

# GMPLS-Based Photonic Burst Switching (PBS) Architecture for Optical Networks

Shlomo Ovadia, Christian Maciocco, and Mario Paniccia  
Intel Corporation, 350 E. Plumeria Drive, M/S CHP3-109, San Jose CA 95134  
[shlomo.ovadia, christian.maciocco, mario.paniccia@intel.com](mailto:shlomo.ovadia, christian.maciocco, mario.paniccia@intel.com)

## ABSTRACT

Photonic burst switching (PBS) architecture with variable time slot provisioning supporting high-speed bursty data transmission within optical enterprise networks is presented as follows: First, the edge and switching node architecture for in-band and out-of-band signaling is defined. Second, we introduce: I) a generalized multi-protocol label switching (GMPLS) based PBS software architecture in terms of control and data-plane operations that is extended to enable the PBS optical interfaces with the software building blocks for edge and switching nodes; II) an adaptive PBS MAC layer functionality and framing of multiple generic payloads. The integration of the proposed PBS network architecture with GMPLS-based software architecture should provide the means for robust and efficient optical transport of bandwidth-demanding applications within enterprise networks.

## I. INTRODUCTION

Optical burst switching (OBS) scheme is emerging as a promising solution to support high-speed bursty data

traffic over wavelength-division-multiplexed (WDM) optical networks [1–4]. The OBS scheme offers a practical alternative between the current optical circuit switching and the emerging all-optical packet switching technologies. It has been shown that under certain conditions, the OBS scheme achieves high-bandwidth utilization and class-of-service (CoS) by elimination of electronic bottlenecks as a result of the optical-electric-optical (O-E-O) conversion occurring at switching nodes, and by using a one-way end-to-end bandwidth reservation scheme with variable time slot duration provisioning scheduled by the ingress nodes. Optical switching fabrics are attractive because they offer at least one or more orders of magnitude lower power consumption with a smaller form factor than comparable O-E-O switches. However, most of the recently published work on OBS networks focuses on the next-generation backbone data networks (i.e. Internet wide network) using high capacity (i.e., 1 Tb/s) WDM switch fabrics with large numbers of input/output ports (e.g., 256x256) and optical channels (e.g., 40 wavelengths), and requiring extensive buffering [1–4]. Thus, these WDM switches tend to be complex and very expensive to manufacture. In contrast, there is a growing demand to support a wide variety of bandwidth-demanding applications such as storage area networks (SANs) and multimedia multicasting at a low cost for both local and wide-area networks.

Consequently, in this work, we are proposing to adapt the OBS scheme to future high-speed optical enterprise networks with limited span and number of hops. These networks are based on fast (< 100 ns) optical switch fabrics with limited number of input/output ports (i.e.,  $\approx 8 \times 8$ ) and with no or limited optical buffering. Preliminary analysis indicates that such optical switch fabrics can be implemented using CMOS compatible technology where the cost per switched bit/s is expected to be at least ten fold lower than conventional high-capacity WDM switches. Although conceptually similar to backbone-based OBS networks, the design, operation

