \leq j \leq n$, then*

$$p_{j-j'}(t-j') \leq p_j(t) + j'K, \textit{ for } j' = 1, 2, \ldots, j-1. \quad (44)$$

**Proof.** From (38) and $0 \leq p_{i,j}(t) \leq 1$ for all $i$ and $j$, it is easy to see that if $p_j(t) < M-(j-1)(N+K)$, then $p_{m,j}(t) = 0$ for some $(j-1)K+1 \leq m \leq M-(j-1)N$. From the space-time advancement property in (35), we can see that

$$p_{m,j-1}(t-1) = p_{m,j}(t) = 0.$$

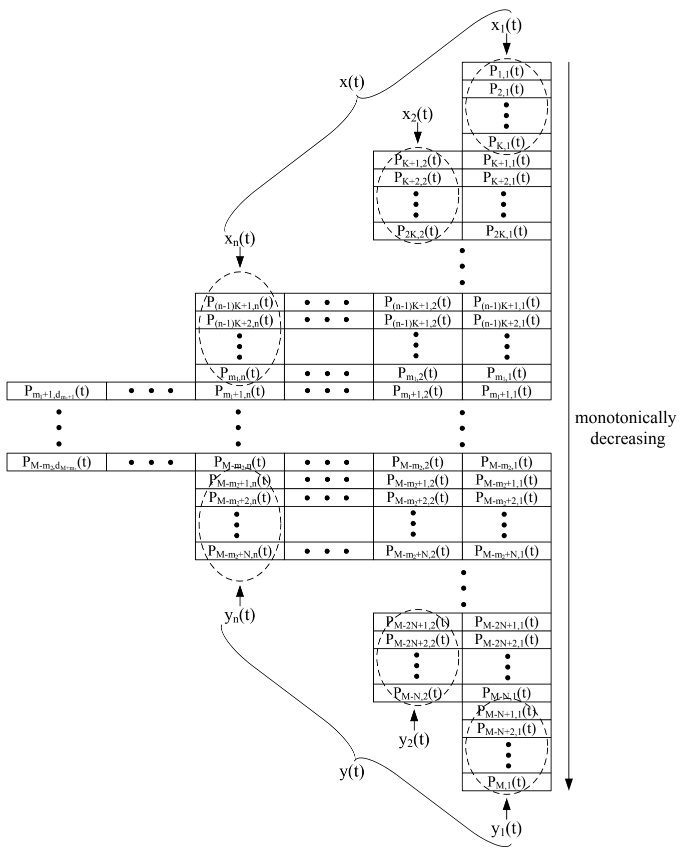


Fig. 10. The definitions of $x_j(t)$ in (40) and $y_j(t)$ in (42) for $j = 1, 2, \ldots, n$, and $x(t)$ in (41) and $y(t)$ in (43).

It follows from the monotonically decreasing property in (37) and $p_{i,j}(t) \geq 0$ for all $i$ and $j$ that

$$\begin{aligned} p_{m,j-1}(t-1) &= p_{m+1,j-1}(t-1) = \cdots \\ &= p_{M-(j-2)N,j-1}(t-1) = 0. \qquad (45) \end{aligned}$$

Thus, we have from (38), $p_{i,j}(t) \leq 1$ for all $i$ and $j$, the space-time advancement property in (35), and (45) that

$$\begin{aligned} & p_{j-1}(t-1) \\ &= \sum_{i=(j-2)K+1}^{M-(j-2)N} p_{i,j-1}(t-1) \\ &= \sum_{i=(j-2)K+1}^{(j-1)K} p_{i,j-1}(t) + \sum_{i=(j-1)K+1}^{M-(j-1)N} p_{i,j-1}(t-1) \\ &\quad + \sum_{i=M-(j-1)N+1}^{M-(j-2)N} p_{i,j-1}(t-1) \\ &\leq K + \sum_{i=(j-1)K+1}^{M-(j-1)N} p_{i,j}(t) + 0 \\ &= p_j(t) + K. \qquad (46) \end{aligned}$$

From (46) and the assumption that $p_j(t) < M-(j-1)(N+K)$, we have

$$\begin{aligned} p_{j-1}(t-1) &\leq p_j(t) + K \\ &< M-(j-1)(N+K) + K \\ &< M-(j-2)(N+K). \end{aligned}$$

By repeating the above argument that leads to (46), we obtain

$$p_{j-2}(t-2) \leq p_{j-1}(t-1) + K \leq p_j(t) + 2K.$$

It should be clear that by repeating the above argument for $j'$ times, we obtain (44) for $j' = 1, 2, \ldots, j-1$ and the proof is completed. ■

In the following lemma, we give an upper bound on $y(t)$ in terms of $x(t)$.

**Lemma 10**

$$y(t) \leq \begin{cases} x(t) + m_2 - m_1, & \textit{if } N \geq K, \\ x(t), & \textit{if } N < K. \end{cases} \qquad (47)$$

**Proof.** If $N \geq K$, then from (43), the monotonically decreasing property in (37), $p_{i,j}(t) \leq 1$ for all $i$ and $j$, (41), $m_1 = nK$, and $m_2 = nN$, we have

$$\begin{aligned} y(t) &= \sum_{j=1}^{n} \sum_{i=M-jN+1}^{M-(j-1)N} p_{i,j}(t) \\ &= \sum_{j=1}^{n} \left( \sum_{i=M-jN+1}^{M-jN+K} p_{i,j}(t) + \sum_{i=M-jN+K+1}^{M-(j-1)N} p_{i,j}(t) \right) \\ &\leq \sum_{j=1}^{n} \left( \sum_{i=(j-1)K+1}^{jK} p_{i,j}(t) + N - K \right) \\ &= x(t) + n(N-K) = x(t) + m_2 - m_1. \end{aligned}$$

On the other hand, if $N < K$, then it is clear from (43), the monotonically decreasing property in (37), $p_{i,j}(t) \geq 0$ for all $i$ and $j$, and (41) that

$$\begin{aligned} y(t) &= \sum_{j=1}^{n} \sum_{i=M-jN+1}^{M-(j-1)N} p_{i,j}(t) \leq \sum_{j=1}^{n} \sum_{i=(j-1)K+1}^{(j-1)K+N} p_{i,j}(t) \\ &\leq \sum_{j=1}^{n} \sum_{i=(j-1)K+1}^{jK} p_{i,j}(t) = x(t). \end{aligned}$$

The proof is completed. ■

In the following lemma, we derive the key inequalities that will be useful in finding a lower bound on the total number of packets stored in the fiber delay lines with priorities higher than or equal to that of the tagged packet.

**Lemma 11**

*(i)* $x(t) \leq x(t-1) + K$.

*(ii)* $x(t-1) - K \leq p_1(t)$

$$\leq \begin{cases} 2x(t-1) + M - 2m_1 + N, & \textit{if } N \geq K, \\ 2x(t-1) + M - m_1 - m_2 + N, & \textit{if } N < K. \end{cases}$$

*(iii)* $x(t-1) \leq p_j(t) + (2j-1)K$ *for* $j = 2, 3, \ldots, n$.

