

Contention Resolution by Retransmission in Single-hop OPS Metro Networks

Akbar Ghaffar Pour Rahbar*

Computer Networks Research Lab, Sahand University of Technology, Sahand New Town, Tabriz, Iran
e-mail: ghaffarpour@sut.ac.ir

Oliver W.W. Yang

CCNR Lab, University of Ottawa, Ottawa, ON, Canada
e-mail: yang@site.uottawa.ca

Abstract— We consider a single-hop slotted all-Optical Packet-Switched (OPS) metro network where contention is the major problem. Since contention resolution hardware such as optical buffers and wavelength converters are expensive, we do not use any of these contention resolution hardware inside a core switch. Instead, we consider the retransmission in the optical domain which appears to be the cheapest solution. We study the prioritized retransmission schemes where dropped traffic has a priority when retransmitted from edge switches. Our study shows that the number of retransmissions can be significantly reduced. We present the analysis in a single-fiber network where additional wavelength channels are used to carry the same traffic. In addition to verifying our analysis, our simulation results demonstrate the capability and the advantage of PR over the traditional RR (Random Retransmission) as well as the advantage of using multiple wavelengths over using multiple fibers.

Index Terms—slotted OPS, retransmission in the optical domain, additional wavelengths, contention avoidance.

I. INTRODUCTION

Future metro networks should provide more capacity in order to cope not only with the current traffic loads but also with the unexpected future demand growth [1]. Due to the high traffic dynamics, Optical Packet Switching (OPS) is necessary for metro networks in order to use network resources efficiently [2]. Therefore, all-optical packet-switched networks appear to be the sole approach to provide such capacity. A common all-optical OPS network architecture maintains data payload in the optical domain from source to destination. However, the control signaling may be processed in the electronic domain from source to destination. Note optical time division multiplexing can provide a finer granularity in OPS and can improve bandwidth usage so that bandwidth can be shared by many source-destination pairs.

One common network topology for a metro network is the single-hop star topology. Apart from the design simplicity, the synchronization required in a TDM network can be easily achieved in an all-optical star-based network [3]. Since a star network suffers from the central node failure problem, the overlaid star topology [4-6] has been considered to provide robustness and reliability. Path reliability has been studied, e.g., [7], and traffic protection problem studied in [8].

Contention is the major problem in an OPS network. Optical buffering, deflection routing, and wavelength conversion are the basic contention resolution techniques [9]. However, optical buffering is expensive and bulky [10]. The wavelength conversion technique is a much more feasible solution, but also very expensive. When a very low loss rate is required, the required number of wavelength converters [11] and optical fiber buffer lengths [12] will drastically increase, and clearly this will result in a very high network cost. Deflection routing is the cheapest technique. However, it cannot be applied for a single-hop metro network because there is only one core switch in the network.

One can use less expensive contention resolution hardware to reduce optical switch cost. However, the lost traffic must be recovered by retransmission at the optical layer. This is because the retransmission of the lost traffic by higher layers may cause the false TCP congestion detection problem even in lower loads [15]. With electronic buffers decreasing in price, it may be worthwhile to revisit retransmission issue in the optical layer to see if a proper management of the dropped traffic can be found.

The conventional retransmission method, called Random Retransmission (RR) in this paper, transmits dropped data until a successful transmission has already been completed. RR has been used in optical networks

* Corresponding author

Manuscript received May 3, 2007; accepted Aug. 8, 2007.

[16-18]. Since the number of retransmissions is not limited, the multiple retransmissions usually lead to retransmissions at higher levels, which may in turn increase the traffic load on the network. The prioritization of the retransmissions is a technique that was first proposed in the wireless and TCP domain [19, 20]. We have for the first time applied and analyzed in [21, 22] the Prioritized Retransmission (PR) scheme in slotted all-optical OPS networks. The scenario of a single-fiber network without using any contention avoidance is analyzed in [21], while a multi-fiber architecture, as a contention avoidance technique [13], is analyzed in [22]. It was shown that PR can improve TCP throughput by limiting the number of retransmissions [22]. One should note, however, that the improvement in performance using multi-fiber architecture along with using PR comes with a price. This is because in a multi-fiber architecture, a larger switch size ($nf \times nf$ instead of $n \times n$, where n is the number of network nodes and f is the number of fibers used between each edge switch and the core switch) is required to support a large number of nodes in the metro environment. This may not be economically suitable because a large number of port counts is required.

This paper extends the results of our previous work in [21, 22] by applying the Prioritized Retransmission (PR) scheme in a slotted-OPS metro network environment. Instead of using extra fibers in [22], we investigate the improvement by using additional wavelength channels in a single-fiber OPS architecture. Our contributions are 1) on the analysis and the use of Prioritized Retransmission in a slotted single-hop all-optical OPS metro network where no contention resolution mechanism (such as wavelength converters and optical buffers) is used at the core switch, 2) performance evaluation and comparison with a single-fiber network without using any contention avoidance as well as the multi-fiber architecture that uses more fibers to avoid loss in the network.

For the remainder of this paper, the following general symbols and notations pertain.