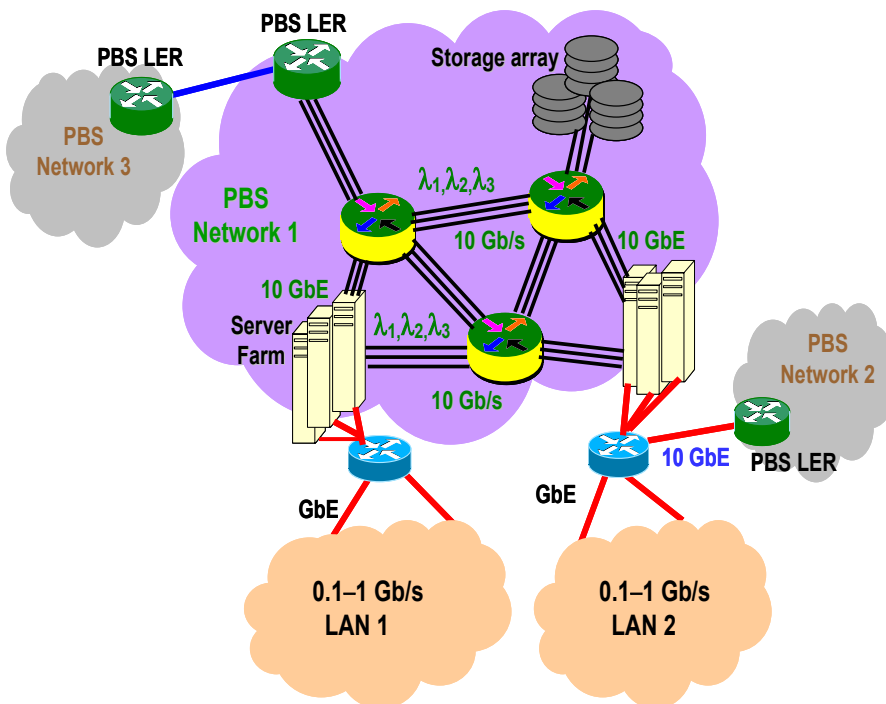
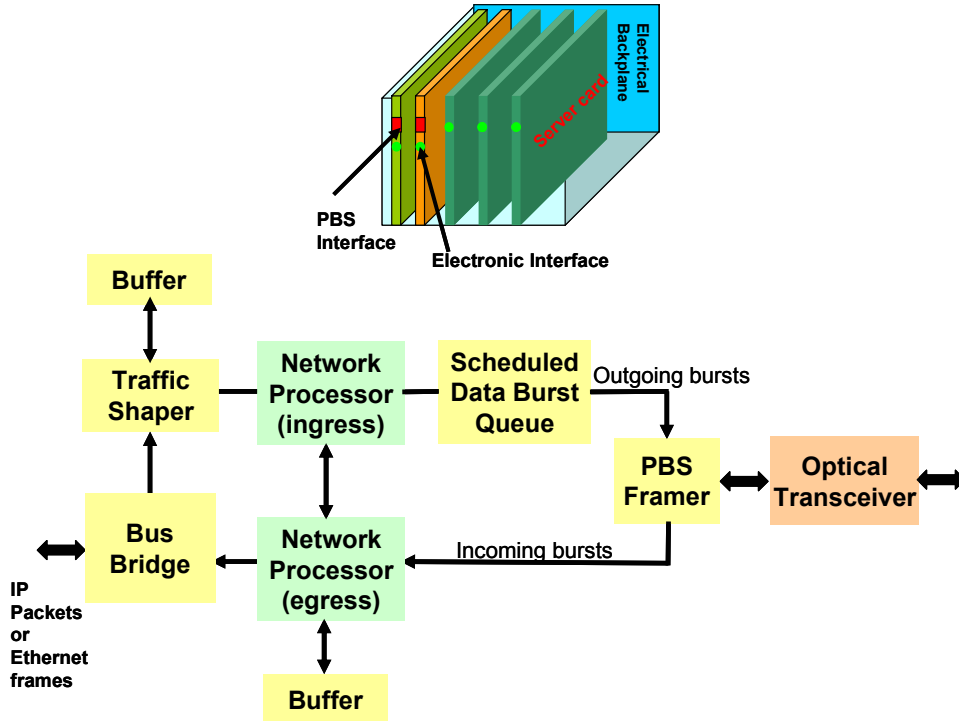


Figure 1. Photonic burst switched network architecture.



**Figure 2.** Multi-server system architecture with block diagram of PBS I/O interface.

constraints, and performance requirements of these high-speed hop and span-constrained optical networks are different. Thus, in this paper we refer to these optical enterprise networks as photonic burst-switched (PBS) networks to distinguish them from conventional OBS networks. The paper is organized as follows: First, we introduce the PBS architecture for hop and span-constrained optical network, including the PBS edge and switching node architecture. Then, we discuss the GMPLS-based PBS software architecture for both ingress/egress nodes and switching nodes to support bursty traffic. This is followed by a discussion on the PBS medium access control (MAC) layer functionality and framing to transport multiple generic data payload.

## II. PHOTONIC BURST SWITCHING ARCHITECTURE

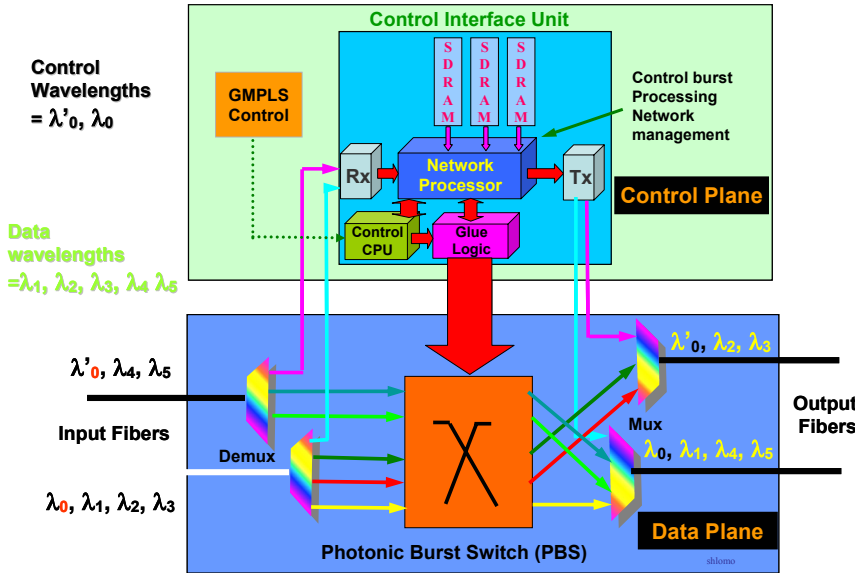
We propose to segment an enterprise network into small islands of high-performance mesh-architecture PBS networks with peer-to-peer signaling where network performance is balanced between implementation and complexity. Such network segmentation into PBS islands considerably simplifies the design, implementation, operation, and management of the optical switching nodes, resulting in reduced overall PBS network costs. Figure 1 shows the proposed PBS network architecture, where an edge device, which can be either a label-edge router (LER) or a multi-server system located at the edge node, is optically connected through a PBS interface to other

