Dynamic Congestion-Based Load Balanced Routing in Optical Burst-Switched Networks

Guru P.V. Thodime, Vinod M. Vokkarane, and Jason P. Jue
The University of Texas at Dallas, Richardson, TX 75083-0688
{vgt015000, vinod, jjue}@utdallas.edu

Abstract— **In optical burst-switched networks, data loss may occur when bursts contend for network resources. There have been several proposed solutions to resolve contentions in order to minimize loss. These localized contention resolution techniques react to contention, but do not address the more fundamental problem of congestion. Hence, there is a need for network level contention avoidance using load balanced routing techniques in order to minimize the loss. In this paper, we propose two dynamic congestion-based load balanced routing techniques to avoid congestion. Our simulation results show that the proposed contention avoidance techniques improve the network utilization and reduce the packet loss probability.**

I. INTRODUCTION

The explosive growth of Internet traffic in the last decade has resulted in the deployment of DWDM in the backbone networks. With increase in high bandwidth applications such as HDTV and video-on-demand, DWDM, which offers multi-gigabit rates per wavelength, is going to become the core technology for the next-generation Internet. Optical burst switching (OBS) is an approach used for transmission of data over DWDM networks.

In OBS, the data to be transmitted is assembled into bursts. Each burst has an associated control packet called the burst header packet (BHP). The BHP carries information about the burst such as, source, destination, offset time, and the burst duration. In OBS networks, besides the data channels, each link has one or more control channels to transmit BHPs. The BHP is transmitted ahead of the burst, while the burst is buffered at the source for an offset time. The offset time is large enough for the BHP to be processed at each intermediate node before the arrival of the burst. As the BHP propagates along the route from source to destination, the intermediate nodes process the BHP and configure the optical switches accordingly. Then, the burst cuts-through the optical layer of the intermediate nodes, avoiding any further delays. A reservation technique known as, just-enough-time (JET) [1], reserves the bandwidth on a channel only for the duration of the burst. In this paper, we will consider an OBS network that uses JET.

One of the primary objectives in the design of an OBS network is to minimize packet loss. Packet loss occurs primarily due to the contention of bursts in the bufferless core. Approaches for resolving contention include, wavelength conversion, optical buffering, and deflection routing. In wavelength conversion, if multiple bursts try to use the same wavelength on the same output port at the same time, then the bursts are shifted to another free wavelength on the same link. In buffering, fiber delay lines are used to provide the required delay to resolve the contention [2]. In deflection routing, the burst is deflected to an alternate port in case of a contention on the primary port [3]. Deflection in the network results in several side effects including looping of bursts, and out-of-order packet arrival at the destination.

The above contention resolution techniques are reactive techniques that attempt to resolve contentions rather than avoiding the contentions. Also, these contention resolution techniques attempt to minimize the loss based on the local information at that node. An alternative to resolving contention when it occurs, is to prevent contention before it happens. In contention avoidance, the goal is to reduce the number of contentions, by policing the traffic at the source, or by routing traffic in a way that the congestion in the network is minimized. In this paper, we propose two dynamic congestion-based load balanced routing techniques to avoid congestion.

This paper is organized as follows. Section II describes contention avoidance techniques. Section III describes the proposed dynamic congestion-based routing techniques. Section IV provides simulation results, and Section V concludes the paper.

II. CONTENTION AVOIDANCE TECHNIQUES

The goal of contention avoidance is to reduce the congestion at the bottleneck links in the network. Contention avoidance techniques may be implemented by utilizing source policing or load balanced routing.

Source policing techniques avoid contention by buffering or dropping data at the source, in order to reduce the arrival rate into the network. The policing at the source may be controlled by feedback information that indicates congestion in the network. Such schemes have been applied in standard congestion control protocols such as those in IP, ATM, or TCP.

On the other hand, there are several techniques to avoid contention by balancing the load in the network. Most routing-based techniques involve two stages; route calculation and route selection. The route calculation can be divided into two categories, namely static and dynamic. In static-route calculation, one or more routes are calculated ahead of time, based on some static metric, such as physical distance or number of hops. For example, in fixed alternate path routing, one or more routes can be computed using Dijkstra's

0-7803-7975-6/03/$17.00 (C) 2003

shortest-path algorithm using the chosen metric. In general, these static techniques are suitable when the traffic is fairly steady; however they may suffer if traffic is fluctuating over time. Dynamic route calculation techniques usually compute a single route periodically. The routes are computed periodically based on certain transient (dynamic) traffic information such as link congestion or number of contentions. The information necessary to make the route computation can be obtained in two ways, namely probe-based or broadcast approaches.

