JUST-IN-TIME OPTICAL BURST SWITCHING FOR MULTIWAVELENGTH NETWORKS

John Y. Wei, Jorge L. Pastor, Ramu S. Ramamurthy, and Yukun Tsai *
Telcordia Technologies
331 Newman Springs Road
Red Bank, NJ 07701-5699
U. S. A.
{ wei,jorel,ramu,ykt } @research.telcordia.com

Abstract   We describe the architecture, performance analysis and simulation result of a novel switching paradigm for optical WDM networks called Just-In-Time Optical Burst Switching (JIT-OBS) designed for ultra-low-latency transport of data-bursts across an optical WDM network. It combines the desirable features of circuit-switching and packet-switching, and features an out-of-band signaling scheme on a separate control channel with explicit feedback on delivery of data-bursts. We provide a performance analysis and simulation of the JIT-OBS approach, and compare its performance with those of circuit-switching and packet-switching approaches. We find that it has the best latency performance among the different switching mechanisms, and it has a better throughput performance than circuit-switching, and its performance is insensitive to network propagation delays.

Keywords: WDM, Optical Network, Optical Burst Switching Just-In-Time Signaling, Signaling Protocol, Packet Switching, Circuit Switching

1.   INTRODUCTION

The emergence of broadband communications has increased the needs of bulk transport of high capacity signals and services. Multi-wavelength reconfigurable optical networks offer such a capability beyond current transport technologies such as SONET. The Multi-wavelength Optical Networking Program (MONET) [1] [2] [3] [4] sponsored by the U.S. Government's Defense Advanced Research Project Agency (DARPA) is a research consortium aimed at addressing the technology, architecture, and the management and control issues for this new emerging technology.

In this paper we describe the Just-in-Time Optical Burst Switching (JIT-OBS) paradigm developed in MONET. The JIT-OBS paradigm is designed for ultra-low-latency unidirectional transport of data-bursts across an optical network. It combines the desirable features of circuit-switching and packet-switching, and features an out-of-band signaling scheme on a separate control channel and provides explicit feedback on delivery of data-bursts.

This paper is organized as follows. In Section 2, we outline a generic WDM switch architecture used for subsequent discussion. We then describe the different WDM switching paradigms, including the JIT switching paradigm we developed. Section 3 then presents an analysis of the performance of the JIT-OBS approach. Section 4 presents and discusses our initial simulation results. Section 5 concludes the paper, and comments on the MONET JIT signaling protocol implementation and the WDM switching experiments to be performed in the MONET Washington DC testbed.

2. WDM SWITCHING PARADIGMS

In optical WDM, the tremendous bandwidth of a fiber (potentially a few tens of terabits per second) are demultiplexed into many independent non-overlapping wavelength channels. Within certain restrictions, the wavelength channels are *transparent* in that they can transport data at different bit rates and modulation formats.

In this section, we outline a functional architecture for a generic WDM switch used for our subsequent discussion. We then describe the different switching paradigms for optical WDM in more detail. We note that each switching paradigm makes different assumptions on the WDM switch hardware, and requires different signaling schemes. The manner in which the header/control info is exchanged and the manner in which the path-setup and data-transfer are performed distinguishes the different schemes.

2.1 WDM SWITCH FUNCTIONAL ARCHITECTURE

A functional architecture of a generic WDM switch is illustrated in Fig. 1(a). Wavelength channels in an input fiber that enter the WDM switch are demultiplexed into individual wavelengths. They are then switched by the crossconnect to a specific output port. All wavelength channels destined to an output port are then multiplexed into the output fiber. The wavelength channels may be amplified and/or gain stabilized before they exit the switch. WDM switches [5] [6] differ in the extent to which they keep the optical signals transparent within the switch. In an all-optical WDM switch, the wavelength channels remain entirely in the optical domain. Such WDM switches may also perform wavelength conversion within the switch. Other WDM switches have optical-to-electronic (O/E), and electronic-to-optical (E/O) conversions performed on each wave-
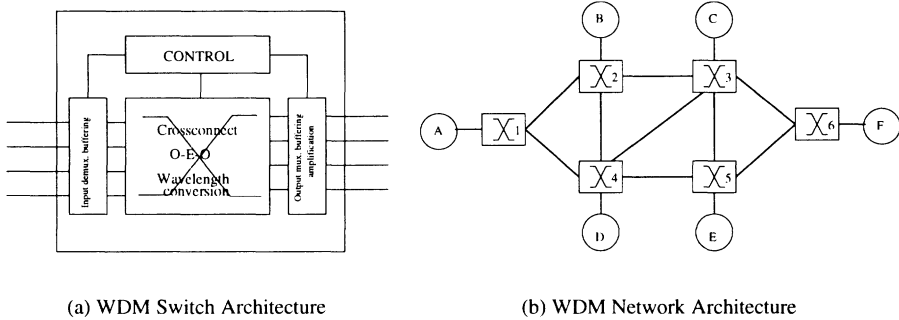
(a) WDM Switch Architecture   (b) WDM Network Architecture