network devices such as other server systems, LERs, and storage arrays, which are also equipped with PBS interfaces, via PBS switches. Each PBS switching node can be connected to multiple edge devices. PBS-to-PBS network connectivity is done through an edge node LER using either a conventional interface or a PBS interface. (Section III discusses PBS-to-PBS network routing). Specifically, PBS networks have limited physical span, typically less than 10 km, and a limited number of optical channels, typically less than 8, from ingress to egress nodes. The limited span plays to the advantage of the PBS network as it reduces the guard band uncertainty due to propagation and processing time delays [3]. PBS networks are also hop-constrained due to

limited optical power budget for lower-cost network implementation. Although the maximum size of a PBS network is still under investigation, analysis indicates that a typical PBS network has about 5–15 switching nodes with about 3–4 hops along a given optical label-switched path (OLSP). Preliminary simulation results show that high bandwidth utilization ( $\approx 70\%$ ) on a given OLSP can be achieved at high traffic loads ( $\geq 0.6$ ), depending on the burst assembly scheme and specific parameters [5]. Various PBS network simulation results are currently underway, and are planned to be published later. A GMPLS-based PBS signaling and operation is discussed in section III. Next, the edge and switching node architectures of the PBS network are discussed.

### II.A Edge Node Architecture

Figure 2 shows the architecture of a high-speed optical input/output interface within a modular reconfigurable multi-server system located at the edge node. Internal data communications between each server card and the PBS interface as well as among the different server cards is done using the electrical backplane fabric of the multi-server system. At the ingress node, the server system receives multiple data flows from local or wide-area networks (LANs/WANs) via its electrical 1 Gb/s Ethernet (GbE) interface. It classifies these flows and statistically-multiplexes them to form photonic control and data bursts. A data burst is a collection of IP packets and/or Ethernet frames with the same classification, e.g. the same destination address, quality-of-service (QoS) parameters,



**Figure 3.** PBS switching node architecture for out-of-band signaling.

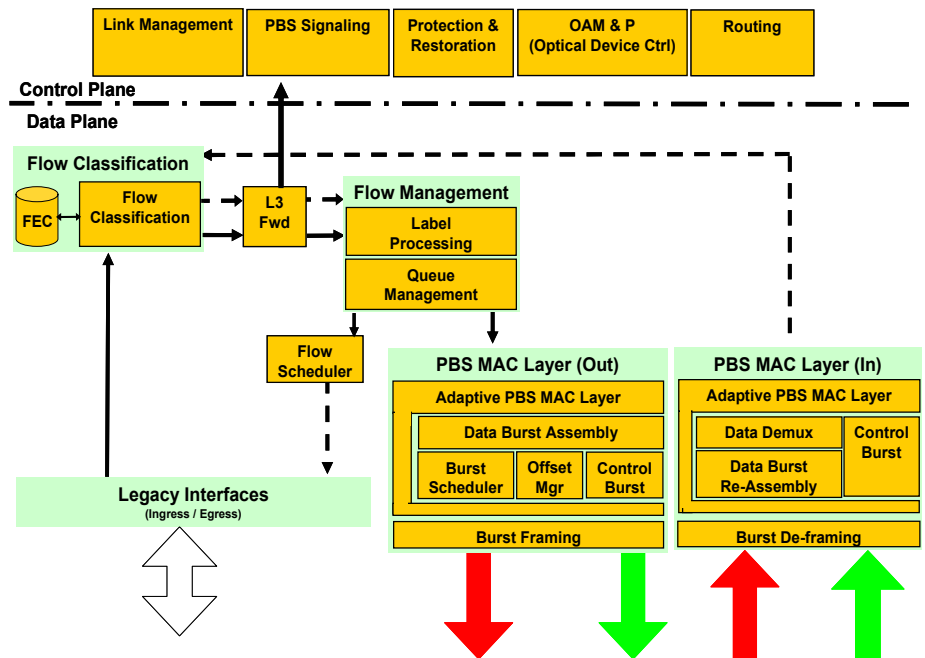
and transmission time window. To enable low-cost PBS networking, for example, a modified 10 GbE interface card at the ingress/egress node can be used. The modified 10 GbE card consists of dual high-end network processors (NPs), one for the outgoing bursts and one for the incoming bursts, on-board memory, a physical layer framer, and a 10 GbE optical transceiver. The on-board DRAM cluster allows one to temporarily buffer multiple scheduled data bursts in case of data burst loss and re-transmit request. The burst assembly, framing, burst scheduling, and control, which are part of the PBS MAC layer and related tasks, are performed by the NPs. NPs are very powerful processors with flexible micro-architectures that are suitable to support a wide-range of packet processing tasks, including classification, metering, policing, congestion avoidance, and traffic scheduling. For example, the Intel® IXP2800 NP, which has 16 microengines, can support the execution of up to 1493 microengine instructions per packet at a packet rate of 15 million packets per second for 10 GbE and a clock rate of 1.4 GHz [6].

