A Scheduled Approach to Optical Flow Switching in the ONRAMP Optical Access Network Testbed

Bishwaroop Ganguly, Vincent Chan
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
ganguly,chan@mit.edu

Abstract: Optical Flow Switching (OFS) creates short-duration optical connections between users on-demand. We have designed and implemented OFS in the ONRAMP optical access network. We demonstrate scheduled optical connections, achieveing 90% efficiency for one-second optical transactions.

1 Introduction

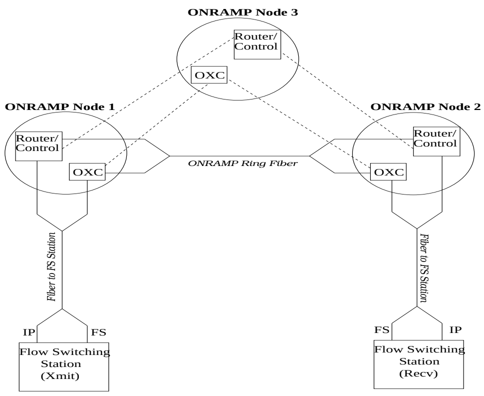
Optical flow switching (OFS) in large-scale wavelength division multiplexing (WDM) networks is an emerging technology enabled by rapid advancement in optical switch and transceiver technology [1]. It is a connection oriented service involving dynamic creation of relatively short (one second or less) duration optical lightpaths between users that desire a high-rate, high quality of service (QoS) connection. A flow switched transaction *bypasses* electronic routers along its path alleviating the potential load on the routers. This provides a twofold benefit. First, a larger amount of data is switched by each node using cheap, off the shelf optical and control computer technology. Second, high QoS is provided to the flow switching user. This can be contrasted with a non-OFS approach which would use a QoS-capable electronic router at an ostensibly higher cost.

We have designed and implemented OFS in the NGI ONRAMP testbed. NGI ONRAMP is an acronym for Next Generation Internet Optical Network for Regional Access using Multi-wavelength Protocols. It is an optical access area network deployed in the metropolitan Boston area. For OFS in such large-scale networks, a critical issue is that of connection setup. In ONRAMP, we have taken the view that it is vital for the network management and control system to have reliable network state information to make decisions about routing and wavelength assignment. Previous work [2] has shown that inaccurate information leads to inefficiency, when connections are brief and arrive in rapid succession.

To solve this problem in ONRAMP OFS, we employ a *scheduled* approach to on-demand connection setup. We divide time into fixed-length timeslots and schedule transmissions which span single or multiple timeslots. This is efficient but requires synchronization of optical network control processes. These on-demand transactions efficiently utilize the Gigabit Ethernet link layer. Our implementation uses standard (i.e. non real-time) OS and widely available processor clock synchronization infrastructure, Network Timing Protocol (NTP)[3]. These results show both the viability and scalability of a scheduled connection setup approach, and is the groundwork for a number of connection scheduling algorithms for OFS in a WAN scenario. This paper is organized as follows. Section 2 describes the ONRAMP testbed, and discusses the network layer for OFS. Section 3 describes the results of the OFS demonstration.

2 OFS in the ONRAMP Testbed

Figure 1(a) shows the ONRAMP network testbed consists of three nodes in a two fiber ring. A passive, remotely pumped Distribution Network connects send and receive stations to the access ring. The two fibers each carry 8 1540-1560nm channels in opposite directions. Each node in the ring has an Optical Cross Connect (OXC) and a node controller/IP router. The controller stations are Alpha processor based multiprocessor computers, running the Linux operating system. A dedicated 1510nm control channel connects all controller stations in the ring. ONRAMP employs an integrated optical/IP network control strategy. The control station/router switches both control and data packets for non flow-switched communication. It also contains software to configure and manage its local OXC.



(a) ONRAMP Topology

X => Output in use at corresponding timeslot

	Slot1	Slot2	Slot3	Slot5	Slot6	Slot7	...
Channel 1			X			X	
Channel 2		X		X			
Channel 3	X						
Channel 4			X		X		
Channel 5	X			X			
...							

(b) Example Node Schedule

Fig. 1. ONRAMP Testbed

2.1 *OFS description*

Figure 1(a) also shows Flow Switching Stations (FSS) in ONRAMP. Each FSS (Xmit and Recv) have dual connections to the network, using two Gigabit Ethernet cards. The connection labeled *IP* is fixed-tuned to a specific wavelength, and is the default communication channel for lower rate data and control packets. It is terminated at a port in the local controller/router. The other connection, labeled *FS* is the wavelength-tunable flow switching connection. FS feeds directly to the local OXC, and thereby onto the access ring. A network process runs on each FSS station to send, receive and process network control packets.

Given network timing information, many on-demand scheduled connection setup approaches are possible. As an example, we present connection setup in our current implementation in ONRAMP. (see Figures 1(a)(b)).