H	Maximum number of retransmissions needed to transmit a slot successfully
L	Normalized traffic load on wavelength channels
N	Average number of non-empty slots in an n -slot-set
N_k	Number of slots arrived to a tagged output link at retransmission level $\geq k$
$P_a\{N_k=a\}$	Arrival probability of a slots to a tagged output link at retransmission level $\geq k$
$P_{i,drop}$	Probability of slot drop at i -th retransmission level
$R_{L,i}\{n_r, k\}$	Slot loss rate for more than k slots among n_r tagged-priority slots
W	Number of required data wavelengths on a fiber link
W_a	Number of additional wavelengths on a fiber link
n	Number of core switch inputs ports
n -slot-set	Set of slots on the same wavelength coming from the n input ports of a core switch
n_r	Average number of tagged-priority slots in an n -slot-set
$n_{r,i}$	Average number of tagged-priority slots in an n -slot-set at retransmission level i
w_i	Wavelength number i on a fiber
π_i	Probability that a slot is retransmitted for the i -th time

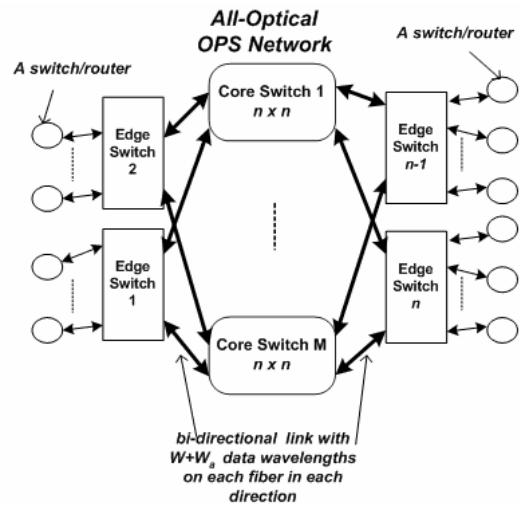


Figure 1: The Single-hop OPS Network Model

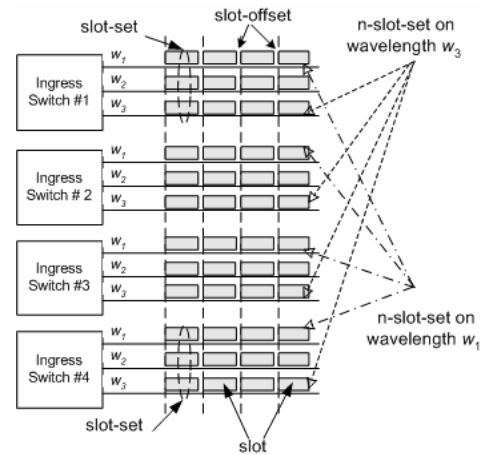


Figure 2: Diagram of Slot Transmission

II. NETWORK MODEL, DEFINITIONS, AND ASSUMPTIONS

We consider an all-optical overlaid star network with a time-slotted operation (see Fig.1) that includes M all-optical wavelength-selective cross-connect switches (hereafter each referred to as a core switch), each located at the star center. Each core switch is connected to n edge switches (each with an electronic buffer). Each edge switch may be connected to a number of switches/routers. Each core switch is an $n \times n$ switch. To design an inexpensive network, we do not use any contention resolution hardware such as optical buffers or wavelength converters inside the core switches. The operation of each core switch in switching traffic is independent of other core switches, which means the overlaid topology can be divided into independent single-star networks. One advantage of the overlaid star topology is to support load balancing by equally dividing its network traffic among the several parallel (overlaid) single-star networks.

Each connection link needs W wavelengths to carry all of its traffic. However, additional W_a wavelengths are allowed on each fiber so that an ingress switch (i.e., a transmitting edge switch) can have up to $W + W_a + 1$ wavelengths on a fiber to transmit its traffic where the

one extra wavelength is for control purposes.

Slotted operation is used to support packet switching. Each wavelength channel is divided into fixed-interval optical time-slots as shown in Fig.2. Each ingress switch can transmit traffic in the time-slots on any wavelength. In each time-slot, an integer number of IP packets can be carried. Traffic in each time-slot is referred to as a *slot* from now on (see *slot* in Fig.2). When this traffic is transmitted/retransmitted to the network we refer to it as transmitting/re-transmitting a slot. Each edge switch can transmit up to $W + W_a$ slots (called a *slot-set*) at the same time to the core switch. Fig.2 shows slot-sets for Ingress Switches #1 and #4 where each ingress switch has three wavelengths. An *empty slot* is a slot with no traffic to any egress (receiver) switch. Each time-slot is separated by a small time gap (called slot-offset). This gap includes the guard time (for timing uncertainties), the processing time at the core switch, and the switching time.

Each ingress switch saves a transmitted slot for further retransmission in its electronic retransmission buffer that is organized by linked lists. Whenever, a slot is dropped at the core switch, the core switch sends a NACK (Negative ACK) command carrying the slot ID back to its source ingress switch. Then the ingress switch will retransmit the backup of the dropped slot at a random time and a random wavelength to the network. If no NACK is received within a timeout period (with a minimum duration equal to twice the one-way propagation delay to the core switch plus some processing time), an ingress switch can remove the backup slot from its retransmission buffer. Both the PR and RR to be compared use this mechanism.

Before transmitting each slot-set and during the offset time, the slot-set header has to be sent over the control channel. The header may include traffic information for each slot in the slot-set such as the ingress switch address, the egress switch address, the slot ID, information about the traffic that is carried in the slot, and the slot priority. The slot ID and the slot priority are used for retransmission purpose. Note that the slot priority is only used under the PR scheme. Each transmitted slot carries a unique slot ID used for sequencing purposes at egress switches and retransmission purposes at ingress switches. The slot ID is fixed and never changed during transmission and all retransmissions. The number of bits in a slot ID should be long enough, e.g., 32 bits, to avoid the wrap-around problem of slot numbers.