## II.B Switching Node Architecture

A simplified block diagram of the PBS switching node architecture is shown in Figure 3. The intelligent node consists of a strictly non-blocking optical switch fabric, a NP, glue logic, optical multiplexers/demultiplexers, and optical

transceivers. The optical switch fabric has a strictly non-blocking space switch architecture with fast ( $< 100$  ns) switching times and with a limited number of input/output ports (i.e.,  $\approx 8 \times 8, 12 \times 12$ ). Each of the incoming or outgoing fiber links typically carries only one data burst wavelength. The switch fabric, which has no or limited optical buffering fabric, performs statistical burst switching within a variable-duration time slot between the input and output ports. If needed, the optical switch buffering can be implemented using fiber-delay-lines (FDLs) on several unused ports. The specific optical buffering architecture such as feed-forward or feedback that is needed for the PBS switch fabric is still under investigation [7]. However, it is expected to be relatively small compared to a conventional packet switching fabric

since these FDLs can carry multiple data burst wavelengths. Other possible contention resolution schemes include deflection routing and using tunable wavelength converters [8]. The PBS network can operate with a relatively small number of control wavelengths ( $\lambda'₀, \lambda₀$ ) since they can be shared among many data wavelengths. Furthermore, the PBS switch fabric can also operate with a single wavelength and multiple fibers, but this case will not be discussed here. The control bursts can be sent either in-band (IB) or out of band (OOB) on separate optical channels (Fig. 3). For the OOB case, the optical data bursts are statistically switched at a given



**Figure 4.** PBS software architecture and building blocks at ingress/egress nodes.

wavelength between the input and output ports within a variable time duration by the PBS fabric based on the reserved switch configuration as set dynamically by the NP in the control interface unit. The NP is responsible for extracting the routing information from the incoming control bursts, providing fixed-duration reservation of the PBS switch resources for the requested data bursts, and forming the new outgoing control bursts for the next PBS switching node on the path to the egress node. In addition, the NP provides overall PBS network management functionality based on an extended generalized multi-protocol labeled-switching (GMPLS) framework (see Section III). For the IB case, both the control and data bursts are transmitted to the PBS switch fabric and control interface unit. However, the NP ignores the incoming data bursts based on the burst payload header information (see Section V). Similarly, the transmitted control bursts are ignored at the PBS fabric since the switch configuration has not been reserved for them. The advantage of this approach is that it is simpler and costs less to implement since it reduces the number of required wavelengths. However, it also leads to lower bandwidth utilization since there are larger timing gaps between successive control bursts required to be processed by the NP at each of the PBS switching nodes. Another approach for IB signaling is to use different modulation formats for the control bursts and the data bursts. For example, the control bursts are non-return to zero (NRZ) modulated while the data bursts are return to zero (RZ) modulated. Thus, only the NRZ control bursts are demodulated at the receiver in the PBS control interface unit while the RZ data bursts are ignored [8]. The preferred OOB or IB control signaling scheme, which maybe application dependent, has not been determined yet.

### III. GMPLS-BASED PBS SOFTWARE ARCHITECTURE

To enable PBS networking within enterprise networks, it is advantageous to extend the GMPLS-based protocols suite to recognize the proposed PBS optical interfaces at both ingress/egress nodes and switching nodes [9, 10]. Under the GMPLS framework, the PBS MAC layer is tailored to perform the different PBS operations while still incorporating the MPLS-based traffic engineering features and

functions for control burst switching of coarse grain (from seconds to days or longer) optical flows established using a reservation protocol and represented by a PBS label. Next, the data and control-plane operations are explained.

#### III.A Data-Plane Operation

Figure 4 shows an integrated data and control-plane PBS software architecture with the key building blocks at ingress/egress nodes. On the data path, packets from legacy interfaces (i.e., IP packets or Ethernet frames) are classified based on n-tuples classification into forward-equivalent classes (FECs) at the ingress/egress node. Specifically, the adaptive PBS MAC layer at the ingress node typically performs data burst assembly and scheduling, control burst generation, and PBS logical framing, while de-framing, de-fragmentation and flow demultiplexing are performed at the egress node.

Figure 5 illustrates the PBS software architecture with the key building blocks at the switching nodes. The transmitted PBS control bursts, which are processed electronically by the PBS NP, undergo the following steps:

- The control burst is de-framed, classified according to its priority, and the bandwidth reservation information is processed. If an optical flow has been signaled and established this flow label is used to look up the relevant information.
- The PBS switch configuration settings for the reserved bandwidth on the selected wavelength at a specific time are either confirmed or denied.
- PBS contention resolution is processed in case of PBS switch configuration conflict. One of the

