Distributed Algorithms and Architectures for Optical Flow Switching in WDM Networks[1]

Bishwaroop Ganguly[2] and Eytan Modiano
Massachusetts Institute of Technology
Laboratory for Information and Decision Systems
Cambridge, Massachusetts
ganguly,modiano@mit.edu

Abstract

This paper is about the design and quantitative analysis of distributed approaches for optical flow switching (using dynamic lightpath setup) in a wide area WDM backbone network. The major contribution of this work is to design realistic integrated approaches for flow routing, wavelength assignment, and connection setup in a truly distributed setting, and assess the relative performance of these on a nationwide wide area network (WAN) backbone. A simulation model is used which models timing, wavelength, and fiber constraints. Results are presented in terms of backbone utilization for a given blocking probability of optical flows.

1 Introduction

Optical flow switching in large-scale wavelength division multiplexing (WDM) networks is an emerging technology whose aim is to take advantage of futuristic, dynamic optical backbone networks. It involves the dynamic setup and use of optical lightpaths between users that want to communicate at high data rates. The potential benefits of flow switching are improved user quality of service (QoS), and bypass of electronic processing nodes in a backbone network, increasing the effective capacity of electronic backbone routers. Our work addresses *architectural* issues involved in flow-switching in a backbone network. These issues include routing, wavelength assignment and connection setup for lightpaths. Routing involves choosing a path on which to send, while the related wavelength assignment problem addresses how to assign a wavelength for lightpaths. Finally, connection setup involves how to signal the network to actually create lightpaths, given a route and a wavelength.

We have designed *integrated* and *distributed* approaches to address the architectural issues of flow switching listed above. Our approaches are integrated in that they address routing, wavelength assignment and connection setup together. They are distributed, in that they do not rely on centralized, perfect network state information, which is impractical for a WAN network. Both of these aspects are departures from previous work.

In this paper, we present and analyze several approaches to flow switching architecture. We present simulation results for our architectures applied to a dynamic WDM backbone network with realistic latency characteristics. Our goal is to assess the relative performance of our approaches, and identify which aspects of them critically affect optical flow switching performance. In addition, we would like to draw conclusions about the feasibility regime of optical flow switching in a nationwide backbone scenario. Our results show that a flow-switching network architecture must overcome distribution of network state information for good performance, even for modest granularity (1 second) flows.

The rest of the paper is organized as follows. Section 2 introduces the concept of optical flow switching in detail and discusses related work in the area. Section 3 details and justifies our candidate integrated approaches. Section 4 highlights the important details of our simulation study, and Section 5 presents our simulation results. Finally, Section 6 discusses our results and future work.

2 Background

A backbone network employing optical flow switching differs from today's backbone by having the capability to set up end-to-end optical lightpaths dynamically. These lightpaths are used to route flows of data all-optically between users. Optical flow switching provides several of benefits over a statically configured approach. First, flow switching provides *optical bypass* of electronic routers in the backbone. A large transaction sent over an optical path

1. B. Ganguly is supported by the Defense Research Projects Agency (DARPA) under the Next Generation Internet (NGI) initiative. Opinions, interpretations, recommendations and conclusions are those of the author and are not necessarily endorsed by the United States Air Force

2. Currently a Lincoln Scholar at MIT Lincoln Laboratory.

0-7695-0722-0/00 $10.00 © 2000 IEEE

bypasses all electronic routers in that path, thus lessening the burden on the intermediate routers, and avoiding a mismatch between electronic processing speeds and optical transmission rates (i.e. the opto-electronic bottleneck). Second, end-to-end optical lightpaths can be used to give users high quality of service (QoS), in terms of bandwidth and delay, since the lightpath is dedicated to that user. Finally, dynamic optical flow switching allows the backbone optical resources to be shared among users.

