

Performance Evaluation of TCP over Optical Burst-Switched (OBS) WDM Networks

Sunil Gowda¹, Ramakrishna K Shenai², Krishna M Sivalingam², & Hakki Candan Cankaya³

¹*Expedia.com, Seattle, WA 98006*

²*University of Maryland Baltimore County (UMBC), Dept. of CSEE, Baltimore, MD 21250*

³*Alcatel USA, Plano, TX 75075*

Abstract—This paper studies the performance of TCP transport protocol over an optical burst-switched (OBS) wavelength division multiplexed (WDM) wide-area network. Typically, an OBS network consists of optical core routers and electronic edge routers connected by WDM links. At the network ingress nodes, IP packets are assembled into bursts that are routed through the core network and disassembled at the network egress nodes. This paper studies the effects of OBS network characteristics and parameters on TCP's delay and throughput performance: (i) Burstification (burst-assembly and disassembly) delays, (ii) Data-burst scheduling and (iii) Variation of burst packet parameters (i.e. burst size, burst timeouts), and (iv) burst drop probability. Detailed results based on an *ns2*-based simulator, that has been extended to incorporate WDM and OBS networking, are presented.

I. INTRODUCTION AND MOTIVATION

Optical wavelength division multiplexed networks, that partition the multi-terabit bandwidth per fiber, into individual multi-gigabit channels have the potential to meet the demands of future generation networks [1], [2]. Current optical technology demonstrations have shown a feasibility of 160 channels, each operating at 10 gigabits per second (Gb/s); and future networks are expected to operate at 40 Gb/s per channel or higher.

In this paper, we focus on the performance of TCP/IP over an optical WDM network using the burst-switching paradigm. Getting IP onto a wavelength (i.e. IP to wavelength mapping) requires an intermediate step of encapsulation which is discussed in [3]–[5]. However carrying IP-over-WDM, without any intermediate encapsulation, is considered attractive due to elimination of redundant functionality introduced by layers such as SONET/SDH and ATM. Since processing of IP packets in the optical domain is still not a reality and electronic switches that operate at such high speeds are still not available, optical burst switching (OBS) was proposed to harness the potential of WDM, without the limitations of circuit-switched networks. OBS is employed at the optical WDM network backbone, where several IP packets having common properties (such as common destination paths, quality-of-service parameters) are assembled into data bursts at the network ingress that

are routed through the optical WDM backbone mesh network and disassembled back into IP packets at the network egress.

The optical channels are reserved only for the data-burst, i.e. the reservation is neither on a per-packet basis (as in packet-switching) nor on a per-session basis (as in circuit-switching). The expectation is that the burst duration will be significantly shorter than the session duration and much longer than a single packet's duration. Thus, channel utilization can potentially be improved as compared to circuit-switching, while avoiding the per-packet processing overhead encountered in packet-switching.

There have been a number of studies on different OBS mechanisms, such as burstification, scheduling, and switch architectures. However, the impact of the burstification process on individual TCP sessions has not been studied earlier. This paper attempts to present such a study by considering the impact of data-burst drops, burstification delays, and burst-switching parameters, such as data-burst size and burst-timeouts. An *ns2* based optical WDM simulator [6] has been used to model the OBS network and gather detailed TCP-level delay and throughput results.

II. GENERAL ARCHITECTURE OF AN OPTICAL BURST-SWITCHED NETWORK

This section provides a brief introduction to optical burst-switching (OBS). For a detailed introduction to burst-switching, related architectures and protocols, please refer to [7]–[10]. Our analysis of TCP incorporates the control architecture proposed in [11].

A. Network Model

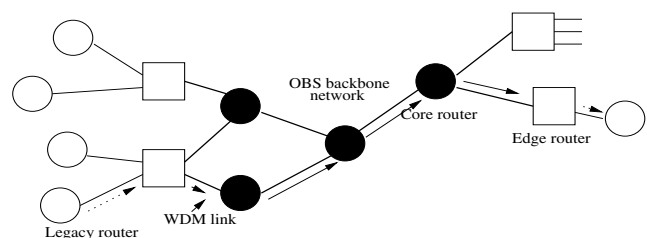


Fig. 1. A sample optical burst-switched network

Corresponding author: K. Sivalingam (krishna@csee.umbc.edu; tel: +1 410-455-3961; fax: +1 410-455-3969). This work was supported in part by a grant from Alcatel USA.

