Message Sequencing Techniques for On-Line Scheduling in WDM Networks

Babak Hamidzadeh
University of British Columbia
Ma Maode, and Mounir Hamdi
Hong Kong University of Science and Technology

Abstract: *Message sequencing and channel assignment are two important issues that need to be addressed in scheduling variable-length messages in a Wavelength Division Multiplexing (WDM) network. Channel assignment addresses the problem of choosing an appropriate data channel via which a message is transmitted to a node. This problem has been addressed extensively in the literature. On the other hand, message sequencing which addresses the order in which messages are sent, has rarely been addressed. In this paper, we propose a set of scheduling techniques for single-hop WDM passive star networks which address both the sequencing aspect and the assignment aspect of the problem. In particular, we develop two priority schemes for sequencing messages in a WDM network in order to increase the overall performance of the network. We evaluate the proposed algorithms, using analytical modeling and discrete-event simulations, by comparing their performance with state-of-the-art scheduling algorithms that only address the assignment problem. We find that significant improvement in performance can be achieved using our scheduling algorithms where message sequencing and channel assignment are simultaneously taken into consideration.*

1. Introduction

Wavelength Division Multiplexing (WDM) is an effective way of utilizing the large bandwidth of an optical fiber. By allowing multiple messages to be transmitted in parallel, on a number of channels, this technique has the potential to improve the performance in optical networks significantly. Several topologies have been proposed for WDM networks [1][2], a popular one of which is the single-hop, passive star-coupled topology [3].

To unleash the potential of single-hop, WDM passive star networks, efficient access protocols and scheduling algorithms are needed to allocate and coordinate system resources optimally while satisfying message and system constraints [1]. Most of these protocols and algorithms can be divided into two main classes, namely preallocation-based [3][4][5][6] and reservation-based [7][8][9][10][11] techniques. Preallocation-based techniques use all channels of a fiber to transmit messages. These techniques assign transmission rights to different nodes in a static and pre-determined manner. Reservation-based techniques allocate a channel as the control channel to transmit global information about messages to all nodes in the system. Once such information is received, all nodes invoke the same scheduling algorithm to determine when to transmit/receive a message and on which data channel. Reservation-based techniques have a more dynamic nature and assign transmission rights based on run-time requirements of the nodes in the network. In this paper, we focus our attention on reservation-based techniques.

Most of the scheduling algorithms proposed for reservation-based techniques can only schedule fixed-length packets for transmissions. Recently, many researchers have relaxed this constraint by allowing their scheduling algorithms to schedule variable-length messages [9] [12] [13] [14]. As a result, these variable-length scheduling algorithms are more general than fixed-length scheduling algorithms and can adapt better to various traffic characteristics (e.g., bursty). We adopt the same strategy in this paper by allowing our scheduling algorithm to handle variable-length messages. There are two fundamental aspects that a variable-length message scheduling algorithm should efficiently solve namely, channel assignment and message sequencing. The assignment aspect of a scheduling algorithm addresses the problem of selecting an appropriate channel and a time slot on that channel to transmit a message. The sequencing aspect, addresses the order in which messages are selected for transmission. The assignment aspect of this problem has been addressed extensively in the literature. The sequencing aspect of this problem, however, has not received much attention. In particular, all the above proposed variable-length messages scheduling algorithms schedule messages individually and independently of one another [9][12][13][14].

In this paper we propose and evaluate a set of scheduling techniques that address the sequencing, as well as the assignment aspect of the scheduling problem. Our techniques are more globally optimizing than the existing approaches, since they not only share global information about each message among receiving and transmitting nodes, but they also consider multiple messages from different transmitting nodes simultaneously, when scheduling.

We have developed a theoretical model to analyze the performance of the techniques discussed in this paper. In addition, we evaluated our techniques by comparing their performance with a recently proposed scheduling algorithm [9] using extensive discrete-event simulations. The results of these experiments demonstrate the significant improvements that can be obtained by using techniques that address sequencing and assignment simultaneously.

The remainder of this paper is organized as follows. Section 2 specifies our WDM system model and the scheduling problem to be addressed. Section 3 discusses our scheduling techniques and Section 4 provides an analytical model of the proposed techniques. Section 5 provides an experimental evaluation of these technique's performance. Finally, Section 6 concludes the paper with a summary of the results and a discussion of our future work.