**Proof.** (i) For $j = 2, 3, \ldots, n$, we have from (40), the monotonically decreasing property in (37), and the space-time advancement property in (35) that

$$\begin{aligned} x_{j-1}(t-1) &= \sum_{i=(j-2)K+1}^{(j-1)K} p_{i,j-1}(t-1) \\ &\geq \sum_{i=(j-2)K+1}^{(j-1)K} p_{i+K,j-1}(t-1) \\ &= \sum_{i=(j-1)K+1}^{jK} p_{i,j-1}(t-1) \\ &= \sum_{i=(j-1)K+1}^{jK} p_{i,j}(t) = x_j(t). \end{aligned} \tag{48}$$

It follows from (41), (40), $p_{i,j}(t) \leq 1$ for all $i$ and $j$, and (48) that

$$\begin{aligned} x(t) &= \sum_{j=1}^{n} x_j(t) = x_1(t) + \sum_{j=2}^{n} x_j(t) \\ &= \sum_{i=1}^{K} p_{i,1}(t) + \sum_{j=2}^{n} x_j(t) \\ &\leq K + \sum_{j=2}^{n} x_{j-1}(t-1) = K + \sum_{j=1}^{n-1} x_j(t-1) \\ &\leq K + \sum_{j=1}^{n-1} x_j(t-1) + x_n(t-1) = x(t-1) + K. \end{aligned}$$

(ii) Note that $p_1(t) = \sum_{i=1}^{M} p_{i,1}(t)$ is the total number of packets stored in the cells $(1,1), (2,1), \ldots, (M,1)$ at time $t$ with priorities higher than or equal to that of the tagged packet. These $M$ packets can only come from the $M$ packets stored in the last cells of the $M$ fiber delay lines at time $t-1$, i.e., the cells $(1,d_1), (2,d_2), \ldots, (M,d_M)$, the $N$ packets arriving from the $N$ arrival links at time $t$, and the $K-N$ packets arriving from the $K-N$ idling links at time $t$ (in the case that $N < K$). Let $n_1(t-1)$, $n_2(t)$, and $n_3(t)$ be the number of packets stored in the last cells of the $M$ fiber delay lines at time $t-1$, the number of packets arriving from the $N$ arrival links at time $t$, and the number of packets arriving from the $K-N$ idling links at time $t$ (in the case that $N < K$), respectively, with priorities higher than or equal to that of the tagged packet. Also, let $n_4(t)$ be the number of packets leaving the system at time $t$ with priorities higher than that of the tagged packet. Clearly, we have

$$\begin{aligned} &p_1(t) \\ &= \begin{cases} n_1(t-1) + n_2(t) - n_4(t), & \text{if } N \geq K, \\ n_1(t-1) + n_2(t) + n_3(t) - n_4(t), & \text{if } N < K. \end{cases} \end{aligned} \tag{49}$$

From the definition of $p_{i,j}(t)$, we immediately see that $n_1(t-1) = \sum_{i=1}^{M} p_{i,d_i}(t-1)$. It follows from (41) and (43) that

$$\begin{aligned} n_1(t-1) &= \sum_{i=1}^{M} p_{i,d_i}(t-1) \\ &= \sum_{i=1}^{m_1} p_{i,d_i}(t-1) + \sum_{i=m_1+1}^{M-m_2} p_{i,d_i}(t-1) \\ &\quad + \sum_{i=M-m_2+1}^{M} p_{i,d_i}(t-1) \\ &= x(t-1) + \sum_{i=m_1+1}^{M-m_2} p_{i,d_i}(t-1) + y(t-1). \end{aligned} \tag{50}$$

Also, it is clear that $0 \leq n_2(t) \leq N$.

We consider the two cases $N \geq K$ and $N < K$ separately.

*Case 1:* $N \geq K$. From the operation rules in Section II-B, we see that if $c(t) = 0$, then we have $n_4(t) = 0$ (as in this case there are no packets departing from the departure links at time $t$, and the $N$ packets dumped through the loss links at time $t$ are the $N$ lowest priority packets in $\tilde{q}(t-1) \cup \tilde{a}(t)$ so that they have priorities lower than that of the tagged packet), and if $c(t) = 1$, then we have $n_4(t) = K$ (as in this case the $K$ packets departing from the departure links at time $t$ are the $K$ highest priority packets in $\tilde{q}(t-1) \cup \tilde{a}(t)$ so that they have priorities higher than that of the tagged packet, and the $N-K$ packets dumped through the loss links at time $t$ are the $N-K$ lowest priority packets in $\tilde{q}(t-1) \cup \tilde{a}(t)$ so that they have priorities lower than that of the tagged packet). It follows that $0 \leq n_4(t) \leq K$.

Thus, we have from (49), (50), $p_{i,j}(t-1) \geq 0$ for all $i$ and $j$, $n_2(t) \geq 0$, and $n_4(t) \leq K$ that

$$\begin{aligned} &p_1(t) \\ &= x(t-1) + \sum_{i=m_1+1}^{M-m_2} p_{i,d_i}(t-1) + y(t-1) + n_2(t) - n_4(t) \\ &\geq x(t-1) - K, \end{aligned}$$

and we have from (49), (50), $p_{i,j}(t-1) \leq 1$ for all $i$ and $j$, (47), $n_2(t) \leq N$, and $n_4(t) \geq 0$ that

$$\begin{aligned} &p_1(t) \\ &= x(t-1) + \sum_{i=m_1+1}^{M-m_2} p_{i,d_i}(t-1) + y(t-1) + n_2(t) - n_4(t) \\ &\leq x(t-1) + (M - m_2 - m_1) + (x(t-1) + m_2 - m_1) + N \\ &= 2x(t-1) + M - 2m_1 + N. \end{aligned}$$

*Case 2:* $N < K$. From the operation rules in Section II-C, we see that if $c(t) = 0$, then we have $n_3(t) = 0$ (as in this case the $K-N$ fictitious packets from the idling links depart from the last $K-N$ departure links at time $t$, they cannot be the tagged packet and hence they have priorities lower than that of the tagged packet) and $n_4(t) = 0$ (as in this case the $K-N$ packets departing from the departure links at time $t$ are the $K-N$ fictitious packets from the idling links and the $N$ packets dumped through the loss links at time $t$ are the $N$ lowest priority packets in $\tilde{q}(t-1) \cup \tilde{a}(t)$ so that they all have priorities lower than that of the tagged packet), and if

$c(t)=1$, then we have $0 \le n_3(t) \le K-N$ (as in this case the $K-N$ fictitious packets from the idling links at time $t$ are routed to the fiber delay lines) and $n_4(t)=K$ (as in this case there are no packets dumped through the loss links at time $t$, and the $K$ packets departing from the departure links at time $t$ are the $K$ highest priority packets in $\tilde{q}(t-1)\cup\tilde{a}(t)$ so that they have priorities higher than that of the tagged packet). It follows that $-K \le n_3(t)-n_4(t) \le 0$.

