Survey of QoS issues for TCP connections in Optical Burst Switched Networks

Terrance Frederick Fernandez
Department of Computer Science and Engineering
Pondicherry Engineering College
Puducherry – 605 014, INDIA

S.Ramachandiran
Department of Computer Science and Engineering
Pondicherry Engineering College
Puducherry – 605 014, INDIA

ABSTRACT
Optical Burst Switching (OBS) can cater the requirements of bandwidth intensive applications like Voice-over-IP, video conferencing, interactive video on demand. It is switching technique for the next generation optical networks. This paper deals with merits and the protocols used for the Optical Burst Switching. A brief survey of QoS open research issues for TCP/OBS connections is also discussed.

General Terms
Optical Burst Switching (OBS)

Keywords
Multiplexing, AON, Light path, FTO, Contention resolution

1. INTRODUCTION
Internet Protocol (IP) dominates network technology because of the growth of internet applications like Voice over Internet Protocol (VoIP), Video Conferencing; Interactive Online Entertainment etc…The inception of Wavelength Division Multiplexing (WDM) allows multiplexing hundreds of channels per fiber. To efficiently transport IP over WDM three architectures are proposed namely,

- Optical Circuit Switching (OCS)
- Optical Packet Switching (OPS)
- Optical Burst Switching (OBS)

OCS is a mature technology. They use Light Path to setup and configure switches. An all - optical path is called as Light Path. The destination acknowledges the source after receiving the data using ACK packet. The bandwidth utilization is very low.

OPS split data into smaller entities called packets. The header is attached to each packet. At each node header is processed electronically and the payload is delayed by Fiber Delay Lines (FDLs). The bandwidth utilization of OPS is better than OCS.

OBS carries both advantages of OCS and OPS. An OBS network consists of electronic edge nodes and optical core nodes. The input edge node is called as "Ingress node" and the output edge node is called as "Egress node". Data is broken into "variable sized packets" called Data Bursts (DB) thus overcoming the demerit of the OPS architecture. A burst is defined as the Digitized talk spurt or the data message [2] that has intermediate granularity between a packet and circuit. Control packet (CP) is sent on a separate wavelength with an "Offset Time (OT)". This OT allows the processing of the control packet until the arrival of data [1].

1.1 Merits of Optical Switching
- The switching speeds of electronics cannot keep up with the transmission capacity offered by optics [1].
- Optical switches handles large number of switching ports when compared with electronic switches.

1.2 OBS merits over other switching
- Buffers are absent at core nodes and so headers can be processed at slower speeds. Thus, synchronization requirements are relaxed in OBS [2].
- Optical Circuit Switching architecture is suitable for constant rate traffic (voice traffic) but unsuitable for dynamic network traffic [1], but OBS does good for both.
- Faster header processing and strict synchronization are required in OPS due to lack of optical buffers and also the FDLs, which are used at the core nodes of OPS are costly.
- Only company that offers commercial OBS products is *"Matisse networks"* as the technology is still immature [1] and thus offers an open research area.
- OBS can cut-through optical switches i.e.., the burst may be transmitted before it is completely assembled or dissembled at intermediate nodes and thus transmission speed is high compared to OPS and OCS.

1.3 Motivation
- OBS suffers from Bandwidth Delay Product (BDP) and so suffers from speed mismatch with TCP. Even if the TCP Scaling option is used to reach cwnd to 4MB from 64KB longer time is consumed.
- The delayedACK must be used in TCP/OBS as in reality all TCP segments cannot be included in a single burst which causes further delay. .
- High Speed TCP (HSTCP) was proposed for high BDP networks that offers bad throughput for Burst losses.

2. SURVEY ON OBS TECHNIQUES
2.1 Assembly Mechanisms
The two categories of burst assembly algorithms are (a) timer and (b) threshold burst size. Timer based algorithms signal the burst formation after some time interval. Since

the arrival of packets is inherently burst, the lengths of bursts in the OBS network are variable. Threshold based algorithms signal the formation of a burst once the size of the packets accumulated exceeds a minimum threshold size. If all the ingress nodes use a certain threshold level, the burst sizes in the OBS network will be more or less the same [3].

