

# Latency in Grid over Optical Burst Switching with Heterogeneous Traffic

Yuhua Chen<sup>1</sup>, Wenjing Tang<sup>2</sup>, and Pramode K. Verma<sup>3</sup>

<sup>1</sup> Department of Electrical and Computer Engineering, University of Houston, Houston, TX, USA

yuhua.chen@mail.uh.edu

<sup>2</sup> Optronic Systems, Inc., Houston, TX, USA

Wenjing.Tang@optronicsys.com

<sup>3</sup> Electrical and Computer Engineering, Telecommunications Systems Program, University of Oklahoma-Tulsa, USA

pverma@ou.edu

**Abstract.** Optical burst switching (OBS) has been proposed as the next generation optical network for grid computing. In this paper, we envision a heterogeneous Grid served by an Optical Burst Switching framework, where grid traffic co-exists with IP and/or a 10 GE based traffic to achieve economy of scale. This paper addresses the latency that Grid jobs experience in OBS networks. The injection of jumbo size grid jobs can potentially affect the latency experienced by IP/10GE traffic. Simulation results have shown that in Grids served by an optical burst switch, grid jobs consistently have lower latency than co-existing IP/10GE traffic, with a slightly elevated latency of IP/10GE traffic when the size of grid jobs increases. We conclude that given the fact that OBS can efficiently handle the enormous amount of bandwidth made available by DWDM technology, Grid over Optical Burst Switching is a cost effective way to provide grid services, even for latency sensitive grid computing applications.

## 1 Introduction

Grid computing [1][2] is a potential approach to making the rapidly growing world wide information resources available when and where they are needed, without the need to physically duplicate those resources. Information resources can be characterized as processing and storage resources. The vehicle that makes these resources available remotely is, of course, the communication facility. Design issues associated with communication facility are critically important because the facility must meet the requirements of the specific application; for example, in terms of the needed bandwidth, while at the same time meeting the latency and availability requirements within the price parameters of the intended application.

The global telecommunication infrastructure is rapidly evolving to make grid computing a reality. The *Dense Wavelength Division Multiplexing* (DWDM) technology provides enormous bandwidth at declining transmission costs making it ubiquitous for long haul transmission. Optical switching techniques, especially, *optical wavelength switching* or *lambda switching* techniques are emerging for large

bandwidth switching or cross connect applications. Because of the enormous bandwidth (10 gigabits per second or higher) associated with a wavelength, lambda switching by itself is not suitable for most end-user applications. The two complementary switching techniques — *optical burst switching* (OBS) [3][4] and *optical packet switching* — will fill in the needs of lower bandwidth granularity applications. Of these two, optical burst switching, has reached a degree of maturity and is the target technology for complementing lambda switching in the near term.

Optical burst switching has been proposed as the next generation optical network for grid computing [5][6][7]. However, in order to achieve economy of scale, grid jobs will co-exist with IP and/or 10GE (*Gigabit Ethernet*) traffic in the OBS networks. It is important to understand the latency that grid jobs experience in the OBS networks since grid jobs are usually associated with a job completion time, especially for latency sensitive grid computing applications. In addition, the injection of jumbo size grid jobs can potentially affect the latency experienced by IP/10GE traffic.

The rest of the paper is organized as follows. Section 2 provides the background of OBS networks. We present the grid over optical burst switching network architecture in Section 3. In Section 4, the latency of grid jobs and how grid jobs affect the latency of co-existing IP/10GE traffic are discussed in detail. We conclude our work in Section 5.

## 2 Optical Burst Switching (OBS) Background

We first give a brief introduction to optical burst switching (OBS). Figure 1 illustrates the basic concept for an optical burst switching network. The network consists of a set of OBS routers connected by DWDM links. The transmission links in the system carry tens or hundreds of DWDM channels, any one of which can be dynamically assigned to a user data burst. One (or possibly more than one) channel on each link is used as a control channel to control the dynamic assignment of the remaining channels to data bursts.