Thus, we have from (49), (50), $p_{i,j}(t-1)\ge 0$ for all $i$ and $j$, $n_2(t)\ge 0$, and $n_3(t)-n_4(t)\ge -K$ that

$$\begin{aligned} p_1(t) &= x(t-1)+\sum_{i=m_1+1}^{M-m_2} p_{i,d_i}(t-1)+y(t-1) \\ &\quad +n_2(t)+n_3(t)-n_4(t) \\ &\ge x(t-1)-K, \end{aligned}$$

and we have from (49), (50), $p_{i,j}(t-1)\le 1$ for all $i$ and $j$, (47), $n_2(t)\le N$, and $n_3(t)-n_4(t)\le 0$ that

$$\begin{aligned} p_1(t) &= x(t-1)+\sum_{i=m_1+1}^{M-m_2} p_{i,d_i}(t-1)+y(t-1) \\ &\quad +n_2(t)+n_3(t)-n_4(t) \\ &\le x(t-1)+(M-m_2-m_1)+x(t-1)+N+0 \\ &= 2x(t-1)+M-m_1-m_2+N. \end{aligned}$$

(iii) Let $2\le j\le n$. If $p_j(t) < M-(j-1)(N+K)$, then from (44) with $j'=j-1$, the lower bound in Lemma 11(ii), and Lemma 11(i), we have

$$\begin{aligned} p_j(t)+(j-1)K &\ge p_1(t-(j-1)) \\ &\ge x(t-j)-K \\ &\ge x(t-j+1)-2K \\ &\vdots \\ &\ge x(t-1)-jK. \end{aligned}$$

It follows that

$$x(t-1)\le p_j(t)+(2j-1)K. \tag{51}$$

On the other hand, if $p_j(t)=M-(j-1)(N+K)$, then we have

$$\begin{aligned} p_j(t)+(2j-1)K &= M-(j-1)(N+K)+(2j-1)K \\ &= M-(j-1)(N-K)+K. \end{aligned} \tag{52}$$

For $N\ge K$, we have from (52), $2\le j\le n$, $m_1=nK$, $m_2=nN$, and $M\ge m_1+m_2$ that

$$\begin{aligned} p_j(t)+(2j-1)K &= M-(j-1)(N-K)+K \\ &\ge M-(n-1)(N-K)+K \\ &= M-m_2+m_1+N \\ &\ge m_1. \end{aligned} \tag{53}$$

For $N<K$, we also have from (52), $2\le j\le n$, and $M\ge m_1+m_2$ that

$$\begin{aligned} p_j(t)+(2j-1)K &= M-(j-1)(N-K)+K \\ &\ge M-(2-1)(N-K)+K \\ &= M+2K-N \\ &\ge m_1. \end{aligned} \tag{54}$$

It follows from (41), (53), and (54) that

$$x(t-1)=\sum_{i=1}^{m_1} p_{i,d_i}(t-1)\le m_1\le p_j(t)+(2j-1)K. \tag{55}$$

The proof is completed by combining (51) and (55). ■

We are now in a position to show that the total number of packets stored in the fiber delay lines at time $t_0$ with priorities higher than or equal to that of the tagged packet is at least $\alpha_{i_0}$. Clearly, it suffices to show that the total number of packets stored in the cells of the first $n$ columns in Figure 6 at time $t_0$ with priorities higher than or equal to that of the tagged packet is at least $\alpha_{i_0}$, i.e., $\sum_{j=1}^n p_j(t_0)\ge\alpha_{i_0}$. Indeed, we have from Lemma 11(iii), the upper bound in Lemma 11(ii), the fact that $p_{i,j}(t)$ is an integer, and $p_1(t_0)=i_0$ that

$$\begin{aligned} &\sum_{j=1}^n p_j(t_0) \\ &= p_1(t_0)+\sum_{j=2}^n p_j(t_0) \\ &\ge p_1(t_0)+\sum_{j=2}^n (x(t_0-1)-(2j-1)K)^+ \\ &\ge \begin{cases} p_1(t_0)+\sum_{j=2}^n \lceil (\frac{p_1(t_0)-(M-2m_1+N)-(4j-2)K}{2})^+\rceil, & \text{if } N\ge K, \\ p_1(t_0)+\sum_{j=2}^n \lceil (\frac{p_1(t_0)-(M-m_1-m_2+N)-(4j-2)K}{2})^+\rceil, & \text{if } N<K, \end{cases} \\ &= \begin{cases} i_0+\sum_{j=2}^n \lceil (\frac{i_0-M+2m_1-N-(4j-2)K}{2})^+\rceil, & \text{if } N\ge K, \\ i_0+\sum_{j=2}^n \lceil (\frac{i_0-M+m_1+m_2-N-(4j-2)K}{2})^+\rceil, & \text{if } N<K. \end{cases} \\ &= \alpha_{i_0}. \end{aligned} \tag{56}$$

*A-2. The second half of the proof*

In the second half of the proof, we show that there are at least $\beta_{i_0}$ packets stored in the fiber delay lines at time $t_0$ with priorities lower than or equal to the priority of the tagged packet.

Let $p'_{i,j}(t)=1$ if the priority of the packet stored in the $(i,j)^{\text{th}}$ cell at time $t$ is lower than or equal to that of the tagged packet, and let $p'_{i,j}(t)=0$ otherwise. Then we have the following *space-time advancement property*:

$$\begin{aligned} &p'_{i,j}(t)=p'_{i,j-1}(t-1)=\cdots=p'_{i,1}(t-(j-1)), \\ &\qquad \text{for } j=1,2,\ldots,d_i \text{ and } i=1,2,\ldots,M. \end{aligned} \tag{57}$$

and we have the following *monotonically increasing property*:

$$\begin{aligned} &p'_{(j-1)K+1,j}(t)\le p'_{(j-1)K+2,j}(t)\le\cdots\le p'_{M-(j-1)N,j}(t), \\ &\qquad \text{for } j=1,2,\ldots,n. \end{aligned} \tag{58}$$

Let $p'_j(t)$ be the total number of packets stored in the cells of the "$j^{\text{th}}$ column" in Figure 6 at time $t$ with priorities lower than or equal to that of the tagged packet for $j=1,2,\ldots,n$, i.e.,

$$p'_j(t)=\sum_{i=(j-1)K+1}^{M-(j-1)N} p'_{i,j}(t), \text{ for } j=1,2,\ldots,n. \tag{59}$$

As $0 \le p'_{i,j}(t) \le 1$ for all $i$ and $j$, it follows from (59) that

$$0 \le p'_j(t) \le M-(j-1)(N+K), \ \text{for } j=1,2,\dots,n. \quad (60)$$

Let $x'_j(t)$ be the total number of packets stored in the last cells of the $((j-1)K+1)^{\text{th}}$, the $((j-1)K+2)^{\text{th}}$, …, and the $(jK)^{\text{th}}$ fiber delay lines at time $t$ with priorities lower than or equal to that of the tagged packet for $j=1,2,\dots,n$, i.e.,

$$x'_j(t) = \sum_{i=(j-1)K+1}^{jK} p'_{i,d_i}(t) = \sum_{i=(j-1)K+1}^{jK} p'_{i,j}(t), \quad \text{for } j=1,2,\dots,n, \quad (61)$$

and let $x'(t)$ be the total number of packets stored in the last cells of the first $m_1$ fiber delay lines at time $t$ with priorities lower than or equal to that of the tagged packet, i.e.,

$$x'(t) = \sum_{i=1}^{m_1} p'_{i,d_i}(t) = \sum_{j=1}^{n} \sum_{i=(j-1)K+1}^{jK} p'_{i,d_i}(t) = \sum_{j=1}^{n} x'_j(t). \quad (62)$$