In the probe-based approach, the source node initiating the data transfer can send a probe packet into the network. The probe collects the necessary information from the core, and returns to the sender with network information. In the broadcast approach, the core nodes perform an active role by transmitting relevant congestion information periodically to all the edge nodes. In wavelength-routed networks, the probe can either be sent once for every connection request or periodically based on some interval τ. In OBS networks, since the duration of the data transmitted is short, the probe can be sent periodically based on some interval τ. Additionally, in order to reduce the control packet traffic in the broadcast approach, the feedback information about a link can be sent to all edge nodes only if there is a change in the congestion status of the link from the previous value. By doing so, the core nodes can eliminate sending the status packets in certain intervals altogether, thereby ensuring that there is minimal feedback to all the edge nodes. In order to implement such an improvisation, additional memory overhead is incurred at the core nodes, which need to maintain the load status of each of its output links.

In the route-selection stage, once the routes are computed either statically or dynamically, one of the routes is selected for the data transmission. If the route calculation technique computes only a single path, the route selection stage is omitted; thus, route selection primarily applies to static route computation techniques that calculate multiple routes. In static route-selection techniques, a fixed fraction of traffic is sent on each of the alternate paths. The amount of traffic sent on each alternate path is decided based on certain static traffic information. Dynamic route-selection policies are based on feedback information, and operate similar to the dynamic route-calculation techniques. The source nodes obtain the necessary information either by using a probe or feedback messages. Using this information and the dynamic route-selection policy, the data is transmitted on the selected route.

A congestion-based fixed alternate routing technique is proposed for wavelength-routed networks in [4]. The congestion information consists of the number of wavelengths available on each link and the information is obtained by sending a probe message along each alternate route. Hence, the approach employs a static route-calculation with dynamic route-selection technique. A feedback-based scheme for congestion control in traditional networks is proposed in [5]. Here, the congestion information is communicated back to the source by setting a congestion-indication bit on packets flowing in the reverse direction. The source updates the sender's transmission rate based on the feedback.

Stabilization is a significant issue in dynamic route calculation and selection techniques. It is possible that multiple source nodes reacting to congestion simultaneously, will result in oscillation between congested and uncongested states. Hence, additional constraints have to be incorporated to ensure that the edge node does not keep switching all the traffic from one path to the other path every time the edge node receives a feedback (update).

In this paper, we propose two techniques to avoid contention. The first is a congestion-based dynamic route-selection technique using fixed alternate shortest path routes. Secondly, we propose a least-congested dynamic route calculation technique with different weight functions. The two techniques are discussed in-depth in the following sections. In both the techniques, the core nodes in the network gather the load information on their output links and send feedback to all the edge nodes, so as to enable the edge nodes to balance the load. The proposed approach is distributed and is simple to implement with a low control overhead. The congestion information gathered at each core node is the offered load on a link.

III. DYNAMIC CONGESTION-BASED ROUTING TECHNIQUES

In order to implement the proposed techniques, at each core node, the following information is maintained:

$l_{(i,j)}$: Link between node i and node j in the network.

τ: Fixed interval over which the load of the links are calculated.

τ_s: Duration of the bursts that have been successfully scheduled during the interval, τ.

τ_d: Duration of the bursts that have been dropped during the interval, τ.

L_{node}: Node load information table that indicates the load status of each of the output links at the node.

The edge nodes maintain the following information:

L_{net}: Network load information table that indicates the load status of each of the links in the network.

The core nodes measure the load, $\rho_{(i,j)}$ on each of the node's output link. Load is expressed as the duration of all arriving bursts over the interval, τ, that is, $\rho_{(i,j)} = (\tau_s + \tau_d)/\tau$. The load of each link is calculated every τ units of time.