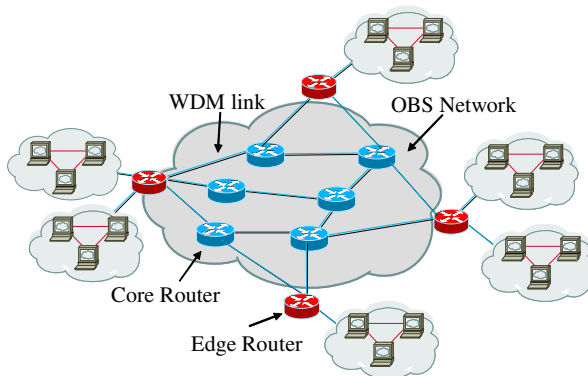


Fig. 1. OBS router architecture

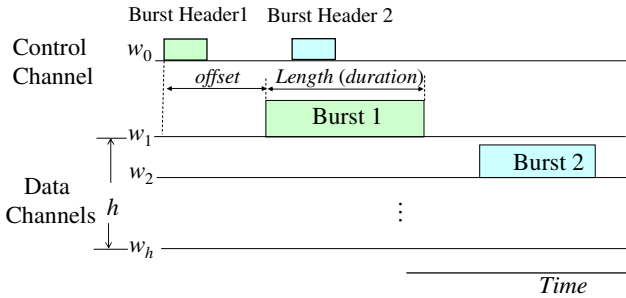


Fig. 2. Bursts and burst headers

An OBS network works as follows. Shortly before the transmission of a data burst on a *data channel*, a *burst header* is sent on the *control channel*, specifying the channel on which the burst is being transmitted and the destination of the burst. The burst header also carries an *offset* field and a *length* field. The offset field defines the time between the transmission of the burst header and the data burst. The length field specifies the time duration of the burst on a DWDM channel. The offset and the length fields are used by the OBS routers to schedule the setup and release of optical data paths *on-the-fly*. Figure 2 shows an example of burst headers sharing the same control channels, while the corresponding data bursts are sent on separate data channels.

An OBS core router, on receiving a burst header, selects an idle channel on the outgoing link leading toward the desired destination. Shortly before the arrival of the data burst, the OBS router establishes a lightpath between the incoming channel that the burst is arriving on and the outgoing channel selected to carry the burst. The data burst can stay in optical domain and flow through the OBS router to the proper outgoing channel. The OBS router forwards the burst header on the control channel of the outgoing link, after modifying the channel field to specify the selected outgoing channel. This process is repeated at every OBS router along the path to the destination.

