A Layered Architecture for Supporting Optical Burst Switching

Farid Farahmand[1], Jason Jue[2]
[1]*Dep of Elect. Engineering*
[2]*Dep. of Computer Science*
University of Texas at Dallas
P.O. Box 830668, EC 31
Richardson, Tx 75083-0688
{ffarid, jjue}@utdallas.edu

Vinod Vokkarane
CIS, University of Massachusetts at Dartmouth
285 Old Westport Road
North Dartmouth, Massachusetts 02747-2300
vvokkarane@umassd.edu

Joel J. P. C. Rodrigues, Mário M. Freire
Department of Informatics, University of Beira Interior
R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal
{mario, joel}@di.ubi.pt

Abstract

This paper defines a new layered architecture for supporting optical burst switching in an optical core network. The architecture takes into account both the control plane as well as the data plane. The paper describes the functionality and the primary protocols that are required at each layer. This paper also explains how the layers interact with each other in the proposed architecture.

1. Introduction

The amount of raw bandwidth available on fiber optic links has increased dramatically with advances in dense wavelength division multiplexing (DWDM) technology; however, existing optical network architectures are unable to fully utilize this bandwidth to support highly dynamic and bursty traffic.

An optical transport network consists of a collection of edge and core nodes as shown in Figure 1. The traffic from multiple client networks is accumulated at the ingress edge nodes and transmitted through high capacity DWDM links over the core. The egress edge nodes, upon receiving the data, provide the data to the corresponding client networks. The three prominent optical transport networks architectures proposed to carry traffic over the optical core are optical circuit switching (OCS) (or wavelength-routed networks), optical packet switching (OPS), and optical burst switching (OBS). These switching techniques primarily differ based on how resources are allocated in the core and the degree of granularity for the resource allocations.

In OCS networks, an all-optical connection, referred to as a lightpath [1], is established to create a logical circuit between two edge nodes across the optical core. These lightpaths may be established dynamically as connection requests arrive to the network, or they may be provisioned statically based on estimated traffic demands. While OCS is suitable for constant rate traffic such as voice traffic, it may be unsuitable for highly dynamic traffic. Furthermore, as lightpaths must be established using a two-way reservation scheme that incurs a round-trip delay, the high overhead of connection establishment may not be well-suited for short bursts of traffic. Also, under bursty traffic, sufficient bandwidth must be provisioned to support the peak traffic load, leading to inefficient network utilization at low or idle loads.

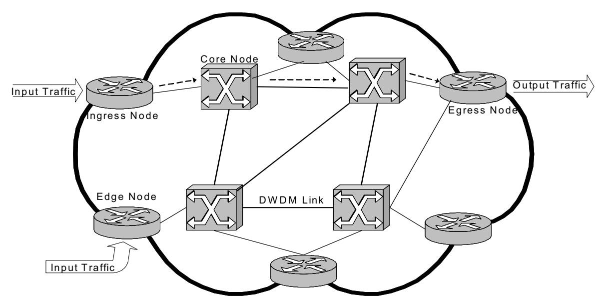

Figure 1. Optical transport network.

In OPS networks [2], data is transmitted in the form of optical packets which are transported across the optical core without conversion to electronics at intermediate core nodes. OPS can provide dynamic bandwidth allocation on a packet-by-packet basis. This dynamic allocation leads to a high degree of statistical multiplexing which enables the network to achieve a higher degree of utilization when the traffic is variable and bursty. However, there are many technical challenges to implementing a practical OPS system. One of the limitations of OPS networks is that it is difficult to implement optical buffers. Furthermore, the requirement for fast header processing, and strict

Proceedings of the Advanced Industrial Conference on Telecommunications/Service Assurance with Partial and Intermittent Resources Conference/E-Learning on Telecommunications Workshop
0-7695-2388-9/05 $20.00 © 2005 **IEEE**

synchronization makes OPS impractical using current technology.

OBS [3, 4] was proposed as a new paradigm to achieve a practical balance between coarse-grained circuit switching and fine-grained packet switching. In OBS networks, incoming data is assembled into basic units, referred to as data bursts (DB), which are then transported over the optical core network. Control signaling is performed out-of-band by control packets (CP) which carry information such as the length, the destination address, and the QoS requirement of the optical burst. The control packet is separated from the burst by an offset time, which allows for the control packet to be processed at each intermediate node before the data burst arrives. OBS provides dynamic bandwidth allocation and statistical multiplexing of data, while having fewer technological restrictions than OPS. By aggregating packets into large sized bursts and providing out-of-band signaling, OBS eliminates the complex implementation issues of OPS. For example, no buffers are necessary at core nodes, headers can be processed at slower speeds, and synchronization requirements are relaxed in OBS. On the other hand, OBS incurs higher end-to-end delay and higher packet loss per contention compared to OPS, due to packet aggregation. Basic architectures for core and edge nodes in an OBS network have been studied in [5].