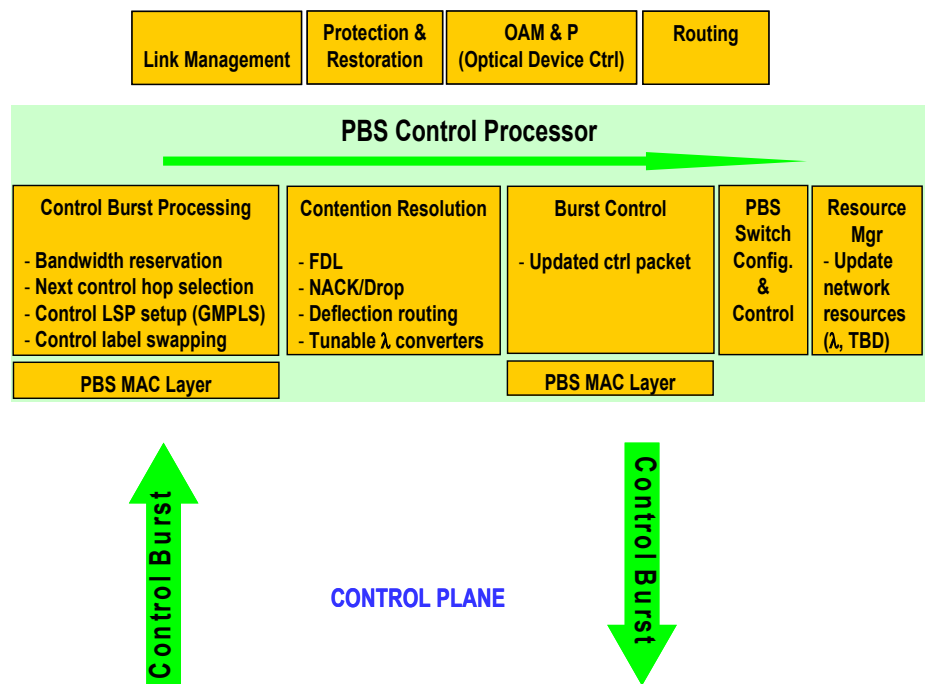


Figure 5. GMPLS-based PBS switching node control-plane software architecture.

three possible contention resolution schemes, namely optical buffering, tunable wavelength converters, and deflection routing can be selected. If none of these schemes are available, the incoming data bursts are dropped until the PBS switch becomes available and a negative acknowledgement message is sent to the ingress node to retransmit.

- A new control burst is generated based on the updated network resources from the resource manager and scheduled for transmission.
- The new control burst is framed and placed in the output queue for transmission to the next node.

### III.B Control-Plane Operation

The key control-plane software components, interacting with the PBS network on a control channel are:

- Label Signaling – coarse grain lightpaths are signaled end-to-end and assigned a unique PBS label. The PBS label has only lightpath segment significance and not end-to-end significance. The signaling of PBS labels for lightpath set-up, tear down, and maintenance, is done through an extension of the IETF resource reservation protocol (RSVP-TE). The PBS label identifying the data burst input fiber, wavelength, and lightpath segment, is used on the control path to enable one to make a soft reservation request of the network resources (through RESV message). If the request is fulfilled (through the PATH message), each switching node along the selected lightpath commits the requested resources and the lightpath is established with the appropriate segment-to-segment labels. Each switching node is responsible for updating the initial PBS label through the signaling mechanism. This indicates to the previous switching node, the label for its lightpath segment. If the request cannot be fulfilled or an error occurred, a message describing the condition is sent back to the originator to take the appropriate action (i.e., to select other lightpath characteristics). Thus, the establishment of a PBS label through signaling enables an efficient MPLS type lookup for the control burst processing. This processing improvement of the control burst at each switching node reduces the required offset time between the control and data bursts. This results in improved PBS network throughput.
- Link Management – this component is responsible for providing PBS network transport link status information such as link up/down, loss of light, etc. The component runs its own link management protocol on the control channel. (IETF link management protocol (LMP) is extended to support PBS interfaces).

- Link protection and restoration – this component is responsible for computing alternate optical paths among the various switching nodes based on user-defined criteria when a link failure is reported by the link management component.
- Routing – this component provides routing information to establish the route for control and data burst paths to their final destination. For PBS networks with bufferless switch fabrics, this component also plays an important role in making PBS a more reliable transport network by providing backup route information that is used to reduce contention. Each PBS network behaves like an autonomous system, employing an internal gateway protocol (IGP) such as a modified routing information protocol (RIP) or open shortest path first (OSPF). The PBS-to-PBS network routing within a larger enterprise network is done using a modified external gateway protocol (EGP) to determine the best available route to a particular PBS network when multiple lightpaths are available. The route selection by the EGP is done via the associated attributes of the specific PBS network.
- Operation, administration, management & provisioning – this component is responsible for performing administrative tasks such as device provisioning.

Thus, the GMPLS architecture is extended to recognize the PBS interface for edge and switching nodes. It also defines the PBS label space and associated signaling, enabling the deployment of IP/Ethernet over WDM in this all-optical network.