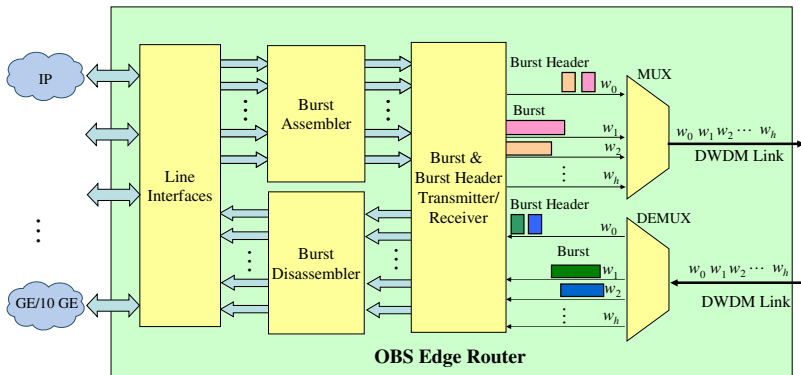


Fig. 3. OBS edge router architecture

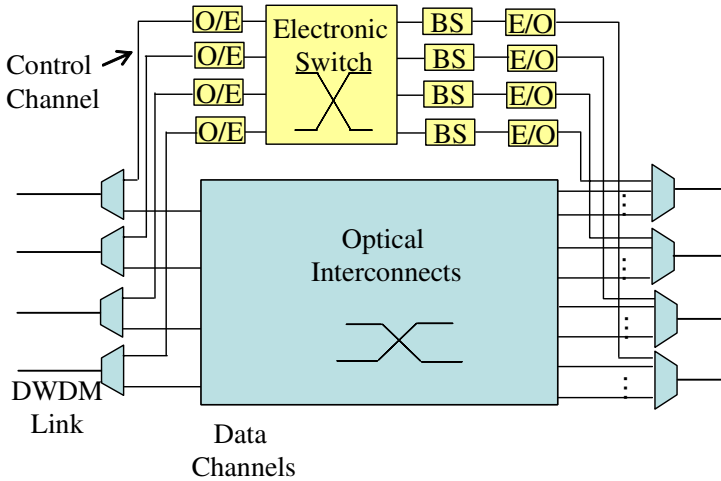


Fig. 4. OBS core router architecture

Figure 3 illustrates the architecture of an OBS edge router. In the ingress direction, packets received on different line interfaces such as IP and *Gigabit Ethernet* (GE)/10 GE are sent to the *Burst Assembler*. The burst assembler classifies the data according to their destinations and QoS levels, and assembles data into different bursts. Once a burst is formed, the burst assembler generates a burst header, which is transmitted on the control channel. After holding the burst for an offset time, the burst assembler releases the data burst to be transmitted on one of the data channels. The control channel and the data channels are combined onto the outgoing DWDM link using a passive optical *multiplexer* (MUX). The outgoing DWDM link is connected to the OBS core router. In the egress direction, the wavelengths on the incoming DWDM link are separated using an optical *demultiplexer* (DEMUX). The burst headers received on the control channel and the data bursts received on data channels are forwarded to the *Burst Disassembler*. The burst disassembler converts bursts back to packets and forwards them to the appropriate line interfaces.

Figure 4 shows the key components of an OBS core router. The architecture consists of two parts, an optical datapath and an electrical control path. The datapath has optical interconnects with/without wavelength conversion capability. The control path includes O/E/O conversion, an electronic switch and a set of *Burst Schedulers* (BSs). Each BS is responsible for making DWDM channel scheduling decisions for a single outgoing link. The electronic switch routes the burst headers received on the control channels of the incoming DWDM links to the corresponding BS according to the destination of the data burst. The BS selects an outgoing channel for the burst and configures the optical switching matrix such that bursts arriving on incoming data channels can pass through to the desired outgoing channels directly without buffering.

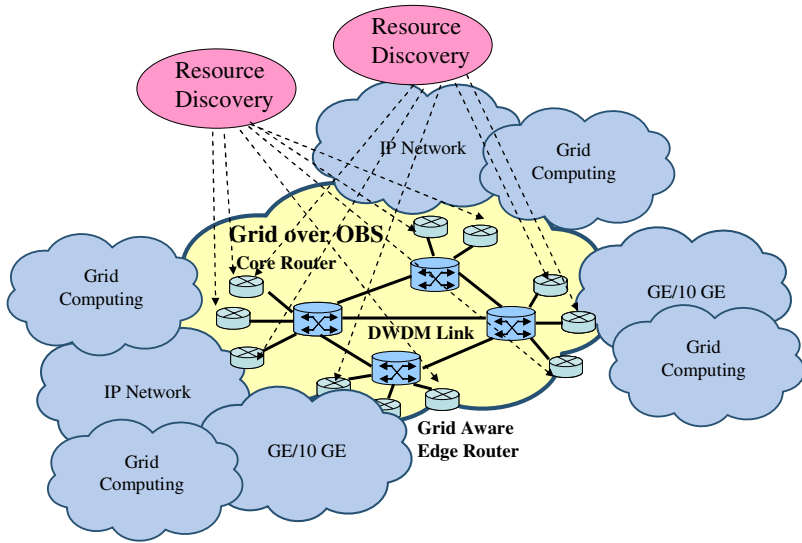


Fig. 5. Grid over optical burst switching architecture

### 3 Grid over Optical Burst Switching Network Architecture

In this section, we describe a decentralized *Grid over Optical Burst Switching* architecture that consists of grid aware edge routers and transparent core routers as shown in Fig. 5. In the decentralized architecture, grid resources can be connected to the *grid aware edge router* directly, or they can be connected through intermediate networks such as IP network or GE/10GE. Distributed grid resource discovery is achieved as an overlay that can potentially affect the states in all or a subset of the edge routers.