2.2 Channel Reservation in OBS

Different schemes in Optical Burst Switching are Tell-And-Go (TAG), Just-In-Time (JIT) and Just-Enough-Time (JET). TAG is an immediate reservation scheme. In TAG, the CB is transmitted on a control channel followed by a DB, which is transmitted on a data channel with zero offset. The CB reserves the wavelength and FDL at each intermediate node along the path for the DB. When the DB reaches a core node, it is buffered using the reserved FDL until the CB processing is finished. Then the DB is transmitted along the reserved channel. If no wavelength is available for reservation, the burst is dropped and a negative acknowledgement (NAK) is sent to the source. The source node sends another CB after transmitting the DB for releasing the reserved wavelengths along the path. Here, the burst size is not fixed in advance. FDLs are expensive and they can only buffer data optically for a very short time. Optical buffering is the main drawback of this scheme. Furthermore, if the "release" CB which is sent to release the reserved bandwidth along the path is lost, then these wavelengths will not be released and this creates bandwidth wastage [4].

JIT scheme also comes under immediate reservation. Here, an output wavelength is reserved for the DB when the CB processing is finished. The source transmits the DB after an offset time which is greater than the total CB processing time. If the wavelength is not available, then the burst is dropped. The difference between JIT and TAG is that the buffering of the DB at each node is eliminated by inserting a time gap between the CB and the DB. Since the bandwidth is reserved immediately after the CB processing, the wavelength will be idle from the time the reservation is made till the first bit of the DB arrives at the node. This is because of the offset between the CB and the DB. Since the offset value decreases as the CB gets close to the destination, the idle time decrease. An in-band-terminator is placed at the end of each burst which is used by each node to release the reserved wavelength after transmitting the DB [4].

JET is a delayed reservation scheme. Here, the size of the burst is decided before the CB is transmitted by the source. The offset between CB and DB is also calculated based on the hop count between the source and destination. At core nodes, if bandwidth is available, the CB reserves wavelength for the burst for a fixed duration of time. The reservation is made from the time when the first bit of DB reaches the node till the last bit of DB is transmitted to the output port. This eliminates the wavelength idle time which is the main difference between JET and JIT. Since the wavelength is reserved for a fixed duration, there is no need for explicit signal for releasing the reserved wavelength along the path. Since there is no wastage of bandwidth in this scheme, the network utilization for this scheme is higher than with the other schemes. But, this scheme involves complex scheduling when compared to other schemes [4].

2.3 Burst Scheduling

On arrival of a control packet at a core node, a wavelength channel scheduling algorithm is used to determine a wavelength channel on an outgoing link for the corresponding data burst. The information required by the scheduler and its duration are obtained from the control packet. The scheduler keeps track of the availability of time slots on every wavelength channel. It selects one among several idle channels. The selection of wavelength channel needs to be done to reduce the burst loss. At the same time, the scheduler must be simple and should not use any complex algorithm, because the routing nodes operate in a very high speed environment handling a large amount of burst traffic. A complex scheduling algorithm may lead to the early data burst arrival situation wherein the data burst arrives before its control packet is processed and eventually the data burst is dropped [4].

A channel is said to be unscheduled at time *t* when no data burst is using the channel at or after time *t*. Algorithms that consider unscheduled channels are called Horizon algorithm. A channel is said to be unused for the duration of voids between two successive data bursts and after the last data burst assigned to the channel. Those which consider voids within channels are called void filling algorithm. According to scheduling strategy used scheduling algorithms can be classified as Horizon or without void filling [5] and with void filling [6].

Representative of Horizon algorithms are: First Fit Unscheduled Channel (FFUC) [7,6,8], Latest Available Unused Channel (LAUC) [8, 9] and that of void filling algorithms are: First Fit Unscheduled Channel with Void Filling (FFUC-VF) [7], Latest Available Unused Channel with Void Filling (LAUC-VF) [9,10,11] and Minimum End Void (Min-EV) [11].