Each of the three types of optical transport network architectures (OCS, OPS, OBS) may support different services. Packet traffic can be supported by any of the three architectures in either a connectionless or connection-oriented manner. OPS and OBS support these different types of packet services through different signaling protocol implementations. In order to support connection-oriented services on OBS, a two-way reservation protocol such as, TAW can reserve the end-to-end path for the requested duration, prior to data transmission. Connectionless services on OBS can be supported by various one-way reservation protocols, such as, JET, JIT, and TAG [3, 4, 6]. Similarly, OPS may support connectionless services by routing packets on an individual basis, and may support connection-oriented services by assigning packets to flows and switching the flows based on labels applied to the packets. OBS differs from OPS primarily in that the signaling is done out-of-band in OBS networks, while signaling is done in-band via packet headers in OPS networks. OCS supports packet traffic by establishing a logical topology consisting of lightpaths, and then switching or routing packets electronically over this logical topology. Signaling for establishing lightpaths in OCS networks is typically done out-of-band. In this article, we will focus on the connectionless mode of operation of OBS; however, the framework for the control plane will be general enough to support any out-of-band signaling scheme, including those for establishing OCS lightpaths.

OBS network architecture can be represented in a layered manner as a set of protocols that provide services and exchange data with one other. A well-defined architecture with well-defined interfaces between the layers is essential for the practical implementation of OBS, as well as for the inter-operability of OBS with other networks. Furthermore, the layered hierarchy representation can provide a detailed insight into various implementation techniques, specifications, and functionalities of an OBS network. This article attempts to provide a layered view of different OBS protocols as well as some brief insight into the functionality of each protocol.

The reminder of this paper is organized as follows. Section 2 discusses the layered architecture of IP-over-OBS. Section 3 describes each layer of the OBS layered architecture, separating them into a data plane and a control plane. Finally, Section 4 concludes the article.

2. IP-over-OBS Layered Architecture

An important objective in the design of OBS networks is the large-scale support of different legacy services, as well as emerging services. In this article, without loss of generality, we will discuss the OBS network as it supports IP traffic; however, the OBS architecture described here is general enough such that it is capable of supporting most types of higher-layer traffic.

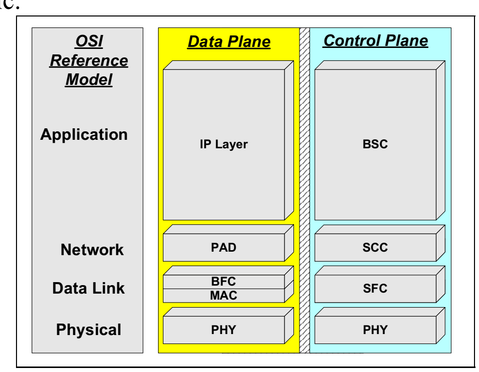

Figure 2. IP-over-OBS hierarchical layered architecture.

Figure 2 shows our proposed OBS layered architecture, which follows the OSI reference model. In this representation we separate the control plane functionalities and protocols from those of the data plane. Such separation appears natural since the control information is transmitted out-of-band in OBS networks. Note that, in this model, we are ignoring the

Proceedings of the Advanced Industrial Conference on Telecommunications/Service Assurance with Partial and Intermittent Resources Conference/E-Learning on Telecommunications Workshop
0-7695-2388-9/05 $20.00 © 2005 **IEEE**

management plane, since the management plane communicates with all other layers and has no hierarchical relationship with them.

The control plane is responsible for transmitting control packets (CPs) while the data plane constructs and processes the data bursts (DBs). The CPs contain the information necessary for switching and routing DBs across the OBS network. The CPs are used for establishing the proper path prior to the arrival of the corresponding DBs, which arrive after some offset time. The CPs can also provide network management signaling.

Having two distinct planes suggests that each plane can operate independently of the other, using its own layers and protocols. Thus, it is conceivable to imagine that the DBs and CPs are encoded and routed on different transmission media.