Figure 1   Multiwavelength Optical Networking Architecture

length channel within the WDM switch. Input and/or output buffering may also be provided in an optical switch.

The control component in a WDM switch controls the state of the cross-connect, the wavelength converters, and output buffers. based on the control information present at the controller. An important parameter of a WDM switch is the *switching-time,* defined as the time it takes for the cross-connect to change state, and for the output channel to stabilize.

Fig. 1b depicts an optical network consists of WDM switches (labeled 1 through 6) interconnected by bidirectional fiber links. Technological constraints dictate that the number of WDM channels that can be supported in a fiber be limited to W (whose value is typical a few tens today.) Access stations are attached to a WDM switch via bidirectional fiber links. An access station is capable of sourcing data on any one of the available wavelengths. Data is transferred between access stations as unidirectional variable length optical bursts.

There are different ways to transmit the control information across the WDM network nodes. Control information can be transmitted along with the optical data as an in-band optical header (e.g. by utilizing a scheme such as subcarrier multiplexing.) In this mode, control information travels along with the data-burst, and is analogous to the packet switching. An implication is that the WDM switch must delay/buffer the data-burst until the header is decoded, and the data-path is established. As an alternative is to transmit control information independent of the data-burst on a separate signaling network. In this case, the control information (i.e. signaling) is utilized to setup the optical path prior to transmitting the data-burst, and is analogous to circuit-switching. Intuitively, a packet-switched scheme is suited for short data-bursts, and a circuit-switched scheme is efficient for large data-bursts.
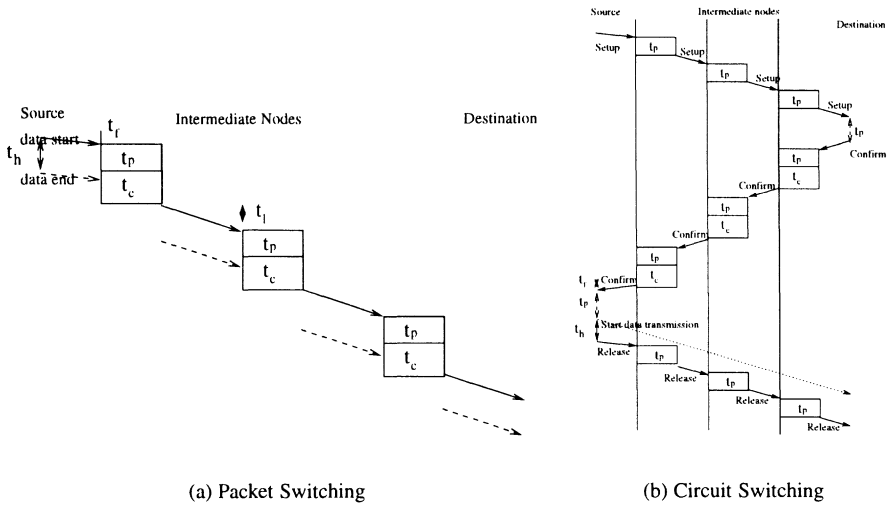
(a) Packet Switching   (b) Circuit Switching

Figure 2   Conventional Switching Paradigms

Current optical technology provides very high bandwidth for transmission, but is limited in its ability to perform optical processing or buffering. Electronics on the other hand allows processing and buffering, but cannot match the bandwidth of optical transmission. In order to maximize the utilization of the optical network, we would like the control mechanism to minimize the time to setup the optical path for a data-burst, under the constraints imposed by the optical WDM switch technology. Just-In-Time switching is designed to combine the desirable features of packet-switching (small setup time) and circuit-switching (out-of-band signaling.)