2.4 Contention Resolution

Contention occurs in OBS if two or more incoming bursts contend for the same output wavelength at the same link and at the same time instant [12]. This contention is to be resolved and is done by:

- Buffering
- Wavelength Conversion.
- Burst Deflection Routing (Alternate Routing).
- Burst Segmentation.
- Burst re-transmission.

Buffering in OBS is done in time domain by the use of the Fiber Delay Lines (FDLs) that limit the amount of time a burst could reside unlike electronic buffers, where a packet can stay in the buffer for an undefined time. Electronic buffers are present at the electronic edge nodes. Optical technology is immature and buffers are not invented for optical core nodes. It is impossible to delay the burst for infinite period of time using Fiber Delay Lines (FDLs). It is done in "Space domain". Wavelength Conversion is the capability of the optical network to convert an input wavelength to a desired output wavelength. It is done in spectral/Wavelength domain. Break the assembled data-burst into a number of segments and the process is called "Segmentation" [13]. "Segments" are basic transport units and are invisible in optical domain [14].

There are two segmentation policies and they are: Head dropping policy and tail dropping policy. In the former,

the head of the contending burst is dropped. In the latter, the tail of the contending burst is dropped.

Deflection Routing is done in a "Space domain". If there is a contention at the preferred link then the burst is forwarded at any available output. This is also called as "Hot Potato Routing" [3]. These deflected bursts may loop multiple times wasting network bandwidth [15].

2.5 TCP Protocol

TCP breaks the data into chunks. These are called segments and are acknowledged by the receiver. However, a TCP sender has a limit on the outstanding data and it is indicated by the value of the send window. Such limit is the minimum of two limits, one imposed by the receiver, the receive buffer, that indicates the amount of data that it is able to buffer, and another by the sender called the congestion window, which is a limit on the outstanding data in order not to overload the network. TCP does not initially transmit at full rate, but it uses the slow start mechanism, by which TCP starts to probe the network. First, the TCP has a low congestion window, typically one segment, and its size is increased by one segment every time an acknowledgement (ACK) is received until a threshold is trespassed.

The next phase is congestion avoidance, where the congestion window is increased at a slower rate, at most one segment per round trip time. TCP also provides reliability by detecting the loss of segments and retransmitting. A TCP sender detects the loss of a segment by means of the reception of three duplicate ACKs, or by the triggering of a retransmission timeout. TCP was created with the idea that the loss of a segment is a clear indication of congestion. Thus, when a segment loss is detected, the TCP congestion window is reduced in order to transmit fewer segments, thus lightening the network load and helping to stop the congestion.

There are a number of TCP versions for congestion avoidance and they are:

- Tahoe
- Reno
- New Reno
- SACK and
- Vegas.

2.6 Delayed ACK in TCP

The TCP protocol [17], states that a TCP receiver sends an acknowledgement for each incoming segment. This behaviour was later modified by RFC 1122 [18], which specify the delayed ACK algorithm. For an incoming segment, the TCP receiver does not immediately send an acknowledgement. Specifically, the ACK should be generated for at least every second full-sized segment and must be generated within 500 ms from the arrival of the first unacknowledged packet. However, this is not the exact behaviour in real world TCP. The most common implementation is that TCP has a timer that goes off every 200 ms. Thus, the timer can expire anytime within 200 ms since the arrival of an incoming packet. The ACK is therefore sent when a second full-sized segment is received or otherwise when the periodic timer goes off. This is the implementation for the Windows TCP stack. The delayed ACK algorithm was introduced to reduce the load in the network and in the hosts that send and process TCP segments. Due to the cumulative value of the acknowledgements, using delayed ACK has no impact on transmission reliability. Furthermore, it has also been found that using delayed ACK can improve performance in asymmetric networks [16]. However, it has also been demonstrated that delayed ACK reduces performance in certain situations.

As in RFC 2581 [16], TCP increases the congestion window (and so the amount of outstanding data in the network) based on the number of ACKs received. Thus, reducing the number of ACKs received leads to a slower increase of the amount of sent data. For instance, it has been found that the time spent during the slow start phase is doubled when delayed ACK is used [17]. During the congestion avoidance phase, due to the same reason, the increase of the congestion window is also slower.