The major rationale behind the proposed architecture as opposed to the intelligent OBS router described in [3] is scalability. With the grid computing shifting from scientific computing oriented grids to consumer oriented grids, the grid resources will be owned and managed by many entities, each with their own policy and privilege. If grid resource management is implemented as part of the OBS core router functions, the resource discovery and management can easily become the performance bottleneck. In addition, it is relatively cheap to add grid resources such as hundreds of Terabits of storage, or a server bank, than to gain access to the precious control resources in the OBS network.

The general edge router model shown in Fig. 3 still applies to the grid aware edge router, except that the capability of the *Burst Assembler* block is different, which we call the *Grid Aware Burst Assembler*. Figure 6 shows the functional blocks inside the Grid Aware Burst Assembler. The inputs to the burst assembler are from the line card interfaces.

The *Traffic Classifier* block separates traditional IP/10GE traffic, grid jobs and grid management packets. The traditional IP/10GE traffic is forwarded to the *Router*

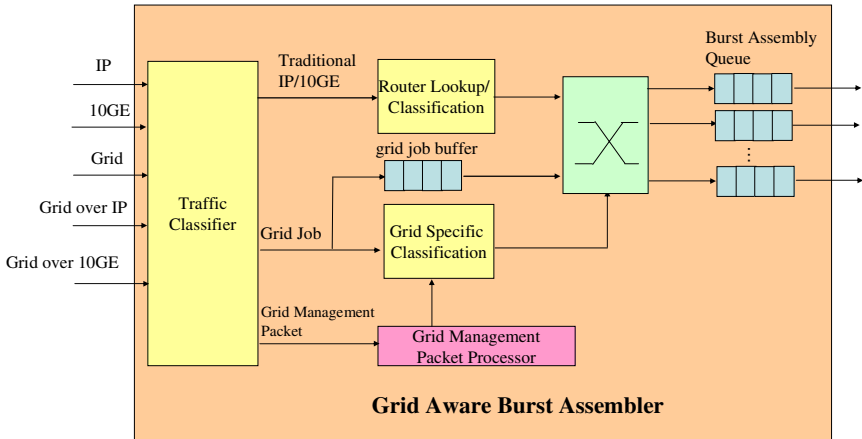


Fig. 6. Grid aware burst assembler

*Lookup/Classification* block. Grid jobs which arrive in jumbo size, whether or not encapsulated in IP/10GE packets, are sent to the grid job buffer while the grid job control information is processed by the *Grid Specific Classification* Block. The *grid management packets* are directed to the *Grid Management Packet Processor*, which in turn configures the states in the *Grid Specific Classification* block when necessary.

When the traditional IP/10GE packets determine the destination edge router and possibly the *Quality-of-Service* (QoS) level, they are routed to proper burst assembly queues. The grid specific classification block will determine the destination of the grid job and forwards the jumbo size grid job to the burst assembly queue. When the conditions to form a burst (maximum burst length or maximum timeout) are met, the burst is sent to the next block for burst header generation.

As we can see, grid jobs will merge with traditional IP/10GE traffic into the burst assembly queue. For simplification purpose, from this point on, we use IP traffic with the understanding that IP and 10GE belong to the same category. In the next section, we will investigate the latency performance of the Grid over OBS model under heterogeneous traffic conditions.

