Minimizing losses in max-min fair-share OBS networks

Tananun Orawiwattanakul[†], *Student Member, IEEE*, Yusheng Ji[††], *Member, IEEE*,

***Abstract*—Preemption is one of the most effective ways to achieve fair bandwidth allocation in OBS networks. Preemption allows ingress edge switches to transmit the traffic of flows over their fairly allocated bandwidth but core switches drop over-used traffic when there is contention. This paper proposes a rate-fairness preemption combined with deflection routing (RFP-DR) scheme to provide service isolation and protection among flows according to the max-min fair bandwidth allocation and minimize total loss probability in OBS networks. Our deflection routing (DR) technique aims to decrease loss probability while preserving the definition of max-min fairness. In addition, our DR can also be applied to rate control approaches. Our simulation results prove that our proposed combination of RFP and DR works well in terms of preserving max-min fair-share and minimizing loss.**

***Index Terms*—optical burst switching, wavelength preemption, deflection routing, max-min fairness**

I. Introduction

Preemption is one of the most popular approaches in OBS networks. It is a well-known solution for QoS provisioning, e.g. the probabilistic preemption-based (PPB) mechanism proposed in [1] and [2]. In addition, preemption is an effective way of providing distance fairness (fairness in terms of burst loss probability with respect to hop counts between the source and destination) in OBS networks as proposed in [3]-[5]. Finally, preemption has been used for achieving fair bandwidth allocation (FBA) in OBS networks. Apart from preemption, FBA can be achieved in OBS networks by using rate control, proposed in an integrated congestion control mechanism (ICCM) [6], and a differentiated-available bit rate (D-ABR) mechanism [7]. In rate control, edge switches adjust the input traffic according to the optimum fairly allocated bandwidth. However, losses in OBS are high due to the non-buffering characteristics of OBS. If the scheme for rate control allows edge switches to inject input traffic up to the link capacity into the network, the loss probability may be too high for some applications. In contrast, if the rate control scheme limits the amount of input traffic to be much less than the network capacity, this may result in low network utilization.

[†] The author is with the Graduate University for Advanced Studies, Tokyo 101-8430, Japan (e-mail: tananun@nii.ac.jp).

[††] The author is with the Graduate University for Advanced Studies and National Institute of Informatics, Tokyo 101-8430, Japan (e-mail: kei@nii.ac.jp).

The preemption method, proposed in the max-min fairness preemption (MMFP) scheme [8] and the rate fairness preemption (RFP) scheme [9], achieves max-min FBA in OBS networks without degrading network utilization. Preemption does not aim to control input traffic according to the max-min rate but aims to provide service isolation and protection among flows. In this paper, the term "flow" is defined as a connection between the same pair of ingress and egress edge switches. Moreover, we define the term "misbehaved flow" as the flow that sends traffic over its fairly allocated rate, e.g., max-min rate, and the term "well-behaved flow" as the flow that its sending rate does not exceed its fairly allocated rate. In RFP, edge switches can send input traffic over the flows' max-min rate but the core switches drop traffic transmitted over the max-min rate where there is contention. This can isolate flows and provide a low loss level for well-behaved flows, e.g., by setting a small value of e in RFP (see Section II), while ensuring efficient network utilization.

Deflection routing (DR) [10-12] is an effective alternative to solve contention in OBS networks. Based on DR, when contention occurs at a node, the corresponding node redirects the new incoming burst to alternative routes or drops the burst where all links are fully occupied. However, conventional DR is not effective under high traffic loads; in fact, DR may result in higher loss probability as presented in [12]. Several DR techniques have been proposed to improve DR's performance under high traffic loads, e.g, assigning low priority to the deflected burst [12] and selecting DR or burst retransmission according to performance criteria [11]. However, simply deflecting the burst to other paths in max-min fair-share OBS networks may destroy the max-min fair-share semantics because deflected flows gain more network resources at the expense of degraded flows using the deflected links.