In general, the OBS data plane architecture must take full advantage of DWDM technology and must support high capacity data transport links with no optical-to-electrical conversions. On the other hand, the design objective in the control plane is to make it flexible with low complexity. One way to achieve this goal is by processing CPs electronically. This approach offers high flexibility but limited processing capacity (a few tens of gigabits per second). Thus, simple encoding techniques and short frame lengths with minimum control overhead are required to allow fast and efficient CP processing. Transmitting CPs free of contention and in a highly reliable manner is also critical, since any error or loss of CPs results in higher data burst loss.

3. OBS Layered Architecture

In the following sections we describe basic functionalities of each layer in the data and control planes. We start with the data plane, which interconnects the OBS network with other client networks. For clarity, we describe the layered architecture of each plane in an order consistent with packet flow.

3.1. Data plane layers

The data plane transports incoming packets from the source edge node to a single or multiple destination nodes. Line cards in the edge node provide an interface with packets arriving from various client networks. The line cards can perform error detection and error correction on incoming IP packet headers. Since in this article we only consider IP-based OBS networks, we assume that all packets entering and leaving the OBS network are IP packets, and that these packets maintain their original format and structure.

3.1.1. Packet Aggregation and De-aggregation (PAD) Layer

The PAD layer aggregates incoming IP packets of the same properties into data bursts. This layer also de-aggregates received data bursts into individual IP packets and assigns the packets to the proper outgoing link.

Transmitting IP packets at the ingress path of an OBS network requires determining individual packet properties and aggregating the packets together. Packet properties include packet Quality-of-Service (QoS) and its client destination address. After each incoming IP packet is decoded, its destination address must be translated to an OBS equivalent edge node address. Packets with similar properties are then aggregated to form the burst payload.

The most common burst assembly techniques are timer-based and threshold-based. In timer-based burst assembly approaches [7], a burst is created and sent into the optical network at periodic time intervals; hence, the network may have variable length input bursts. In threshold-based burst assembly approaches [8], a limit is placed on the maximum number of packets contained in each burst. Hence, fixed-size bursts will be generated at the network edge. A threshold-based burst assembly approach will generate bursts at non-periodic time intervals. A combination of timer and threshold-based approaches has been proposed in order to reduce the variation in the burst characteristic due to the variations of load [9]. In addition, a composite burst assembly approach [10] can be adopted in order to support QoS. A composite burst is created by combining packets of different classes into the same burst. The packets are placed from the head of the burst to the tail of the burst in order of decreasing class.

In the egress path, the PAD disassembles data bursts into IP packets. Each packet's header must be processed for its destination address and the type of service it requires. The destination address is translated to identify which line card the IP packet must be sent to. Line cards, in turn, forward packets to the appropriate interfaced client network such as a LAN or WAN.

The PAD layer contains various flow control mechanisms and offers sequence verification of incoming data bursts. The flow control protocols can pace the rate at which DBs are placed on a link. If data burst deflection routing is allowed throughout the OBS network, then DB re-sequencing at the destination node may be required to ensure ordered delivery of IP packets.

3.1.2. Burst Framing Control (BFC) Layer

The function of the burst framing control layer is to receive the aggregated packets from the higher layer (PAD) and to encapsulate them into proper frame structures. This layer also decodes incoming data burst frames and extracts the data field. Figure 3 represents a generic framing format of a data burst. Data burst frames have two characteristics: they can have variable length, and, by nature, they are non-periodic, meaning that they can arrive at any random time. Therefore, a framing pulse is necessary to indicate the beginning of each optical data burst frame. Framing pulses are typically isolated from the data-field by using a preamble to ensure data integrity.

Guard bands are normally a stream of fixed pulses used to separate consecutive frames. They are mandatory for reasons such as link length error, precision of clock distribution, and thermal effects. The checksum field may be required when data burst retransmission from the source to destination edge nodes is supported. In this case, edge nodes must be designed with considerable storage capacity. Use of the checksum may be considered especially when the medium does not offer the required transmission error rate.

The data field in the data burst frame can be further subdivided into fixed or variable sized segments. In this technique, which is referred as segmentation [10], the BFC inserts extra control information in each segment, containing multiple IP packets.