A. Congestion-Based Static-Route Calculation Technique

The congestion-based contention avoidance technique statically computes link-disjoint alternate paths and dynamically selects one of the paths based on the collected congestion information. Let ρ^{MAX} be the maximum load threshold, which, if exceeded, will signal congestion on the link. Therefore, if the load of a link, $l_{(i,j)}$ is above the threshold value, that is, $\rho_{(i,j)} \geq \rho^{MAX}$, then the load status of the link, $LS_{(i,j)}$, is set to one. Once the load status of all the links at a core node have been determined, this information is sent to all edge nodes in the form of load status packet. The edge nodes, upon receiving the load status packet, parse the packet and retrieve each entry. If the entry does not have a corresponding matching entry in the edge node's load information table, L_{net}, then an entry is added to L_{net}. Otherwise, the corresponding entry in L_{net} is updated with new information from the received load status packet.

When the edge node has a burst ready to be transmitted, the node determines whether the burst has to be sent on the primary path or on the alternate path. The edge node calculates the load status of the primary path, $LS^p_{(s,d)}$, where $LS^p_{(s,d)} = \sum_{(i,j)\in r^p_{(s,d)}} (LS_{(i,j)})$ and $r^p_{(s,d)}$ represents the primary path between source, s and destination, d. If $LS^p_{(s,d)}$ is greater than zero, then at least one of the links on the path has its load status set to one, indicating that the primary path is congested. In this case, the edge node calculates the load of the alternate path, $LS^a_{(s,d)}$, where $LS^a_{(s,d)} = \sum_{(i,j)\in r^a_{(s,d)}} (LS_{(i,j)})$ and $r^a_{(s,d)}$ represents the alternate path between source, s and destination, d. If $LS^a_{(s,d)}$ is zero, then the burst is sent on the alternate path, otherwise, the burst is sent on the least congested path, that is, the path corresponding to $MIN(LS^p_{(s,d)}, LS^a_{(s,d)})$. Instead of sending the burst on the already congested primary or alternate route, an alternative is to buffer the burst at the edge node for a fixed duration or to drop the burst at the edge node.

B. Least-Congested Dynamic Route Calculation Technique

In this section, we describe a dynamic route calculation technique. The dynamic route calculation can be based on many different metrics such as the physical distance, number of hops, congestion information, and link utilization. The routes are recomputed every τ units of time.

The weight, $W_{(i,j)}$, is based on a single metric or a combination of metrics. One option is to set the weight function equal to the congestion metric, resulting in the least congested path. The issue with the above metric, is that some of the resultant routes will have a high number of hops. Therefore, while sending the burst on the least congested route results in low packet loss probability at lower loads, under higher loads, longer paths will result in higher overall network loads, thereby increasing the probability of contention. In order to avoid this situation, we consider a weighted function based on congestion as well as hop distance:

$$W_{(i,j)} = \rho_{(i,j)} + 1 \tag{1}$$

where $\rho_{(i,j)}$ is the offered load on the link (i,j).

An other option is to define the weight function based on congestion as well as physical distance:

$$W_{(i,j)} = \rho_{(i,j)} + \frac{d_{(i,j)}}{d^{MAX}}, \tag{2}$$

where $d_{(i,j)}$ is the physical distance of the link (i,j) and d^{MAX} is the maximum physical distance of any link in the network.

In general, the hop-based metric (Eq. 1), results in better performance in terms of loss, since minimal number of nodes are selected in a path, thereby reducing the probability of contention. On the other hand, the distance-based metric (Eq. 2), results in better performance in terms of delay, since minimal link distances are selected in a path, thereby reducing the propagation delay.

C. Parameter Selection

In this section, we describe several issues related to the selection of parameters.

The duration over which the offered load on a link is computed, τ, significantly affects the performance of the load balancing algorithm. The three important considerations in selecting the value of τ are, the amount of control overhead, the accuracy of algorithm, and the effect of outdated information. We see that, if the duration over which the offered load is computed is very short, there will be a large number of feedback packets, thereby increasing the control overhead in the network. Also, the value of the load status computed during this period may not be very accurate, compared to the average load over a longer time duration. Hence, the load value obtained in case of a larger τ value is more accurate. At the same time, if we have larger τ value, the feedback information sent to the edge nodes in the previous round will be outdated, the longer the information is not updated.