Figure 1 illustrates a sample OBS backbone mesh network. As shown in the figure, the burst-switched network consists of:

Edge routers: These represent the electronic transit point between the burst-switched backbone and the legacy network. IP packets incident on a common ingress edge node that have a common egress edge node are aggregated together to form a *data-burst*, the basic data unit in a burst-switched network. This process is termed *burst-assembly* or *burstification*. *Burst-disassembly* or *de-burstification* is the reverse process at the egress node where the data-burst is disassembled into individual IP packets that routed individually to the destination. The data-bursts are built by buffering the packets at the ingress edge nodes until burst generation is initiated based on a temporal burstification algorithm presented in [12]. Burst generation is triggered either when the burst size reaches a pre-configured limit or upon the firing of a timer set on the arrival of the first packet in the burst.

Every data-burst, has an associated control packet called the burst header packet (BHP). The BHP is responsible for reserving the wavelengths for the data-burst. The wavelengths are reserved by scheduling the links and configuring the switches on the intermediate nodes of the path.

Core routers: These are connected to either edge or core routers, and form the heart of the OBS network. Core routers are responsible for forwarding data-bursts within the optical core. The opto-electronic core router has an electronic control plane that processes BHPs, which maintain enough control information needed to configure the data plane to switch the data-bursts optically at each hop. This independent switching between the control packets and their respective data-bursts reduces the router overhead, in addition to providing a transparent optical path to data-bursts.

The wavelengths on a link are partitioned into two sets – data channels and control channels for carrying data packets and BHPs respectively. Our study uses the *latest available unused channel - with void filling* (LAUC-vf) data-channel scheduling algorithm [11] to determine the data channel for a burst request. To improve the schedule success rate, we consider the use of intermediate optical buffers realized by fiber-delay-lines (FDLs).

III. TCP OVER AN OPTICAL BURST-SWITCHED WDM LAYER

Section I provides substantial reason for considering TCP over an optical burst-switched WDM layer. In this section we identify certain key parameters inherent in the optical burst-switched WDM layer that will affect TCP performance. As discussed in [13], TCP reacts to one or more of the following changes in the network (i) Packet drops (ii) End-to-end delay variations and (iii) Throughput changes. Previous performance studies of optical burst switching have primarily focused on the switching layer itself, and assume the functionality provided

by the transport layer especially in terms of handling re-transmissions. Some of the OBS parameters that may affect the TCP performance include maximum burst-size, burst timeout, delays introduced by input and output FDLs, and the number and granularity of the FDLs.

Impact of burst-switching

There are many ways in which transport layer may be affected by the underlying OBS layer.

Effect of burst drops: The probability of a burst drop depends upon the network load and the level of burst contentions in the network. In an IP network, packet loss probability of each packet is independent of other packets and is largely due to overflow of buffers at the routers. As mentioned earlier, burst drops also cause packet losses. In most cases, several packets from different sessions are included in a burst and a burst drop could result in loss of many packets per session, resulting in a network wide drop in throughput.

Effect of burst-assembly (burstification) delay: Another factor affecting performance is the delay that a packet experiences during the burstification process before its associated burst is transmitted. For high arrival rates of packets at the interfaces, burstification is triggered only when the burst reaches the maximum burst size. If λ is the arrival rate and the maximum burst size is B , the average time for the burst to be assembled is B/λ . For low packet arrival rates, burst generation is triggered due to the occurrence of the burst timeout.

Effect of fiber-delay-lines (FDLs): Other factors such as the number of FDLs at each node, the granularity of these FDLs also have an impact on the delay and drop probability of the TCP packets.

IV. PERFORMANCE ANALYSIS

This section presents an evaluation of the performance of TCP over an OBS network, obtained via simulation. Our simulations were conducted on *ns-2* by incorporating modules required to implement optical burst-switching [6].