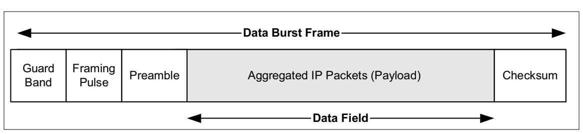

Figure 3. IP-over-OBS hierarchical layered architecture

3.1.3. Medium Access Control (MAC) Sublayer

The MAC sublayer in data plane includes the reservation and scheduling protocols, the contention resolution schemes, and the offset time assignment protocols. The MAC layer can also provide class differentiation in order to provide higher protection for DBs with QoS requirements. The actual signaling process by which a node requests the network to setup or release a connection is performed in the control plane.

An OBS network is inherently a point-to-point network in which adjacent nodes are interconnected to each other through direct physical links. However, asynchronous data bursts entering a core node from different links may need to access the same outgoing link. The MAC sublayer provides a way to control access to the outgoing links among these data bursts. In general, access control schemes proposed for OBS networks can be categorized as centralized or distributed.

In a *centralized* OBS network [11] a single node (called the request server) will be in charge of data burst transmission throughout the entire network. Clearly, this mode of operation makes medium access straightforward since the request server provides a single point of coordination that eliminates contention and packet loss. However, centralized scheme is very complex and considered to have low reliability and robustness.

In a *distributed* OBS network each node operates autonomously. This scheme suffers from lack of any centralized coordination. Consequently, the number of DBs entering a node and attempting to access the medium may exceed the number of available channels of the outgoing port. This is the primary source of contention in distributed OBS networks. Therefore, efficient and reliable algorithms in the MAC sublayer are required to simultaneously minimize contention as well as expected end-to-end delay of DBs.

Based on the type of service requested by an application, such as connection-less or connection oriented services, the OBS MAC needs to assign sufficient bandwidth and resources. Such assignments are obtained through the appropriate reservation protocols. Reservation protocols indicate the mechanisms in which a burst allocation starts and ends. Various out-of-band reservation approaches have been proposed for OBS networks. The most widely considered examples of such schemes are the Just-In-Time (JIT) reservation scheme (also know as tell-and-go), and the Just-Enough-Time (JET) reservation scheme. Although the JET reservation scheme provides a more efficient use of bandwidth, its implementation requires higher complexity. Different variations of the JIT reservation scheme have been described in [6].

Various scheduling disciplines can be implemented in the MAC depending on the reservation protocols employed in the system. An OBS scheduling discipline determines the manner in which available outgoing data channels are found for DBs. Scheduling algorithms must be fast and efficient in order to lower the processing time and to minimize the data burst loss. Current data channel scheduling algorithms include first-fit unscheduled channel (FFUC) [5], latest available unscheduled channel (LAUC) or Horizon Scheduling [4, 5], and latest available unscheduled channel with void filling (LAUC-VF) [5].

FFUC and LAUC scheduling disciplines can be considered for JIT whereas LAUC-VF is more practical for the JET reservation scheme. Note that scheduling protocols in the MAC layer should support class differentiation and provide a greater degree of protection and transmission reliability for high priority data bursts.

0-7695-2388-9/05 $20.00 © 2005 **IEEE**

A major concern in distributed medium access control scheme is high contentions. Packet transmissions in this scheme can only be statistically guaranteed. Many different techniques and algorithms have been introduced to improve OBS reliability and to reduce the DB drop ratio. In general, such solutions can be divided into four basic categories: space deflection (such as deflection routing), time deflection (such as buffering and delaying the data) [12], wavelength conversion, and soft-contention resolution policies [10, 13].

The MAC sublayer can also support establishing multipoint multicast connections. In these schemes any edge node can transmit its DBs to multiple destination edge nodes. Efficient signaling protocols can be implemented in the application layer of the control plane to support multipoint multicast.

3.1.4. Physical (PHY) Layer

The physical layer of OBS is responsible for the actual transport of DBs and CPs from one node to another. It includes converting signals into appropriate electrical or optical format and uploading DBs into appropriate transmission frames. The physical layer also defines the actual physical interfaces between nodes in OBS. The PHY layer is divided into two sublayers: data transport component and medium dependent component. We describe these sublayers briefly in the following paragraphs.

Data Transport Component (DTC): This is the medium independent sublayer of the physical layer. In the ingress direction it encodes data bits into specific pulse transmission called line codes (such as NRZ, AMI, HDB3, etc) and performs electrical/optical conversions. This sublayer also specifies transmission capacity.