1. Flow Switching Station (Xmit) makes a flow request to local ONRAMP Node (Node 1). Assume transmission target is Recv, and duration of transmission is known.
2. Node 1 computes the earliest timeslot with a free route, wavelength pair for Xmits transmission (t0). This scheduling is on-demand.
3. Node 1 communicates t0 to Nodes 2 and 3, which update OXC schedules (see Figure 1(b)).
4. At time t0, Nodes 1, 2 and 3 configure OXCs (as per their local schedule) and inform Recv and Xmit to tune to the chosen channel. Xmit is then informed of the newly created connection.
5. Upon expiration of flow time, resources are released for next transmission.

This particular scheme maintains network state information at a central location (Node 1). However, in a more general scheme where state information is distributed, resources can be scheduled on demand using *accurate* information about the future network state. For a non-scheduled approach, these resources may have to be queried for availability at transmission time, leading to inefficiency.

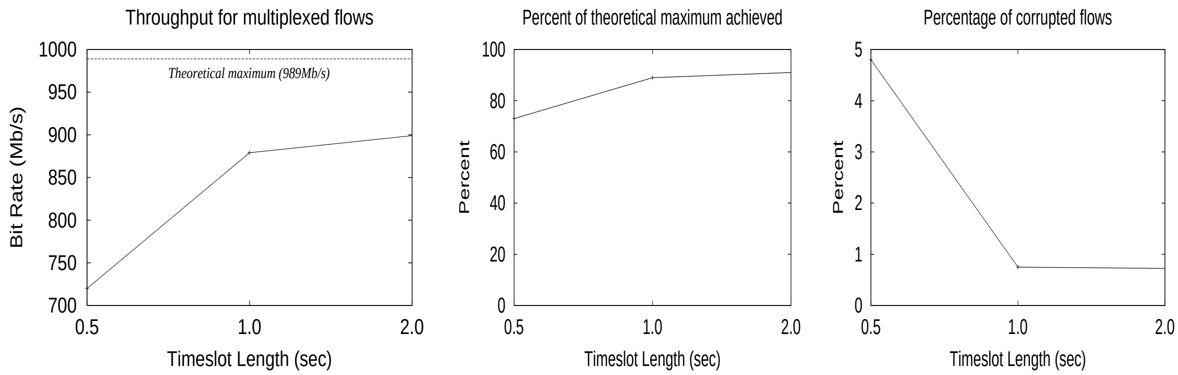
Both FSSs are ignorant of synchronization, promoting scalability. The time that the network must wait between timeslots for the network to reconfigure is the *network settling time*. This is represented by the thick vertical lines in the OXC schedule, Figure 1(b). Settling time is overhead in a scheduled connection setup approach and should be minimized for high efficiency.

2.2 *OFS Transport Layer*

OFS transactions are unidirectional high-rate data connections. For transport of data, we have chosen to use the connection-less UDP datagram delivery protocol. We have implemented a simple but effective *send-side rate limited* transport layer implementation. This implementation allows these short-lived transactions to fully utilize the Gigabit Ethernet link layer, achieving the theoretical maximum of 989Mb/s.

There are a number of advantages of this network layer over a two-way connection-based approach (e.g. TCP) First, this transport implementation matches the maximum send rate of the link layer immediately, as opposed to an approach such as TCP slow start. This is especially important for short duration connections. Second, there is no need for a reverse path such as that needed for TCP acknowledgments. This avoids the need to allocate a dedicated path for protocol feedback, and averts the risk of loss of acknowledgments in a feedback path such as the Internet. Note that unidirectional flows still require an acknowledgment of a completed transaction.

3 OFS Demonstration Results



(a) Throughput (b) Pct. of max. throughput (c) Percent of flows corrupted

Fig. 2. Demonstration Results

The performance results of our demonstration are shown in Figures 2(a)(b)(c). These results are for more than 200 single timeslot transactions, sent over the entire channel space in both directions of the ring from Xmit to Recv (see Figure 1(a)). Figure 2(a) shows that the scheduled connection setup achieved very high bit rates for the multiplexed flows, close to the theoretical maximum. Figure 2(b), shows that for one-second flows, we achieve 90% efficiency, though we are changing routing and wavelength assignment on a second-by-second basis. Figure 2(c) shows that the limiting factor for performance for 0.5 second timeslots was the corruption of flows. We believe this is due to unresponsiveness of the link layer to reconfiguration and wavelength switching.

References

1. Vincent W.S. Chan et al., *Architectures and Technologies for High-Speed Optical Data Networks*, (Journal of Lightwave Technology), Vol. 15, No. 12, pp 2146-2168.
2. Bishwaroop Ganguly and Eytan Modiano, *Distributed Algorithms and Architectures for Optical Flow Switching in WDM networks*, in (Proceedings of ISCC 2000), pp 134-139.
3. D.L. Mills, *On the accuracy and stability of clocks synchronized by the Network Time Protocol in the Internet system*, (ACM Computer Communication Review), Vol. 20, No. 1, pp 65-75.