In the congestion-based static-route calculation technique, the selection of ρ^{MAX} is also critical, since the value determines if a link is congested or not. Hence, ρ^{MAX} should be chosen based on the desired operating load range of the network. Setting a low value to ρ^{MAX}, will lead to better route selection decisions when the load is low; however, when the operating loads are much higher, ρ^{MAX} will be ineffective, since ρ^{MAX} will signal congestion on all the alternate paths, thereby not providing any useful information for the edge node. On the other hand, setting a high value of ρ^{MAX}, will result in good decisions at high loads; however, at lower loads, due to a high threshold value, all the routes between the source and destination will not be congested. Hence, all the traffic will be sent on the primary path, leading to a congested primary path.

IV. SIMULATION

In order to evaluate the performance of the proposed dynamic congestion-based routing techniques, a simulation model is developed. Burst arrivals to the network are Poisson. Burst length is an exponentially generated random number rounded to the nearest integer multiple of the fixed sized packet length. Mean burst length, is 100 μs, with a packet length of 1250 bytes. The transmission rate is 10 Gb/s, the switching time is 10 μs, and the burst header processing time at each node is 2.5 μs. The primary paths are computed using

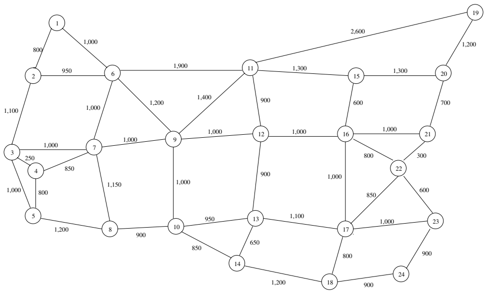

Fig. 1. 24-Node mesh network.

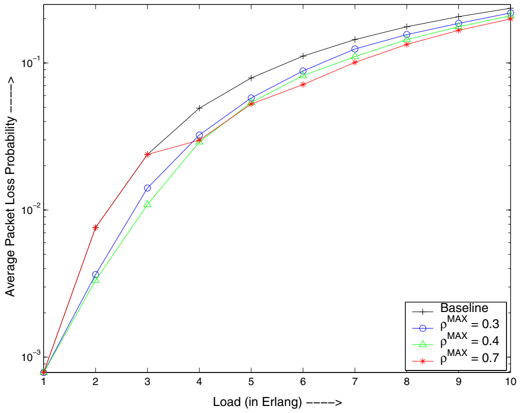

Fig. 2. Packet loss probability versus load for the congestion-based static-route calculation technique with different ρ^{MAX} values.

the shortest-path routing algorithm, while the alternate paths are the link-disjoint next shortest paths for all node pairs. All the simulation results are obtained for a 24-node mesh network, with 43 bi-directional link, an average hop distance of 2.992, an average nodal degree of 3.583, and 4 data channels on each link (Fig. 1). We assume full wavelength conversion at every node. We adopt the latest available unscheduled channel (LAUC) algorithm to schedule data bursts at the core nodes [6]. The LAUC algorithm aims to minimize voids by selecting the latest available unscheduled data channel for each arriving burst.

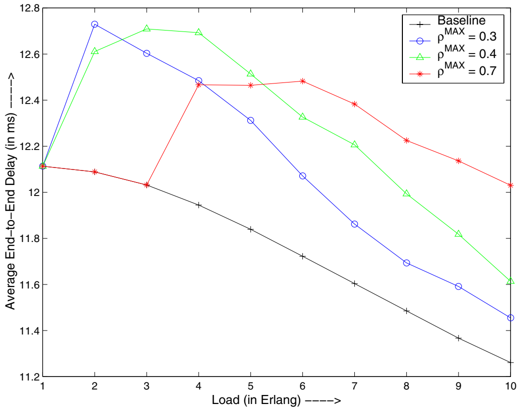

Fig. 3. Average end-to-end delay versus load for the congestion-based static-route calculation technique with different ρ^{MAX} values.