We proposed RFP to achieve max-min FBA in our previous work [9]. In this paper, we aimed to use DR to reduce loss probability in max-min fair-share OBS networks. We propose rate fairness preemption combined with deflection routing (RFP-DR) in this paper to allocate max-min fair bandwidth while minimizing losses. RFP-DR is a unique integration of our previous proposed RFP scheme with DR. By combining with RFP, our purposed DR technique can preserve the max-min fair-share semantics. RFP-DR does not require additional control messages among switches for updating arrival and allocated rates, and the core switches do not need to monitor the arrival rates of all flows as required in other FBA schemes. In addition, RFP-DR does not increase total burst loss probability under high traffic loads. Although our scheme is based on preemption, our DR technique can also be applied to the OBS core network to enable rate control schemes to minimize losses.

Digital Object Identifier: 10.4108/ICST.BROADNETS2009.7196
http://dx.doi.org/10.4108/ICST.BROADNETS2009.7196

We describe the RFP-DR scheme in Sections II. We present the simulation results in Section III and Section IV is the conclusion.

II. RFP-DR ALGORITHM

OBS networks are composed of edge and core switches. The wavelengths are divided into two groups, control and data channels. IP packets are assembled into a burst, called a data burst (DB), at the ingress edge switch. The ingress edge switch sends out a burst control packet (BCP) over the control channels to reserve the DB bandwidth and after the offset time, the ingress edge switch sends a corresponding DB over the data channels. When the BCP arrives at the intermediated core switches, the core switch converts the BCP to an electronic signal for the process of reserving the DB bandwidth. Upon arrival at the intermediate-core switches along the path, where the bandwidth has been successful reserved, the DB is routed over the reserved channel. However, if the bandwidth reservation fails, the conventional OBS core switch will drop the DB.

Table I lists the parameters we used for RFP-DR. To support RFP-DR, we modified BCP formats and the max-min rate calculation in three respects.

- We create a rate priority parameter F in the BCP field format to indicate burst types (under-rate, over-rate, or deflected bursts).
- We modify the BCP field to contain all previous traveled node addresses to protect loop routing. If the core switch finds that the address of the next hop node is in the traveled nodes listed in BCP, the switch will drop the BCP to avoid loop routing.
- The RFP-DR-based core switches use the effective link capacity, $e \times C$ ($0<e\leq 1$), to allocate the max-min rate instead of the actual link capacity C due to the high loss and incomplete link-utilization characteristics of OBS networks. This modification is the same as that done in [6], [7], and [9]. A properly set value for e can be found in [7]. Note that the value of e does not affect the total loss probability in RFP [9] and RFP-DR but a small value for e can protect well-behaved flows well, resulting in a lower loss level for well-behaved flows in RFP and RFP-DR.

TABLE I
PARAMETERS USED IN RFP-DR

Parameters	Description
i	Flow id
A_i	Arrival rate of flow i at ingress switch
T_i	Allocated max-min rate of flow i
e	Effective capacity ratio
F	Rate priority
C	Link capacity
$P_{i\text{-}O}$	Over-rate burst marking probability
$P_{i\text{-}U}$	Under-rate burst marking probability