0-7803-4198-8/97/$10.00 © 1997 IEEE

2. WDM System Model and the Scheduling Problem

As mentioned previously, in this paper we consider message transmission in a single-hop, WDM optical network whose nodes are connected via a passive star coupler. The star coupler supports C channels and there exist N nodes in the network. C-1 channels, referred to as data channels, are used for message transmission. The other channel, referred to as the control channel, is used to exchange global information among nodes about the messages to be transmitted. The control channel is the basic mechanism for implementing the reservation scheme. Each node in the network has two transmitters and two receivers. One transmitter and one receiver are fixed and are tuned to the control channel. The other transmitter and receiver are tunable and can tune into any of the data channels to access messages on those channels. This is similar to the network proposed in [9]. The nodes are divided into two non-disjoint sets of source (transmitting) nodes s_i and destination (receiving) nodes d_j. A queue for the messages to be transmitted is assumed to exist at each source node s_i.

A Time Division Multiple Access (TDMA) protocol is used on the control channel to access that channel. According to this protocol, each node can transmit a control packet during a predetermined time slot. The basic time interval on the control channel is the transmission time of a control packet. N control packets make up one control frame on the control channel. Thus, each node has a corresponding control packet in a control frame, during which that node can access the control channel. The length of a control packet is a system design parameter and depends on the number of messages l about which each node is allowed to broadcast control information, and the amount of control information about each message (e.g., the address of the destination node, message length). Values greater than one for the parameter l signify a situation in which a source node s_i can transmit information about multiple messages in its queue to all nodes through a control packet i in a control frame. In our model, we have ignored the transmitter and receiver tuning times.

3. Scheduling Techniques

In this section, we discuss the basic steps of our scheduling techniques and discuss some of their performance tradeoffs. During the transmission of a control frame, each source node s_i sends a control packet during time slot i on the control channel to all other nodes. The control packet contains information about one (at the head of s_i's message queue) or more messages that it intends to transmit. The larger the number l of messages that are represented in a control packet, the more globally optimizing our scheduling algorithms will be. Larger values of l result in longer durations but less frequent invocations of the scheduling algorithms.

After $R+F$ time units, where R is the round-trip propagation delay between a node and the star coupler and F is the time duration of a control frame, all nodes in the network will have information contained in a control frame about messages to be transmitted. At this point, an identical copy of a distributed scheduling algorithm is invoked by all nodes to assign the messages represented in the control frame to appropriate data channels to be transmitted at a point in time. The technique for assignment of data channels and transmission time may vary based on different models. Examples of such techniques that are receiving attention are EATS, CDS and TTAS as proposed in [9]. For the sake of simplicity and to be able to clearly illustrate the importance of sequencing messages in these networks, we adopt EATS (*Earliest Available Time Scheduling*) as our basic channel assignment mechanism. However, the choice of channel assignment technique in our approach is independent of our sequencing algorithms. EATS assigns a message to the data channel that has the earliest available time among all channels. Once the data channel is assigned, the message is scheduled to transmit as soon as that channel becomes available.

We have selected two priority schemes for sequencing messages in our scheduling algorithms, namely the Shortest Job First (SJF) and the Longest Job First (LJF) schemes. By scheduling shorter messages first, SJF is expected to reduce average delays. SJF's ability to reduce average delays has been demonstrated. In an environment where messages can be transmitted in parallel on different data channels, however, SJF is expected to result in a poorly balanced load among different channels. This is because the larger messages that are scheduled last may have large differences in size which will lead to a coarser schedule with uneven loads among channels. This is why we have chosen LJF as an alternative priority scheme to see the trade-off between load balancing and reducing average delays in our algorithms. LJF is expected to balance the load by first scheduling long messages on data channels and then filling the uneven loads with smaller messages.

Sequencing messages at the source-node message queues or messages represented in the control frame, or both, can lead to a number of different scheduling policies some of which we have adopted and evaluated. In the following sub-sections, we discuss some of these strategies and their characteristics.

3.1. Frame Scheduling

Using this strategy, each message queue, at the source nodes, is maintained as a First-Come-First-Served (FCFS) queue. During each time slot i, control information about the message at the head of s_i's queue is placed in packet i of a frame. After all packets of a frame reach all nodes in the network, a sequencing algorithm based on a priority scheme (e.g., SJF or LJF) is called to sort the messages represented in that frame according to their priorities. Once the order of message transmissions is determined, a channel assignment algorithm (e.g., EATS) is invoked to assign the channel and time of transmission. The source nodes will then know on which channel to transmit the message at the head of their message queues and at what time. The receiver nodes will also know to which channel they should tune and at what time to receive the appropriate message.

