Dynamic Contention Resolution Protocol for OBS Networks

Wael Hosny Fouad Aly, Halima El-Biaze

Abstract— Optical burst switching (OBS) is a switching technique that was proposed as a hybrid switching technology to support the next generation Internet. In OBS, incoming IP packets are assembled into super-sized packets called *data bursts*. Burst contention is a well-known challenging problem in Optical Burst Switching (OBS) networks. Burst contention can be resolved using several approaches, such as wavelength conversion, buffering based on fiber delay line (FDL) or deflection routing Retransmission technique is used to reduce the Burst Loss Ratio (BLR) by deflecting dropped bursts. Segmentation also resolves contention by dividing the contended burst into smaller parts called *segments*. Combining deflection routing technique and retransmission technique outperforms both pure deflection and pure retransmission techniques to improve the performance. Previous work uses only static combination of retransmission and deflection of bursts to reduce contention. This paper proposes a dynamic protocol to resolve contention based on combining deflection, retransmission and delaying bursts to improve the OBS performance. Experiments were conducted to test the proposed protocol. The proposed technique was tested on complex models such as NSFNET and COST238 topologies. Results show that the proposed protocol outperforms existing techniques in terms of burst lost ratio.

Index Terms— OBS networks, contention resolution techniques, retransmission techniques.

◆

1 INTRODUCTION

Optical Burst Switching (OBS) [1] is a promising technology that is used to handle bursty and dynamic Internet Protocol traffic in optical networks effectively.

In OBS networks, user data that could be sent in terms of IP packets is assembled as a huge segment called *data bursts,* which is sent using one-way resource reservation. The burst is preceded in time, called offset time, by a control packet, called *Burst Header Packet (BHP).* BHP is sent on a separate control wavelength and requests resource allocation at each switch.

When the control packet arrives at a switch, the capacity is reserved in the cross-connect for the burst. If the needed capacity can be reserved at a given time, the burst can then pass through the cross-connect without the need of buffering or processing in the electronic domain.

Since data bursts and control packets are sent out without waiting for an acknowledgment, the burst could be dropped due to resource contention or due to insufficient offset time if the burst catches up the control packet. Thus, it is clear that burst contention resolution approaches play an essential role to reduce the Burst Loss Ratio (BLR) in OBS networks [3].

Burst contention can be resolved using several approaches, such as wavelength conversion, buffering based on fiber delay line (FDL) or deflection routing. Another approach, called burst segmentation, resolves contention by dividing the contended burst into smaller parts called *segments*, so that a segment is dropped rather than the entire burst.

Deflection routing is an attractive solution to resolve the contention in OBS networks since it does not require extra cost in terms of physical components and uses the available spectral domain. However, as the load increases, deflection routing could lead to performance degradation and network instability [4]. Since deflection cannot eradicate the burst loss, retransmission at the OBS layer has been suggested by Torra et al. [5].

Several studies have performed to improve OBS performance, an implementation of TCP Vegas for OBS was experimented [17][14][16]. A congestion detection scheme for TCP over OBS is studied in [20][10]. Several issues, solutions and challenges related to TCP over OBS networks are studied in [12]. Some collaboration was performed to improve the TCP based on a dynamic approach for contention loss notification[2][13]. Different TCP characteristics were studied in [15] and their effect on OBS including a responsive rate control for the TCP when used over OBS network. Improving fairness for optical burst switching networks is studied in [16]. Studying cluster processing for OBS taking into consideration the signal processing receivers are discussed in [18].

A static combination of deflection and retransmission has been proposed by Son-Hong Ngo et. al.[6] [6]. They have proposed a Hybrid Deflection and Retransmission (HDR) algorithm [6] which combines deflection routing and retransmission. Simulation results show that HDR gives bad performance in terms of BLR since it first applies deflection even if the load is high. To overcome this shortcoming, Son-Hong Ngo et. al have developed another mechanism called *Limited Hybrid Deflection and Retransmission (LHDR)* that controls the deflection.

This paper introduces a dynamic protocol called *Dynamic Contention Resolution Protocol (DCRP)* to combine deflection routing, retransmission or delay dynamically. A decision is made dynamically to select whether to use deflection, retransmit, or delay of bursts. The decision is based on a local knowledge about network condition. The offset time is also adapted by using an adaptive decision threshold. In order to make the local knowledge feasible, DCRP algorithm exploits sending and receiving of Positive Acknowledgement (ACK) and Negative Acknowledgement (NACK) messages to advertise useful statistics about network conditions stored by all nodes.

This paper is organized as follows. Section 2 describes the proposed *Dynamic Contention Resolution Protocol (DCRP)*. Section 3 presents the experimental results. Finally, Section 4 contains the conclusion and the future work.