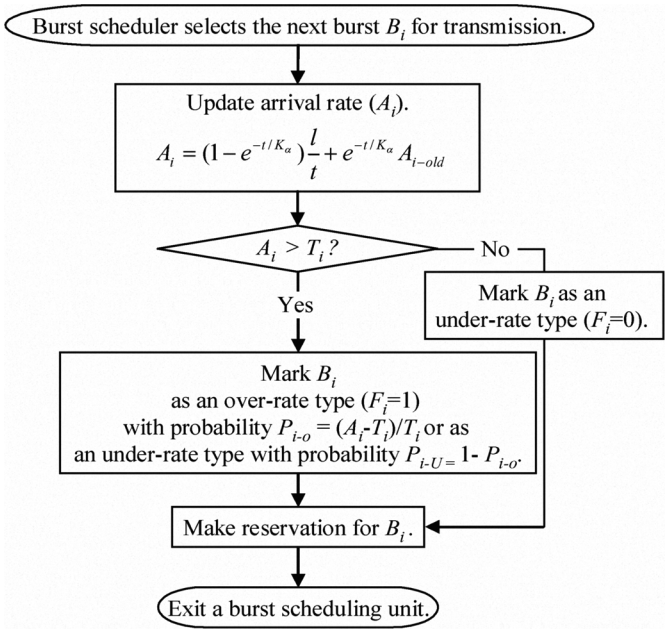

Fig. 1 Burst scheduling at edge switch

Figures 1 and 2 show the flows corresponding to the burst-scheduling process at an edge switch and a core switch. We divide the functions of the edge and core switches regarding RFP-DR as follows.

A. Function of edge switches

One of the functions of edge switches in RFP-DR is to classify input traffic into an under max-min rate type or an over max-min rate type before injecting traffic into the core network. First, the network allocates the max-min rate (T_i) to each flow. We proposed two methods of allocating max-min rate in [9], i.e., adaptive and fixed. In this paper, we selected the fixed max-min rate allocation because of its simplicity. Fixed allocation does not require control messages for updating arrival rates or the max-min rate among switches as required by the adaptive method. Note that the adaptive method can also be used and its details are presented in [9]. The fixed allocated max-min rate is based on the progressive filling algorithm (PFA) [13]. The route is based on the static shortest path selection. The route used for PFA calculation is named the primary path. The max-min rates of all flows begin at zero, and they increase together at the same pace until the total input traffic in one or more links reaches the effective link capacity $e \times C$. PFA stops increasing the max-min rates for sources that use these full-capacity links (bottle-neck links) and continues increasing the rates for other flows. PFA repeats this process until all flows have bottleneck links.

Next, the ingress edge switch monitors the arrival rate, A_i, of all flows by using a rate estimation scheme, e.g., the exponential moving average of the arrival rate. The ingress edge switch sets F to equal 0 to indicate that the burst type is under-rate and to equal 1 to indicate that it is over-rate. When $A_i > T_i$, the ingress edge switch marks flow i's bursts as over-rate or under-rate with probabilities $P_{i\text{-}O}$ and $P_{i\text{-}U}$. In contrast, when $A_i \leq T_i$, the ingress edge switch marks all flow i's bursts as under-rate. $P_{i\text{-}O}$ and $P_{i\text{-}U}$ are calculated as

Digital Object Identifier: 10.4108/ICST.BROADNETS2009.7196
http://dx.doi.org/10.4108/ICST.BROADNETS2009.7196

$$P_{i-O} = \begin{cases} (A_i - T_i)/A_i, & when\ A_i > T_i \\ 0, & when\ A_i \le T_i \end{cases} \quad (1)$$

$$P_{i-U} = 1 - P_{i-O}. \quad (2)$$

According to (1) and (2), the amount of input traffic under the max-min rate is marked as under-rate while that over the max-min rate is marked as over-rate. Finally, the edge switch injects traffic into the core network.

B. *Function of core switches*

Bursts in the core network are classified into three groups, i.e., under-rate (i.e., the rate priority F in BCP equals 0), over-rate (F=1), and deflected (F=2). We denote a new incoming burst as B_N and the original scheduled burst as B_O. Each switch is installed the primary paths (paths used for PFA-based max-min rate calculation) and all alternative paths from itself to all possible destination nodes. When a new BCP arrives at a core switch and intends to reserve a wavelength for B_N, the core switch first searches the free wavelength for B_N in the primary link (the link connected from the corresponding node to the next hop according to the primary path). If there is free wavelength, the core switch reserves a wavelength for B_N. In contrast, the following steps will be taken if there is no available wavelength in the primary link.