For the remainder of this paper, the following definitions and assumptions pertain:

- L is the normalized traffic load, normalized with respect to the channel bandwidth, on each wavelength channel when no additional wavelength channel is used on a fiber.
- L_e is the effective normalized traffic load, normalized with respect to the channel bandwidth, on each wavelength channel when there are W_a additional wavelength channels on a fiber.
- Arrival of slots is synchronized at the core switch.

- There is no error on the transmission/receiving links.
- Each ingress switch transmits its traffic on its wavelengths in a balanced manner in order to equalize the traffic load on all wavelength channels.

III. PRIORITIZED RETRANSMISSION (PR)

We use the now-established Prioritized retransmission (PR) scheme to recover the dropped slots and to limit the number of retransmissions. In PR, the priority of a newly transmitted slot is set to zero. Then, whenever a slot is dropped at the core switch, its priority is increased by one at its source ingress switch. So, even if a newly transmitted slot has a little chance to pass through the core switch at the first transmission during heavy traffic, the higher priority it would acquire in its subsequent retransmission(s) will help it to pass through the core switch eventually, thus cutting down the number of retransmissions. This mechanism can also be used to increase the fairness among different traffic streams as opposed to the RR scheme where one slot may pass through the core switch at the first transmission and another slot may be (theoretically) retransmitted forever.

The core switch first receives the slot information from all edge switches at the same time. After receiving a header, the core switch evaluates and resolves the potential contention during the slot-offset. Finally, the core switch is made ready to switch the incoming slots to their desired egress switches. During contention resolution, the retransmitted slots are given a higher priority to pass through the core switch. In general, an n -slot-set has slots with different numbers of retransmission times when competing for a tagged output link. So the available output ports of an output link are first given to those slots that have the highest number of retransmissions so far, and progressively retransmit those with less numbers down to those retransmitted for the first time, and finally the newly transmitted slots. It is understood that the retransmission are carried out subject to the availability of the remaining ports.

For a successful slot transmission, no ACK is required to be sent back to the source ingress switch. However, when a slot is dropped, the relevant slot code is encapsulated in a NACK command and sent back to the source ingress switch over control channel to identify the blocked slot. The ingress switch is responsible for retransmitting any blocked slot while increasing its priority number.

A. Slot Loss Rate

Let n -slot-set be the set of n slots from n ingress switches on the same wavelength in a given time-slot (see Fig.2). At each n -slot-set, at most n non-empty slots may arrive at the core switch on the same wavelength channel from n ingress switches. Considering the definition of the effective traffic load, the parameter L_e can also represent the probability of a non-empty slot arriving on each wavelength. Therefore, there are $N = nL_e$ non-empty slots in the n -slot-set on the average.

Let H be the maximum number of retransmissions

required to transmit a tagged slot successfully so that a slot with priority H is dropped with a probability of less than X where $X \ll 1$. The parameter X is a rare event probability defined to be the probability that a slot has not been successfully transmitted after it has been retransmitted for more than H times. Similar to [22], we start our analysis of the drop probability at level H and determine all slot retransmission probabilities by repeating the analysis process one level down at a time until we reach level zero. At level j during the analysis, we refer to the level- j retransmitted slots as the *tagged-priority* slots, and all the other non-empty slots in the n -slot-set are referred to as *non-tagged-priority slots*.

Define n_r to be the average number of non-empty tagged-priority slots at any level. Clearly, the remaining $N - n_r$ slots are the average number of non-tagged-priority slots. Define slot loss rate $R_L\{n_r, k\}$ to be the drop probability among the n_r tagged-priority slots destined to the tagged output link given that only k slots capacity ($k = 0$ or 1) are available at the tagged output link. Then, $R_L\{n_r, k\}$ can be obtained as [22]:

$$R_L\{n_r, k\} = \frac{\sum_{c=k+1}^{n_r} (c-k) \binom{n_r}{c} (n-1)^{n_r-c}}{n^{n_r-1} n_r} \quad \text{for } n_r \geq 1. \quad (1)$$

$$R_L\{n_r, k\} = 0, \quad \text{Otherwise.}$$

In order to reduce the slot loss rate in an OPS network, one can use additional wavelength channels on a fiber to transmit the same traffic. This is equivalent to lowering the traffic load in the network because the number of empty slots is now increased. Suppose there exists some traffic that can occupy all W slots in a slot-set. By using W_a additional wavelength channels, there are now $W+W_a$ slots for the transmission of W slots in each slot-set. Let the slots be uniformly distributed over the $W+W_a$ wavelengths. Therefore, the probability of one non-empty slot on a wavelength channel is $W/(W+W_a)$. Considering the normalized traffic load L in an ingress switch and using W_a additional wavelength channels on a fiber, the effective normalized traffic load (L_e) on each wavelength channel is given by:

$$L_e = \frac{W}{W+W_a} L$$

B. Retransmission Distribution at Steady-State

In this analysis, we make the assumption that there may not always be enough capacity to transmit all slots requiring retransmission at level j , and some of them have to be dropped. Their priority will increase by one, and they are retransmitted as level $j+1$ priority slots. Let $\Pi_H = \{\pi_0 \pi_1 \pi_2 \dots \pi_H \pi_{H+1}\}$ denote the probability vector that at most H retransmissions are required in an n -slot-set to send a slot to the tagged output link in steady state, where π_0 is the transmission probability of new slots; and π_i is the probability that a slot is retransmitted for the i -th time. Let $P_{i,drop}$ denote the probability of dropping a slot at the i -th retransmission level. Similar to [22], we can

derive the following equations to obtain Π_H :