The work presented here addresses the three architecture layer issues of flow routing, wavelength assignment and connection setup in a WAN. The contribution of this work is to present and analyze performance of *integrated* architectural approaches in a *distributed* setting. Past work has examined the individual architectural problems of flow switching, but has not presented a stand-alone, robust solution that encompasses all of the aforementioned problems. In addition, most of the related work in this area assumes centralized network state information that is maintained in a single, globally-accessible data structure. Our work is a departure from this, as it explores distributed approaches where no such data structure exists in the network. We believe de-centralization of network state information will have a significant impact on both relative and absolute performance achievements of flow-switching schemes.

2.1 Related Work

The general problem of routing in data networks has been widely studied, and is treated in several data network textbooks [1]. For WDM optical networks, the wavelength routing problem is important. Most previous studies involve static or quasi-static networks, that do little dynamic lightpath setup [7]. Wavelength assignment [9] in WDM networks deals with dynamic, per-call wavelength setup in optical networks.

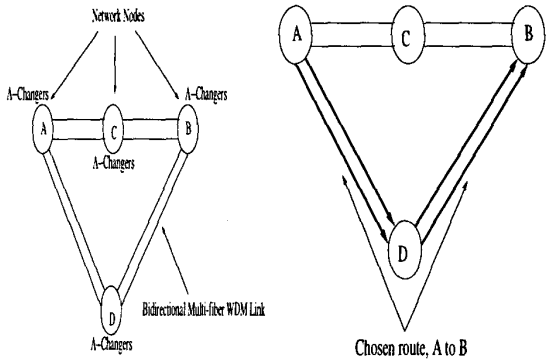
For dynamic optical networks of the type discussed in this paper, the routing and wavelength assignment problem are highly related. Recent work has addressed the combined routing and wavelength assignment problem. Both [6] and [2] present a number of combined algorithms for the routing and wavelength assignment problem in WDM networks, with the latter comparing performance of algorithms with and without wavelength conversion. While this work presents and compares a number of approaches, it assumes a centralized network state database to make routing decisions.

Connection setup in electronic networks has been studied. Efforts such as IP Switching over ATM [4] and MPLS [8] have defined standard connection setup protocols for electronic flows. For optical networks, Terabit Burst Switching [10] is an approach in which a lightpath is set up for an optical burst of data by an electronic setup packet that precedes the burst. Two optical connection establishment approaches are described in [5] which essentially use control messages to reserve lightpaths through an optical network between two nodes. Joint routing, wavelength assignment, and *path discovery*, which attempts to find an available path to route a lightpath, are not addressed by current connection setup work.

3 Integrated Flow-Switching Approaches

In this section, we present three possible approaches to optical flow switching. Each of these approaches combine the problems of wavelength assignment, flow routing and connection setup. They all assume an optical network structure with optical switches at each node, that allow for dynamic lightpath setup. They also assume availability of a reliable control network for signaling network entities as needed. Flow requests are generated by end-user nodes, and are received by the network infrastructure. If the network finds insufficient resources for a flow, the flow is dropped. Assume for the remainder of this section that a flow is requested from node A to node B in the network pictured in Figure 1(a).

3.1 Ideal



(a) Network (b) Routing

Figure 1. Ideal Approach Illustration

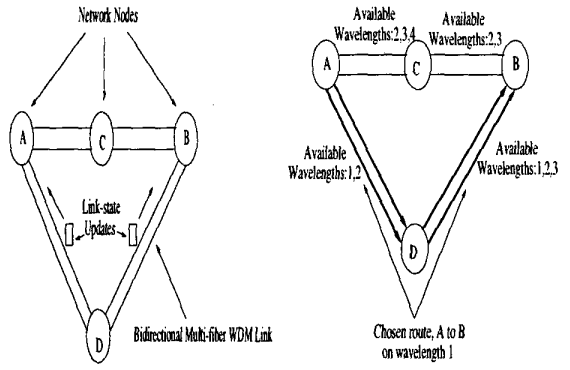
The **Ideal** architecture for flow switching is illustrated in Figure 1(a). Shown is a multi-hop optical network with *full wavelength changers* at each node. Thus, an input to a node on any wavelength can be switched to any output wavelength on any output fiber. This eliminates the issue of wavelength assignment for flows in the Ideal scenario.