2.2   PACKET-SWITCHING

In optical packet-switching as illustrated in Fig. 2(a), the control information associated with a data-burst travels with the data-burst as the packet header. At each intermediate node, the header is separated from the data-burst, and is processed to determine the output-port. A routing protocol may be used to determine the next-hop given the destination. The WDM switch controller sets up the crossconnect (along with wavelength conversion.) During the period of header processing, and cross-connect setup, the data-burst is buffered. If an output port is not available, the data-burst is dropped, and lost. No feedback is sent to the source access station.

2.3   CIRCUIT-SWITCHING

In optical circuit-switching, the data transfer path is setup prior to the transmission of the data burst. The access station that has a data-burst to transmit initiates an out-of-band distributed signaling procedure to determine the path, wavelengths, and setup cross-connects. When a data transfer path is available, the data-burst is transmitted by the access station.

An example of the circuit-switched signaling procedure is depicted in Fig. 2(b). A *SETUP* message is sent from source to destination. On its way, wavelengths are reserved at each link along the path. If at some intermediate node no wavelengths are available on the output port, then a *BLOCKED* message is sent back to the source. If the setup message reaches the destination access station successfully, then the destination responds with a *CONFIRM* message back to the source along the reverse path, and at each intermediate node, cross-connects are setup. When the confirm message reaches the source, the source transmits the data-burst. After the data-burst is transmitted, the source sends a *RELEASE* message which releases wavelengths along the path. We assume that the routing is performed by a routing control protocol, and is independent of the signaling protocol. In this work, we assume that the shortest-hop path is utilized for routing.

The setup time for a data-burst in circuit switching can be improved by pipelining the cross-connect setup times with the propagation time. Two variations of pipelining are possible:

- *Cut-at-Confirm:* where the crossconnect is installed (cut-through) after the CONFIRM message is sent (see Fig. 3(a)) on the reverse path.

- *Cut-at-Setup:* where the cut-through is performed right after the SETUP message is sent (see Fig. 3(b)) on the forward path.

2.4   JUST-IN-TIME OPTICAL BURST SWITCHING

Just-in-time optical burst switching combines desirable features of circuit-switching (explicit feedback, separate signaling network) while minimizing the setup time. The detailed design of the JIT signaling protocol is reported in [8]. Fig. 4 depicts a simplified abstract form of signaling for JIT-OBS that captures the salient features for our present purpose. An access station initiates JIT signaling by sending a *JIT_SETUP* message to its attached WDM switch. The WDM switch responds with a *JIT_CALL_PROCEEDING* to indicate that connection setup is on its way to the destination. In the reply message there is a *delay* parameter, which indicates how long the access station should wait before launching its data-burst. This delay parameter is estimated by the WDM switch (e.g. by a suitably developed routing algorithm) from the number of hops to the destination and associated setup time of the crossconnects along the
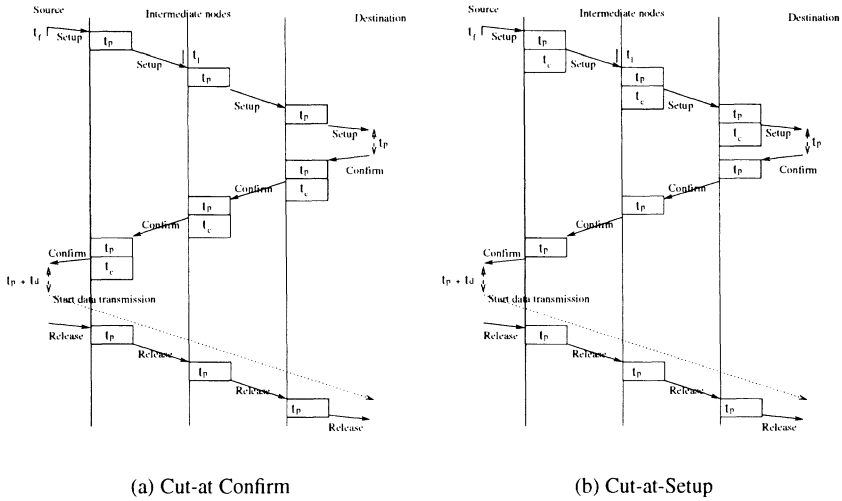
(a) Cut-at Confirm   (b) Cut-at-Setup