Prioritizing message transmissions in frame scheduling does not lead to starvation, since this prioritization takes place in batches all of whose messages receive service before the next batch of messages is scheduled/serviced.

3.2. Frame-and-Queue Scheduling

Using this strategy, sequencing is done at two points; once at the message queues of the source nodes and once at the time the messages of a control frame are scheduled. The message queues at the source nodes are maintained according to some priority scheme (e.g., SJF or LJF). Thus, the head of each queue contains the message m with the highest priority among all messages that have arrived at a source for transmission. During time slot i, therefore, control information about message m will be placed in the appropriate control packet of the control frame. Once all packets of the frame have reached all nodes, a sequencing algorithm (based on the same priority scheme as the one at the message queues) is applied again to sequence the messages represented in that frame.

A point to note is that the frame-and-queue scheduling technique may lead to starvation for some messages at the message queues. This is because a higher-priority message can always arrive into a message queue and replace existing lower-priority messages at the head of the queue. Thus, some form of aging mechanism should be adopted for this kind of technique to increase the priority of messages as they stay longer in the message queues.

3.3. Multiple-Messages-per-Node Scheduling

This technique attempts to do scheduling at a more global level than the previous two approaches. To do this, it represents a number l ($l \geq 1$) of messages in each control packet. Thus, control information about multiple messages at each source node's message queue can be placed in each of the corresponding control packets of a frame. This technique performs sequencing once at the time the frame has reached all nodes. The sequencing algorithm is applied to all messages of the frame and an order is imposed on the messages according to some priority scheme as discussed before. After scheduling, each source node will know which message in its queue is to be transmitted next. The source nodes will also know on which channel to transmit the message and at what time.

Like the frame scheduling technique, the multiple-messages-per-node technique is free from starvation. This is again attributed to the fact that this technique schedules messages in independent batches. Since this technique schedules more messages in each scheduling phase than the previous two approaches, its scheduling time is higher. The frequency of scheduling invocations, on the other hand, is lower since a larger number of messages is scheduled each time.

4. Analytical Model

In this section we present a simple analytical for our WDM network that follows the analytical model originally proposed in [9]. The performance metric of our interest in this model is the average message delay in the network. In order to make our WDM model mathematically manageable, several assumptions have been adopted as follows: 1) The tuning time is negligible. 2) We have a finite message population, M, at the head of each node's queue. 3) The message arrival process at each node is a Poisson process with a mean arrival rate of l. 4) A message transmitted by a node is destined to every other node with equal probability. 5) For each of the nodes i, the message length is exponentially distributed with a mean value of 1/m'i. 6) For each of the nodes, the probability that each node has one message is approximately 1/N.

The frame scheduling algorithms introduced in the previous sections provide mechanisms to sequence the messages according to length-based priorities assigned to each message (i.e., SJF or LJF). As a result, a WDM system that adopts one of these algorithms can be modeled as an M/G/1 with a priority queuing system [17][18][19]. The population of the system queue in this model is bounded by the number of the nodes, since we consider the system queue as being composed of every first message at each node (i.e., the head of every node's queue). The servers of the queue can be considered as the set of data channels in the system with different service rates. The service rate of a channel depends on the messages it serves combined with the restriction on message destinations.

The message population is limited to one per node with the same arrival rate λ and probability $1/N$. The arrival rate of the system can be approximated by $(N-k)\times\lambda/N$, where the system state k is the number of messages in the system and $k \in \{0,1,\ldots, N\text{-}1\}$.

The service rate, which is the inverse of the service time of the system server can be considered as a function of two factors. One is the mean message service rate μ'_i of the message with the ith priority. The other is $a_i(k)$ which denotes the probability that out of k messages, i messages are destined to different nodes. This term signifies that the destination of a message plays a role in determining the service rate of the system server. The service rate μ_k of the server, when kth priority message is served in the system, can be expressed by:

When $k < C$ - 1:

$$\mu_k = \sum_{i=1}^{k} \mu'_i \times i \times a_i(k)$$

When $k \geq C$ - 1:

$$\mu_k = \sum_{i=1}^{C-1} \mu'_i \times i \times a_i(C-1)$$

These formulas demonstrate the effect of the number of channels on the service rate. According to assumption 4, $a_i(k)$ can be computed as follows [9]:

$$a_i(k) = \frac{\binom{N}{i} \times i! \times S(k,i)}{N^k}$$

where $S(k, i)$ is the Stirling number.