The simulations were conducted on three specific scenarios (i) A three node topology with a single TCP source, (ii) A three node topology with multiple TCP sources, and (iii) An eight node topology with multiple TCP sources. The results of (ii) and (iii) are reported here. We set the TCP segment size to 4000 bytes, TCP acknowledgment (ACK) packet size to 40 bytes, and the TCP maximum window size to 300,000 segments. Since a maximum window size of 64 KBytes for basic TCP is not sufficient for high bandwidth-delay networks, the TCP agent was modified to allow arbitrarily large window sizes (This is similar to the TCP window scaling option presented in RFC 1323 [14]). Each TCP session generated traffic at rates up to 1 Gbps, with exponentially distributed inter-arrival times.

The propagation delay of each link was set to 2ms. Although this would imply short links, this enables us to study the impact of the OBS related delays. The input delay experienced

by each data-burst is set to $1 \mu s$. Each node had a set of 5 FDLs on each channel and each set is able to introduce delays of $10 \mu s, 20 \mu s, \dots, 50 \mu s$.

The performance metrics measured are (i) TCP session throughput and (ii) Per-packet delay. The simulations were conducted for varying maximum data-burst sizes (or data-burst length in bytes), burstification timeouts (or burst-assembly time), and data-burst dropping rates. Random data-burst drops are introduced at the edge nodes with a uniform probability. They are introduced only at the edge nodes to avoid cumulative effect on the data-burst drops.

A. 3-node topology with multiple TCP agents

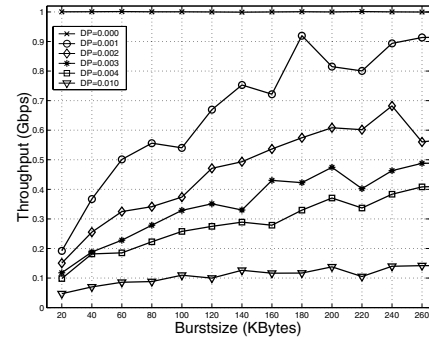
Initial simulations were conducted on a 3-node topology (that consisted of two edge nodes connected through a core node) with three TCP sources generating traffic at 1000 Mbps, 400 Mbps and 200 Mbps. The three sources fed traffic to a common edge node. We chose this to study the effect of session multiplexing (at the burstification module in the edge router) on the performance of the TCP sessions. The results are presented for each individual traffic rate and the obtained measures are the mean of the two sessions of equal traffic rate, one at each of the edge nodes.

1) *Effect of variation in data-burst length:* Figure 2 presents the TCP throughput performance for the three-node topology. The varied parameters were maximum burst sizes and drop probabilities. The burst time-out was set to 1 ms, which is the time required for a steady 1 Gbps traffic source (this excludes the TCP acknowledgments (ACKs)) to fill a data-burst of 125 KBytes.

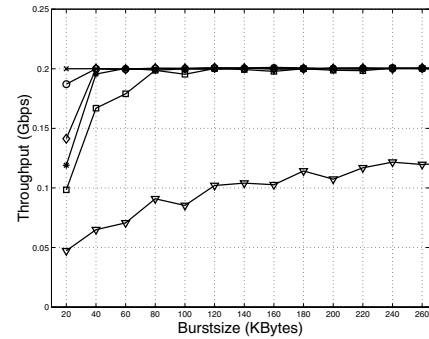
As seen in the Figure 2(a), for a data-burst drop probability of 0, the throughput for each session is a constant that equals the session's traffic arrival rate. Since the maximum data-burst size increases the burstification delay, the per packet end-to-end delay is also shown to increase correspondingly (Figure 3). The TCP session throughput is not affected, as all packets are successfully transmitted although with longer delays.

Lost packets trigger TCP's congestion control that reduces the sender window size; and it is a fact that the average (sender) window size of a session determines the achieved throughput. Hence, higher burst drop probabilities will result in lower throughput, as seen in Figure 2. For example, for the 1000 Mbps session at 260 Kbytes burstsize, the throughput drops by 10% and 50% for drop probabilities of 0.001 and 0.003 respectively, compared to the system with no burst drops.