2 DYNAMIC CONTENTION RESOLUTION PROTOCOL (DCRP)

In this section, we describe the proposed Dynamic Contention Resolution Protocol (DCRP). DCRP optimizes the decision of performing deflection, retransmission or delay. When no contention occurs, the primary path is used as shown in Fig 1. However, when contention occurs, DCRP chooses between the best contention resolution strategy among deflection routing, retransmission and delay as shown in Fig 2.

DCRP enhances the selection of alternate routes. The offset time is determined at the ingress node where it is predicted by considering if a deflection is needed or not.

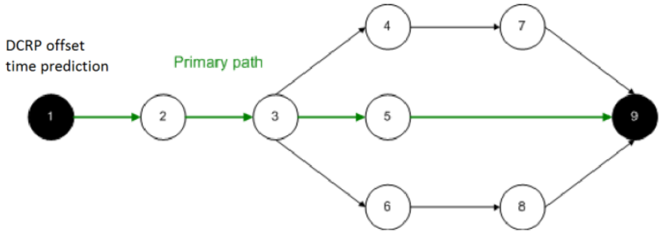

Fig 1: No congestion scenario

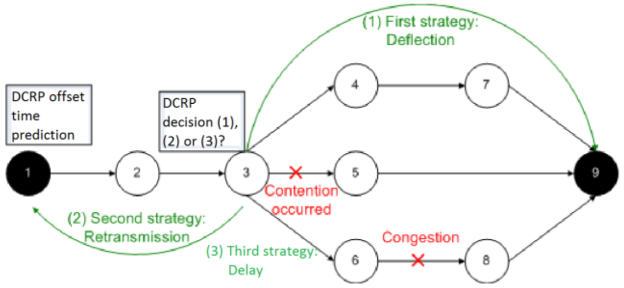

Fig 2: Congestion scenario, contention occurred

2.1 GATHERING NETWORKS STATISTICS

When the control packet reaches the destination, an acknowledgement is sent to the source.

If the control packet is dropped, then the proposed algorithm uses a negative acknowledgment to notify the source for burst retransmission.

The proposed protocol DCRP uses the positive and the negative acknowledgements to perform two functionalities: (1) To perform the notification stating whether the control packet was received by the receiver or re-transmission is required (2) To transmit statistics about links states.

This is implemented using the following procedure. The BLR and the link utilization are measured on each link and this information is integrated into the notification acknowledgement. In the case of a negative acknowledgement, statistics about the current node and the next node is used. In the case of a positive acknowledgement, the BLR and the link utilization between the destination node and the last node before the destination are used. When a node receives a positive or a negative acknowledgement control packet, it collects and analyzes statistics. Thus, statistics of the whole network is eventually updated through acknowledgements.

2.2 ADAPTIVE DECISION THRESHOLD

In order to resolve contention, the main point is to decide whether to deflect, retransmit or delay the burst. As the load increases, deflection routing can destabilize the network [3]. However, we want to maximize the bandwidth utilization. This is maintained by reaching the maximum possible deflections before destabilizing the network. The maximum possible deflections are reached by comparing a metric value by a certain decision threshold:

In this paper the metric value used is *a success probability calculated* with the BLR and the link utilization. Plus, decision threshold has to be adapted that depends on the network conditions.

The following parameters are used:

- *deflengthreshold:* carries the deflection route length threshold,
- *deflength*: a possible deflection route,
- *primaryroutelength:* the primary route,
- *route*: the number of hops of the route

If the deflength satisfies the condition that deflength is less than the product of the primary and the *deflengthreshold*, then the deflength is added as a possible deflection alternative

DCRP protocol drops bursts between two nodes based on dropping probability. The dropping probability is a factor in the burst loss ratio and the link utilization. The dropping probability is the sum of the weighted performance metric used. Thus the dropping probability is measured by the sum of the BLR and the link utilization between 2 consecutive nodes.

$droprob(n_i, n_{i+1})=BLR(n_i, n_{i+1}) *prob_{BLR}+U(n_i, n_{i+1})* prob_U$ where the sum of the probabilities is equal to unity.

Thus, the success probability is the complement of the dropping probability of dropping at each single link. This means that the success probability *successprobability* is calculated as: 1-dropprob for each link. Therefore, $successprob(Route)=\prod_{i=1}^{Route}(1-dropprob(\,n_i,n_{i+1}))$.

This is a simple multiplication of all the success probability links to get a global success probability for the entire route.