$$\pi_i = P_{i-1,drop}, \quad 1 \leq i \leq H \quad (2)$$

$$\pi_{H+1} = P_{H,drop} < X \quad (3)$$

$$\pi_0 = 1 - \sum_{j=1}^H \pi_j \quad (4)$$

To solve these equations, we have to first find $P_{i,drop}$ which must consider the probability that the retransmitted slots at levels $\geq i+1$ have already occupied an output port in the tagged output link. Let N_k denote the number of slot arrivals to the tagged output link at a retransmission level $\geq k$, and let $P_a\{N_k=\alpha\}$ be the corresponding probability of having α slot arrival ($\alpha = 0$ or 1) such that α ports of the tagged output link are occupied. Using the assumptions of uniform slot arrivals to each output link, and equal normalized traffic load L_e on all wavelength channels, then we can obtain

$$P_a\{N_k = \alpha\} = \begin{cases} 1 - L_e \pi_k & , \text{if } \alpha = 0 \\ L_e \pi_k & , \text{if } \alpha = 1 \end{cases}$$

where $L_e \pi_k$ is the probability of one slot arrival (and occupation) to the tagged output link at retransmission level $\geq k$. Let $P_{i,\alpha,drop}$ be the probability that a tagged-priority slot is dropped at retransmission level i , provided that α slots have already arrived and occupied the tagged-output link ports at retransmission level $i+1$. Then,

$$P_{i,\alpha,drop} = \frac{n_{r,i} P_a\{N_{i+1} = \alpha\} R_L\{n_{r,i}, 1-\alpha\}}{N}$$

where the numerator represents the average number of dropped slots among $n_{r,i} = N\pi_i$ tagged-priority slots with priority i in the n -slot-set. The second factor in the numerator indicates the probability of α slot arrivals at levels $\geq i+1$, and the third term, calculated in (1), represents the drop probability among $n_{r,i}$ tagged-priority slots given that a capacity of only $1-\alpha$ slots is available at the tagged output link. The denominator is the average number N of non-empty slots in the n -slot-set. The marginal probability $P_{i,drop}$ is obtained as

$$P_{i,drop} = P_{i,0,drop} + P_{i,1,drop} \quad (5)$$

By simplifying (5), we obtain the loss rate for the retransmission level i , $P_{i,drop}$ as

$$\pi_{i+1} = P_{i,drop} = \pi_i ((1 - L_e \pi_{i+1}) R_L\{nL_e \pi_i, 1\} + (L_e \pi_{i+1}) R_L\{nL_e \pi_i, 0\}) \quad (6)$$

C. Scheduling at the Core Switch

We detail how a core switch resolves the contention at a given output link under both PR and RR. Let set $S_{i,l} = \{s_{i,0}, s_{i,1}, \dots, s_{i,l-1} \mid 0 < l \leq n\}$ denote the subset of l contending slots on wavelength w_i ($i=0, \dots, W+W_a-1$) at the output link. Let vector (Src_j, ID_j, r_j) denote the three parameters for slot s_j carried in SSH, where Src_j is the ingress switch address, ID_j is the slot ID, and r_j is the priority of slot j . The parameter r_j is not used under RR. Note when slot s_j is dropped, the core switch sends a NACK command that includes ID_j to Src_j under both PR and RR.

In PR, the slots in set $S_{i,l}$ are sorted in a descending order according to the priority value. Then the top slot is

picked for transmission first and the contention of the

solve the equations in (2) to (4) by the iteration method

TABLE I: THE PERFORMANCE COMPARISON OF PR AND RR AT $n=32$ AND $W_a=0$

Traffic Load L	Retrans. Scheme	Type of Result	π_0 (%)	π_1 (%)	π_2 (%)	π_3 (%)	π_4 (%)	π_5 (%)	π_6 (%)
0.4	P	Sim.	82.83551	16.60980	0.55405	0.00065	0.00000	0.00000	0.00000
		Ana.	83.48360	16.22550	0.29090	0.00000	0.00000	0.00000	0.00000
	R	Sim.	82.84037	14.21117	2.44124	0.42073	0.07181	0.01249	0.00214
		Ana.	83.4863	13.78668	2.27669	0.41992	0.07221	0.01025	0.00169
0.7	P	Sim.	72.46298	25.12895	2.38792	0.02015	0.00000	0.00000	0.00000
		Ana.	72.69410	25.19990	2.10600	0.00000	0.00000	0.00000	0.00000
	R	Sim.	72.46250	19.93331	5.49789	1.52186	0.42199	0.11771	0.03225
		Ana.	72.7044	19.84510	5.41684	1.47856	0.40358	0.11016	0.03007
1.0	P	Sim.	63.84280	30.52217	5.48361	0.15126	0.00000	0.00000	0.00000
		Ana.	63.82080	30.80280	5.37640	0.00008	0.00000	0.00000	0.00000
	R	Sim.	63.84134	23.02355	8.34649	3.03778	1.10841	0.40605	0.14948
		Ana.	63.7945	23.09712	8.36243	3.02766	1.09618	0.39688	0.14369

remaining $l-1$ slots is resolved using PR or RR. In this way, a slot with a higher priority always finds a higher chance to pass through the core switch. This contrasts with the RR scheme where slots are randomly chosen for transmission.