For low burst drop probabilities, increasing the burst size significantly improves the throughput. With larger burst sizes, fewer bursts are generated for the same input traffic rate and thus fewer bursts are dropped during the simulation duration. Hence, fewer reductions in window size occur that leads to a higher average window size and hence higher throughput. At a session level, the total number of packets lost over a period of time remains unchanged as larger bursts would carry more



(a) Throughput for 1000 Mbps sessions



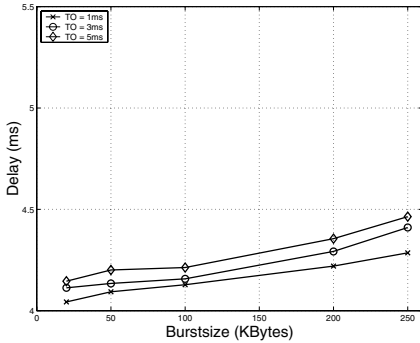
(b) Throughput for 200 Mbps sessions

Fig. 2. TCP throughput performance for the 3-node topology with multiple TCP agents, with session traffic rates of 1000 Mbps and 200 Mbps varying the maximum data-burst size and drop probability, for burst-assembly time of 1 ms

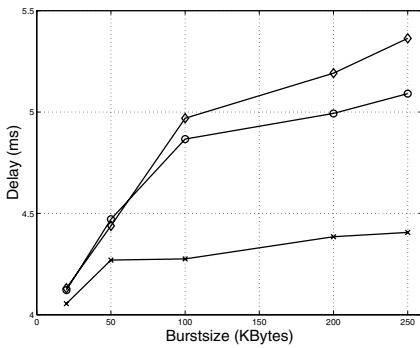
packets and a drop would result in more packets being lost. Hence a better throughput is achieved for larger burst sizes for low burst drop rates. For example, for the 400 Mbps sessions (not shown here), increasing the burstsize from 20 Kbytes to 120 Kbytes increases the throughput by about 167%. For high burst drop rates, the TCP sessions operate continuously at smaller window sizes. Hence, throughput does not increase significantly with increasing burst sizes.

We also observe that lower traffic rate sessions (e.g. 200 Mbps) achieve the maximum possible throughput at much smaller burst sizes. This can be attributed to the fact that for the lower traffic rate sessions, a smaller window is enough to achieve the maximum throughput. As the window size drops down to 1 packet after a packet loss in the TCP implementation used, the 200 Mbps TCP sources get back to the required window size much faster.

The packet delay values behaved as expected with higher delays for large burst sizes and higher drop probabilities. Thus, the key to configuration of the burst size is to strike a balance between achieving higher throughput and keeping



(a) Delay for 1000 Mbps sessions



(b) Delay for 200 Mbps sessions

Fig. 3. Average packet delay, for the 3-node topology, varying maximum data-burst size and timeout values, for data-burst drop probability of 0.002.

the delay within acceptable limits. For higher burst drop probabilities, the throughput drops for the same burst sizes. Hence, to identify a suitable operating burstsize, a metric such as (*Throughput/Delay*) may be considered.

2) *Effect of variation in burst-assembly times (Burstification timeouts)*: Figures 3 present the performance of the network obtained by varying data-burst timeout values and burst sizes, for a drop probability of 0.002. The TCP sessions generated traffic at 1000 Mbps, 400 Mbps or 200 Mbps.

Burstification timeouts are used to limit the burstification delay experienced by the TCP segments in the edge router. When the arrival rate of TCP segments at the burstification module is low, the bursts are generated based on a timeout resulting in bursts smaller than the maximum burst size. We only present the delay performance here since the throughput values did not vary much with burst timeout values. As expected, the delays experienced by all sessions for a 1 ms burst time-out reaches a maximum at 200 KBytes and the increase is relatively marginal thereafter. The maximum increase in delay observed by increasing the timeout from 1 ms to 5 ms, was only 10%.