## IV. PBS MAC LAYER FEATURES

The PBS MAC layer, at the ingress nodes, performs data burst assembly, scheduling, control burst generation, and PBS MAC layer framing. It should be pointed out that the adaptive burst assembler at the ingress nodes should be designed to guarantee acceptable limits on data burst end-to-end latency and throughput at various network traffic loads to satisfy different CoS. Recent studies have shown that the assembled burst traffic follows a Gaussian distribution with long-range dependence characteristics where the mean burst size and its variance can affect the network performance [11]. Thus, burst length predictions, burst time-out period, and the need to dynamically monitor the lightpath bandwidth utilization at different traffic loads are keys to enable efficient PBS network operation. The PBS MAC scheduling component schedules the outgoing data burst using a variety of known algorithms such as just-in-time and just-enough-time [2, 1]. However, these burst scheduling algorithms can further reduce the burst loss probabilities at the switching nodes in hop- and

span-constrained PBS networks if the network topology and the bandwidth utilization of different lightpath segments are taken into account. Consequently, from a practical point of view, an optimized PBS network design can be achieved by balancing network complexity, implementation, and operating costs with acceptable limits on end-to-end latency, burst loss probabilities, and throughput.

The role of the adaptive PBS MAC layer framing is to enable the following features: I) concatenation of multiple payloads (i.e., IP packets or Ethernet frames) within the same PBS data burst frame, II) adaptive PBS data burst segmentation and re-assembly based on available network processing resources and transport protocol characteristics. As an example, for TCP/IP traffic the adaptive burst assembler sets the maximum PBS burst assembly period to match the TCP window size allowing a higher throughput,

especially until TCP reaches its optimal window size [12]. Figures 6 and 7 below show the generic PBS framing format for both control and data bursts. The generic PBS burst frame has the following fields:

- A PBS generic payload header common for all types of PBS payloads, i.e. control burst or data burst. The PT (Payload Type) field of this header identifies the payload carried by the burst.
- A PBS burst payload with either a control or data payload having:
  - A specific payload header
  - Payload data
  - An optional payload frame check sequence (FCS) – set to 0 if not used

Figure 7 also illustrates the encapsulation of existing LAN/WAN traffic such as Ethernet (10/100 Mb/s, 1Gb/s, 10GbE) over the PBS Network. When framing Ethernet