The system traffic intensity or the load P_k, which is the ratio of the messages arrival rate to the service rate of the system, when kth priority message is served by the server, can be expressed by the following relationship:

$$P_k = \frac{\lambda_k}{\mu_k} = \frac{(N-k)\times\lambda/N}{\sum_{i=1}^{k} \mu'_k \times i \times a_i(k)}$$

Applying Little's result to our M/G/1 priority queuing system, we can obtain the relationships between the average delay D_k of the kth priority message, and the average waiting time DW_k of the kth priority message in the queue. Finally, we can derive the average delay time D of all messages in the system. In particular, the waiting time D_k of the kth priority message can be expressed as follows:

$$DW_k = \frac{\sum_{i=1}^{k} \frac{p_i}{u_i}}{2\times\left(1-\sum_{i=1}^{k} p_i\right)\times\left(1-\sum_{i=1}^{k-1} p_i\right)}$$

The delay time D_k of the kth priority message can be obtained from adding that message's service time $1/\mu_k$, to the waiting time DW_k of the kth priority message in the queue.

$$D_k = \frac{1}{\mu_k} + DW_k$$

Based on the above formulae, we can calculate the average delay time of all messages in the system as:

5. Experimental Evaluation

In this section, we discuss the results of a set of experiments to evaluate the performance of the proposed scheduling techniques and also compare them with the scheduling scheme adopted in [9]. The experiments were conducted using a discrete-event simulator.

5.1. Experiment Design

The parameters involved in the design of our WDM system include the number of nodes, which was chosen to be 100, and the number of channels, which ranges from 50 to 100. Round-trip propagation delay is another system parameter which was set to 10 in the experiments. Message lengths vary according to a normal distribution with a mean of 100 time units and a standard deviation of 50. A uniform message arrival rate across all nodes was considered which ranges from 0.004 to 0.007 messages per unit time for each node in the network. Destination nodes for messages were chosen according to a uniform probability distribution. The behavior of the candidate algorithms was observed over a simulation period of 100,000 time units to make sure that our results are stable. Each point in the performance graphs is the average of 10 independent runs. Metrics of performance in the experiments are *average delay*, defined as the average time a message spends in the WDM network, and *throughput* defined as the number of packets that are transmitted per unit of time.

The channel assignment strategy chosen for all candidate algorithms is the EATS technique as proposed in [9]. This technique assigns a message to the data channel with the earliest available time. The candidate algorithms for the performance-comparison experiments were First-Control-Packet-First-Served (FCPFS), Frame scheduling with SJF (F-SJF) and with LJF (F-LJF) priority schemes, Frame-and-Queue scheduling with SJF (FQ-SJF) and with LJF (FQ-LJF) priority schemes, and Multiple-Messages-per-Node scheduling with SJF (MMN-SJF) and with LJF (MMN-LJF) priority schemes. The operation of F-SJF, F-LJF, FQ-SJF, FQ-LJF, MMN-SJF, and MMN-LJF were discussed in previous sections. The value of parameter l in MMN algorithms was set to 7. The FCPFS algorithm is the basic algorithm against which our proposed algorithms are compared which was originally given in [9]. This algorithm does not sequence messages in any particular order and assigns them to the data channels according to the index of their control packets in the control frame. This means that a message originated at source node s_1 whose corresponding control packet in the control frame is packet 1, will be scheduled before a message originated at source node s_2 with the second control packet as its corresponding slot in the control frame.

5.2. Experimental Results

Figures 1 and 2 show the performance of the different algorithms under varying loads (arrival rates) per node. Figure 1 compares the average delay of the algorithms. As the figure shows, the algorithms which perform both sequencing and assignment (e.g., F, FQ and MMN) significantly outperform those which perform only assignment (e.g., FCPFS), as arrival rates increase. The figure also reveals that as the degree of globally optimizing behavior increases, the algorithms' performance consistently improves (i.e., MMN outperforms FQ and FQ outperforms F).