Figure 3   Pipelined Circuit Switching Paradigms

path. When the source access station receives the *JIT_CALL_PROCEEDING* message, it waits for the delay parameter, and then transmits its data-burst. When a WDM switch receives a *JIT_SETUP* message, it will attempt to reserve the wavelength on the output port and forward the *JIT_SETUP* message to its next hop. Cross-connect setup is performed in parallel with the next hop propagation. If the switch has no output wavelengths available on the next-hop output port, it sends back a *JIT_BLOCKED* message to the source access station. When a WDM switch gets the JIT_BLOCKED message it releases the reserved wavelength on the output port. When the destination access station receives the *JIT_SETUP* message, it responds with a JIT_CONNECT message which travels back to the source access station. After transmitting the data-burst, the source access station transmits a *JIT_RELEASE* to release all reserved wavelengths.

We note that the setup time for the circuit switched approach is significant, and therefore it is efficient only for data-bursts which are much longer than the setup time. The packet switched scheme on the other hand has shorter setup times since the control information travels with the data burst. However, in an all-optical WDM switch the header processing requirements are significant, and the data-burst has to be buffered during the time that the header is being processed, For example, a WDM packet switch must have the technology to separate the header from the optical data-burst at the input port and to reimpose it at the output port. Such technology is not required for the circuit-switched schemes. Furthermore, JIT-OBS and circuit switching schemes provide ac-
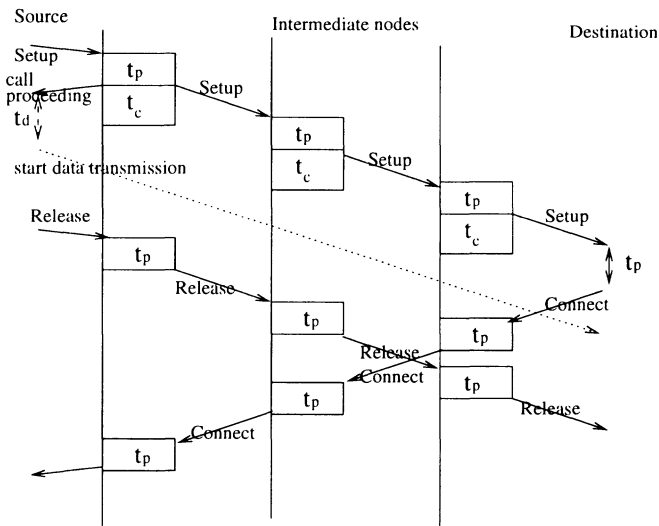
Figure 4 Just-In-Time Optical Burst Switching

knowledgement of delivery of the data-burst, while packet-switched schemes do not provide any such acknowledgements.

3. PERFORMANCE ANALYSIS

In this section we present a simple analytical model that examines the performance of each of the above schemes. In analyzing the performance of a WDM optical network, we are interested in the throughput, the latency, and their sensitivity to different network parameters.

3.1 ASSUMPTIONS AND NOTATION

We assume that the propagation delay on each link is identical (i.e., the fiber links are of the same lengths.) We assume that each WDM switch is capable of full wavelength conversion. We assume that the traffic distribution among node pairs is uniform. For each signaling scheme, we obtain the formulas for the channel holding time (i.e., the time duration from the instant a channel is reserved to the instant the channel is released) for a successful data-burst.

We now outline the notation utilized in the rest of this paper.

M: Number of access stations.

W: Number of wavelengths on the fiber that are used to transfer data-bursts.

t_f: Propagation delay from an access station to its attached WDM switch.

t_p: Protocol processing time at a WDM switch.

t_c: Crossconnect switching and stabilization time at a WDM switch.

t_l: Propagation delay on a fiber link between WDM switches.

t_h: Average burst duration.

γ: An upper bound on the throughput achievable on a single wavelength channel

C, S_1, S_2: Let there be C fiber links on a minimum cut of of the graph that represents the WDM network topology. Let the cut divide the network access stations into two sets, of sizes S_1 and S_2.

λ: Offered load to the network (in data-bursts/unit time)

λ_m: Maximum rate of traffic that can be sourced by any pair of access stations.

3.2 ANALYTICAL RESULTS

Channel Holding Time. We define the *channel reserve time* to be the instant at which an outgoing channel is reserved at a WDM switch for a call. *Channel release time* is the instant at which an outgoing channel that was reserved for a data-burst is released. *Channel hold-time* is the duration for which an outgoing channel is reserved by a WDM switch for a data-burst. Let n be the number of intermediate WDM switches along a path. From Figures 2, 3, and 4, we make the following observations.