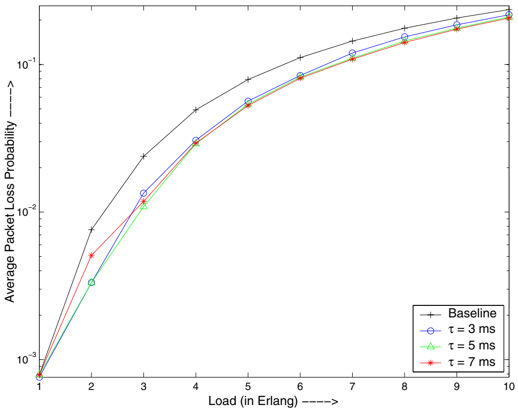

Fig. 4. Packet loss probability versus load for the congestion-based static-route calculation technique with different τ values.

We consider a *baseline* case for the congestion-based static route calculation technique in which there is no alternate routing, to compare the performance of our approaches. We consider average packet loss instead of average burst loss. The packet loss gives the actual measure of data lost in the network, as the bursts vary in size. Fig. 2 plots the average packet loss probability versus load for the congestion-based static route calculation technique for different values of ρ^{MAX}, with τ equal to 5 ms. We observe that the network employing the proposed congestion-based alternate routing approach performs better than the network without any alternate routing scheme. We observe that at low loads, ρ^{MAX} equal to 0.3 performs better than higher values of ρ^{MAX}, while at high loads, a high ρ^{MAX} value performs better than lower values of ρ^{MAX}. Since at low loads, most of the links are under utilized, a low ρ^{MAX} value will trigger the route selection policy earlier than with high ρ^{MAX} values. Therefore, the former case (low ρ^{MAX}) would make use of the alternate path for diverting the incoming traffic, resulting in lower loss. As the load increases, the former case will deflect most of the incoming traffic and hence would result in going a greater number of hops. Since at high loads, most of the links on the alternate path would also be congested, the probability of contention in the network is increased. Hence, the latter (high ρ^{MAX}) performs better at high loads.

Fig. 3 plots the average end-to-end packet delay versus load for the congestion-based static route calculation technique for different values of ρ^{MAX}, with τ equal to 5 ms. We can observe that, at low loads, low ρ^{MAX} values have higher delay when compared to the case in which ρ^{MAX} equal to 0.7, since with low ρ^{MAX} values the route selection policy is triggered at low loads. Hence, selecting the alternate path with a greater number of hops accounts for higher delay. At high loads, a high ρ^{MAX} value incurs higher delay.

Fig. 4 plots the average packet loss probability versus load for the congestion-based static route calculation technique for different values of τ, with ρ^{MAX} equal to 0.4. We observe that, at low loads, low values of τ, such as 3 ms, perform better than higher values of τ. As the load increases, higher values of τ perform better, since at low loads most of the links are under-utilized. Also at low loads, as the time

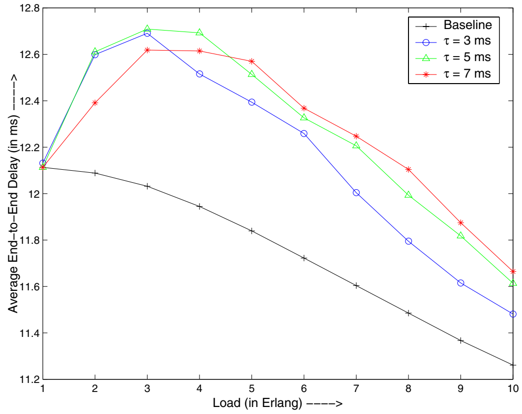

Fig. 5. Average end-to-end delay versus load for the congestion-based static-route calculation technique with different τ values.

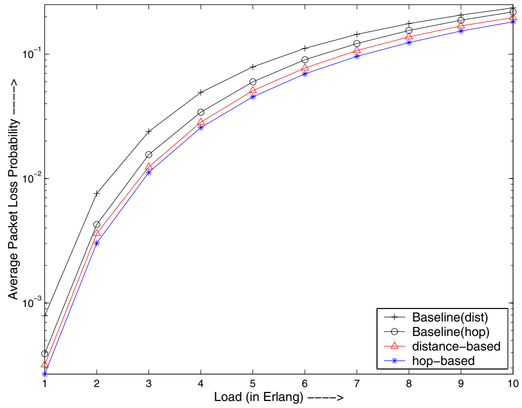

Fig. 6. Packet loss probability versus load for the least-congested dynamic route calculation technique.

interval, τ, is short, most of the links on the primary path would signal congestion, and would deflect the traffic onto the alternate path, resulting in a successful transmission. At high loads, even the links on alternate path would be congested, thereby increasing the probability of dropped packets, since the deflected traffic would have to traverse a longer path.