Step 1) ***Preemption***

B_N can preempt a channel from B_O in the primary link with a different flow ID and a higher value for F. The original scheduled burst with the largest value of F will be the first selection to be preempted. If there are many preempted candidates, the burst with the longest residual time (LRT) (measured from the end of B_O to the beginning of B_N) will be preempted. It has been found in [14] that the LRT preempted burst selection rule is the most effective approach to reduce the burst loss probability. In addition, the corresponding switch sends a resource-cancellation packet to release the reserved wavelengths of the preempted burst upstream and downstream. Then, scheduling ends. In contrast, if no original scheduled burst with a different flow ID and a higher value for F exists, go to step 2.

Step 2) ***Deflection Routing***

The core switch searches a free channel in alternative links for B_N If there are free links, the switch sets the type of B_N to deflected ($F(B_N)$=2) and deflects B_N randomly to a free alternative shortest-path link. If all links are fully occupied, the core switch drops B_N. Scheduling then ends.

Note that next hop must not be in the previous passed-nodes field contained in BCP for protecting loops. RFP-DR gives the highest preemptive priority to traffic transmitted under the max-min rate (F=0). Consequently, RFP-DR prevents quality degradation for traffic transmitted under max-min rate. In addition, before deflecting the burst to an alternative path, the RFP-DR based core switch changes the burst type to deflected even the original of the deflected burst is under-rate type. Deflected traffic does not degrade the quality of traffic using the deflected link as their primary path attained by assigning the lowest preemptive priority to deflected traffic (F=2). Therefore, our DR technique can preserve the max-min fair-share

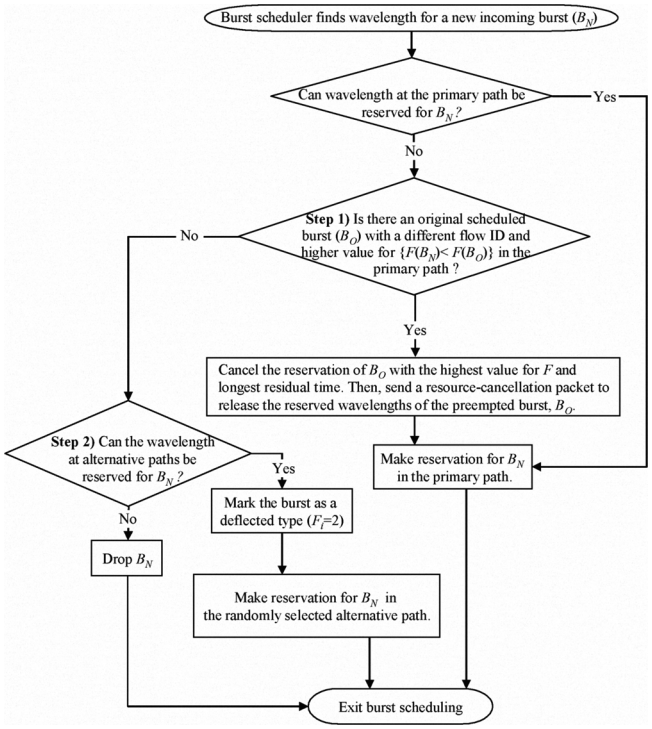

Fig. 2 Burst scheduling at core switch

semantics. Besides, conventional DR tends to increase the burst loss probability in high-traffic-load environments as described in [12] because deflected traffic increases congestion in deflected links. Giving lowest priority to deflected traffic can also avoid the degraded quality caused by DR under high loads. To avoid using fiber delay lines (FDLs) to extend the offset time of the deflected burst, we set the minimum offset time according to the proposal in [12].