For the circuit-switched scheme, a setup request arrives at the kth WDM switch on its path at time $t_f + kt_p + (k-1)t_l$. This is when an outgoing channel is reserved at the kth node. A release message arrives at the kth node at time $5t_f + (2n + 2 + k)t_p + nt_c + t_h + (2n - 3 + k)t_l$. Therefore the duration for which the channel is reserved for a call at the kth node is: $4t_f + (2n + 2)t_p + nt_c + 2(n - 1)t_l + t_h$

Similarly, for the Cut-at-Confirm circuit-switching scheme, we observe the following: (a) channel reserve time = $t_f + kt_p + (k-1)t_l$, (b) channel release time = $5t_f + (2n + 2 + k)t_p + t_c + t_h + (2n - 3 + k)t_l + t_d$, and therefore, (c) channel hold time = $4t_f + (2n + 2)t_p + t_c + 2(n - 1)t_l + t_h + t_d$, where the data delay, t_d, is determined as: $t_d \geq t_c - t_p - 2t_f, t_d \geq 0$.

Similarly, for the Cut-at-Setup circuit-switching scheme, we observe the following: (a) channel reserve time = $t_f + kt_p + (k-1)t_l$, (b) channel release time = $5t_f + (2n + 2 + k)t_p + t_h + (2n - 3 + k)t_l + t_d$, and therefore, (c) channel hold time = $4t_f + (2n + 2)t_p + 2(n - 1)t_l + t_h + t_d$, where the data delay, t_d, is determined as: $t_d \geq (t_c + nt_p) - (4t_f + 2(n - 1)t_l + (2n + 2)t_p), t_d \geq 0$.

For the packet switched scheme the channel holding time is $t_c + t_h$. For the JIT-OBS scheme, the data delay t_d is determined from the fact that, for any switch on path, its crossconnect must be setup before the data-burst arrives at the switch, i.e., for any $k \leq n$, where n is the number of WDM switches on the path,

$$t_d \geq (k - 2)t_p + t_c - 2t_f \qquad (1)$$

At k^{th} WDM switch on a n node path, (a) channel reserve time $= t_f + (k)t_p + (k-1)t_l$, (b) channel release time $= t_f + (n+k)t_p + (k-1)t_l + t_h + t_c$, and (c) channel hold duration $= nt_p + t_h + t_c$

Latency. The latency for a data-burst is defined as the duration from the instant the data-burst arrives at the source access-station, to the instant it arrives at the destination access station. For a route with n intermediate WDM switches (n+1 hops), the latency for the different schemes are the following:

- *Circuit Switching:* $6t_f + 3(n-1)t_l + (2n+2)t_p + nt_c$

- *Cut-at-Confirm Circuit Switching:* $6t_f + 3(n-1)t_l + (2n+2)t_p + t_c + t_d$

- *Cut-at-Setup Circuit Switching:* $6t_f + 3(n-1)t_l + (2n+2)t_p + t_d$

- *Packet Switching:* $2t_f + n(t_p + t_c) + (n-1)t_l$

- *JIT Optical Burst Switching:* $2t_f + nt_p + t_c + (n-1)t_l$

These equations are derived by summing up the different delay components in the end-to-end switching scenarios depicted in Figures 2, 3, and 4. We observe that JIT-OBS has the lowest latency.

Analysis of a single WDM switch. The switching time for a WDM switch imposes an upper bound on the achievable throughput on any channel.

$$\gamma = t_h/(t_h + t_c) \tag{2}$$

This arises from the simple fact that on any channel, data-bursts must be spaced apart by t_c.

Analysis of a WDM path. Let b_p be the probability that a data-burst arriving at a WDM switch is blocked due to output contention. Let there be k burst types arriving to the output link at a WDM switch with arrival rates, $\lambda_1, \lambda_2, \ldots, \lambda_k$ respectively, and with channel holding times $t_1, t_2, \ldots, t_k$, (which implies that the loads for each burst type is $\rho_1, \rho_2, \ldots, \rho_k$ where $\rho_i = \lambda_i/(1/t_i)$, respectively.) Then, the probability that all W wavelength channels are busy on the link is given by the M/M/m/K Markov chain with k customer arrival classes is determined from [7] as:

$$b_p = \sum_{n_1, n_2, n_k \leq W, \sum_{i=1}^{k} n_i = W} \prod_{i=1}^{k} (\rho_i)^{n_i}/G \tag{3}$$

where G is

$$G = \sum_{n_1,n_2,n_k \le W, \sum_{i=1}^{k} n_i = W} \sum_{i=1}^{k} (\rho_i)^{n_i} \qquad (4)$$