2.7 TCP over OBS networks

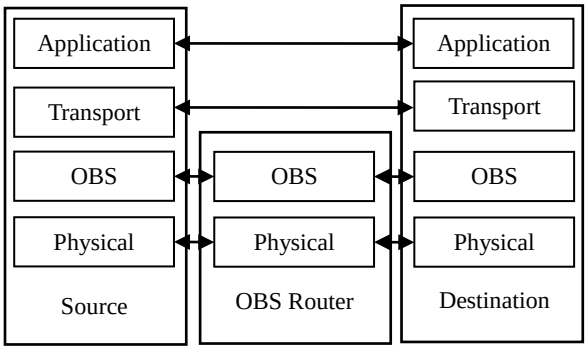

Fig 1: TCP over OBS Stack

In a TCP/IP network, IP layer is involved in routing of packets, congestion control and addressing the nodes. When OBS is introduced in the network, it takes care of routing of data and congestion control. The routing information computed by IP layer need not be considered by OBS routers. It is because, the routes are computed based on number of hops and wavelength availability by the OBS routers. However, the addressing of the various nodes in the network is not taken care by OBS by default. Hence the functionality of IP may be limited to addressing and packet formation. Due to the above reason, we consider the functionality of OBS layer decimates the functionality of IP layer as shown in Fig 1.

For TCP over OBS networks, some OBS features do have an impact on TCP performance. These are mainly due to:-

- The burst assembly/disassembly process in the ingress/egress router [3] and
- The contention at the intermediate nodes when bursts compete for the same channel leads to burst loss due to the lack of practical optical buffers [18].

3. OPEN RESEARCH ISSUES IN OBS

During Burst Contention, if the tail dropping segmentation policy is applied to resolve it then the network suffer from "Shadow Contention" [19], i.e.., the header contains the total burst length even if the tail is dropped. This causes the downstream nodes to be unaware of size of the Burst due to truncation and this would make the sender resend the data wasting network resources.

In head dropping segmentation scheme, out-of-order delivery will be more in contrast to the tail dropping

policy where the sequence is maintained. Long bursts that pass through different switches experience contention at many switches. So, it is essential to create Burst with fewer packets. Creating shorter bursts also results in "fragmentation" and will be deflected as new bursts. These fragments carry low priority [20]. So, the segmentation policy is not truly advised with very long or tiny bursts.

In Slotted OBS, to implement high-precision clocks at the routers is a major challenge. Here, the edge routers must have the full knowledge regarding the timeslot in which a router needs to send data [21].This proves lethal especially in multicast environment where data flow is high.

In a TCP over OBS network, out of order burst delivery may occur if TCP segments of a certain TCP session if they are assembled in two or more bursts and sent in a different order. This issue is still to be dealt with proper assembly mechanism.

In a TCP over OBS network, "Slow Convergence" is much more pronounced due to longer RTT of each burst caused by burstification process and BCL delay [22]. This is an important issue to be sorted out to improve assembly delay.

In Time Sliced OBS, bursts are prioritized with higher Offset times. This offers a good architecture. But the edge routers must have the knowledge of all core routers which is practically impossible.

In TCP over OBS, during contention if decoupling approach is used for higher loads then packet dropping will be more [23] thus leading to higher Burst losses. This area is still left undone for dynamic loads.

Time Out event due to a single Burst loss is called as "False Time Out (FTO)". It causes the network to start with "Slow start" and thus a lead to performance degradation. It happens especially for Fast TCP flows [24].

To minimize burst loss we need to balance traffic among multiple paths [25]. Even if the traffic is balanced the network still suffers from linear effects due to channel shifting and the data at the receiver is noisy.

In a TCP over OBS network, lack of global scheduling and One-way signaling leads to random contention losses even if the network is lightly loaded. The network treats this contention loss as permanent network congestion and further reduced the TCP window size thereby going to slow start phase thereby affecting the throughput.