Similarly, let $y'_j(t)$ be the total number of packets stored in the last cells of the $(M-jN+1)^{\text{th}}$, the $(M-jN+2)^{\text{th}}$, …, and the $(M-(j-1)N)^{\text{th}}$ fiber delay lines at time $t$ with priorities lower than or equal to that of the tagged packet for $j=1,2,\dots,n$, i.e.,

$$y'_j(t) = \sum_{i=M-jN+1}^{M-(j-1)N} p'_{i,d_i}(t) = \sum_{i=M-jN+1}^{M-(j-1)N} p'_{i,j}(t), \quad \text{for } j=1,2,\dots,n, \quad (63)$$

and let $y'(t)$ be the total number of packets stored in the last cells of the last $m_2$ fiber delay lines at time $t$ with priorities lower than or equal to that of the tagged packet, i.e.,

$$y'(t) = \sum_{i=M-m_2+1}^{M} p'_{i,d_i}(t) = \sum_{j=1}^{n} \sum_{i=M-jN+1}^{M-(j-1)N} p'_{i,d_i}(t) = \sum_{j=1}^{n} y'_j(t). \quad (64)$$

The following three lemmas (Lemma 12–Lemma 14) are the counterparts of the three lemmas (Lemma 9–Lemma 11) in the first half of the proof. They can be obtained from Lemma 9–Lemma 11 by replacing $p_j(t)$, $x(t)$, and $y(t)$ with $p'_j(t)$, $y'(t)$, and $x'(t)$, respectively, interchanging $K$ and $N$, and interchanging $m_1$ and $m_2$. As the proofs for Lemma 12–Lemma 14 are very similar to those for Lemma 9–Lemma 11, we only give the proof of Lemma 14(ii) here.

**Lemma 12** *If $p'_j(t) < M-(j-1)(N+K)$ for some $2 \le j \le n$, then*

$$p'_{j-j'}(t-j') \le p'_j(t) + j'N, \ \textit{for } j'=1,2,\dots,j-1. \quad (65)$$

**Lemma 13**

$$x'(t) \le \begin{cases} y'(t), & \textit{if } N \ge K, \\ y'(t)+m_1-m_2, & \textit{if } N < K. \end{cases} \quad (66)$$

**Lemma 14**

*(i)* $y'(t) \le y'(t-1)+N$.

*(ii)* $y'(t-1)-N \le p'_1(t) \le \begin{cases} 2y'(t-1)+M-m_1-m_2+K, & \textit{if } N \ge K, \\ 2y'(t-1)+M-2m_2+K, & \textit{if } N < K. \end{cases}$

*(iii)* $y'(t-1) \le p'_j(t)+(2j-1)N$ *for* $j=2,3,\dots,n$.

**Proof.** (ii) Let $n'_1(t-1)$, $n'_2(t)$, $n'_3(t)$, and $n'_4(t)$ be the number of packets stored in the last cells of the $M$ fiber delay lines at time $t-1$, the number of packets arriving from the $N$ arrival links at time $t$, the number of packets arriving from the $K-N$ idling links at time $t$ (in the case that $N<K$), and the number of packets leaving the system at time $t$, respectively, with priorities lower than or equal to that of the tagged packet. As in the proof of Lemma 11(ii), we can see that

$$p'_1(t) = \begin{cases} n'_1(t-1)+n'_2(t)-n'_4(t), & \text{if } N \ge K, \\ n'_1(t-1)+n'_2(t)+n'_3(t)-n'_4(t), & \text{if } N < K. \end{cases} \quad (67)$$

From the definition of $p'_{i,j}(t)$, (62), and (64), we see that

$$\begin{aligned} n'_1(t-1) &= \sum_{i=1}^{M} p'_{i,d_i}(t-1) \\ &= \sum_{i=1}^{m_1} p'_{i,d_i}(t-1) + \sum_{i=m_1+1}^{M-m_2} p'_{i,d_i}(t-1) + \sum_{i=M-m_2+1}^{M} p'_{i,d_i}(t-1) \\ &= x'(t-1) + \sum_{i=m_1+1}^{M-m_2} p'_{i,d_i}(t-1) + y'(t-1) \end{aligned} \quad (68)$$

Also, it is clear that $0 \le n'_2(t) \le N$.

We consider the two cases $N \ge K$ and $N < K$ separately.

*Case 1:* $N \ge K$. From the arguments in Case 1 in the proof of Lemma 11(ii), we can see that if $c(t)=0$, then we have $n'_4(t)=N$, and if $c(t)=1$, then we have $n'_4(t)=N-K$. It follows that $N-K \le n'_4(t) \le N$.

Thus, we have from (67), (68), $p'_{i,j}(t-1) \ge 0$ for all $i$ and $j$, $n'_2(t) \ge 0$, and $n'_4(t) \le N$ that

$$\begin{aligned} &p'_1(t) \\ &= x'(t-1) + \sum_{i=m_1+1}^{M-m_2} p'_{i,d_i}(t-1) + y'(t-1) + n'_2(t) - n'_4(t) \\ &\ge y'(t-1) - N, \end{aligned}$$

and we have from (67), (68), $p'_{i,j}(t-1) \le 1$ for all $i$ and $j$, (66), $n'_2(t) \le N$, and $n'_4(t) \ge N-K$ that

$$\begin{aligned} &p'_1(t) \\ &= x'(t-1) + \sum_{i=m_1+1}^{M-m_2} p'_{i,d_i}(t-1) + y'(t-1) + n'_2(t) - n'_4(t) \\ &= y'(t-1) + (M-m_2-m_1) + y'(t-1) + N - (N-K) \\ &= 2y'(t-1) + M - m_1 - m_2 + K. \end{aligned}$$

*Case 2:* $N < K$. From the arguments in Case 2 in the proof of Lemma 11(ii), we can see that if $c(t)=0$, then we have

$n_3'(t) = K - N$ and $n_4'(t) = K$, and if $c(t) = 1$, then we have $0 \le n_3'(t) \le K - N$ and $n_4'(t) = 0$. It follows that $-N \le n_3'(t) - n_4'(t) \le K - N$.

Thus, we have from (67), (68), $p_{i,j}'(t-1) \ge 0$ for all $i$ and $j$, $n_2'(t) \ge 0$, and $n_3'(t) - n_4'(t) \ge -N$ that

$$\begin{aligned} p_1'(t) &= x'(t-1) + \sum_{i=m_1+1}^{M-m_2} p_{i,d_i}'(t-1) + y'(t-1) \\ &\quad + n_2'(t) + n_3'(t) - n_4'(t) \\ &\ge y'(t-1) - N, \end{aligned}$$

and we have from (67), (68), $p_{i,j}'(t-1) \le 1$ for all $i$ and $j$, (66), $n_2'(t) \le N$, and $n_3'(t) - n_4'(t) \le K - N$ that

$$\begin{aligned} p_1'(t) &= x'(t-1) + \sum_{i=m_1+1}^{M-m_2} p_{i,d_i}'(t-1) + y'(t-1) \\ &\quad + n_2'(t) + n_3'(t) - n_4'(t) \\ &\le (y'(t-1) + m_1 - m_2) + (M - m_2 - m_1) \\ &\quad + y'(t-1) + N + (K - N) \\ &= 2y'(t-1) + M - 2m_2 + K. \end{aligned}$$