III. Simulation Results

We conducted simulations on a modified optical burst switching simulator originally developed at the optical internet research center (OIRC) [15] on the basis of ns-2 [16]. The simulations were conducted on a 4×4 TORUS topology consisting of 16 core switches as shown in Fig. 3. Each core switch was attached to an edge switch. The distance between adjacent switches was equal to 200 km and the transmission delay was 0.1 ms in each link. We assumed that the number of control wavelengths would be sufficiently large to ensure no losses for BCPs and there were no DB losses on the link between edge and core switches. There were 16 wavelengths for DB between adjacent core switches and the capacity of each wavelength was 1 Gbps. We set e to equal 0.5. We assumed that the network had full wavelength conversion capabilities and it did not employ FDLs. The BCP processing time was 0.1 msec. The primary path was selected based on the shortest path routing. Edge switches generated bursts with exponentially distributed inter-arrival times and burst lengths with an average burst length of 1MB. We set the input traffic rate of flow i to the designated value by adjusting the average burst inter-arrival time. Edge switches use an exponential moving average to

Digital Object Identifier: 10.4108/ICST.BROADNETS2009.7196
http://dx.doi.org/10.4108/ICST.BROADNETS2009.7196

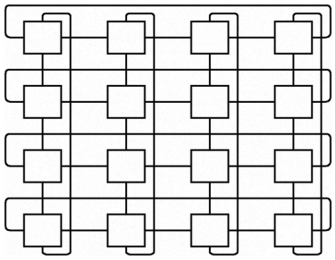

Fig. 3 4×4 TORUS network

estimate the arrival rate of each flow. The arrival rate is calculated as

$$A_{new} = (1 - e^{-T/K_\alpha})\frac{l}{T} + e^{-T/K_\alpha} A_{old}, \quad (3)$$

where A_{new} is the estimated arrival rate, T is the inter-arrival time between the current and the previous burst, and l is the burst size. Here, A_{old} is the previous value of the arrival rate before updating and K_α = 0.1. Note that the normalized rate of 1 in all simulations was equivalent to the link capacity C (16 Gbps) in the core network. For instance, the normalized sending rate, 0.035, was equal to 0.56 Gbps (0.035x16 Gbps).

We compared RFP-DR with RFP and conventional OBS. The simulated 4×4 TORUS network consisted of 240 flows with four different hop-counts: three hops, four hops, five hops, and six hops. The ratio of three-hop, four-hop, five-hop, and six-hop flows was 0.67:1:0.67:0.17. In our simulation, flows were classified into two groups: well-behaved and varied-sending-rate groups. Although the sending rate fluctuates, the term "sending rate" in this section refers to the average sending rate for simplicity. All flows in the well-behaved group sent the fixed sending rate at the rate under their max-min rate. The sending rate of flows in the varied-sending-rate group was varied. Some flows in the varied-sending-rate group are well-behaved, i.e., flows that send traffic under the max-min rate, where some flows are misbehaving, i.e., flows that send traffic over the max-min rate. We divided the simulations into two cases as follows.

A. Misbehaved flows: long-path selection

In this simulation, we assumed that long-path flows, i.e., all five-hop and six-hop flows, tend to be misbehaving and they were in the varied-sending-rate group while short-path flows, i.e., all three-hop and four-hop flows, were in the well-behaved group. Thus, the ratio of flows in the well-behaved group and the varied-sending-rate group was 2:1. Although the number of flows in the varied-sending-rate group was half of the well-behaved flows, their input rate had high impacts to the network because they traveled through many links.

In our simulated TORUS network, there were 13 flows in the most congested link. Therefore, the minimum PFA normalized max-min rate (PFA_{min}) assigned to flows traveling through this most congested link was equal to 0.038 (PFA_{min} = e×{normalized rate of C}/{number of flows in the most congested link} =0.5×1/13). We fixed the normalized sending rates for each flow in the well-behaved group to be slightly less than PFA_{min} at 0.035. Therefore, all flows in the well-behaved group sent traffic under the max-min rate. We varied the normalized sending rate per flow in the varied-sending-rate group from 0.04 to 0.2. Therefore, some flows in the varied-sending-rate group are well-behaved where some flows are misbehaving.