IV. PERFORMANCE EVALUATION

We would like to compare the performance of PR and RR based on our network model. Since parameter n_r in function $R_L\{n_r, k\}$ in (6) is not necessarily an integer value (due to multiplying N by a probability), we have to use the Gamma function to calculate both $factorial(n)$, i.e., $n! = \Gamma(n+1)$, and the *choose* function $\binom{m}{n} = \frac{m!}{n!(m-n)!}$ in

our analysis. In our network, each ingress switch transmits slots to other egress switches with equal probabilities. Fixed-length packets are generated according to a Poisson arrival process with a mean rate of L packets per time-slot. Each slot carries one packet. In the following simulations, more than 1,000,000 *n-slot-sets* are generated from all ingress switches. We have used C language and OPNET [14] to implement our computations and simulations respectively. For each scenario, enough replications are run to achieve a 95% level of confidence intervals to within 1% of the mean values shown.

To show the analytical results for RR, we also need to obtain the parameters π_i ($i=0,1,2,\dots$) for this scheme. Note that there is no difference between newly generated slots and previously retransmitted slots in RR. Using a geometric distribution, the probability of k transmissions until a success is $P\{k \text{ transmissions until success}\} = \pi_{k-1} = \pi_0(1-\pi_0)^{k-1}$, where π_0 is the probability of a successful transmission. The parameter π_0 can be obtained from $\pi_0 = 1 - R_L\{N, 1\} = 1 - R_L\{nL_e, 1\}$, in which the second term can be obtained in Section III.A.

We would also like to obtain a maximum value for H in order to compute π_i in our analysis. To do this, we can

solve the equations in (2) to (4) by the iteration method for different values of H , and then determine the maximum H . Note that by increasing the loss rate, the number of slot retransmissions is also increased. Therefore, the worst case happens when the loss rate in (1) is maximized. Since the function inside the summation in this equation is an increasing function of n_r , the maximum loss rate occurs when $n_r = n_{r,i} = nL_e\pi_i$ (at any retransmission level) is

maximized. The maximum value for n_r is obtained when n and L_e are maximized. By assuming $L=1.0$, $W_a=0$, $X=10^{-7}$, and $n=1000$ as the largest possible core switch dimension, and solving the $H+1$ equations, we find that $\pi_0=0.63230406$, $\pi_1=0.30779933$, $\pi_2=0.05816581$, $\pi_3=0.00172963$, $\pi_4 = 0.00000064$ and $\pi_5 \ll X$. Therefore, we have $H=4$ that will guarantee us the required loss level. Hence, we obtain the analysis results for the PR scheme using $X=10^{-7}$ and $H=4$.

A. Performance Comparison under Additional Wavelength Channels

To verify the correctness of our analysis, we consider a single-hop single-fiber OPS network with $n=32$ edge switches where each fiber (between an edge switch and the core switch) has $W=4$ wavelengths. Table I compares the retransmission analysis and simulation results (in percentage) under traffic load $L=0.4$, $L=0.7$, and $L=1.0$ when using no additional wavelength channel ($W_a=0$) are used. As discussed in Section III.A, the effective traffic load is the same as traffic load in this case, i.e., $L_e = L$. One can see that the analysis and the simulation results agree well for both PR and RR.

One can see that under the same retransmission scheme, the volume of traffic retransmission goes up by increasing traffic load L at each retransmission level. This is because the number of slot collisions goes up with increasing L . Therefore, the volume of traffic retransmission is the lowest at $L=0.4$ according to Table I. Under PR, for instance, $\pi_1 \approx 0.16$, $\pi_2 \approx 0.25$, and $\pi_3 \approx 0.30$ for $L=0.4$, $L=0.7$, and $L=1.0$ respectively. A similar behavior is observed for other retransmission levels.

Both the analysis and simulation results indicate that most parts of the dropped slots can pass through the core switch within two retransmissions under PR, and only less than 0.2% would require the third retransmission. For the worst case of $L=1.0$, for instance, Table I shows that almost 64% of the traffic at the core switch are new arrivals, and almost 31% of the slots will be retransmitted once. While only about 5.5% and 0.1% of the slots require the second and the third retransmissions

TABLE II: THE PERFORMANCE COMPARISON OF PR AND RR AT $n=32$ AND $W_a=2$

Traffic Load L	Retrans. Scheme	Type of Result	π_0 (%)	π_1 (%)	π_2 (%)	π_3 (%)	π_4 (%)	π_5 (%)	π_6 (%)
0.4	P	Sim.	88.09640	11.72374	0.17981	0.00005	0.00000	0.00000	0.00000
		Ana.	88.98840	11.01160	0.00000	0.00000	0.00000	0.00000	0.00000
	R	Sim.	88.09996	10.48239	1.24879	0.14862	0.01778	0.00216	0.00026
		Ana.	88.99666	9.79265	1.07753	0.11856	0.01305	0.00144	0.00016
0.7	P	Sim.	80.36900	18.79065	0.83869	0.00167	0.00000	0.00000	0.00000
		Ana.	80.90700	18.54660	0.54640	0.00000	0.00000	0.00000	0.00000
	R	Sim.	80.36987	15.77077	3.09954	0.61004	0.12027	0.02372	0.00465
		Ana.	80.90665	15.44788	2.94954	0.56317	0.10753	0.02053	0.00392
1.0	P	Sim.	73.52772	24.34137	2.11599	0.01492	0.00000	0.00000	0.00000
		Ana.	73.79330	24.38070	1.82600	0.00000	0.00000	0.00000	0.00000
	R	Sim.	73.52653	19.45219	5.15588	1.36908	0.36405	0.09717	0.02585
		Ana.	73.8024	19.33446	5.06516	1.32695	0.34763	0.09107	0.02386