Let there be n WDM switches between two access stations. Then the probability that a burst from an access station destined to another access station is blocked, B_p, is given as: (assuming that each WDM NE blocks the burst independently.)

$$B_p = 1 - (1 - b_p)^n \qquad (5)$$

Analysis of a WDM Network. The maximum load that the network can carry is determined by the bandwith of the minimum cut of the network as $C\gamma W$. Since the minimum cut separates the network into two sets of node with S_1, and S_2 nodes respectively, we obtain:

$$S_1 \times S_2 \times \lambda_m \le C \times \gamma \times W \qquad (6)$$

From the blocking probability of a path given in Eqn. 5, we can determine, B, the probability that a data-burst arriving to thé network is blocked, B, by averaging the blocking probabilities of each node pair (assuming shortest path routing) weighted with the load between the node pair.

The normalized throughput ρ, when the offered load is λ is defined as follows:

$$\rho = \frac{\lambda}{(\lambda_m \times M \times (M-1))}(1-B) \qquad (7)$$

4. ILLUSTRATIVE SIMULATION RESULTS

Simulation Parameters. We simulate two representative networks, one with a 7 node bidirectional ring topology and another one with a 5×5 two-dimensional torus topology. We assume that the number of wavelengths was 8. For the simulations, we assumed that each node-pair was equally loaded, and the applied load was a fraction of the maximum load determined from Eqn. 6. We assumed Poisson arrivals and fixed data-burst sizes. We simulated two values of burst-sizes, one small value of 0.01ms (corresponding to 12500 bytes at 10Gbps), and a large value of 1000s. We simulated two values for propagation time on a link, one value of 0.27ms (corresponding to a link length of about 50 miles) and another value of 1ms. We assumed that the propagation delay to the first WDM switch is $0.0025ms$ corresponding to a fiber distance of about 1 mile. We assumed that the switching time at a WDM switch is $0.1ms$, and the protocol processing time at a WDM switch is $0.1ms$.
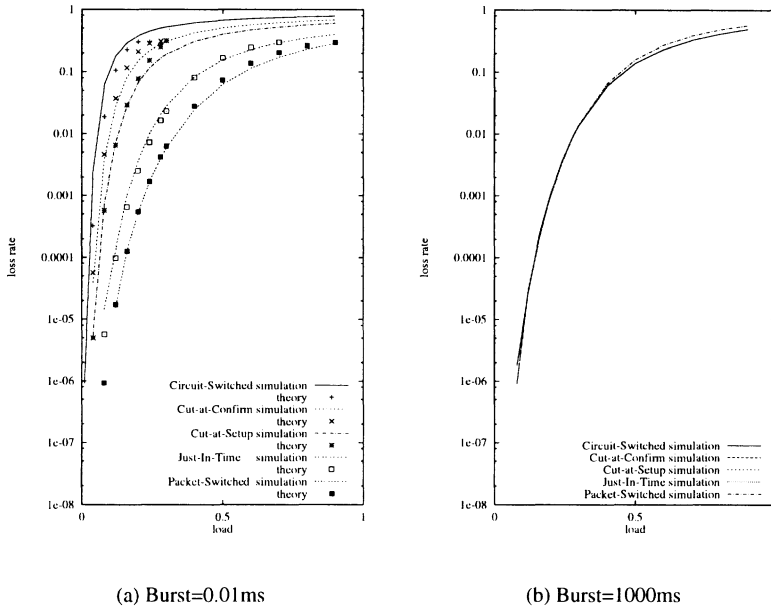
(a) Burst=0.01ms   (b) Burst=1000ms

Figure 5   Packet Loss versus Applied load

Simulation Results and Discussion. We observed similar performance be-haviors for both topologies. Below we illustrate the results using those of the torus network. Fig. 5(a) illustrates the packet loss rate versus load when the data-burst is of size 0.01ms. First, we observe that the analytical values are in good agreement with the simulation values. Then, we observe that for a given load, packet-switching has the least blocking probability, followed by JIT-OBS, pipelined circuit-switching and circuit-switching. The intuitive reason is the channel-hold duration for the schemes increases in the same order. Fig. 5(b) illustrates the packet loss rate versus load when the data-burst is of size 1000ms. We observe that in this case, all the schemes have similar performance. This is because, since the data-burst transmission is large compared to the setup time, the channel hold duration for each scheme is approximately the same.