The architecture uses a centralized, globally-accessible data structure to maintain network state. This state information is accurate and is used in all individual flow routing decisions. Routing is performed using a method which maximizes the capacity of the minimum capacity link on the selected route. This type of routing is similar to Least-Loaded Route routing, and is shown to give good performance in [2]. Connection setup is achieved using *tell-and-go*, in which a control packet precedes the optical flow along the chosen route (here, A-D-B). In this case, the switch at node D will configured by the "Tell" packet sent from A to set up a lightpath for the trailing optical flow. If, at any hop resources are not available on the outgoing link, the control packet and flow are terminated.

The need for global network information at each node makes the Ideal approach not practically realizable in a WAN. We assume that it performs better than any approach without wavelength conversion, and coherent global network state information. It serves as an upper bound for performance of flow switching architectures in our study.

3.2 Tell-and-Go

Tell-and-Go (TG) is a distributed algorithm with no wavelength conversion, and is based on link-state updates as shown in Figure 2(a). Received updates are processed at each node into a table of global network state. Note that these tables can contain inaccurate information due to latency involved in broadcast updates.



(a) Update

(b) Wavelength, Route Selection

Figure 2. Tell-and-Go Approach Illustration

Given the network state information, TG uses a combined routing and wavelength assignment strategy. For any routing decision, the K shortest paths are considered, where K is a variable parameter in our study. Routing proceeds by using the *First-fit* wavelength assignment strategy over the K paths. First-fit numbers all wavelengths in sequential order, and routes a flow over the path with the lowest-numbered wavelength available at every hop. Figure 2(b) shows that the path from A to B through D has wavelength 1 available at all hops, whereas the lowest number wavelength available from A to B through C is 2. Thus, the path from A to B through D using wavelength 1 is chosen to route the flow. If no route with an available wavelength is found, the flow is dropped. As in Ideal, tell-and-go is used for connection setup.

Our goal in implementing TG is to assess the viability and performance of a link-state (or update-based) protocol in the context of optical flow switching. If link-state updates provide information similar to a theoretical centralized information database, then TG can make good routing decisions, and will result in high efficiency. If, on the other hand, link-state updates cannot provide sufficiently accurate information, TG becomes a strategy in which flows are routed along random routes.

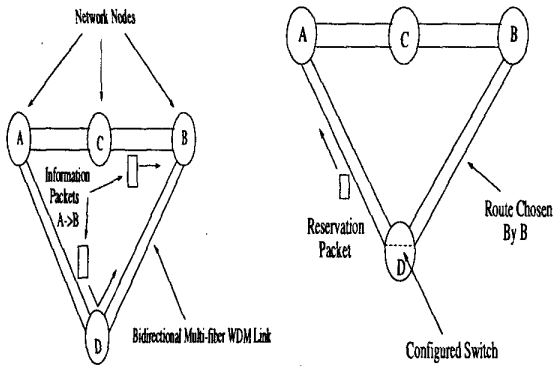
3.3 Reverse Reservation

An alternative to maintaining global network state information is to do on-demand *path discovery*. In the **Reverse Reservation (RR)** flow switching architecture, the initiator of a flow sends information-gathering packets (info-packets) to the flow destination, on the K shortest paths. These packets record link-state information at each hop, and upon arrival at the destination, contain information about the link state of all links along the specified route. Figure 3(a) shows node A sending info-packets along two paths to node B.

Routing and wavelength assignment is then performed by the *destination* node of the flow, once all K info-packets from the sender have arrived. The calculation is done using the *First-fit* strategy as described above. As in TG, if no route with an available wavelength is found, the flow is dropped. Once a route has been selected by the receiver, a reservation control packet is sent along the chosen route in reverse, as shown in Figure 3(b), establish the connection. This control packet configures switches along the chosen route for the chosen wavelength, arrives at the flow sender, and informs the sender that the lightpath is configured and to send the flow.

If the reservation packet finds insufficient available resources anywhere on its journey, the reservation is terminated, and all resources held by the reservation in progress are released by sending additional control packets. A successful reservation sends the flow along the reserved lightpath, in this case A-D-B.