The proof is completed. ■

Now we can show that the total number of packets stored in the fiber delay lines at time $t_0$ with priorities lower than or equal to that of the tagged packet is at least $\beta_{i_0}$. It is clear that it suffices to show that the total number of packets stored in the cells of the first $n$ columns in Figure 6 at time $t_0$ with priorities lower than or equal to that of the tagged packet is at least $\beta_{i_0}$, i.e., $\sum_{j=1}^{n} p_j'(t_0) \ge \beta_{i_0}$. This can be seen from Lemma 14(iii), the upper bound in Lemma 14(ii), the fact that $p_{i,j}'(t)$ is an integer, and $p_1'(t_0) = M - i_0 + 1$ as follows:

$$\begin{aligned} &\sum_{j=1}^{n} p_j'(t_0) \\ &= p_1'(t_0) + \sum_{j=2}^{n} p_j'(t_0) \\ &\ge p_1'(t_0) + \sum_{j=2}^{n} (y'(t_0-1) - (2j-1)N)^+ \\ &\ge \begin{cases} p_1'(t_0) + \sum_{j=2}^{n} \lceil (\frac{p_1'(t_0) - (M-m_1-m_2+K) - (4j-2)N}{2})^+ \rceil, & \text{if } N \ge K, \\ p_1'(t_0) + \sum_{j=2}^{n} \lceil (\frac{p_1'(t_0) - (M-2m_2+K) - (4j-2)N}{2})^+ \rceil, & \text{if } N < K, \end{cases} \\ &= \begin{cases} M - i_0 + 1 + \sum_{j=2}^{n} \lceil (\frac{m_1+m_2-i_0+1-K-(4j-2)K}{2})^+ \rceil, & \text{if } N \ge K, \\ M - i_0 + 1 + \sum_{j=2}^{n} \lceil (\frac{2m_2-i_0+1-K-(4j-2)K}{2})^+ \rceil, & \text{if } N < K. \end{cases} \\ &= \beta_{i_0}. \end{aligned} \tag{69}$$

## Appendix B
## Proof of Theorem 6

Let $d_1 \le d_2 \le \cdots \le d_M$ be the delays of the $M$ fiber delay lines in the $N$-to-$K$ priority queue. First note that $d_i = 1$ for $i = 1, 2, \ldots, N + K$, i.e., there are at least $N + K$ fibers with delays equal to 1. To see this, assume that there are $B$ packets stored in the priority queue at time $t$. Then it is easy to see that the $K$ highest priority packets must be stored in the last cells of $K$ fibers at time $t$. Otherwise, if at least one of the $K$ highest priority packets is not stored in the last cell of a fiber at time $t$, then when the control input is enabled at time $t+1$ and there are no arriving packets at time $t+1$, that packet will still be stored in its fiber at time $t+1$ and it cannot depart from the priority queue at time $t+1$, and we will have a contradiction to the priority departure property in (P4) in Definition 1. Similarly, the $N$ lowest priority packets must be stored in the last cells of another $N$ fibers at time $t$. Otherwise, if at least one of the $N$ lowest priority packets is not stored in the last cell of a fiber at time $t$, then when the control input is disabled at time $t+1$ and there are $N$ arriving packets at time $t+1$ with priorities higher than those of the $N$ lowest priority packets, that packet will still be stored in its fiber at time $t+1$ and it cannot be dumped from the priority queue at time $t+1$, and we will have a contradiction to the priority loss property in (P5) in Definition 1. Furthermore, assume that the control input is disabled at time $t+1$ and there are no arriving packets at time $t+1$. Then it is clear that the $B$ packets stored in the priority queue at time $t$ remain stored in the priority queue at time $t+1$, and hence the $K$ highest priority packets and the $N$ lowest priority packets that appear at the outputs of $N+K$ fibers at the beginning of the $(t+1)^{\text{th}}$ time slot must be routed to the inputs of $N + K$ fibers and stored in the "first" cells of those $N + K$ fibers at time $t + 1$. As there are $B$ packets stored in the priority queue at time $t + 1$, it follows from the above argument that the $K$ highest priority packets and the $N$ lowest priority packets must be stored in the "last" cells of those $N + K$ fibers at time $t + 1$. Apparently, this is only possible if those $N + K$ fibers have delays equal to 1. This shows that there are at least $N + K$ fibers with delays equal to 1.

Let

$$j = \max\left\{ N + K \le j' \le M : d_{k+1} \le \left\lceil \frac{\sum_{i=1}^{k} d_i + 1}{K} \right\rceil \text{ for all } k = 0, 1, \ldots, j' - 1 \right\}.$$

In other words, if $N + K \le j \le M - 1$, then $j$ is the unique positive integer in $\{N+K, N+K+1, \ldots, M-1\}$ such that $d_{k+1} \le \lceil \frac{\sum_{i=1}^{k} d_i + 1}{K} \rceil$ for all $k = 0, 1, \ldots, j - 1$ and $d_{j+1} \ge \lceil \frac{\sum_{i=1}^{j} d_i + 1}{K} \rceil + 1$. On the other hand, if $j = M$, then $d_{k+1} \le \lceil \frac{\sum_{i=1}^{k} d_i + 1}{K} \rceil$ for all $k = 0, 1, \ldots, M - 1$.

We claim that

$$B \le \sum_{i=1}^{j} d_i. \tag{70}$$

If $j = M$, then (70) holds trivially as the $M$ fibers with delays $d_1, d_2, \ldots, d_M$ can only accommodate a maximum of $\sum_{i=1}^{M} d_i$ packets at any time. On the other hand, if $N+K \le j \le M-1$, then we prove (70) by contradiction. Suppose on the contrary that $B \ge \sum_{i=1}^{j} d_i + 1$. Assume that the priority queue is empty at time $t$, there is an arriving packet at time $t+1, t+2, \ldots, t + \sum_{i=1}^{j} d_i$, and the control input is disabled at time $t + 1, t +$