Figure 4 presents the loss probabilities of flows in the well-behaved group and the varied-sending-rate group for conventional OBS, RFP, and RFP-DR. The loss probability of flows in the varied-sending-rate group in conventional OBS is higher than that of flows in the well-behaved group because flows in the varied-sending-rate group are long-path flows. In addition, when flows in the varied-sending-rate group send large amounts of traffic into the network, the loss probability of flows in the well-behaved group in conventional OBS also increases rapidly. This is because the conventional OBS does not isolate services or protect them. In contrast, both RFP and RFP-DR effectively protect against quality degradation in flows in the well-behaved group. The loss probabilities of flows in the well-behaved group in RFP and RFP-DR do not increase to a high level even if many misbehaved flows send input traffic over their max-min rate. Consequently, RFP-DR can provide service isolation and protection even when DR is implemented. RFP-DR aims to use the DR technique to decrease losses. We can see that RFP-DR can decrease the loss probabilities of flows in both groups.

In terms of total burst loss probabilities (Fig. 5), we can see that RFP and RFP-DR do not degrade the total burst loss probability because it drops overload traffic at the beginning of transmission before it contends with other bursts in the remaining paths. RFP and RFP-DR also uses certain strategies to prevent the total burst loss probability from becoming high, e.g., the LRT wavelength selection rule and resource-cancellation packets, as described in Section II.B. RFP-DR gives the lowest total burst loss probability because of the effectiveness of DR. Besides, RFP-DR can perform well even under high traffic loads. This is because RFP-DR assigns the low preemptive priority to deflected traffic.

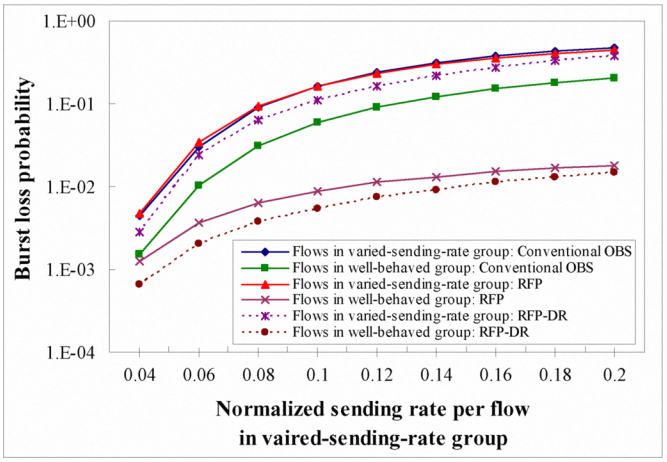

Fig. 4 Burst loss probabilities of flows in the well-behaved and varied-sending-rate groups when long-path flows are misbehaving

Digital Object Identifier: 10.4108/ICST.BROADNETS2009.7196
http://dx.doi.org/10.4108/ICST.BROADNETS2009.7196