In DCRP, an adaptive decision threshold is used to make the decision whether to deflect, retransmit or delay the burst in case of contention

```
if (successprob(defl)>successprobthreshold(BLR, U, delay))
then Selection=deflection
else
if (successprob(retransmission)>successprobthreshold(BLR, U, delay))
then Selection= retransmission
else
if (successprob(delay)>successprobthreshold(BLR, U, delay))
then Selection= delay
```

Fig 3 shows the linear correlation in the NSF Network (NSFNET) between the BLR and the decision threshold. We found that the R^2 that measures the goodness of the model reaches 96% [7][19][19].

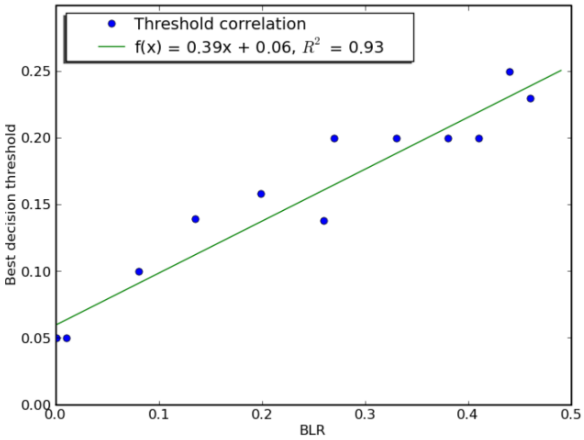

Fig 3: Correlation between BLR and decision threshold

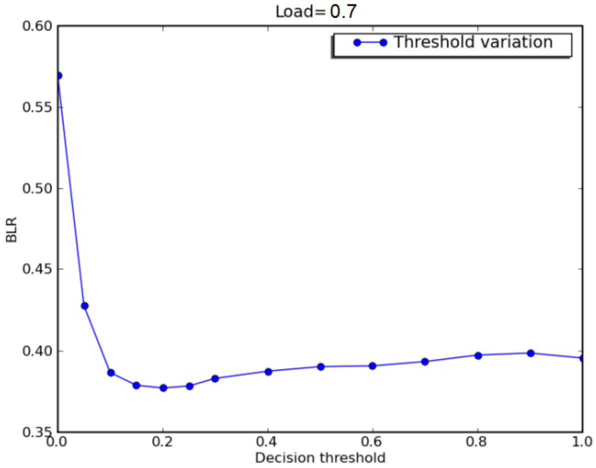

Fig 4: Decision threshold variation with a fixed load of 0.7 in NSFNET

As the BLR increases, the decision threshold also increases in order to reduce the number of deflections. All permutations of several loads and various decision thresholds are already conducted during the simulation phase.

In Fig 4, the load is measured at 0.7, we note that the best decision threshold is near 0.2. If we use the regression line with BLR= 0.37, we get f(0.37) = 0.2043, which is approximately 0.2. The squared correlation coefficient (R^2) quantifies the goodness of the fit. R^2= 93% which is an accepted match. R^2= 93% indicates that the correlation between the BLR and the best decision threshold for an effective decision is very high.

$$Successprobthreshold=W_{BLR}*BLR+W_{Utiliz}*Utiliz+W_{delay}*delay$$

where W_{BLR} is the weight applied for the BLR, W_{Utiliz} is the weight applied for the utilization and W_{delay} is the weight applied for the delay. $W_{BLR} + W_{Utiliz} + W_{delay}=1$. The success probability of the deflection route is then compared to the adaptive decision threshold.

If the success probability of a given deflection alternative is greater than or equal the adaptive decision threshold, then it means that this alternative should currently be tried.

Let *Selection* denotes a variable to select whether to deflect, retransmit, or delay the burst

The algorithm is used to determine if the current burst should be deflected, retransmitted or delayed by considering current network conditions. Those formulas are pre-calculated periodically and a typical routing table is periodically updated so that the forwarding process is not penalized.
Several techniques could be used to find the weights (W_{BLR}, W_{Utiliz}, W_{delay}) for BLR, utilization and delay respectively.

Probabilistic graphical models have been extensively studied in machine learning [8] and could be used to find good weights. Neural Network is also a good learning model to fine tune an output metric by considering inputs metrics.

2.3 DCRP FORWARDING PROCESS

When a Burst Header Packet (BHP) arrives at a core node, the next hop has to be selected from a routing table in the electrical domain in order to reserve bandwidth for the data burst. The approach used for the forwarding process is as follows:

- When a control packet is received, the current node is compared to the destination node. If the BHP arrives at the destination, then an ACK is sent to the source
- Then, the offset time is checked in order to verify if it is still sufficient. If it is not sufficient, a NACK is sent to the source and the burst is retransmitted after an idle time.
- The shortest path output port is selected. In case of resource contention, it is solved by deflection, retransmission, or delay. The burst is not retransmitted, dropped or delayed after a certain number of retransmissions.
- DCRP successively extracts best deflection alternatives in order to minimize the BLR and the number of retransmissions. The best output port is found by extracting the next hop in the route, where it carries the maximum success probability value.
- The success probability of the deflection route is then compared to the adaptive decision threshold
- If the success probability of the current alternative is smaller than the adaptive decision threshold, then a NACK is sent to the source and the burst is retransmitted

after a delay time. Otherwise, the current output port alternative is scheduled.