Furthermore, DTC is responsible for implementing mechanisms to resolve synchronization issues between nodes including transmission techniques (slotted or unslotted). For example, in an unslotted asynchronous network where each node has its own internal clock, DTC ensures sufficient inter-frame gap and defines the maximum allowable clock variation. The DTC also specifies the buffering requirement to alleviate any clock jitters among nodes.

Medium Dependent Component (MDC): This sublayer deals with the actual type of the medium used to transmit CPs and DBs including, coax, radio frequency, or optical fiber. Selections of connectors, transmitters, receivers, etc., are considered as parts of the MDC sublayer. In an OBS network, as a special category of burst switching, the MDC is transparent to the photonic (WDM) sublayer, which provides lightpaths to the network. A lightpath is an end-to-end connection established across the optical network, and the lightpath uses a wavelength on each link in the path between the source and destination. Consequently, tasks such as optical amplification and wavelength conversion are defined in the MDC sublayer.

3.2. Control plane layers

We now turn our attention from the data plane to the control plane. As we mentioned earlier, separation of planes in the OBS network architecture was inspired by the need to provide practical and reliable medium access protocols at high speeds. Due to current technological limitations in all-optical packet switching, it is not practical to implement MAC protocols in the data plane without interrupting data by optical-electrical converters. In OBS networks, implementing the MAC sublayer as the application layer of the control plane allows arbitration protocols to be performed in a domain (electrical) independent of data (optical).

3.2.1. Burst Signaling Control (BSC) Layer

The BSC layer contains the data plane MACs' scheduling, contention resolution, and offset control protocols through its signaling protocols. Data burst properties including destination address, quality-of-service, etc., are passed to the BSC layer from the MAC sublayer. The BSC layer determines the type of the control packet to be transmitted to the next hop. Typical examples of the control packet types are burst header packets (BHP), burst cancellation packets (BCP), or network management packets (NMP). BHPs contain their associated data burst properties, BCPs can be used to cancel an existing reservation in downstream nodes, and NMPs provide network status information. Other types of control packet can be considered to support multipoint multicasting connections.

Each received control packet on the incoming port is identified by its type and its BSC functions accordingly. For example, if a BHP is received, its data burst reservation request is checked for adequate bandwidth and, upon verification of availability availability, the request is scheduled. New changes in data burst reservations must be communicated to the switch fabric control unit to update its scheduling table.

3.2.2. Signaling Connection Control (SCC) Layer

Similar to the PAD layer in the data plane, the SCC layer includes the routing algorithms for control packets in order to establish the physical path for incoming data bursts. Hence, the actual data burst routing also takes place in this layer. Note that, in general, since the data and control planes can be implemented on separate mediums, it is possible that the physical routing paths for CPs and DBs are different. Various routing protocols can be considered for implementation in the SCC layer.

3.2.3. Signaling Frame Control (SFC) Layer

The main purpose of the SFC layer is to provide reliable transmission of control packets. The SFC layer can be considered as a pure data link protocol operating between adjacent nodes. The SFC layer receives bit streams containing the control packet type and its associated data burst properties, and it constructs CP frames by attaching overhead bits. Many popular framing mechanisms such as High-Level Data Link Control (HDLC) may be considered for the data link protocol. However, the protocol complexity and cost are critical as interface speed increases. Figure 4 (a) shows a generic framing format of a control packet frame.

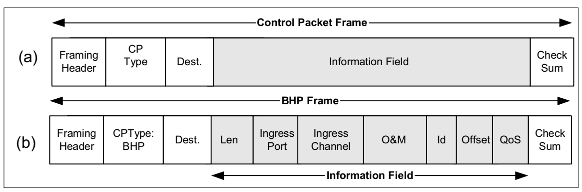

Figure 4. OBS framing structure of the control packet; (a) a generic control packet frame, (b) BHP frame.

To guarantee fast processing of control packets at each node, CPs must contain limited information, yet, it is crucial to protect control packets from errors on each link. Transmission errors in control packets can result in bits being changed in the information field. Incorrect bits will be misinterpreted in the downstream core node and result in, for example, dropping high priority bursts, incorrect switch fabric setup, or even burst misrouting. To protect the CP from error, a cyclic redundancy check (CRC) can be implemented in the checksum field. CRC codes can provide a large selection of error correcting capacity. Each CP must also have a destination field indicating its destination node. Furthermore, all CPs must have a type indication specified in the CP-type field. Different CP types were described in the Burst Signaling Control section. Contents of the information field vary depending on the CP type.