RR is an attractive approach, because it does not rely on periodic or event-driven updates to maintain global network



(a) Path Discovery

(b) Reservation (Conn. Setup)

Figure 3. Reverse Reservation Approach Illustration

state and do path discovery. Our goal is to study its effectiveness in a WAN scenario, where end-to-end latency of control packets is a factor.

4 Simulation Overview

Simulations of our integrated flow-switching approaches were conducted using the Opnet [3] network modeling tool on a fixed topology. The topology used closely resembles the vBNS data backbone network and is a US nationwide backbone network, comprised of around 10 nodes (cities), approximately 3000 miles in diameter. The links connecting the nodes are bi-directional multi-fiber WDM optical links, and there is an implicit bi-directional control network link per WDM link. The delays between nodes are assumed to be the actual distances between nodes times propagation delay of typical optical fiber. Each node has a fully configurable optical switch, which can connect any input wavelength and fiber to any output fiber (using the same wavelength). We assume each node has a large number of transmitters and receivers.

The traffic model used is a Poisson arrival distribution of flows with exponentially distributed durations. Each node has a flow arrival process, and each process is independent. By varying the parameters of the flow arrival and duration random processes, we change the nature of the flow traffic arriving to the overall network. Flow destinations are randomly chosen upon arrival.

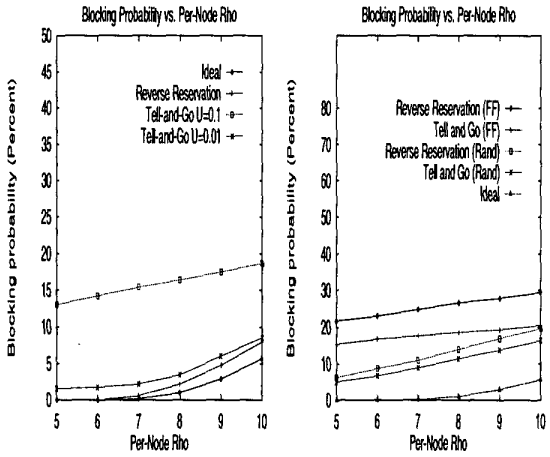
4.1 Simulation Parameters

A number of parameters can be varied in our study, and each plays a role in determining relative and absolute performance of our approaches. A list follows: **F** - The number of fibers per link. Assumed to be the same for all links. **L** - The number of wavelengths *per fiber*. Assumed to be the same for all links. **K** - Number of routes considered by the routing approaches, for each routing decision. **U** - Update interval (seconds). For TG, this is the interval at which broadcast updates of link state are sent by all nodes. λ - The arrival rate of flows (flows/second). Assumed to be the same for all nodes. μ - Average service rate of flows (flows/second). Assumed to be the same for all nodes. ρ - Traffic intensity. Equal to λ/μ.

Each of these parameters can be varied in our simulation to answer various questions about the merits of flow switching architectures.

5 Simulation Results

In this section, we present results of our simulations, and discuss the relative and absolute performance of our architectures for flow switching. All simulations were run as described in Section 4, with parameters set as described below.



(a) Zero Latency Control Network

(b) Realistic Latency Control Network

Figure 4. Average One Second Flow Results

5.1 Zero Latency Control Network

We first simulated our approaches using a latency-free control network in which all control packets are instantly delivered. For TG and Ideal, this means that the tell packets make immediate reservations for trailing flows. For RR, this implies that information is gathered and reservations are made instantly. These results show performance for our approaches under ideal conditions, but without wavelength conversion for TG and RR.

Figure 4(a) plots Blocking Probability vs. ρ per-node for flows that one seconds. For these figures, K=10, F=1, L=16, and U=0.1 or U=0.01. The flow graphs show that performance is ordered Ideal, RR, TG (U=0.01) and TG (U=0.1) from best to worst. For shorter duration flows, TG (U=0.1) performs significantly worse than the two others, due to the update interval resulting in inaccurate network state information. Even though the updates suffer no latency, the state of the network changes significantly between updates. Reducing the update interval, U, to two orders of magnitude less than average flow duration seems to alleviate the problem greatly.