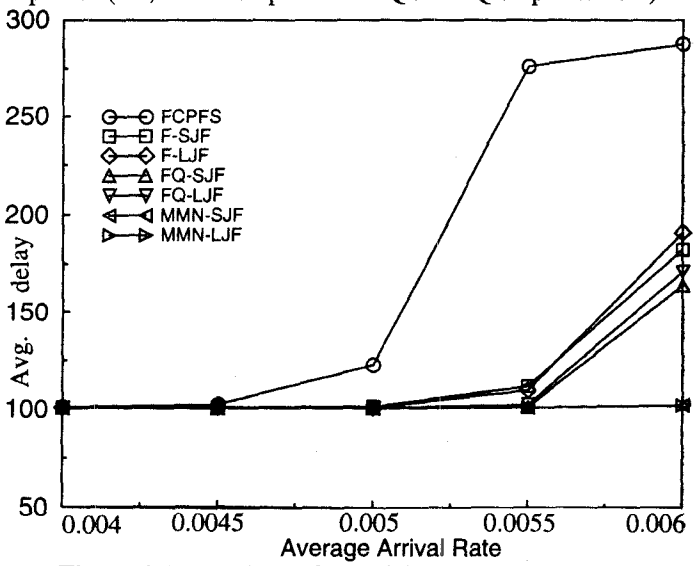

Figure 1. Comparison of avg. delay vs. arrival rates.

The F algorithms can outperform FCPFS by as much as 40%. The FQ and MMN algorithms outperform FCPFS by as much as 45%

and 66%, respectively. This is an affirmation on the importance of efficiently sequencing messages in variable-length message scheduling algorithms. As a general trend, algorithms using the SJF priority scheme perform slightly better, in terms of reducing average delay, than those employing LJF. Figure 2 compares the throughput of the candidate algorithms. We observe that the sequencing-and-assignment techniques result in similar throughputs that saturate at significantly higher values than EATS's throughput at high loads.

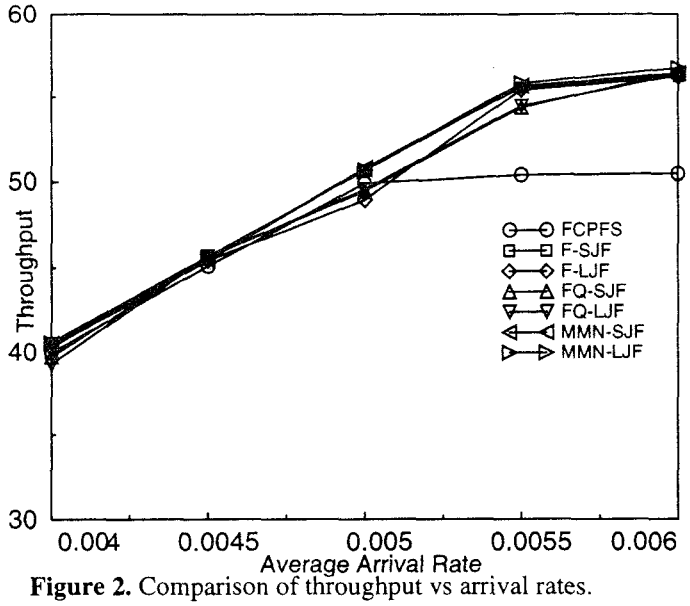

Figure 2. Comparison of throughput vs arrival rates.

Figure 3 compares the results of the analytical model developed in the previous section to those obtained through extensive discrete-event simulations. The analytical results agree well with the experimental result and confirm that F-SJF demonstrates improved performance with respect to FCPFS. These results also verify the accuracy of the analytical model and its usefulness in adopting various priority schemes for sequencing messages.

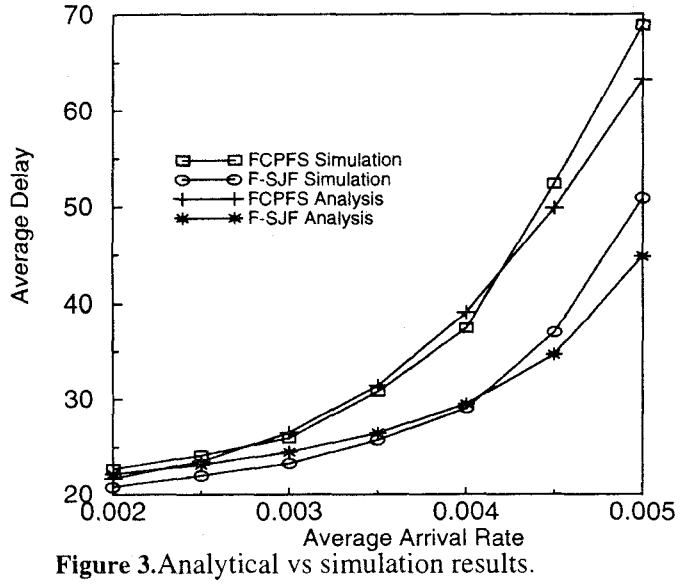

Figure 3. Analytical vs simulation results.