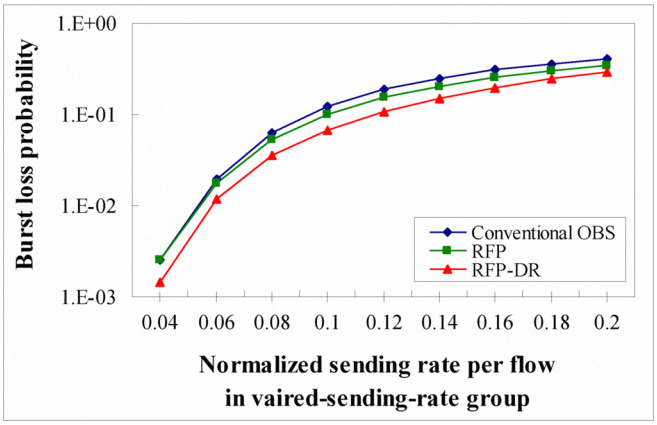

Fig. 5 Total burst loss probability when long-path flows are misbehaving

B. *Misbehaved flows: random selection*

We randomly classified 240 flows into well-behaved (120 flows) and varied-sending-rate (120 flows) groups. The normalized sending rate per flow in the well-behaved group was fixed at 0.035 while the normalized sending rate per flow in the varied-sending-rate group ranged from 0.04 to 0.2. Figure 6 plots the loss probabilities of flows in the well-behaved and varied-sending-rate groups for conventional OBS, RFP, and RFP-DR. In conventional OBS, the loss probability of flows in the well-behaved group is slightly lower than that of in the varied-sending-rate group and it increases when the input rate of flows in the varied-sending-rate group increases because flows are not isolated or protected. We evaluated the performances of RFP and RFP-DR in terms of service isolation and protection and found they had the same tendency as that discussed in Section III.A. Both RFP and RFP-DR can prevent degraded quality in well-behaved flows and RFP-DR decreases the loss probabilities of flows in both groups.

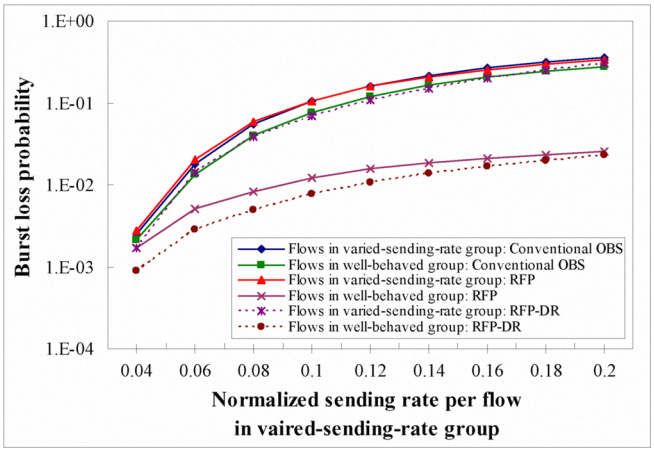

Fig. 6 Burst loss probabilities of flows in the well-behaved and varied-sending-rate groups when misbehaved flows are randomly selected

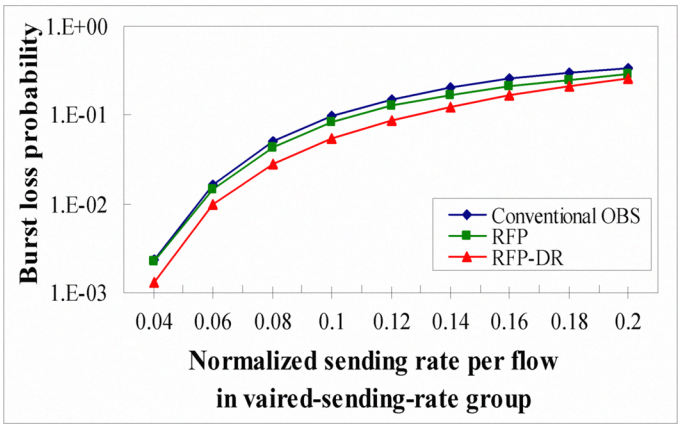

Fig. 7 Total burst loss probability when misbehaved flows are randomly selected

The results in Fig. 7 indicate that by selecting the proper preempted candidates, i.e., using the LRT selection rule, RFP can decrease the total burst loss probability. RFP-DR performs the best in terms of total burst loss probability. The deflected traffic can use only the unused bandwidth in the deflected link. Therefore, it does not degrade total burst loss probability under high traffic loads as the conventional DR does.

IV. Conclusion

We proposed a rate fairness preemption (RFP) scheme in our previous work to allocate max-min fair bandwidth in OBS networks. This paper proposed a deflection routing (DR) strategy to minimize losses in max-min fair-share OBS networks. We integrated DR with RFP and derived a new scheme, i.e., rate fairness preemption with deflection routing (RFP-DR). We demonstrated through simulation that RFP-DR could isolate and protect services as the original RFP does, but RFP-DR effectively decreased the total burst loss probability. In addition, RFP-DR did not degrade total loss probability even under high traffic loads.