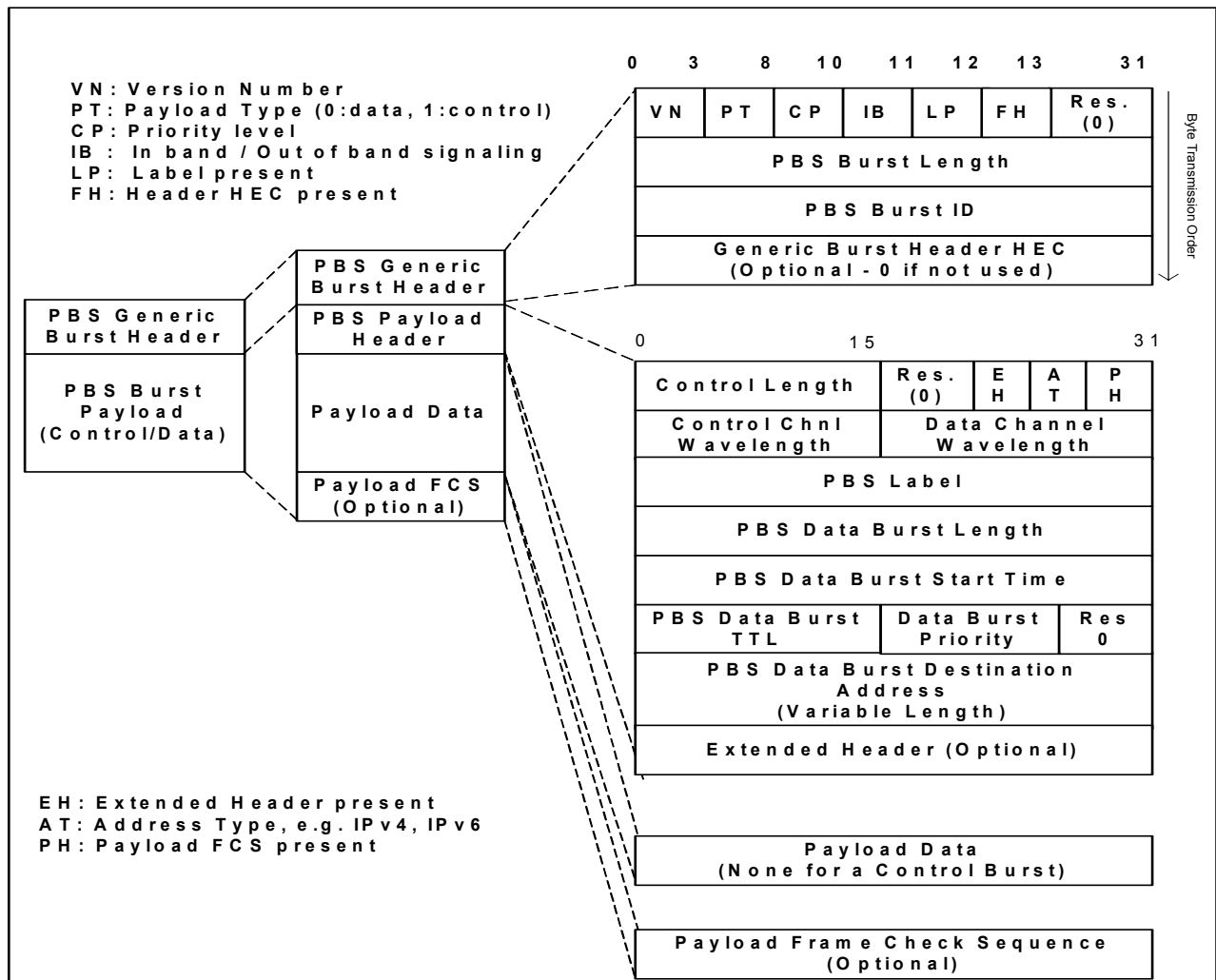


Figure 6 PBS MAC layer framing for control and data bursts.

MAC frames, one must be careful to take into account the inter frame gap (IFG) requirement. It is usually a 12-byte ( $\approx 9.6$  ns for 10GbE) timing gap between frames to allow the receiving MAC device to update its internal counters, calculate the frame FCS, etc.

There are two ways to take the IFG into account:

- Include the IFG bytes in front of the encapsulated Ethernet frame at the expense of wasted network bandwidth
- Rely on the receiving device to offset the data at the expense of extra processing

The selected IFG method is signaled in the PBS Control Burst. Thus, a key advantage of the PBS MAC layer framing is that it takes care of the PBS requirements such as segmentation/re-assembly, scheduling and control, and enables flexible mapping of multiple generic payloads such as Ethernet frames and/or IP packets within PBS frames in order to satisfy different CoS requirements under various traffic loads.

## V. CONCLUSIONS

We have proposed a new intelligent GMPLS-based PBS architecture for hop and span-constrained optical enterprise networks to support bursty traffic. The overall control and operation of PBS switching nodes and the ingress/egress nodes are performed by the NPs for both IB and OOB signaling. The GMPLS-based software architecture, with the key building blocks for both ingress/egress nodes as well as switching nodes, was introduced to recognize the PBS interfaces at the edge nodes, and to define the PBS label space with the associated signaling. This enables the development of IP/Ethernet over fiber integration. Under the GMPLS framework, the adaptive PBS MAC layer is tailored to perform various PBS operations such as data burst assembly, segmentation, re-assembly, concatenation of multiple payloads within the same PBS data burst frame, control burst generation, scheduling, and framing while still incorporating the MPLS-based traffic engineering