TABLE III: THE PERFORMANCE COMPARISON OF PR AND RR AT $n=32$ AND $W_a=4$

Traffic Load L	Retrans. Scheme	Type of Result	π_0 (%)	π_1 (%)	π_2 (%)	π_3 (%)	π_4 (%)	π_5 (%)	π_6 (%)
0.4	P	Sim.	90.87974	9.04059	0.07966	0.00001	0.00000	0.00000	0.00000
		Ana.	91.88780	8.11220	0.00000	0.00000	0.00000	0.00000	0.00000
	R	Sim.	90.89203	8.27745	0.75475	0.06884	0.00632	0.00056	0.00005
		Ana.	91.9391	7.41112	0.59740	0.04816	0.00388	0.00031	0.00003
0.7	P	Sim.	84.76053	14.85502	0.38419	0.00026	0.00000	0.00000	0.00000
		Ana.	85.59860	14.40130	0.00005	0.00000	0.00000	0.00000	0.00000
	R	Sim.	84.75985	12.91605	1.96935	0.30054	0.04591	0.00694	0.00109
		Ana.	85.4963	12.40013	1.79848	0.26085	0.03783	0.00549	0.00080
1.0	P	Sim.	79.16905	19.81999	1.00839	0.00257	0.00000	0.00000	0.00000
		Ana.	79.65180	19.63120	0.71690	0.00000	0.00000	0.00000	0.00000
	R	Sim.	79.17264	16.48128	3.43709	0.71840	0.15010	0.03177	0.00688
		Ana.	79.6575	16.20433	3.29637	0.67056	0.13641	0.02775	0.00564

respectively. There is no loss after the fourth retransmission. On the other hand, the number of retransmissions is not limited in RR so that we can even see a slot that is retransmitted for the 18-th time.

Under RR, the rate of the reduction of retransmission probability at each retransmission level is smoother than what we can observe for PR. For example at $L=1.0$, the third retransmission occurs with probability 0.0015 under PR, where as this event happens with almost probability 0.03 under RR. A similar behavior can be observed for $L=0.4$ and $L=0.7$.

Table II compares the retransmission analysis and simulation results at different traffic loads when using two additional wavelength channels on each fiber. Since in this case the effective traffic load on each wavelength channel is reduced to two-third of the traffic load (see Section III.A), the slot loss rate is less than the previous case (with no additional wavelength channels on a fiber) as expected. Performance behavior similar to Table I can be observed in Table II. However, the volume of traffic retransmission at each retransmission level and under the same retransmission scheme is lower than the case $W_a=0$ (see Table I). Therefore, under the same traffic load and the same retransmission scheme, π_0 is higher when using $W_a=2$ additional wavelength channels than using no

additional wavelength channel (i.e., $W_a=0$). For instance, under $L=0.7$ and using RR, $\pi_1 \approx 0.20$, $\pi_1 \approx 0.15$ at $W_a=0$ and $W_a=2$ respectively. Also, by using additional wavelength channels more traffic can pass through the core switch at the first transmission, i.e., $\pi_0 \approx 0.72$, $\pi_0 \approx 0.80$ at $W_a=0$ and $W_a=2$ respectively.

Table III illustrates the retransmission analysis and simulation results when using four additional wavelength channels on each fiber. Here, the effective traffic load on each wavelength channel is half of the traffic load, i.e., $L_e = 0.5 L$ (see Section III.A). So, we expect to have a lower slot loss rate than using $W_a=0$ and $W_a=2$. A similar behavior as discussed for Tables I and II can be observed

for Table III.

Comparing Table III with Tables I and II reveals that:

- 1) The case $W_a=4$ provides the best performance in reducing the volume of traffic that should be retransmitted as expected because using additional wavelength channels reduces traffic load and slot loss rate as a result. This leads to a lower number of retransmissions as a result.
- 2) The difference between the performance of PR and RR is reduced when traffic load is decreased or more additional wavelength channels are used on a fiber. This is because at a lower traffic load slots mostly pass the core switch due to a lower loss rate, and a small percentage of them require further retransmission. Therefore, a fewer number of retransmissions is required even for RR.
- 3) The rate of increase in π_0 and decrease in π_i ($i>0$) reduces when W_a goes up. Under PR at $L=1.0$, for example, using $W_a=2$ leads to $\pi_0 \approx 0.73$ and $\pi_1 \approx 0.24$. Comparing these results to the results of $W_a=0$ shows an increase of almost 15% for π_0 and a decrease of almost 20% for π_1 . On the other hand, using $W_a=4$, we have $\pi_0 \approx 0.79$ and $\pi_1 \approx 0.20$. When comparing the results of $W_a=4$ with the results of $W_a=2$, an increase of almost 7.7% for π_0 and a decrease of almost 18.5%

for π_1 can be observed. These two rates are lower than the previous rates (i.e., %15 and %20).

B. Performance Comparison under Additional Wavelength Channels and More Fibers

Here, we compare the performance of a single-hop OPS metro network with $n=32$ edge switches at traffic load $L=0.7$ under two network architecture scenarios: 1) a single-fiber network that uses W_a additional wavelength channels on a fiber on each connection link; and 2) a multi-fiber architecture with f fibers on each connection link. Both network scenarios use hardware-based contention avoidance techniques (i.e., either using additional wavelengths or using more fibers) in order to reduce collision in the network. In both scenarios, each fiber between and edge switch and the core switch has $W=4$ wavelengths.