Many of fair bandwidth allocation schemes, e.g., schemes proposed in [6]-[9], including the case of using the adaptive max-min rate allocation for RFP, use control plane to update arrival and fairly allocated rates of flows among switches. This information can be used for predicting the status of the network. Our future work is to develop an adaptive path selection scheme based on the resource available in the network.

References

[1] L. Yang, Y. Jiang, and S. Jiang, “A probabilistic preemptive scheme for providing service differentiation in OBS networks,” in Proc. IEEE Globecom, Dec. 2003, vol. 5, pp. 2689 - 2693.

[2] J. Phuritatkul, Y. Ji, and Y. Zhang, "Blocking probability of a preemption-based bandwidth-allocation scheme for service differentiation in OBS networks," IEEE/OSA J. Lightwave Technology, Vol.24, No.8, pp.2986-2993(2006).

[3] M. Udeda, T. Tachibana, and S. Kasahara, “Intermediate-hop preemption to improve fairness in optical burst switching networks”, IEICE Trans. Commun., vol. E91-B, no. 3, pp. 710-721, 2008.

[4] T. Orawiwattanakul and Y. Ji, "Resource consumption based preemption for providing fairness in optical burst switching networks," Proc. of

Digital Object Identifier: 10.4108/ICST.BROADNETS2009.7196
http://dx.doi.org/10.4108/ICST.BROADNETS2009.7196

International Workshop on Optical Burst/Packet Switching (WOS), 2007.

[5] B. Zhou and M. A. Bassiouni, "Concurrent enhancement of network throughput and fairness in optical burst switching environments," J. Photonic Network Communications, vol. 14, no. 2, pp. 199-207, 2007.

[6] S. Kim, Y.-C. Kim, B.-Y. Yoon, and M. Kang, "An integrated congestion control mechanism for optimized performance using two-step rate controller in optical burst switching networks", J. Computer and Telecommunications Networking, vol. 51, no. 3, pp.606-620, 2007.

[7] H. Boyraz and N. Akar , "Rate-controlled optical burst switching for both congestion avoidance and service differentiation," J. Optical Switching and Networking (Elsevier), vol. 2, no. 4, pp. 217-229, 2005.

[8] Y. Liu, K. C. Chua, and G. Mohan, "Max–min fairness in WDM optical burst switching networks," J. high speed networks, vol. 16, no. 4, pp. 379-398, 2007.

[9] T. Orawiwattanakul, Y. Ji, Y. Zhang, and J. Li, "Fair bandwidth allocation in optical burst switching networks," accepted by IEEE/OSA J. Lightwave Technology (2009).

[10] C. Hsu, T. Liu, and N. Huang., "Performance analysis of deflection routing in optical burst-switched networks", in Proc. of Infocomm, vol. 1, pp. 66-73, 2002.

[11] S. K. Lee, K. Sriram, H. S. Kim, and J. S. Song., "Contention based limited deflection routing in OBS networks", in Proc. IEEE Globecom, Dec. 2003.

[12] C. Cameron, A. Zalesky and M. Zukerman, "Prioritized Deflection Routing in Optical Burst Switching Networks", IEICE Trans. Commun., vol. E88-B, no. 5, pp. 1861-1867, May 2005.

[13] D. P. Bertsekas and R. Gallager, Data Networks, Englewood Cliffs, NJ: Prentice-Hall, 1992.

[14] T. Tachibana and S. Kasahara, "Two-Way Release Message Transmission and Its Wavelength Selection Rules for Preemption in OBS Networks," IEICE Trans. Commun., vol. E90-B, no. 5, pp. 1079-1089, 2007.

[15] The Optical Internet Research Center Optical Burst Switching NS Simulator. [Online]. Available: http://wine.icu.ac.kr/~obsns/index.php

[16] "ns-2 network simulator," 2000. [Online]. Available: http://www.isi.edu/nsnam/ns/

Digital Object Identifier: 10.4108/ICST.BROADNETS2009.7196
http://dx.doi.org/10.4108/ICST.BROADNETS2009.7196