If a control packet is associated with an incoming DB, it is referred to as a BHP. A typical BHP frame is shown in Figure 4 (b). Note that the BHP information field is divided into several fields including length, ingress port, and ingress channel, which refer to DB's duration, its edge node source, and the wavelength on which it is expected to arrive, respectively. The id field can be useful for checking data burst sequencing when deflection routing is allowed. The QoS and offset fields indicate the incoming data burst priority level and the offset time between a BHP-type control packet and its associated data burst, respectively. The O&M field contains network management related signaling information, such as loop-back requests, protection switching, or link failure notification.

4. Conclusions

A layered architectural representation of the OBS network can be used as a baseline for understanding protocol requirements as well as their future development and design. In this article we provided an organized decomposition of the different layers for supporting OBS networks. Detailed descriptions of each layer along with their functionalities and related protocols were presented.

References

[1] I. Chlamtac, A. Ganz, and G. Karm, "Lightpath Communications: An Approach to High Bandwidth Optical WAN's", *IEEE Trans. on Communications*, Vol. 40, no. 7, July 1992, pp. 1171-1182.

[2] D. K. Hunter, *et al.*, "WASPNET: A Wavelength Switched Packet Network", *IEEE Communications Magazine*, Vol. 37, no. 3, March 1999, pp. 120-129.

[3] C. Qiao and M. Yoo, "Optical burst switching (OBS) - A new paradigm for an optical Internet", *Journal of High Speed Networks*, Vol. 8, no. 1, January 1999, pp. 69-84.

[4] J. S. Turner, "Terabit burst switching", *Journal of High Speed Networks*, Vol. 8, no. 1, January 1999, pp. 3-16.

[5] Y. Xiong, M. Vandenhoute, and H. C. Cankaya, "Control architecture in optical burst-switched WDM networks", *IEEE J. on Sel. Areas in Com.*, Vol. 18, no. 10, October 2000, pp. 1838-1851.

[6] L. Xu, H. G. Perros, and G. N. Rouskas, "Techniques for Optical Packet Switching and Optical Burst Switching", *IEEE Communications Mag.*, Vol. 39, no. 1, January 2001, pp. 136-142.

[7] A. Ge, F. Callegti, and Lakshman, "On optical burst switching and Self-similar traffic", *IEEE Communications Letters*, Vol. 4, no. 3, March 2000, pp. 98-100.

[8] V. M. Vokkarane, K. Haridoss, and J. P. Jue, "Threshold-Based Burst Assembly Policies for QoS Support in Optical Burst-Switched Networks", Proc. SPIE OptiComm 2002, Vol. 4874, Boston, MA, pp. 125-136, July, 2002.

[9] V. M. Vokkarane, Q. Zhang, J. P. Jue, and B. Chen, "Generalized burst assembly and scheduling techniques for QoS support to optical burst-switched networks", Proc. IEEE Globecom 2002, Taipei, Taiwan, November 17-21, 2002.

[10] V. M. Vokkarane, J. P. Jue, and S. Sitaraman, "Burst Segmentation: An Approach For Reducing Packet Loss In Optical Burst Switched Networks", Pro. IEEE ICC 2002, IEEE, 5, New York, NY, pp. 2673-2677, April, 2002.

[11] E. Kozlovski, M. Duser, I. de Miguel, and P. Bayvel, "Analysis of Burst Scheduling for Dynamic Wavelength Assignment in Optical Burst-Switched Networks", Proc. IEEE LEOS 2001, Vol. 1, San Diego, CA, pp. 161-162, 2004.

[12] I. Baldine, G. Rouskas, H. Perros, and D. Stevenson, "JumpStart - A Just-In-Time Signaling Architecture for WDM Burst-Switched Networks", *IEEE Com. Mag.*, Vol. 40, no. 2, Feb. 2002, pp. 82-89.

[13] S. Yao, B. Mukherjee, S. J. B. Yoo, and S. Dixit, "All-Optical Packet-Switched Networks: A Study of Contention Resolution Schemes in an Irregular Mesh Network with Variable-Sized Packets. Proceedings", Proc. SPIE OptiComm 2000, Dallas, TX, pp. 235-246, Oct., 2000.

Proceedings of the Advanced Industrial Conference on Telecommunications/Service Assurance with Partial and Intermittent Resources Conference/E-Learning on Telecommunications Workshop
0-7695-2388-9/05 $20.00 © 2005 **IEEE**