Fig.3 compares the retransmission probabilities (in a logarithmic scale) of PR and RR under two contention avoidance schemes; only 6 retransmission probabilities (i.e., up to π_6) are shown. Under PR, there is no observation for more than three retransmissions. Since under PR, there is no observation for more than two retransmissions (i.e., $\pi_3=0$) in a multi-fiber architecture (i.e., $f > 1$), nothing is showed (due to the logarithmic scale) at the right hand side of the performance curve of π_3 in Fig.3.a. Since there is no retransmission observed for more than three times under PR, there is also no curve displayed for $\{\pi_i, i > 3\}$ in Fig.3.a. However, there are observations up to 14 retransmissions under RR.

In the diagrams, the case ($f=1, W_a=0$) at the center of the diagrams denotes to a single-fiber network ($f=1$) that uses no additional wavelength channels ($W_a=0$) as contention avoidance technique. Both diagrams show that by using more contention avoidance hardware, the probability of the arrival of newly generated traffic (i.e., π_0) in the network is increased. For example, under PR, π_0 at $f=1$ and $W_a=0$ has increased from almost 0.72 to almost 0.83 and 0.88 in a multi-fiber architecture with $f=2$ and $f=3$ fibers, respectively. The value of π_0 at $f=1$ and $W_a=0$ has also risen to almost 0.80 and 0.85 when using, respectively, $W_a=2$ and $W_a=4$ additional wavelength channels in a single-fiber network architecture. A similar behavior can also be observed in Fig.3.b. The diagrams also show that by using more contention avoidance hardware, the probability of retransmission at all levels (i.e., $\pi_i, i > 0$) is decreased. For example, under PR, π_2 at $f=1$ and $W_a=0$ has decreased from almost 0.023 to almost 0.0013 and 0.00006 in a multi-fiber architecture with $f=2$ and $f=3$ fibers, respectively. The value of π_2 at $f=1$ and $W_a=0$ has gone down to almost 0.0084 and 0.0038 at $W_a=2$ and $W_a=4$ additional wavelength channels in a single-fiber architecture, respectively. A similar behavior can also be observed under RR.

Among the four cases of contention avoidance, a multi-fiber architecture with $f=3$ fibers provides the highest π_0 , and the lowest $\pi_i, i > 0$. However, such an architecture needs a core switch of size 96×96 , which

is a large optical switch. When using additional wavelength channels, we still need a 32×32 core switch, but with a higher number of wavelengths on a fiber, say eight wavelengths on each fiber at $W_a=4$.

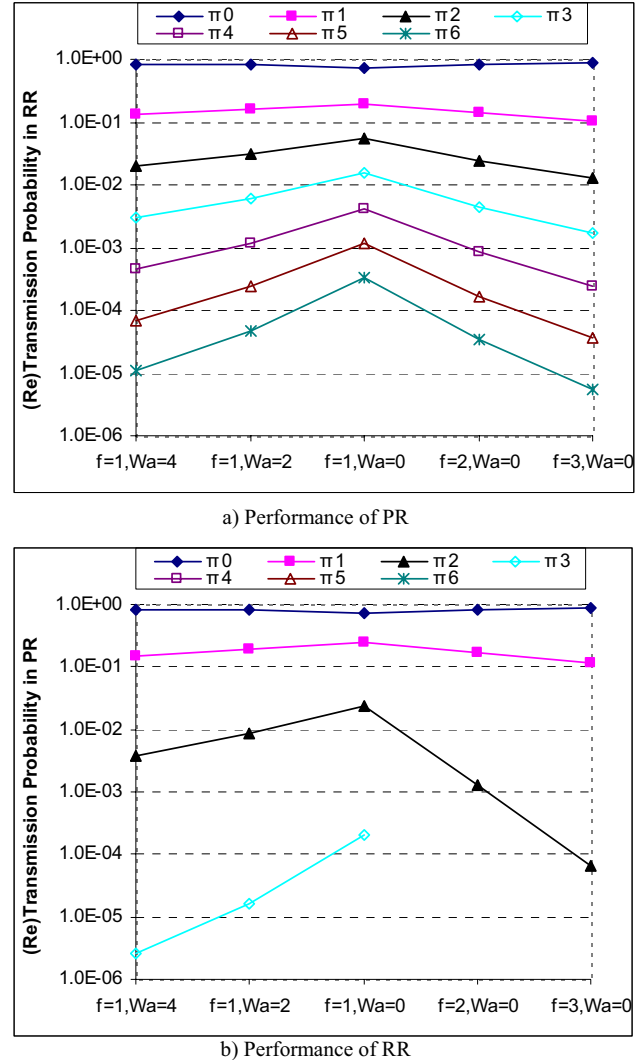


Figure 3: Performance Comparison of PR and RR under two Contention Avoidance Schemes

C. Discussion

The proposed PR protocol is simple since it only requires a few bits in SSH to keep the priority of each slot. For each slot a priority field is assigned that can only be updated at its source ingress switch. The SSH carries the priority of each slot. This field is the criterion by which the core switch decides to prioritize the slots and then finds the eligible slots for switching. On the other hand, each ingress switch sets the slot priority field with the retransmission number when (re)transmitting a slot. The proposed scheme can also be used in the asynchronous OPS and Optical Burst Switching (OBS) networks.

The PR scheme is much more effective than the RR scheme whenever the loss rate in an OPS network is not low. We expect such a loss rate whenever: 1) traffic load is high; and 2) a lower number of contention resolution/avoidance hardware has been used in a core switch. Therefore, we would recommend using PR in

networks with a medium or higher loss rate. In an OPS network with a lower loss rate (this low loss rate network can be achieved when using so many expensive contention resolution hardware in core switches), there may be no significant difference between the performance of PR and RR, although PR is always better than RR. On the other hand, RR has a lower computational complexity than PR. Thus, we recommend RR for such a network.