However, to not penalize the forwarding process physically, the routing table (TABLE I) is updated periodically. The cost is expressed by:

Cost(Next hop, Dest)=1-successprob(Route(Next hop, Dest)) where Route(Next hop, Dest) is the route of a given next hop to the destination so that next hops having a high success probability will result in a low cost. We note that next hops are sorted as follows:

$$\forall_{i=1}^{N}\forall_{i=j}^{M-1}Cost\left(Hop_j, Dest_i\right) \le Cost\left(Hop_{j+1}, Dest_i\right)$$

The decision threshold is updated periodically using the success probability threshold formula

2.4 ADAPTIVE OFFSET TIME

In OBS networks, the data burst follows the control packet after a predetermined offset time calculated at the ingress node. The offset time has to be large enough so that bursts arrive at each switch after the control packet. The minimum offset time t_{offset} must consider the BHP processing time t_p at each hop, the node switching and the configuration time t_{conf}. However, the minimum offset time is expressed by:

$t_{offset}=t_{conf}+N_{hops}*t_p$

where N_{hops} is the number of hops.

The main key to find the best offset time is to predict the number of hops because t_{conf} and t_p are fixed values. However, if deflection occurs, a longer route could be used which increases N_{hops}.

The number of hops (N_{hops}) to be used in the offset takes one of two values. The protocol will test for the best path number of hops, otherwise, the number of hops used for the shortest path is used instead.