6. Conclusion

In this paper, we proposed a set of reservation-based techniques for scheduling variable-length messages in a single-hop, WDM passive star network. Unlike many existing reservation-based techniques, the proposed techniques address both message sequencing and channel assignment aspects of the scheduling problem simultaneously. We formulated a mathematical model to study the performance of the proposed techniques. We also evaluated the performance of the proposed techniques and the tradeoffs between the priority schemes in a number of experiments. These experiments compared the proposed algorithms with another scheduling technique which only addresses channel assignment problem (no message sequencing) [9]. The results of our experiments show significant improvements over this channel assignment technique, and the results of our mathematical analysis support these conclusions as well.

References

1. B. Mukherjee, "WDM-based Local Lightwave Networks- Part I: Single-Hop Systems," *IEEE Network*, pp. 12-27, May 1992.
2. B. Mukherjee, "WDM-based Local Lightwave Networks- Part II: Multi-Hop Systems," *IEEE Network*, pp. 20-32, July 1992.
3. K. Bogineni, K. M. Sivalingam, and P. W. Dowd, "Low-Complexity Multiple Access Protocols for Wavelength-Division Multiplexed Photonic Networks," *IEEE Journal on Selected Areas of Communications*, 11 (4), pp. 590-603, May 1993.
4. A. Ganz and Y. Gao, "Time-Wavelength Assignment Algorithms for High Performance WDM Star Based Systems," *IEEE Transactions on Communications*, 42(2/3/4), pp. 1827-1836, Feb/ Mar/April 1994.
5. G. N. Rouskas and M. H. Ammar, "Analysis and Optimization of Transmission Schedules for Single-Hop WDM Networks," *IEEE/ ACM Transactions on Networking*, pp. 211-221, April 1995.
6. M. S. Borella, and B. Mukherjee, "Efficient Scheduling of Nonuniform Packet Traffic in a WDM/TDM Local Lightwave Network with Arbitrary Transceiver Tuning Latencies," *IEEE Journal on Selected Areas in Communications*, pp. 923-934, June 1996.
7. K. Bogineni and P. W. Dowd, "A Collisionless Multiple Access Protocol for a Wavelength Division Multiplexed Star-Coupled Configuration: Architecture and Performance Analysis," *Journal of Lightwave Technology*, 10(11), pp. 1688-1699, November 1992.
8. R. Chipalkatti, Z. Zhang, and A. S. Acampora, "Protocols for Optical Star-Coupler Network using WDM: Performance and Complexity Study," *IEEE Journal on Selected Areas of Communications*, 11(4), pp. 579-589, May 1993.
9. F. Jia, B. Mukherjee, and J. Iness, "Scheduling Variable-Length Messages in a Single-Hop Multichannel Local Lightwave Network," *IEEE/ACM Transactions on Networking*, Vol. 3, No. 4, pp. 477-487, August 1995.
10. C. S. Li, M. S. Chen, and F. F. K. Tong, "POSMAC: A Medium Access Protocol for Packet-Switched Passive Optical Networks using WDMA," *Journal of Lightwave Technology*, 11(5/6), pp. 1066-1077, May/June 1993.
11. N. Mehravari, "Performance and Protocol Improvements for Very High-Speed Optical Fiber Local Area Networks using a Passive Star Topology," *Journal of Lightwave Technology*, 8(4), pp. 520-530, April 1990.
12. J. H. Lee and C. K. Un, "Dynamic Scheduling Protocol for Variable-sized Messages in a WDM-based Local Network," *Journal of Lighwave Technology*, pp. 1595-1600, July 1996.
13. H. Jeon and C. Un, "Contention-based Reservation Protocols in Multiwavelength Optical Networks with a Passive Star Topology," in *Proc. IEEE ICC*, pp. 1473-1477, June 1992.
14. A. Muir and J. J. Garcia-Luna-Aceves, "Distributed Queue Packet Scheduling Algorithms for WDM-Based Networks, in *Proc. IEEE INFOCOM '96*, pp. 938-945, 1996.
15. Kleinrock, Leonard. *Queueing systems*. Wiley, 1975-1976.
16. I. Mitrani. *Modelling of Computer and Communication Systems*. Cambridge University Press,1987.
17. M. K. Molloy. *Fundamentals of Performance Modelling*. Macmillan Publishing Company,1989.