B. 8-node topology

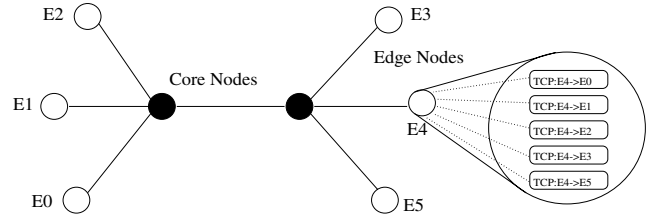


Fig. 4. 8-node network topology.

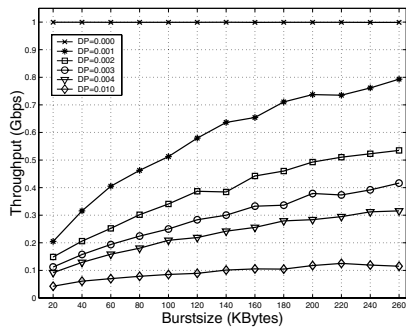
The results presented in this section were obtained for the 8-node the topology, shown in Figure 4, consisting of two core nodes, C_1 and C_2 , and six edge nodes, denoted by E_0 to E_5 . TCP sessions were established between each pair of edge nodes. Since bursts are assembled by aggregating packets with the same ingress-egress edge node pair, each session undergoes burstification independently and hence there is no session multiplexing at the burstification module. As explained earlier, we have to establish two sessions between each node pair in opposite directions, resulting in a total of 30 sessions between the six edge nodes.

1) *Effect of variation in data-burst length*: Figure 5(a) presents the throughput performance for the eight-node topology with multiple TCP agents, with session traffic rates of 1000 Mbps. The varied parameters were maximum data-burst length and data-burst drop rates. The data-burst time-out was set at 1 ms. As earlier, we observe that larger burst sizes yield higher throughput.

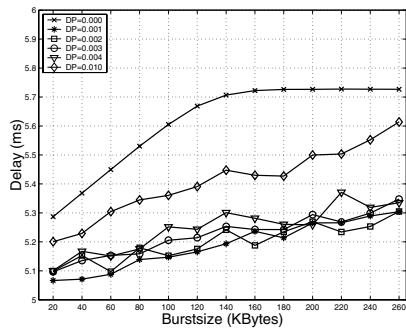
2) *Effect of variation in burst-assembly times (Burstification timeouts)*: Figure 6(a) presents the effect of burstification timeout values on the performance of the TCP sessions. Experiments were conducted for timeouts varying from 1ms to 8ms with drop probabilities in $\{0.001, 0.002, 0.05, 0.1\}$. The maximum burst size was set to 200 KBytes. For low drop probabilities, the throughput decreases as the burstification timeout increases. For instance, the throughput drops by 27% when timeout value is increased from 1 ms to 8 ms, for drop probability of 0.001. For higher drop probabilities, the drop in throughput is not as significant. The delay values increase with increasing timeout values. Again, the increase is seen to be significant for lower drop probability values. For instance, the delay increases by as much as 53% when timeout value is increased from 1 ms to 8 ms, for drop probability of 0.001.

V. CONCLUSIONS

This paper studied TCP-level performance for an optical burst switched wavelength division multiplexed (WDM) network. The architecture and control mechanisms of the burst-switched network were presented. This paper discussed the impact of burst-switching on the performance of the TCP. The simulation based experiments were used to identify the

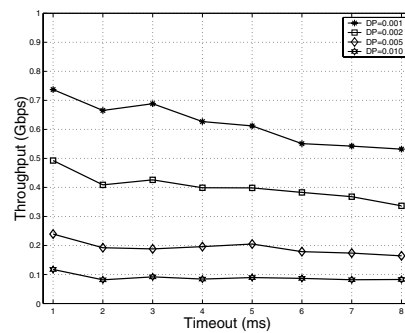


(a) Throughput

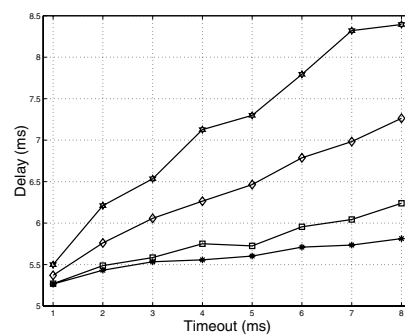


(b) Delay

Fig. 5. Throughput and delay performance for the 8-node topology varying burstsize and drop probability, and burst timeout value set to 1ms.