$2,\ldots,t+\sum_{i=1}^{j} d_i$. Then it is easy to see that there are $\sum_{i=1}^{j} d_i$ packets in the priority queue at time $t+\sum_{i=1}^{j} d_i$ and they must be stored in the $j$ fibers with delays $d_1, d_2, \ldots, d_j$ at time $t+\sum_{i=1}^{j} d_i$. Otherwise, if at least one of the $\sum_{i=1}^{j} d_i$ packets in the priority queue at time $t+\sum_{i=1}^{j} d_i$ is stored in one of the $M-j$ fibers with delays $d_{j+1}, d_{j+2}, \ldots, d_M$ at time $t+\sum_{i=1}^{j} d_i$, say the fiber with delay $d_{j'}$, where $j+1 \le j' \le M$, then that packet must have entered the fiber with delay $d_{j'}$ at some time $t'$ with $t+1 \le t' \le t+\sum_{i=1}^{j} d_i$. In the case that the control input is enabled at time $t'+1, t'+2, \ldots$ and there are no arriving packets at time $t'+1, t'+2, \ldots$, the packet entering the fiber with delay $d_{j'}$ at time $t'$ cannot depart from the priority queue during $[t', t'+d_{j'}-1]$, and we will have a contradiction to the nonidling property in (P2) in Definition 1 requiring that all the $t'-t$ packets stored in the priority queue at time $t'$ should depart from the priority queue during $[t'+1, t'+\lceil \frac{t'-t}{K} \rceil]$ (the contradiction can be seen from $\lceil \frac{t'-t}{K} \rceil \le \lceil \frac{\sum_{i=1}^{j} d_i}{K} \rceil \le \lceil \frac{\sum_{i=1}^{j} d_i + 1}{K} \rceil \le d_{j+1}-1 \le d_{j'}-1$). Furthermore, assume that there is an arriving packet at time $t+\sum_{i=1}^{j} d_i+1$, there are no arriving packets at time $t+\sum_{i=1}^{j} d_i+2, t+\sum_{i=1}^{j} d_i+3, \ldots$, and the control input is disabled at time $t+\sum_{i=1}^{j} d_i+1$ and is enabled at time $t+\sum_{i=1}^{j} d_i+2, t+\sum_{i=1}^{j} d_i+3, \ldots$. As we just showed that there are $\sum_{i=1}^{j} d_i$ packets stored in the $j$ fibers with delays $d_1, d_2, \ldots, d_j$ at time $t+\sum_{i=1}^{j} d_i$, these $j$ fibers are full of packets at time $t+\sum_{i=1}^{j} d_i$ so that there are $j$ packets at the outputs of these $j$ fibers at the beginning of the $(t+\sum_{i=1}^{j} d_i+1)^{\text{th}}$ time slot. Since the control input is disabled at time $t+\sum_{i=1}^{j} d_i+1$, it follows that among the packet arriving at time $t+\sum_{i=1}^{j} d_i+1$ and the $j$ packets that appear at the outputs of these $j$ fibers at the beginning of the $(t+\sum_{i=1}^{j} d_i+1)^{\text{th}}$ time slot, at least one of them must be stored in the first cell of one of the $M-j$ fibers with delays $d_{j+1}, d_{j+2}, \ldots, d_M$ at time $t+\sum_{i=1}^{j} d_i+1$, say the fiber with delay $d_{j''}$, where $j+1 \le j'' \le M$, as there are at most $j$ packets that can enter the $j$ fibers with delays $d_1, d_2, \ldots, d_j$ at time $t+\sum_{i=1}^{j} d_i+1$. It follows that the packet entering the fiber with delay $d_{j''}$ at time $t+\sum_{i=1}^{j} d_i+1$ cannot depart from the priority queue during $[t+\sum_{i=1}^{j} d_i+1, t+\sum_{i=1}^{j} d_i+d_{j''}]$, and we will have a contradiction to the nonidling property in (P2) in Definition 1 requiring that all the $\sum_{i=1}^{j} d_i+1$ packets stored in the priority queue at time $t+\sum_{i=1}^{j} d_i+1$ should depart from the priority queue during $[t+\sum_{i=1}^{j} d_i+2, t+\sum_{i=1}^{j} d_i+\lceil \frac{\sum_{i=1}^{j} d_i+1}{K} \rceil+1]$ (the contradiction can be seen from $\lceil \frac{\sum_{i=1}^{j} d_i+1}{K} \rceil+1 \le d_{j+1} \le d_{j''}$). As such, we have shown that (70) also holds if $N+K \le j \le M-1$.

Let $b = \lceil \frac{N+K+1}{K} \rceil$. As $d_i = 1$ for $i = 1, 2, \ldots, N+K$ and $d_{k+1} \le \lceil \frac{\sum_{i=1}^{k} d_i+1}{K} \rceil$ for all $k = 0, 1, \ldots, j-1$, we can use the inequality that $\lceil x+y \rceil \le \lceil x \rceil + \lceil y \rceil$ for $x, y \ge 0$ to deduce that

$$
\begin{aligned}
d_{N+K+1} &\le \left\lceil \frac{\sum_{i=1}^{N+K} d_i + 1}{K} \right\rceil = \left\lceil \frac{N+K+1}{K} \right\rceil = b, \\
d_{N+K+2} &\le \left\lceil \frac{\sum_{i=1}^{N+K+1} d_i + 1}{K} \right\rceil \\
&\le \left\lceil \frac{\sum_{i=1}^{N+K} d_i + 1}{K} \right\rceil + \left\lceil \frac{d_{N+K+1}}{K} \right\rceil \\
&\le b + \left\lceil \frac{b}{K} \right\rceil \le b + \left( \frac{b}{K} + 1 \right) \\
&= (b+K)\left(1 + \frac{1}{K}\right) - K, \\
d_{N+K+3} &\le \left\lceil \frac{\sum_{i=1}^{N+K+2} d_i + 1}{K} \right\rceil \\
&\le \left\lceil \frac{\sum_{i=1}^{N+K+1} d_i + 1}{K} \right\rceil + \left\lceil \frac{d_{N+K+2}}{K} \right\rceil \\
&\le (b+K)\left(1 + \frac{1}{K}\right) - K \\
&\quad + \left\lceil \frac{(b+K)\left(1+\frac{1}{K}\right) - K}{K} \right\rceil \\
&\le (b+K)\left(1 + \frac{1}{K}\right) - K + \frac{(b+K)\left(1+\frac{1}{K}\right)}{K} \\
&= (b+K)\left(1 + \frac{1}{K}\right)^2 - K, \\
&\ \ \vdots \\
d_j &\le \left\lceil \frac{\sum_{i=1}^{j-1} d_i + 1}{K} \right\rceil \\
&\le \left\lceil \frac{\sum_{i=1}^{j-2} d_i + 1}{K} \right\rceil + \left\lceil \frac{d_{j-1}}{K} \right\rceil \\
&\le (b+K)\left(1 + \frac{1}{K}\right)^{j-N-K-2} - K \\
&\quad + \left\lceil \frac{(b+K)\left(1+\frac{1}{K}\right)^{j-N-K-2} - K}{K} \right\rceil \\
&\le (b+K)\left(1 + \frac{1}{K}\right)^{j-N-K-2} - K \\
&\quad + \frac{(b+K)\left(1+\frac{1}{K}\right)^{j-N-K-2}}{K} \\
&= (b+K)\left(1 + \frac{1}{K}\right)^{j-N-K-1} - K.
\end{aligned}
$$

Therefore, it follows from (70), $b = \lceil \frac{N+K+1}{K} \rceil \le \frac{N+2K}{K}$, and $N+K \le j \le M$ that

$$
\begin{aligned}
B &\le \sum_{i=1}^{j} d_i = \sum_{i=1}^{N+K} d_i + \sum_{i=N+K+1}^{j} d_i \\
&\le (N+K) \\
&\quad + \sum_{i=N+K+1}^{j} \left( (b+K)\left(1+\frac{1}{K}\right)^{i-N-K-1} - K \right) \\
&= (N+K) + (b+K)K\left( \left(1+\frac{1}{K}\right)^{j-N-K} - 1 \right) \\
&\quad - K(j-N-K) \\
&\le (N+K) + (K^2+2K+N)\left( \left(1+\frac{1}{K}\right)^{j-N-K} - 1 \right)
\end{aligned}
$$

$$\leq (K^2 + 2K + N)\left(1 + \frac{1}{K}\right)^{M-N-K}$$
$$= (K^2 + 2K + N)2^{(M-N-K)\log_2(1+\frac{1}{K})}.$$

The proof is completed.

## References


[1] M. J. Karol, "Shared-memory optical packet (ATM) switch," in *Proceedings SPIE : Multigigabit Fiber Communication Systems (1993)*, October 1993, vol. 2024, pp. 212–222.