Figure 6 illustrates the normalized throughput against the applied load. We observe that for a given load, packet-switching achieves the highest throughput, followed by JIT-OBS and circuit-switching.

The average latency is the duration from the instant a data-burst arrives at an access station, to the instant that the data-burst arrives at the destination access station. Fig. 7 illustrates the network-wide average latency against the applied load. We find that JIT has the least latency followed by PKT switching, and pipelined circuit switching and circuit switching. JIT pipelines the crossconnect
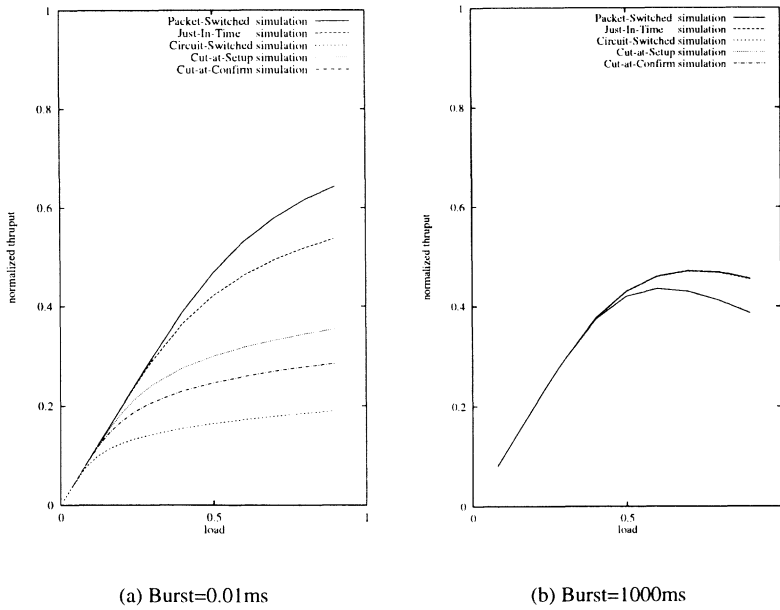
(a) Burst=0.01ms   (b) Burst=1000ms

Figure 6   Normalized Throughput versus Applied Load

setup time with data transmission and as a result is able to achieve the least latency.

Figure 8 illustrates the normalized throughput against load for two different values of the link propagation time: a value of 0.27 ms and a value of 1 ms. When the propagation delay increases, the throughput performance of circuit-switched schemes is adversely affected, while the performance of JIT and packet-switched schemes remain unaffected.

5.   CONCLUSION

JIT-OBS combines the desirable features of circuit and packet switched schemes. We found that when the burst-size is small compared to the other parameters, throughput performance of JIT-OBS was better than circuit-switching and the latency performance of JIT-OBS was better than packet switching on representative network topologies. Furthermore, the performance of JIT-OBS is insensitive to the link propagation delay.

Under the MONET project, a signaling protocol for JIT-OBS has been designed and implemented [8]. This signaling implementation is currently being tested and validated against a laboratory network of Lucent optical WDM network elements. Additional testing and experimentation are scheduled in
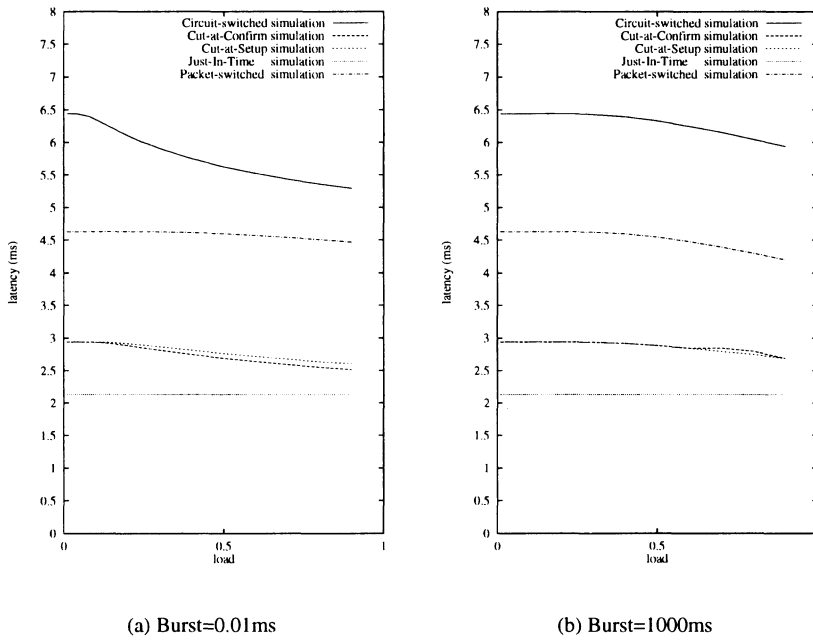
(a) Burst=0.01ms   (b) Burst=1000ms