Fig. 5 plots the average end-to-end packet delay versus load for the congestion-based static route calculation technique for different values of ρ^{MAX}. We observe that, at low loads, lower τ values, such as 3 ms, incur high delay as compared to higher τ values. At high loads, a higher τ value will result in high delay, since there are more successful transmissions on alternate paths.

In order to evaluate the performance of the least-congested dynamic route calculation technique, we consider *min-distance* and *min-hop* techniques, where in a single fixed shortest-path is calculated based on the physical distance and the number of hops, respectively. Fig. 6 plots the average packet loss probability versus load for the least-congested dynamic route calculation technique, for τ equal to 5 ms. We observe that the dynamic route calculation technique performs better than the min-distance and min-hop techniques at all loads. The hop-based metric performs better than the distance-based metric, since the hop-based metric will result in fewer hops as compared to the distance-based metric. As a result, there is a lower probability of contentions.

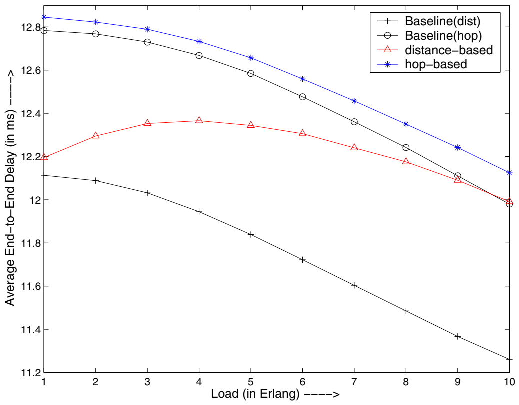

Fig. 7. Average end-to-end delay versus load for the least-congested dynamic route calculation technique.

Fig. 7 plots the average end-to-end packet delay versus load for the least-congested dynamic route calculation technique. We observe that the distance-based metric results in lower delay compared to the hop-based metric, since the distance-based metric computes the shortest distance path from each source to destination.

V. CONCLUSION

In this paper we have proposed two congestion-based routing techniques for optical burst switching. These techniques significantly reduce the packet loss probability in the network, as compared to networks without any contention avoidance techniques. Also, the proposed contention avoidance techniques can be applied to other all-optical networks, such as optical packet-switched networks.

Areas of future work are to maintain more than one alternate paths between source-destinations pair in order to avoid contention in higher nodal degree network topologies. Also, since the proposed techniques are dynamic, we are working on stabilization policies to make the techniques more robust.

REFERENCES

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

[2] S. Yao, B. Mukherjee, S.J.B. Yoo, and S. Dixit, “All-Optical Packet-Switched Networks: A Study of Contention Resolution Schemes in an Irregular Mesh Network with Variable-Sized Packets,” *Proceedings, SPIE OptiComm 2000*, Dallas, TX, pp.235-246, Oct. 2000.

[3] A. Bononi, G. A. Castanon, and O. K. Tonguz, “Analysis of Hot-Potato Optical Networks with Wavelength Conversion, ” *IEEE Journal of Lightwave Technology*, vol. 17, no. 4, pp. 525-534, April 1999.

[4] L. Li and A.K. Somani, “Dynamic Wavelength Routing Using Congestion and Neighborhood Information,” IEEE/ACM Transactions on Networking, vol.1, no.5, pp. 779-786, October 1999.

[5] K.K. Ramakrishnan and Raj Jain, “A Binary Feedback Scheme for Congestion Avoidance in Computer Networks,” ACM Transactions on Computer Systems, vol.8, no.2, pp.158-181, May 1990.

[6] Y. Xiong, M. Vanderhoute, and H.C. Cankaya, “Control Architecture in Optical Burst-Switched WDM Networks,” IEEE Journal on Selected Areas in Communications, vol. 18, no. 10, pp. 1838-1851, October 2000.