[2] Z. Hass, "The "staggering switch": An electronically controlled optical packet switch," *IEEE Journal of Lightwave Technology*, vol. 11, pp. 925–936, May/June 1993.

[3] I. Chlamtac and A. Fumagalli, "Quadro-star: A high performance optical WDM star network," *IEEE Transactions on Communications*, vol. 42, pp. 2582–2591, August 1994.

[4] I. Chlamtac, A. Fumagalli, L. G. Kazovsky, P. Melman, W. H. Nelson, P. Poggiolini, M. Cerisola, A. N. M. M. Choudhury, T. K. Fong, R. T. Hofmeister, C.-L. Lu, A. Mekkittikul, D. J. M. Sabido IX, C.-J. Suh, and E. W. M. Wong, "Cord: contention resolution by delay lines," *IEEE Journal on Selected Areas in Communications*, vol. 14, pp. 1014–1029, June 1996.

[5] R. L. Cruz and J.-T. Tsai, "COD: alternative architectures for high speed packet switching," *IEEE/ACM Transactions on Networking*, vol. 4, pp. 11–21, February 1996.

[6] D. K. Hunter, D. Cotter, R. B. Ahmad, D. Cornwell, T. H. Gilfedder, P. J. Legg, and I. Andonovic, "$2 \times 2$ buffered switch fabrics for traffic routing, merging and shaping in photonic cell networks," *IEEE Journal of Lightwave Technology*, vol. 15, pp. 86–101, January 1997.

[7] D. K. Hunter, W. D. Cornwell, T. H. Gilfedder, A. Franzen, and I. Andonovic, "SLOB: a switch with large optical buffers for packet switching," *IEEE Journal of Lightwave Technology*, vol. 16, pp. 1725–1736, October 1998.

[8] E. Varvarigos, "The "packing" and the "scheduling packet" switch architectures for almost all-optical lossless networks," *IEEE Journal of Lightwave Technology*, vol. 16, pp. 1757–1767, October 1998.

[9] I. Chlamtac, A. Fumagalli, and C.-J. Suh, "Multibuffer delay line architectures for efficient contention resolution in optical switching nodes," *IEEE Transactions on Communications*, vol. 48, pp. 2089–2098, December 2000.

[10] Y.-T. Chen, C.-S. Chang, J. Cheng, and D.-S. Lee, "Feedforward SDL constructions of output-buffered multiplexers and switches with variable length bursts," in *Proceedings IEEE International Conference on Computer Communications (INFOCOM'07)*, Anchorage, AK, USA, May 6–12, 2007.

[11] I. Chlamtac, A. Fumagalli, and C.-J. Suh, "Optimal $2 \times 1$ multi-stage optical packet multiplexer," in *Proceedings IEEE Global Telecommunications Conference (GLOBECOM'97)*, Phoenix, AZ, USA, November 3–8, 1997, pp. 566–570.

[12] C.-S. Chang, D.-S. Lee, and C.-K. Tu, "Recursive construction of FIFO optical multiplexers with switched delay lines," *IEEE Transactions on Information Theory*, vol. 50, pp. 3221–3233, December 2004.

[13] C.-S. Chang, D.-S. Lee, and C.-K. Tu, "Using switched delay lines for exact emulation of FIFO multiplexers with variable length bursts," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 108–117, April 2006.

[14] C.-C. Chou, C.-S. Chang, D.-S. Lee and J. Cheng, "A necessary and sufficient condition for the construction of 2-to-1 optical FIFO multiplexers by a single crossbar switch and fiber delay lines," *IEEE Transactions on Information Theory*, vol. 52, pp. 4519–4531, October 2006.

[15] J. Cheng, "Constructions of fault tolerant optical 2-to-1 FIFO multiplexers," *IEEE Transactions on Information Theory*, vol. 53, pp. 4092–4105, November 2007.

[16] J. Cheng, "Constructions of optical 2-to-1 FIFO multiplexers with a limited number of recirculations," *IEEE Transactions on Information Theory*, vol. 54, pp. 4040–4052, September 2008.

[17] J. Cheng, C.-S. Chang, T.-H. Chao, D.-S. Lee, and C.-M. Lien, "On constructions of optical queues with a limited number of recirculations," in *Proceedings IEEE International Conference on Computer Communications (INFOCOM'08)*, Phoenix, AZ, USA, April 13–18, 2008.

[18] X. Wang, X. Jiang, and S. Horiguchi, "Improved bounds on the feedfoward design of optical multiplexers," in *Proceedings International Symposium on Parallel Architectures, Algorithms, and Networks (I-SPAN'08)*, Sydney, Australia, May 7–9, 2008, pp. 178–183.

[19] S.-Y. R. Li and X. J. Tan, "Fiber memory," in *Proceedings Annual Allerton Conference on Communication, Control, and Computing (Allerton'06)*, Monticello, IL, USA, September 27–29, 2006. Full version of this paper is submitted to *IEEE Transactions on Information Theory*.

[20] C.-S. Chang, Y.-T. Chen, and D.-S. Lee, "Constructions of optical FIFO queues," *IEEE Transactions on Information Theory*, vol. 52, pp. 2838–2843, June 2006.

[21] B. A. Small, A. Shacham, and K. Bergman, "A modular, scalable, extensible, and transparent optical packet buffer," *Journal of Lightwave Technology*, vol. 25, pp. 978–985, April 2007.

[22] P.-K. Huang, C.-S. Chang, J. Cheng, and D.-S. Lee, "Recursive constructions of parallel FIFO and LIFO queues with switched delay lines," *IEEE Transactions on Information Theory*, vol. 53, pp. 1778–1798, May 2007.

[23] N. Beheshti and Y. Ganjali, "Packet scheduling in optical FIFO buffers," in *Proceedings IEEE High-Speed Networks Workshop (HSN'07)*, Anchorage, AK, USA, May 11, 2007.

[24] X. Wang, X. Jiang, and S. Horiguchi, "A construction of shared optical buffer queue with switched delay lines," in *Proceedings IEEE International Conference on High Performance Switching and Routing (HPSR'08)*, Shanghai, China, May 15–17, 2008, pp. 86–91.

[25] A. D. Sarwate and V. Anantharam, "Exact emulation of a priority queue with a switch and delay lines," *Queueing Systems: Theory and Applications*, vol. 53, pp. 115–125, July 2006.

[26] H.-C. Chiu, C.-S. Chang, J. Cheng, and D.-S. Lee, "A simple proof for the constructions of optical priority queues," *Queueing Systems: Theory and Applications*, vol. 56, pp. 73–77, June 2007

[27] H. Kogan and I. Keslassy, "Optimal-complexity optical router," in *Proceedings IEEE International Conference on Computer Communications (INFOCOM'07 Minisymposium)*, Anchorage, AK, USA, May 6–12, 2007.

[28] H. Rastegarfar, M. Ghobadi, and Y. Ganjali, "Emulation of Optical PIFO Buffers," in *Proceedings IEEE Global Communications Conference (GLOBECOM'09)*, Honolulu, HI, USA, November 30–December 4, 2009.

[29] F. Jordan, D. Lee, K. Y. Lee, and S. V. Ramanan, "Serial array time slot interchangers and optical implementations," *IEEE Transactions on Computers*, vol. 43, pp. 1309–1318, 1994.