The results show that RR and TG with short update intervals relative to flow duration are very competitive with Ideal, even though the latter has the significant advantage of full wavelength conversion. The Ideal approach has an overall worst-case blocking probability of 5.7%. RRs blocking probability is no more than 40% worse than the Ideal case for any simulation and is no worse than 8% overall. TG (U=0.01) has a maximum blocking probability around 8.5% for any run, which is only 49% worse than Ideal. It is likely that TGs blocking probability can be made better for shorter flows by further decreasing the update interval, U, but this may be impractical.

5.2 Control Network With Latency

We now analyze results for a control network that models nationwide, end-to-end optical fiber latency. Figure 4(b) shows results for our approaches with real control latency, for one second average duration flows. For this figure, K=10, F=1, L=16, and U=.1.

The results show that latency has a profound impact on relative performance of the approaches. For one-second duration flows, both TG and RR have three to four times worse blocking probabilities compared to Ideal in the highest traffic intensity case. Clearly, the latency of the control messages is having a profound effect on the information being used to make routing decisions. For ten second flows (not shown), TG and RR fared better (50% and 80% worse than Ideal,respectively), due to slower-changing network state.

Another interesting observation is that TG performs better than RR in the network with control latency where just the opposite was true in the zero control latency network. The explanation for this is as follows: Measurements have shown that the average reservation time is 0.06 seconds. When ρ=10, there is a 40% chance that two consecutive flows arrive to a given node within a reservation time of one another. In RR, the reservation does not change the state of the network until the return trip. Therefore, the second arriving flow will gather *the same* information as the previous flow, with 40% probability. This results in the two flows competing for the same resources, and an increased chance of collision. This explanation is strengthened by further results, which showed that for ten second average duration flows, the problem is not apparent. The reversal of TG and RR seems to imply a fundamental limitation of the RR approach as flows arrive more quickly. Strategies to overcome this include introducing a more randomized wavelength selection policy.

Overall, these results show that even relatively small latency (in this case, flight time of control messages), *greatly* impacts the blocking performance of flow switching schemes. Subsection 5.1 and papers listed in section 2 have shown good performance for proposed routing and wavelength assignment algorithms in idealized networks. However, the results shown here demonstrate that this is not enough to assure good blocking performance for flow-switching in a WAN scenario.

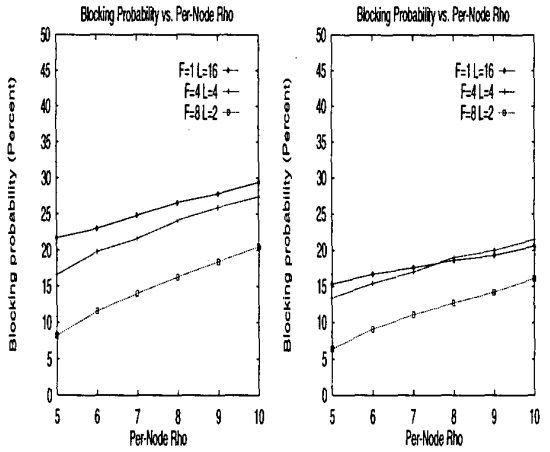
5.3 Multiple Fiber Network

In this subsection, we present results using multiple fibers (F>1) per WDM link. In this case, the aggregate number of channels per link remains the same (F*L = constant), but the number of channels per fiber (L) decreases with increasing F. For the Ideal case, multiple fibers makes no difference, due to full wavelength conversion capability at each node.

For TG and RR, the change to multiple fibers per link is expected to make blocking performance better, because we can map a particular input wavelength onto *any output fiber*, as long as the selected wavelength on that fiber is free. Finally, it is easy to see that one channel per fiber (L=1) is equivalent to having wavelength conversion at each node.

Results of multi-fiber runs are shown in Figures 5(a)(b), for RR and TG, respectively, for flows that average one second in duration. For these plots, K=10, L and F are as shown, and U=0.1.