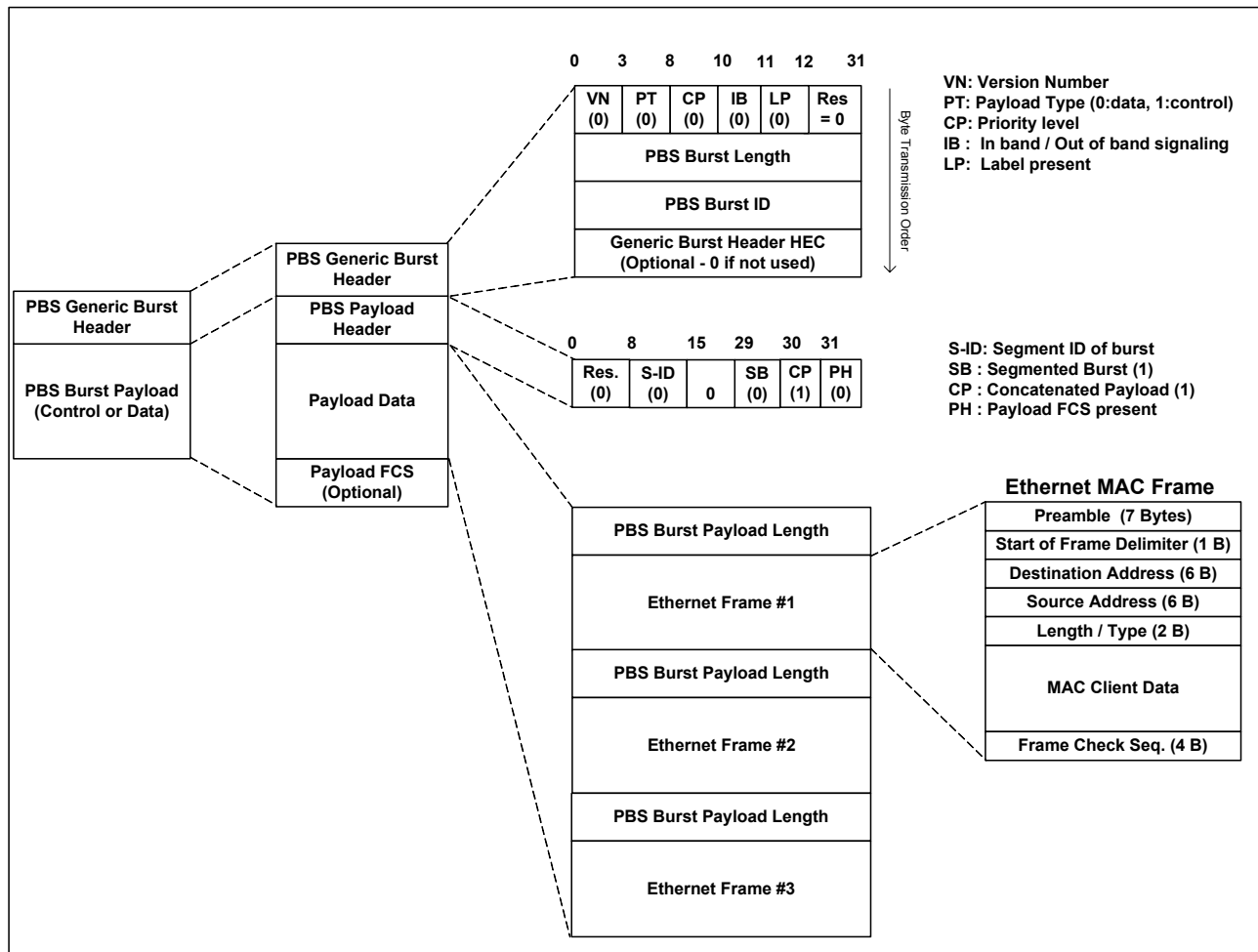


Figure 7. PBS MAC layer framing of Ethernet frames within PBS data burst

features and functions for control burst switching and network management. The PBS MAC layer framing enables the mapping of multiple generic payloads such as Ethernet frames and/or IP packets into PBS data burst frames. Various PBS network simulations are currently underway, and are planned to be published later this year [5].

Looking forward, there are still many technical challenges for these hop- and span-constrained PBS networks that must be solved. For example, the availability and scalability of low-cost fast non-blocking photonic switching fabric that can be integrated with high-speed electronics remains one of the key obstacles. A research program developing a low-cost fast (< 100 ns) optical switch fabric based on CMOS compatible technology is currently underway at Intel. Other key issues such as an optimal optical buffering scheme that can be integrated with the PBS switching fabric as well as the balancing of network complexity and implementation costs with traffic performance must also be addressed. Successful integration of PBS network architecture with low-cost optical switching fabrics based on CMOS compatible technology and GMPLS-based software architecture should provide the means for robust and efficient transport of bandwidth-demanding applications on optical enterprise networks.

## REFERENCES

- [1] C. Qiao, "Labeled Optical Burst Switching for IP-over-WDM Integration," *IEEE Communications Magazine* **9**, 104–114 (2000).
- [2] J. Y. Wei and R. I. McFarland Jr., *IEEE J. of Lightwave Technology* **18**, 2019–2037 (2000).
- [3] J. S. Turner, "WDM Burst switching for Petabit Data Networks," *Technical Digest OFC* (2000).
- [4] M. Düser and P. Bayvel, "Analysis of a Dynamically Wavelength-Routed Optical Burst Switched Network Architecture," *IEEE J. of Lightwave Technology* **20**, 564–585 (2002).
- [5] R. Rajaduray, D. J. Blumenthal and S. Ovadia, "Impact of Burst Assembly Parameters on Edge Router Latency in an Optical Burst Switching Network," submitted to IEEE/LEOS Annual Meeting, October 26–30, Tucson (2003).
- [6] M. Adiletta, M. Rosenbluth, D. Bernstein, G. Wolrich, and H. Wilkinson, "The Next Generation of Intel IXP Network Processors," *Intel Technology Journal* **6**, (no. 3), 6–18 (2002).
- [7] L. Xu, H. G. Perros, and G. Rouskas, "Techniques for Optical Packet Switching and Optical Burst Switching," *IEEE Communications Magazine* **1**, 136–142 (2001).
- [8] D. J. Blumenthal, B. E. Olson, G. Rossi, T. E. Dimmick, L. Rau, M. Masanovic, O. Lavrova, R. Doshi, O. Jerphagnon, J. E. Bowers, V. Kaman, L. Coldren, and J. Barton, "All-Optical Label Swapping Networks and Technologies," *IEEE J. of Lightwave Technology* **18**, 2058–2075 (2000).
- [9] Banerjee, A., Drake, J., Lang, J. P., Turner, B., Kompella, K., and Rekhter, Y., "Generalized Multiprotocol Label Switching: Overview of Routing and Management Enhancements," *IEEE Communications Magazine* **39** (no.1) 144–150 (2001).
- [10] Banerjee, A., Drake, J., Lang, J. P., Turner, B., Awduche, D., Berger, L., Kompella, K., and Rekhter, Y., "Generalized Multiprotocol Label Switching; an Overview of Signaling Enhancements and Recovery Techniques," *IEEE Communications Magazine*, **39** (no. 7), 144–151 (2001).
- [11] K. Laevens, "Traffic Characteristics Inside Optical Burst switched Networks," *Proceedings of SPIE Optical Networking and Communications* Vol. 4874, 137–148 (2002).
- [12] X. Cao, J. Li, Y. Chen, and C. Qiao, "Assembling TCP/IP Packets in Optical Burst Switched Networks," *Proceedings, IEEE GLOBECOM Conference*, Taiwan (2002).