[30] Y.-T. Chen, J. Cheng, and D.-S. Lee, "Constructions of linear compressors, non-overtaking delay lines, and flexible delay lines for optical packet switching," *IEEE/ACM Transactions on Networking*, vol. 17, pp. 2014–2027, December 2009. Conference version appeared in *IEEE INFOCOM* 2006.

[31] C.-S. Chang, J. Cheng, T.-H. Chao, and D.-S. Lee, "Optimal constructions of fault tolerant optical linear compressors and linear decompressors," *IEEE Transactions on Communications*, vol. 57, pp. 1140–1150, April 2009. Conference version appeared in *IEEE INFOCOM* 2007.

[32] T.-H. Chao, C.-S. Chang, D.-S. Lee, and J. Cheng, "Constructions of multicast flexible delay lines and optical multicast switches with 100% throughput," in *Proceedings IEEE Global Telecommunications Conference (GLOBECOM'07)*, Washington DC, USA, November 26–30, 2007.

[33] H.-W. Lan, C.-S. Chang, J. Cheng, and D.-S. Lee, "Constructions and analysis of crosstalk-free optical queues," in *Proceedings IEEE International Conference on High Performance Switching and Routing (HPSR'08)*, Shanghai, China, May 15–17, 2008, pp. 27–32.

[34] D.-S. Lee, K.-J. Hsu, C.-S. Chang, and J. Cheng, "Emulation and approximation of a flexible delay line by parallel non-overtaking delay lines," in *Proceedings IEEE International Conference on Computer Communications (INFOCOM'09)*, Rio de Janeiro, Brazil, April 19–25, 2009.

[35] D.-S. Lee, C.-S. Chang, J. Cheng, H.-S. Chueh, and K.-T. Wang, "Emulation of an optical flexible delay line by parallel variable optical delay lines," *IEEE Communications Letters*, vol. 14, pp. 770–772, August 2010.

[36] C.-S. Chang, J. Cheng, and D.-S. Lee, "SDL constructions of FIFO, LIFO and absolute contractors," in *Proceedings IEEE International Conference on Computer Communications (INFOCOM'09)*, Rio de Janeiro, Brazil, April 19–25, 2009.

[37] H. Kogan and I. Keslassy, "Fundamental complexity of optical systems," in *Proceedings IEEE International Conference on Computer Communications (INFOCOM'07 Minisymposium)*, Anchorage, AK, USA, May 6–12, 2007.

[38] D.-S. Lee, C.-S. Chang, J. Cheng, and H.-S. Yan, "Queueing analysis of loss systems with variable optical delay lines," in *Proceedings IEEE International Conference on Computer Communications (INFOCOM'08)*, Phoenix, AZ, USA, April 13–18, 2008.

[39] J. Liu, T. T. Lee, X. Jiang, and S. Horiguchi, "Blocking and delay analysis of single wavelength optical buffer with general packet size distribution," *IEEE Journal of Lightwave Technology*, vol. 27, pp. 955–966, April 2009.
[40] D. K. Hunter, M. C. Chia, and I. Andonovic, "Buffering in optical packet switches," *IEEE Journal of Lightwave Technology*, vol. 16, pp. 2081–2094, December 1998.
[41] S. Yao, B. Mukherjee, and S. Dixit, "Advances in photonic packet switching: An overview," *IEEE Communications Magazine*, vol. 38, pp. 84–94, February 2000.
[42] D. K. Hunter and I. Andonovic, "Approaches to optical Internet packet switching," *IEEE Communications Magazine*, vol. 38, pp. 116–122, September 2000.
[43] S. Yao, B. Mukherjee, S. J. B. Yoo, and S. Dixit, "A unified study of contention-resolution schemes in optical packet-switched networks," *Journal of Lightwave Technology*, vol. 21, pp. 672–683, March 2003.
[44] S. J. B. Yoo, "Optical packet and burst switching technologies for the future photonic internet," *Journal of Lightwave Technology*, vol. 24, pp. 4468–4492, December 2006.
[45] K. Miklós, "Congestion resolution and buffering in packet switched all-optical networks," Ph.D. Dissertation, Budapest University of Techonology and Economics, Budapest, Hungary, 2008.
[46] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated service networks: the single-node case," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 344–357, June 1993.

**Jay Cheng** (S'00-M'03-SM'09) received the B.S. and M.S. degrees from National Tsing Hua University, Hsinchu, Taiwan, R.O.C., in 1993 and 1995, respectively, and the Ph.D. degree from Cornell University, Ithaca, NY, USA, in 2003, all in Electrical Engineering. In August 2003, he joined the Department of Electrical Engineering at National Tsing Hua University, Hsinchu, Taiwan, R.O.C., where he is currently an Associate Professor. Since October 2004, he has also been affiliated with the Institute of Communications Engineering at National Tsing Hua University, Hsinchu, Taiwan, R.O.C. His current research interests include optical queueing theory, high-speed switching theory, network science, communications theory, and information theory.

**Hsien-Chen Chiu** received the B.S. degree in Electrical Engineering and the M.S. degree in Communications Engineering from National Tsing Hua University, Hsinchu, Taiwan, R.O.C., in 2004 and 2006, respectively. He is now with HTC Corporation, Taipei, Taiwan, R.O.C.

**Cheng-Shang Chang** (S'85-M'86-M'89-SM'93-F'04) received the B.S. degree from National Taiwan University, Taipei, Taiwan, R.O.C., in 1983, and the M.S. and Ph.D. degrees from Columbia University, New York, NY, USA, in 1986 and 1989, respectively, all in Electrical Engineering. From 1989 to 1993, he was a Research Staff Member at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA. Since 1993, he has been with the Department of Electrical Engineering at National Tsing Hua University, Hsinchu, Taiwan, R.O.C., where he is currently a Professor. His current research interests are in high speed switching, communication network theory, and mathematical modeling of the Internet. Dr. Chang received an IBM Outstanding Innovation Award in 1992, an IBM Faculty Partnership Award in 2001, and Outstanding Research Awards from the National Science Council, Taiwan, R.O.C., in 1998, 2000 and 2002, respectively. He also received Outstanding Teaching Awards from both the college of EECS and the university itself in 2003. He was appointed as the first Y. Z. Hsu Scientific Chair Professor in 2002. He is the author of the book "Performance Guarantees in Communication Networks" and a coauthor of the book "Principles, Architectures and Mathematical Theory of High Performance Packet Switches." He served as an Editor for Operations Research from 1992 to 1999 and as an Associate Editor for IEEE/ACM Transactions on Networking from June 2007 to June 2009. Dr. Chang is a member of IFIP Working Group 7.3.

**Duan-Shin Lee** (S'89-M'90-SM'98) received the B.S. degree from National Tsing Hua University, Hsinchu, Taiwan, R.O.C., in 1983, and the M.S. and Ph.D. degrees from Columbia University, New York, NY, USA, in 1987 and 1990, respectively, all in Electrical Engineering. He was a research staff member at the C&C Research Laboratory of NEC USA, Inc., Princeton, NJ, USA, from 1990 to 1998. He joined the Department of Computer Science at National Tsing Hua University, Hsinchu, Taiwan, R.O.C., in 1998, where he is currently a Professor. His research interests are in high-speed switch and router design, wireless networks, performance analysis of communication networks, and queueing theory.