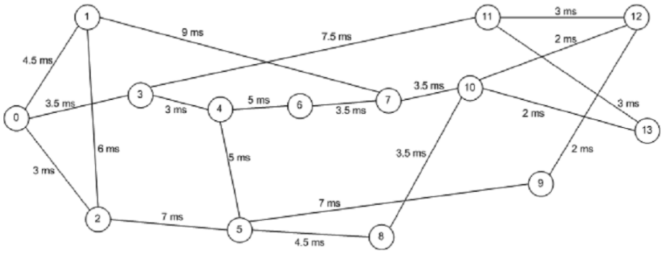

Fig 5: NSFNET topology

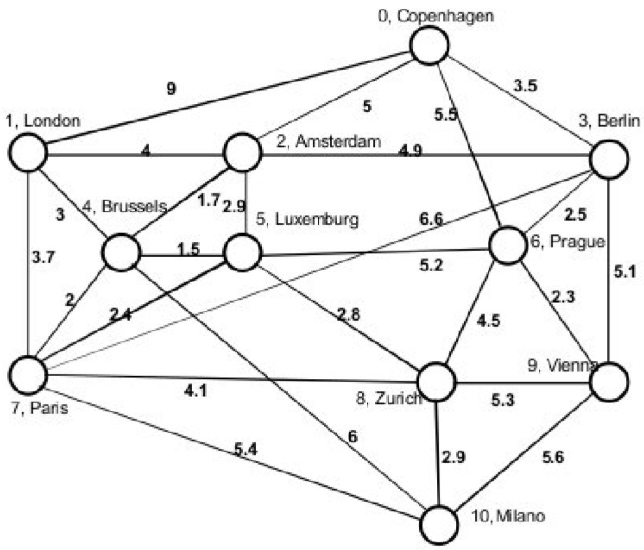

Fig 6: COST239 topology

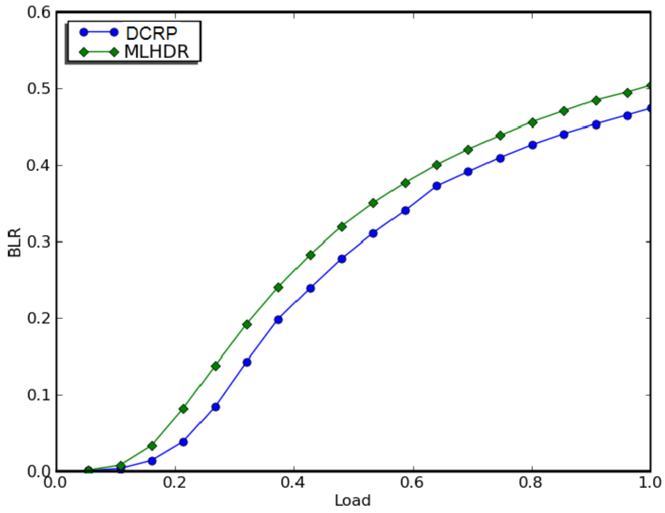

Fig 7: BLR in NSFNET

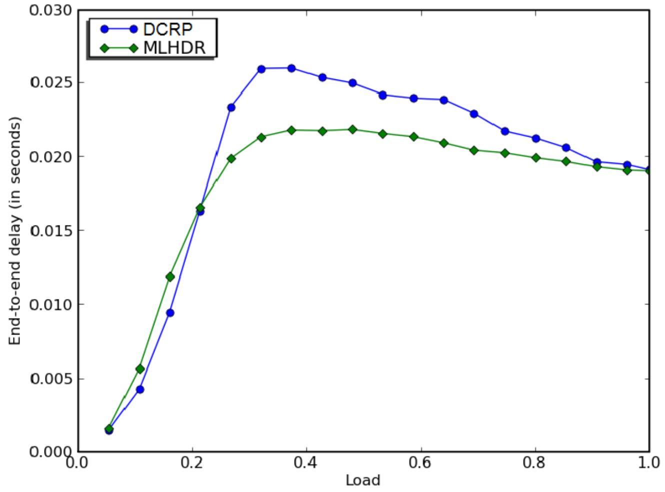

Fig 8: End-to-end delay in NSFNET

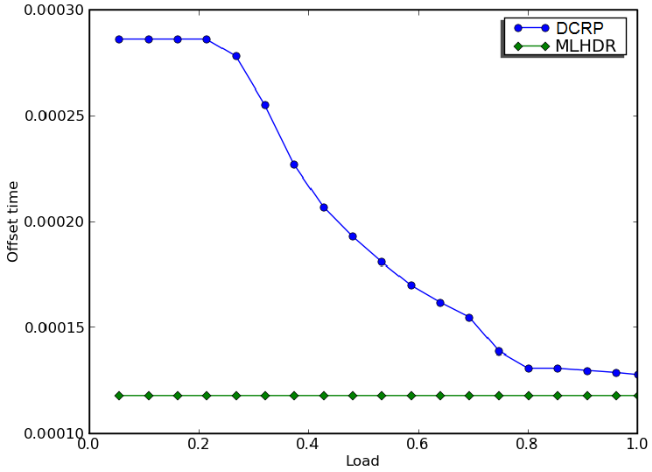

Fig 9: Offset time variation in NSFNET

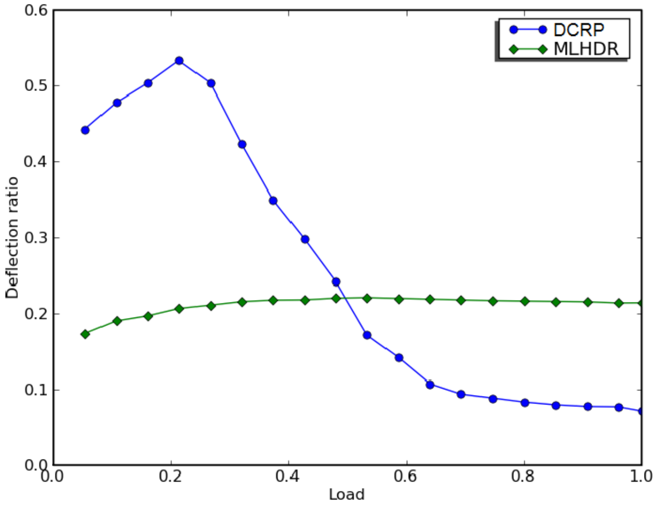

Fig 10: Deflection ratio in NSFNET

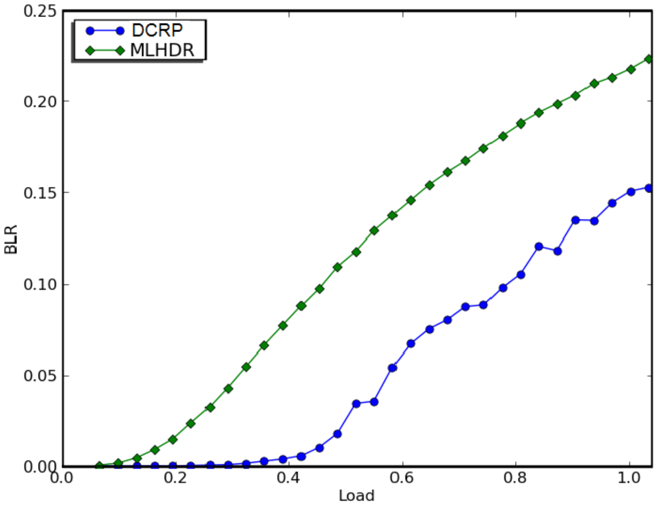

Fig 11: BLR in COST239

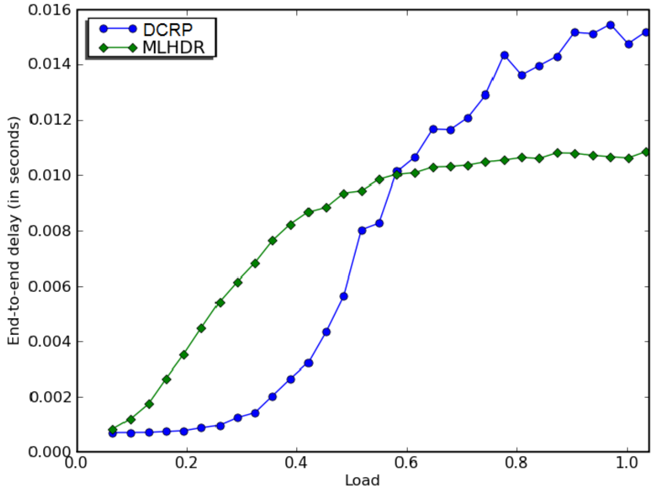

Fig 12: End-to end delay in COST239

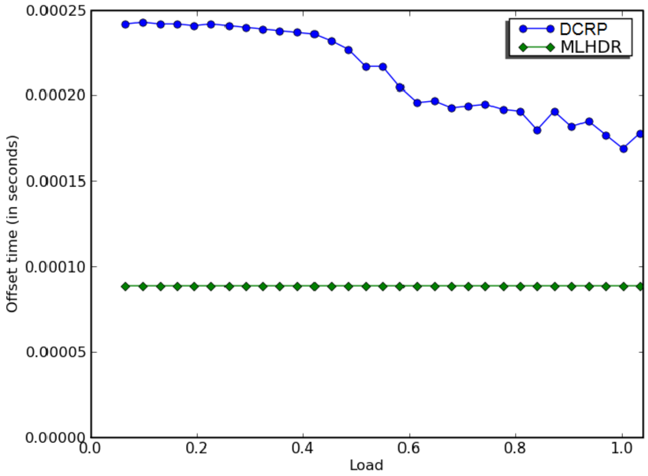

Fig 13: Offset time variation in COST239

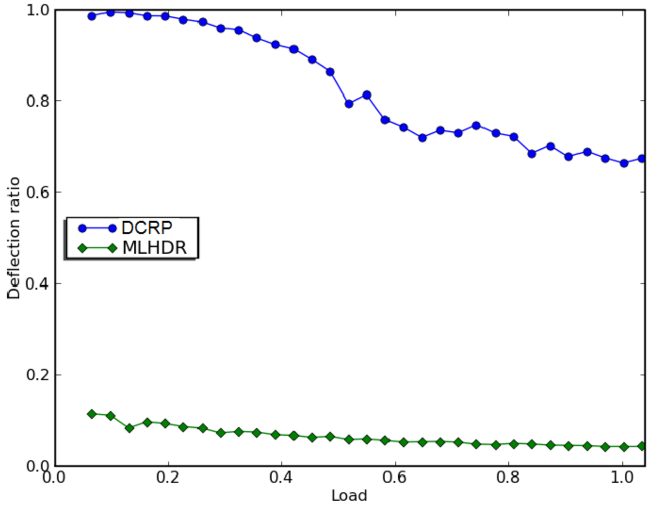

Fig 14: Deflection ratio in COST239

3 SIMULATION RESULTS

Simulations are performed with NSFNET and COST239 topologies using Network Simulator 2 (ns-2) with an extra module for OBS. Because in LHDR there is only one possible deflection alternative and DCRP try several deflection alternatives, LHDR is modified (MLHDR) in order to do proper comparisons. The only difference between LHDR and MLHDR is that when contention occurs, MLHDR applies LHDR for the shortest alternative first, then for the second shortest alternative and so forth as long as long as there is a possible deflection alternative.

The following simulation configuration is used

- Each wavelength has 1 Gbps of bandwidth capacity
- Each link has 2 control channels and 4 data channels.
- The mean burst size equals 400 KB in NSFNET topology and 4 MB in COST239 topology.
- Packet and burst generations follow a Poisson distribution for the input packet rate and for the burst size.
- Traffic generators are distributed randomly over the network.
- Each traffic generator sends bursts to any node (except itself.
- Bursts are indefinitely lost after a certain number of retransmissions N_{ret}(truncated retransmission). N_{ret} is fixed to 1 in order not to increase the end to end delay significantly
- Finding the best N_{ret} is out of the scope of this paper
- Dropped bursts are retransmitted after Rand(0,0.05) seconds so that retransmitted bursts are highly penalized in terms of end-to-end delay.
- Rand returns a random value between a minimum and a maximum value.
- We define the traffic load to be the ratio of the total input source nodes throughput over the capacity of the whole network [9].