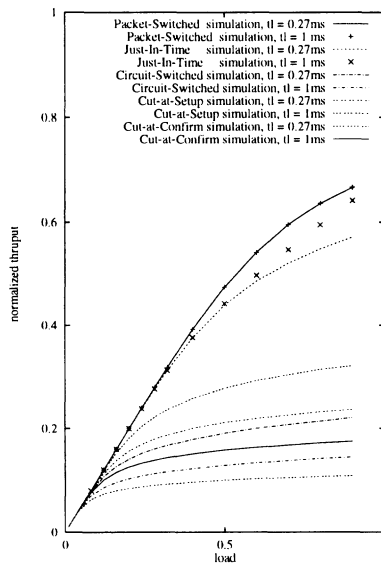*Figure 7*   Latency versus Applied Load

Figure 8   Normalized Throughput versus Applied Load

the upcoming MONET Washington DC testbed network demonstrations [9] with participation from several government agencies. These experiments will provide the opportunity to stress test our implementation and to validate our design.

As it was noted in [10], "...in the future, bandwidth will not be our problem. Latency will be the major challenge to overcome." Towards meeting this challenge, Just-In-Time Optical Burst Switching provides a good mechanism for ultra-low latency transport of variable sized bursts of data across an optical WDM network.

Acknowledgements

References

[1] R.E. Wagner, R.C. Alferness, A.A.M. Saleh, and M.S. Goodman, "MONET: Multi-wavelength Optical Networking," *Journal of Lightwave Technology,* vol. 14, no. 6, pp. 1349–1355, June 1996.

[2] J.Y. Wei, C.-C. Shen, B.J. Wilson, and M.J. Post, "Network Control and Management of a Reconfigurable WDM Network," *MILCOM'96,* vol. 2, pp. 581–586, McLean, Virginia, October 1996.

[3] The MONET Research Team, "The MONET New Jersey Network Demonstration," *IEEE Journal on Selected Areas in Communications,* vol. 16, no. 7, pp. 1199-1219, September 1998.

[4] J.Y. wei, C.-C. Shen, B.J. Wilson, M.J. Post, and Y. Tsai, "Connection Management for Multiwavelength Optical Networking," *IEEE Journal on Selected Areas in Communications,* vol. 16, no. 7, pp. 1097–1108, September 1998.

[5] F. Masetti et. al., "High Speed, High Capacity ATM Optical Switches for Future Telecommunication Transport Networks," *IEEE Journal on Selected Areas in Communications,* vol. 14, no. 5, pp. 979-998, June 1996.

[6] D. Chiaroni et.al., "All-Optical Fast Packet-Switched Networks: Physical and Logical Limits of Operation," *SPIE Conference on All-Optical Networking: Architecture, Control and Management Issues,* vol. 3531, pp. 378–389, Boston, Massachusetts, November 1998.

[7] D. Bertsekas and R. Gallager, *Data Networks,* Prentice-Hall: Englewood Cliffs, New Jersey, 1992.

[8] J.Y. Wei, J.L. Pastor, R. McFarland, and J. Parker, "Design of The MONET Just-In-Time Signaling Protocol." In preparation.

[9] A.V. Lehmen, "Large-Scale Demonstration of Multiwavelength Optical Networking in MONET Washington DC Network," *OFC'99,* San Diego, California, February 1999.

[10] D. Farber, *Keynote Speech at the IEEE LEOS Summer Topical Workshop on Broadband Networking,* Monterey, California, July 1998.