V. CONCLUSION

The Prioritized Retransmission (PR) scheme is shown to be a very simple but efficient protocol for slotted OPS networks. Our analysis and simulation results show that this scheme can limit the number of retransmissions, and can perform better than the conventional RR with any number of additional wavelength channels in a fiber and under any traffic load. In the worst case (i.e., using no additional wavelength channels under full traffic load), PR can pass most of the dropped slots through the core switch in two retransmissions. Clearly, by limiting the number of retransmissions, the extra load injected into the network due to the higher layer retransmissions can be reduced, which in turn helps to increase the network throughput. In addition, we showed that using additional wavelength channels can be a good approach to reduce contention in an OPS network and to reduce the required number of retransmissions both under PR and RR. In summary, employing PR for an OPS network can lower the number of contention avoidance/resolution hardware and therefore can reduce the network cost.

ACKNOWLEDGMENT

This research was financially supported by Sahand University of Technology, and also by the Canadian Natural Sciences and Engineering Research Council (NSERC) and industrial and government partners, through the Agile All-Photonic Networks (AAPN) Research Network.

REFERENCES

- [1] M.Maier and M.Reisslein, "AWG-based metro WDM networking," *IEEE Comm. Magazine*, vol.42, no.11, 2004, pp.S19-S26.
- [2] C.Papazoglou, G.Papadimitriou, and A.Pomportsis, "Design alternatives for optical-packet-interconnection network architectures," *Journal of optical Networking*, vol.3, no.11, 2004, pp.810-825.
- [3] F.J.Blouin, A.W.Lee, A.J.M.Lee, and M.Beshai, "Comparison of two optical-core networks," *Journal of Optical Networking*, vol.1, no.1, Jan.2002, pp.56-65.
- [4] Home page of "Agile All-Photonic Networks (AAPN)", <http://www.aapn.mcgill.ca/eng/index.html>, accessed in Jul. 2006.
- [5] M.Jin and O.Yang, "A TDM solution for all-photonic overlaid-star networks," in *Proc. of Information Sciences and Systems*, Princeton, USA, 2006, pp.1691-1695.
- [6] M.Jin and O.Yang, "APOSN: operation, modeling and performance evaluation" *Computer Networks (COMNET)*, vol.51, no.6, April 2007. pp.1643-1659.
- [7] Zhenxiao Liu and Oliver W. W. Yang, "Terminal-pair reliability analysis of overlaid-star networks", *Proc. CCCT2004*, Austin, Texas, Aug.2004, vol.4, pp.406-411.
- [8] C.Fan, M.Maier, and M.Reisslein, "The AWG||PSC network: a performance enhanced single-hop WDM network with heterogeneous protection," *Journal of Lightwave Technology*, vol.22, no.5, 2004, pp.1242-1262.
- [9] 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, no.3, Mar.2003.
- [10] I.Chlamtac, A.Fumagalli, et. al., "CORD: contention resolution by delay lines," *IEEE J. on Selected Areas in Communications*, vol.14, Jun.1996, pp.1014-1029.
- [11] V.Eramo, M.Listanti, and P.Pacifici, "A comparison study on the number of wavelength converters needed in synchronous and asynchronous all-optical switching architectures," *IEEE J. of Lightwave Technol.*, vol.21, no.2, Feb.2003, pp.340-355.
- [12] D.K.Hunter, M.C.Chia, and I.Andonovic, "Buffering in optical packet switches" *Journal of Lightwave Technology*, vol.16, no.12, Dec.1998.
- [13] Y.Li, G.Xiao and H.Ghafouri-Shiraz, "On the benefits of multifiber optical packet switch," *Microwave and Optical Technology Letter*, 43(5), pp.376-378, Dec. 2004.
- [14] <http://www.opnet.com/products/modeler/home.html>.
- [15] X.Yu, C.Qiao, and Y.Liu, "TCP implementation and false time out detection in OBS networks," *Proc. IEEE Infocom 2004*, Hong Kong, Mar. 2004.
- [16] I.Chlamtac, and A.Fumagalli, "QUADRO-Star: a high performance optical WDM star network", *IEEE Trans. on Communications*, vol.42, no.8, Aug.1994, pp.2582-2590.
- [17] E.Modiano, "Random algorithms for scheduling multicast traffic in WDM broadcast-and-select networks," *IEEE Trans. on Networking*, vol.7, no.3, Jun.1999, pp.425-434.
- [18] K.Samaras, D.C.O'Brien, and D.J.Edwards, "Analytical calculation of throughput of ALOHA based protocols in optical wireless data networks," *IEE Proc. part J-Optoelectronics*, vol.147, 2000, pp.322-328.
- [19] R.Rejaie, and A.R.Reibman, "Design issues for layered quality-adaptive Internet video playback", *Proc. Tyrrhenian International Workshop on Digital Communications*, Taormina, Italy, Sep. 2001.
- [20] C.Ladas, R.M.Edwards, M.Mahdavi, and G.A.Manson, "TCP retransmission prioritization for rapid recovery in slow and lossy networks," *Proc. European Personal Mobile Communications Conf.*, Glasgow, UK, Apr. 2003.
- [21] A.G.P.Rahbar and O.Yang, "Retransmission in slotted optical networks," *Proc. IEEE HPSR2006*, Poznan, Poland, June 2006, pp.21-26.
- [22] A.G.P.Rahbar and O.Yang, "Prioritized retransmission in slotted all-optical packet-switched networks," *OSA Journal of Optical Networking*, vol.5, no.12, Dec.2006.