Weights applied to the success probability threshold input devices varies with several loads. Best results in terms of BLR are obtained using weights where

$$0.4 \leq W_{BLR}+W_{Utilization}+W_{delay} \leq 0.8$$

In this section, we present the obtained simulation results that compare MLHDR and DCRP performance at two complex topologies, NSFNET and COST239 topologies.

1.

2. 3.1 COMPARISON BETWEEN MLHDR AND DCRP AT NSFNET TOPOLOGY

3.

4. Several simulations were conducted with NSFNET topology.

5. Let L be the number of links in NSFNET, N is the number of nodes, NSFNET has a low connectivity of C= 0.23 where $C = \frac{L}{N(N-1)/2}$.

DCRP gives significant improvements in terms of BLR even at high loads (Fig. 7). NSFNET has a low connectivity which is considered as a drawback for the DCRP since DCRP selects low loaded links.

Because of that, when only a few alternatives are available, the gain difference is limited. In this case scenario, the gain comes from two main sources:

1. The adapted offset time (Fig. 9)
2. DCRP forwarding process.

We could expect that the end-to-end delay is highly increases since the forwarding process of DCRP selects longer routes compared to the shortest path. However, the end-to-end delay is similar compared to MLHDR. The highest end-to-end delay for DCRP gives 5 ms higher than MLHDR (Fig. 8). From a client's point of view, 5 ms is acceptable in general since the Internet uses the Best Effort paradigm. The adapted offset time highly influence the number of deflections, the number of retransmissions and also the delay during the simulation phase.

Fig. 10 shows the deflection ratio variation of DCRP and MLHDR. When using DCRP at load that falls in the range of 0 to 0.25, we can clearly observe that DCRP performs higher number of deflections as compared to MLHDR.

However, as the load increases, we can see the benefit of using DCRP compared to a static approach. That is, deflections are performed and are effective as long as it reduces the BLR. We can also observe that MLHDR does not perform enough deflections where at the same load range of 0 to 0.5 but performs deflections at load range of 0.5 to 1.

3.2 COST239 TOPOLOGY

Several simulations are conducted using COST239 topology in order to compare DCRP and MLHDR. The topology is a highly connected topology (C= 0.47). DCRP gives significant improvements in terms of BLR (Fig. 11)

For loads between 0 and 0.6, there is also a gain in terms of end-to-end delay (Fig. 12). For loads greater than 0.6, the end-to-end delay reaches a maximum value of 6 ms higher with DCRP in order to reduce the BLR. The deflection ratio should be high with highly connected topologies like COST239 at least when the load is low. Static hybrid deflection and retransmission mechanisms such as MLHDR under-utilize deflection routing (Fig. 14) where the deflection ratio is hovering around 0.08. DCRP deflect bursts when the load is low.

As the load increases, the offset time is reduced (Fig. 13) in order to reduce the deflection ratio in an adaptive manner.

4 CONCLUSION AND FUTURE WORK

This paper presents a novel algorithm called Dynamic Contention Resolution Protocol (DCRP) that combines contention resolution strategies such as deflection, retransmission and delayed routing. In order to make effective decisions between those contention resolution strategies, DCRP uses an adaptive decision threshold. This decision threshold is adapted using network metrics such as BLR and link utilization since there exists a correlation between the BLR and the best decision threshold to use in order to reduce the BLR. Low connected topologies such as NSFNET needs an adaptive mechanism to balance deflection routing, retransmission and delay in order to reduce the BLR with a small cost in terms of end-to-end delay. Also, static approaches such as MLHDR when using NSFNET does not perform deflection sufficient enough at low load and performs high number of deflections as load increases. Highly connected topologies such as COST239 offer the ability to have a high ratio of deflections over retransmissions. This is to reduce the BLR with a small cost in terms of end-to-end delay when the load is high.

The future work of this research is to combine several contention resolution strategies in a dynamic way because we believe that the feasibility of OBS requires effective and adaptive algorithms to overcome the burst loss issue. We are presently working on a new approach which deploys a probabilistic graphical model used in artificial intelligent in order to make efficient and dynamic decisions among several contention resolution strategies.

ACKNOWLEDGMENT

The authors would like to thank Son-Hong Ngo help and support during the paper simulation

REFERENCES

[1] C. Qiao and M. Yoo, “Optical Burst Switching - A New Paradigm for an Optical Internet”, Journal of High Speed Networks, vol. 8, no. 1, pp. 69–84, 1999.

[2] B. Shihada, P-H. Ho, and Q. Zhang, "A Novel False Congestion Detection Scheme for TCP over OBS Networks", Proc. of the IEEE Global Communications Conference (GLOBECOM), pp. 2428-2433, 2007.