In Figure 5(a) we observe that RRs blocking probability decreases with increasing F. This is to be expected, as larger F gives RR more choices to route flows on a particular channel and path at each hop. However, the improvement falls far short of the Ideal curve shown in Figure 4(b). Even with F=8, which is close to having wavelength conversion at each node, RRs blocking probability is many times worse



(a) RR, Ave. Flow Duration = 1sec.

(b) TG, Ave. Flow Duration = 1sec.

Figure 5. Multi-Fiber Network Results

than Ideals.

Figure 5(b) shows an interesting trend for TG. With F=8, performance improves over F=4 and F=1. However, the F=1 and F=4 cases are extremely close in performance with F=4 actually performing worse for $\rho > 8$. This result is counterintuitive, based on the arguments we presented in favor of having multiple fibers in the network to help with wavelength related blocking.

We have determined that the wavelength selection approach that we have used, first-fit, is the main cause of poor multi-fiber performance. In brief, first-fit has a tendency to *prioritize* lower-numbered wavelengths, over-utilizing them, while others remain underutilized. Previous work has hypothesized that first-fit is a good strategy because of its simplicity and performance. In Subsection 5.1 we found this to be true in a zero latency network. However, for a multi-fiber distributed network scenario, first-fit causes severe blocking problems, and other approaches must be investigated.

6 Conclusions

The overall performance of our flow switching approaches shows that a significant amount of traffic a can be switched optically, despite end-to-end to end latency of control and data messages. A number of key factors affect performance, and each of these factors affect each approach differently.

For TG, the update interval appears to be the key parameter that governs performance. Our quantitative study suggests that U must be two orders of magnitude lower than the average flow duration to achieve reasonable performance in our WAN network. Intuitively, this prevents the average flow duration from being lower than a second as U<.01sec seems unreasonable in any real network, due to control network constraints.

RR seems to suffer more than TG when latency is introduced in the network. Note that *any hop* that changes state in the time intervening between information gathering and reservation can drop an entire reservation. As flows become shorter, the probability of this happening is higher.

In summary, our approaches for dynamic optical flow switching approach the performance of an ideal approach with wavelength changers in certain cases. We have identified several key aspects of distributed implementations that need attention. The most important of these is overcoming the latency of control signaling in order to allow for smaller duration flows to be sent optically. Our future work addresses these concerns and will move us toward a integrated high-performance approach that can be applied to future dynamic optical backbone networks.

References

[1] D. Bertsekas and R. Gallager. *Data Networks*. Prentice Hall, 1992.

[2] E. Karasan and E. Ayanoglu. Effects of wavelength routing and selection algorithms on wavelength conversion gain in wdm optical networks. *IEEE/ACM Transactions on Networking*, 6(2):186–196, April 1998.

[3] I. Katzela. *Modeling and Simulating Communication Networks: A Hands-On Approach Using OPNET*. Prentice Hall, 1998.

[4] S. Lin and N. McKeown. A simulation study of ip switching. *ACM. Computer Communication Review*, 27(4):15–24, 1997.

[5] Y. Mei and C. Qiao. Efficient distributed control protocols for wdm all-optical networks. In *Sixth International Conference on Computer Communications and Networks*, pages 150–153, 1997.

[6] A. Mokhtar and M. Azizoglu. Adaptive wavelength routing in all-optical networks. *IEEE/ACM Transactions on Networking*, 6(2):197–206, April 1998.

[7] R. Ramaswami and K. N. Sivarajan. *Optical Networks: A Practical Perspective*. Morgan Kaufman, 1998.

[8] E. C. Rosen, A. Viswanathan, and R. Callon. Multiprotocol label switching architecture. Technical report, Internet Engineering Task Force, July 1998.

[9] S. Subramaniam and R. A. Barry. Wavelength assignment in fixed routing wdm networks. In *1997 IEEE International Conference on Communications*, volume 1, pages 406–410, 1997.

[10] J. Turner. Terabit burst switching. *Journal of High Speed Networks*, 8(1):3–16, 1999.