Wavelength Conversion is immature and produces linear effects like noise. To avoid this we may use Optical Tunable Wavelength Converter (OTWC) but is expensive [26]. So, using wavelength converters for contention resolution may not always yield a contention-free OBS network and considerable amount of research needed to be done in this designing a Wavelength Converter.

In SOBS, the major cost will be for the FDLs and thus shorter FDLs will result in lower cost of the FDLs. Using Optical buffer (FDLs) for contention resolution results with noise produced by amplifiers in it and also regenerators used are expensive [27].Also, bursts of bigger lengths cannot be stored at the "Fiber Delay Lines (FDLs)".

BTCP can effectively resolve contention in a TCP over OBS environment by distinguishing FTO from TTO and react with re-transmission of bursts [28, 29]. But it uses higher layers like TCP which is not advisable. Thus, an independent contention resolution mechanism is still unavailable for OBS.

4. CONCLUSION

Optical Burst Switching offers promising core architecture for the future of Optical Internet because of its transmission speed. But it suffers from contention issue that is absent in other networks. Although these issues are resolved by various resolution strategies, there are more issues associated with using these as discussed under open research issues.

In the future, when OBS is implemented on Core networks further more new issues may come fertile. Our future work would be to test/simulate an OBS core under TCP with proper load balancing strategies and using no or minimum contention resolution strategies thus avoiding most of the discussed QoS issues.

5. REFERENCES

[1] C. Siva Ram Murthy and Mohan Guruswamy, "WDM Optical Networks – Concepts, Design, and Algorithms", *Prentice-Hall, Inc*, 2002.

[2] Farid Farahmand, Jason Jue, Vinod Vokkarane, "A Layered Architecture for Supporting Optical Burst Switching", *Proceedings of the Advanced Industrial Conference on Telecommunications, pp. 213 – 218, 17 October 2005, Dallas ,Texas, USA*.

[3] Óscar González, Ignacio de Miguel, Noemí Merayo, Patricia Fernández, Rubén M. Lorenzo, Evaristo J. Abril, "The impact of delayed ACK in TCP flows in OBS networks", *Innovation Strategy, Telefonica*.

[4] Tetsuya Miki, "Optical Transport Networks*", Proceedings of the IEEE, Vol. 81, No. 11, pp. 1594 – 1609, November 1993.*

[5] David K. Hunter and Ivan Andonovic, "Approaches to Optical Internet Packet Switching", *IEEE Communications Magazine, Vol.38, No.9, pp. 116-122, September 2000.*

[6] Won-Seok Park, Minsu Shin, Hyang-Won Lee and Song Chong, "A Joint Design of Congestion Control and Burst Contention Resolution for Optical Burst Switching Networks" , *Journal of Lightwave technology, Vol. 27, No. 17, pp. 2209-2214, September 1, 2009.*

[7] T.Venkatesh, A.Jayaraj, and C. Siva Ram Murthy, "Analysis of Burst Segmentation in Optical Burst Switching Networks Considering Path Correlation", *Journal of Lightwave technology, Vol. 27, No. 24, pp. 5563 – 5570, December 15, 2009.*

[8] Yong Liu, Gurusamy Mohan, Senior Member, IEEE, Kee Chaing Chua, and Jia Lu, " Multipath Traffic Engineering in WDM Optical Burst Switching Networks" , *IEEE Transactions on Communications, Vol. 57, No. 4, pp. 1099 – 1108, April 2009.*

[9] Onur Ozturk, Ezhan Karasan and Nail Akar, "Performance Evaluation of Slotted Optical Burst Switching Systems with Quality of Service Differentiation", *Journal of lightwave technology,*

vol. 27, no. 14, pp 2621 - 2633, July 15, 2009.