(a) Throughput



(b) Delay

Fig. 6. Throughput and delay performance for the 8-node topology, with session traffic rates of 1000 Mbps, varying burst timeout and drop probability, and burst size set to 200 KBytes.

impact of data-burst lengths, burst-assembly times, and data-burst drop rates. Higher drop probabilities resulted in poorer performance and the performance degradation was severe for drop probability as low as 0.003 for the studied cases. For low drop probabilities, increasing burst sizes resulted in higher throughput and increased delay. For higher drop probabilities, there was no significant gain with increasing burst sizes. Increasing timeout values did not result in significant performance degradation for the three-node topology with multiple TCP agents, while some degradation was observed for the eight-node topology.

REFERENCES

- [1] P. E. Green, "Optical networking update," *IEEE Journal on Selected Areas in Communications*, vol. 14, pp. 764–779, June 1996.
- [2] K. Sivalingam and S. Subramaniam, eds., *Optical WDM Networks: Principles and Practice*. Boston, MA: Kluwer Academic Publishers, 2000.
- [3] P. Bonenfant, A. Rodriguez-Moral, and J. Manchester, "IP over WDM: The Missing Link," *White Paper, Lucent Technologies*, Dec. 1999.
- [4] B. Doshi, S. Dravida, E. Hernandez, W. Matragi, M. Quershi, P. Langer, J. Anderson, and J. Manchester, "A Simple Data Link (SDL) Protocol for Next Generation Packet Network," *IEEE Journal on Selected Areas in Communication*, vol. 18, pp. 1825–1838, Oct. 2000.

- [5] J. Anderson, J. S. Manchester, A. Rodriguez-Moral, and M. Veeraraghavan, "Protocols and architectures for IP Optical Networks," *Bell Labs Technical Journal*, vol. 4, January-March 1999.
- [6] B. Wen, N. M. Bhide, R. K. Shenai, and K. M. Sivalingam, "Optical Wavelength Division Multiplexing (WDM) Network Simulator (OWNs): Architecture and performance Studies," *SPIE Optical Networks Magazine*, pp. 16–26, Sep/Oct 2001.
- [7] M. Yoo, M. Jeong, and C. Qiao, "High-speed protocol for bursty traffic in optical networks," in *SPIE Proceedings, All Optical Communication systems: Architecture, Control and Network Issues*, vol. 3230, pp. 79–90, Nov. 1997.
- [8] C. Qiao and M. Yoo, "Optical burst switching (OBS) - a new paradigm for an optical internet," *Journal on High Speed Networks*, vol. 8, no. 1, pp. 69–84, 1999.
- [9] C. Qiao and M. Yoo, "Choices, features and issues in optical burst switching," *SPIE Optical Networks Magazine*, Apr 2000.
- [10] J. S. Turner, "Terabit burst switching," *Journal on High Speed Networks*, vol. 8, no. 1, pp. 3–16, 1999.
- [11] Y. Xiong, M. Vandenhoude, and H. Canakya, "Control Architecture in Optical Burst-Switched WDM Networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1838–1851, Oct. 2000.
- [12] Y. Xiong, M. Vandenhoude, and H. Cankaya, "Design and Analysis of Optical Burst-Switched Networks," in *Proc. SPIE '99 Conf. All Optical Networking*, vol. 3843, (Boston, MA), pp. 12–19, Sept. 1999.
- [13] V. Jacobson, "Congestion avoidance and control," in *Proc. ACM SIGCOMM*, (Stanford, CA), Sept. 1988.
- [14] V. Jacobson, R. Braden, and D. Borman, "TCP Extensions for High Performance," *Internet RFC 1323*, May 1992.