[3] Christoph M. Gauger, Martin K ¨ ohn, and Joachim Scharf, “Performance of Contention Resolution Strategies in OBS Network Scenarios”, Proceedings of the 9th Optoelectronics and Communications Conference/3rd International Conference on the Optical Internet (OECC/COIN2004), 2004.

[4] X. Wang, H. Morikawa, and T. Aoyama, “Burst Optical Deflection Routing Protocol for Wavelength Routing WDM Networks”, Optical Networks Magazine, vol. 8, pp. 12–19, 2002.

[5] Anna Agusti-Torra and Cervello-Pastor, “A New Proposal to Reduce Burst Contention in Optical Burst Switching Networks”, 2nd International Conference on Broadband Networks, vol. 2, 2005.

[6] Son-Hong Ngo, Xiaohong Jiang, and Susumu Horiguchi, “Hybrid Deflection and Retransmission Routing Schemes for OBS Networks”, Workshop on High Performance Switching and Routing, 2006.

[7] Douglas C. Montgomery and George C. Runger, Applied Statistics Probability for Engineers, Wiley, 2006

[8] Robert G. Cowell, A.Philip Dawid, Steffen L. Lauritzen, and David J. Spiegelhalter, Probabilistic Networks and Expert Systems, Springer, 2003

[9] Abdeltouab Belbekkouche and Abdelhakim Hafid, “An Adaptive Reinforcement Learning-based Approach to Reduce Blocking Probability in Buffer-less OBS Networks”, Conference on Communications, 2007 (ICC ’07). IEEE International., pp. 2377–2382, June 2007

[10] B. Shihada and P-H. Ho, "Transport Control Protocol (TCP) in Optical Burst Switched Networks: Issues, Solutions, and Challenges",IEEE Communications Surveys and Tutorials, Second Quarter, Vol. 10, No. 2, pp. 70-86, 2008.

[11] B. Shihada, Q. Zhang, and P-H. Ho, "Threshold-based TCP Vegas over Optical Burst Switched Networks", Proc. of the IEEE International Conference on Computer Communications and Networks (ICCCN), pp. 119-124, 2006.

[12] B. Shihada and P-H. Ho, "A Novel TCP with Dynamic Burst-Contention Loss Notification over OBS Networks", Elsevier Journal of Computer Networks, Vol. 52, No. 2, pp. 461-471, 2008.

[13] B. Shihada, P-H. Ho, and Q. Zhang, "TCP-ENG: Dynamic Explicit Congestion Notification for TCP over OBS Networks", Proc. of theIEEE International Conference on Computer Communications and Networks (ICCCN), pp. 516-521, 2007.

[14] B. Shihada, Q. Zhang, and P-H. Ho, "Performance Evaluation of TCP Vegas over Optical Burst Switched Networks", Proc. of theICST Broadnets, the 6th International Workshop on Optical Burst/Packet Switching (WOBS), pp. 1-8, 2006.

[15] B. Shihada, P-H. Ho, F. Hou, X. Jiang, S. Horiguchi, M. Guo, and H. Mouftah, "BAIMD: A Responsive Rate Control for TCP over Optical Burst Switched (OBS) Networks", Proc. of the IEEE International Conference on Communication (ICC), pp. 2550-2555, 2006.

[16] Bin Zhou, Mostafa Bassiouni, and Guifang Li, “Improving fairness in optical-burst-switching networks,” Journal of Optical Networking, Vol. 3, Issue 4, pp. 214-228 (2004)

[17] B. Shihada, P-H. Ho, Q. Zhang, and J. Jue "A Novel Implementation of TCP Vegas for Optical Burst Switched Networks ", Optical Switching and Networking - Elsevier,

Vol. 7, No. 3, pp 115-126, 2010.

[18] Mendinueta JMD,Bayvel P,Thomsen BC, “Cluster Processing for the Study of Optical Burst-mode Digital Signal Processing Receivers and Subsystems for Dynamic Optical Burst Switching Networks”, 5th International Conference on Broadband Communications Networks and Systems, 2008

[19] S. Parekh, N. Gandhi, J. Hellerstein, D.M. Tilbury, Y.Diao, “MIMO Control of an Apache Web Server: Modeling and Controller Design,” 2002.

[20] B. Shihada, P-H. Ho, and Q. Zhang, "A Novel Congestion Detection Scheme for TCP over OBS Networks", IEEE/OSA Journal of Lightwave Technology, Vol. 27, No. 4, pp. 386-395, 2009.

[21] T. H. Wonnacottand, R.J. Wonnacott, “Introductory statistics for business and economics,” FourthEdition, Wiley, 1990.