[10] Basem Shihada and Pin-Han Ho, "Transport Control Protocol in optical Burst Switched Networks: Issues, solutions, and Challenges", *IEEE Communications Surveys & Tutorials, pp- 70-86, 2008.*

[11] Jiangtao Luo , Jun Huang, Hao Chang, Shaofeng Qiu, Xiaojin , " ROBS: A novel architecture of Reliable Optical Burst Switching with congestion control", *Journal of High Speed Networks, vol. 5626, pp.440–447, 2007.*

[12] Vinod M. Vokkarane, Jason P. Jue, and Sriranjani Sitaraman, "Burst Segmentation: An Approach For Reducing Packet Loss In Optical Burst Switched Networks", *2002.*

[13] S.Y. Wang, "Using TCP Congestion Control to Improve the Performances of Optical Burst Switched Networks", *International Conference on Communications, pp. 1438 - 1442 2003, Hsinchu, Taiwan.*

[14] Xiang Yu. Chunming Qiao and Yong Liu, "TCP Implementations and False Time out Detection in OBS Networks", *IEEE INFOCOM, 2004, Hong Kong.*

[15] Pushpendra Kumar Chandra, Ashok Kumar Turuk, Bibhudatta Sahoo , "Survey on Optical Burst Switching in WDM Networks", *International conference on Industrial and Information Systems, pp. 83-88, 2009, Sri Lanka.*

[16] Georgios I. Papadimitriou, Chrisoula Papazoglou and Andreas S. Pomportsis, "Optical Switching: Switch Fabrics, Techniques and Architectures", *Journal of lightwave technology, Vol. 21, No. 2, pp. 384-405, February 2003.*

[17] J. Postel, "Transmission Control Protocol", *RFC 793, IETF, Sep. 1981.*

[18] R. Braden, "Requirements for Internet Hosts – Communication Layers", RFC *1122, IETF, Oct. 1989.*

[19] H. Balakrishnan, V.N. Padmanabhan, R.H. Katz, "The Effects of Asymmetry on TCP Performance, "*Proceedings on ACM/IEEE Mobicom, Sept 1997, Budapest, Hungary.*

[20] M. Allman, V. Paxson and W. Stevens, "TCP Congestion Control", RFC *2581, IETF, April 1999.*

[21] M. Allman, "On the Generation and Use of TCP acknowledgments", *ACM SIGCOMM Computer Communication Review, pp. 4-21, Oct. 1998.*

[22] K. Koduru, "New Contention Resolution Techniques for Optical Burst Switching," *Master's thesis, Louisiana State University, May 2005.*

[23] K. Dozer, C. Gauger, J .Spath, and S. Bodamer, "Evaluation of Reservation Mechanisms for Optical Burst Switching," *AEU International Journal of Electonics and Communications, vol. 55, no.1, pp. 2017- 2022, January 2001.*

[24] M. Ljolje, Robert Inkret, and Branko Mikac, " A Comparative Analysis of Data Scheduling Algorithms in Optical Burst Switching Networks," *In Proceeding of Optical Network Design and Modeling*, pp 493-500, 2005.

[25] M. Yoo, C. Qiao, and S. Dixit, "QoS Performance of Optical Burst Switching in IP-Over-WDM Networks" ,*IEEE Journal on Selected Areas in Communications, vol. 18, no.10, pp. 2062-2071, October 2000.*

[26] W.M. Golab and R. Boutaba. "Resource Allocation in User-Controlled Circuit-Switched Optical Networks," *LNCS Spinger-Verlag, vol. 16, no.12, pp. 2081-2094, December 1998.*

[27] Y. Xiong, Marc Vandenhoute and Hakki C. Cankaya, "Control Architecture in Optical Burst Switched WDM Networks," *IEEE JSAC, vol. 18, no.1, pp. 1838-1851, October 2000.*

[28] M. Yang, S. Q. Zheng, and D. Verchere, "A QoS Supporting Scheduling Algorithm for Optical Burst Switching DWDM Networks," *In Proceeding of GLOBECOM , pp. 86-91, 2001, Richardson, Texas, USA.*

[29] J. Xu, Chunming Qiao, Jikai Li and Guang Xu, "Efficient Channel Scheduling Algorithms in Optical Burst Switching Networks," *In Proceeding of IEEE INFOCOM, vol. 3, pp. 2268-2278, 2003, NY,USA.*