#### 4 Latency Analysis with Heterogeneous Traffic

When grid jobs co-exist with traditional IP traffic, the jumbo size grid job can potentially affect the latency of the IP traffic. It is also important to understand the latency characteristics of grid jobs when grid jobs are transported over OBS networks because a time constraint is usually placed on the grid job, especially for latency sensitive grid computing applications.

The latency that packets (both IP packets and grid job jumbo packets) experience in OBS networks is determined by (1) the *burst formation latency*  $T_{form}$ ; (2) the

transmission latency  $T_{\text{tran}}$ ; (3) the propagation delay  $T_{\text{prop}}$ ; and (4) the offset time  $T_{\text{offset}}$  between the burst header and the burst.

The burst formation latency  $T_{\text{form}}$  is the time that a packet spends at the ingress edge router before the burst of which it is a part is formed. The transmission latency  $T_{\text{tran}}$  is the time that the burst experiences in the process of exiting the ingress edge router. The propagation delay  $T_{\text{tran}}$  is the time that burst takes to traverse the core OBS network and is determined by the fiber miles between the ingress edge router and egress edge router. The offset time  $T_{\text{offset}}$  accounts for the burst header processing time in the core OBS network and is equal to a pre-set per hop offset multiplied by the number of hops a burst is going to traverse in the OBS network.

Therefore, the total latency  $T_{\text{total\_latency}}$  that a packet experiences is

$$T_{\text{total\_latency}} = T_{\text{form}} + T_{\text{tran}} + T_{\text{offset}} + T_{\text{prop}}. \quad (1)$$

In this paper, we focus on the factors that can potentially be affected by the injection of jumbo size grid jobs. Since the propagation delay  $T_{\text{prop}}$  and the offset time  $T_{\text{offset}}$  can be calculated for both IP packets and grid job jumbo packets once the destination edge router is determined, we do not include these two terms in the following discussion. For the rest of the section, we investigate the latency that a packet (IP or grid job) experiences during the burst formation time and the transmission latency that a packet experiences after the burst is launched.

We formulate the problem as follows.

Assume a burst contains  $m$  IP packets and  $n$  grid jobs before it is launched. Denote the  $i$ -th IP packet in the burst  $p_{\text{ip}}^i$  with arrival time  $t_{\text{ip}}^i$  and length  $l_{\text{ip}}^i$ ,  $i = 1, \dots, m$ . Denote the  $j$ -th grid job jumbo packet in the burst  $p_{\text{grid}}^j$  with arrival time  $t_{\text{grid}}^j$  and length  $l_{\text{grid}}^j$ ,  $j = 1, \dots, n$ . Without loss of generality, we have

$$t_{\text{ip}}^1 < t_{\text{ip}}^2 < \dots < t_{\text{ip}}^i < \dots < t_{\text{ip}}^m < t_{\text{ip}}^{m+1}, \quad i = 1, \dots, m, \quad (2)$$

and

$$t_{\text{grid}}^1 < t_{\text{grid}}^2 < \dots < t_{\text{grid}}^j < \dots < t_{\text{grid}}^n < t_{\text{grid}}^{n+1}, \quad j = 1, \dots, n. \quad (3)$$

At an OBS edge router, a burst is formed when the maximum burst length  $L_{\text{max}}$  is reached, or the maximum timeout  $T_{\text{max}}$  is reached [8]. Therefore, based on the above definition, a burst is formed under either of the following conditions:

$$\sum_{i=1}^m l_{\text{ip}}^i + \sum_{j=1}^n l_{\text{grid}}^j \geq L_{\text{max}}, \quad (4)$$

or

$$\begin{cases} \max(t_{ip}^m, t_{grid}^n) - \min(t_{ip}^1, t_{grid}^1) \leq T_{max} \\ \min(t_{ip}^{m+1}, t_{grid}^{n+1}) - \min(t_{ip}^1, t_{grid}^1) > T_{max} \end{cases} \quad (5)$$

For an IP packet  $p_{ip}^i$ , the burst formation latency it experiences is

$$T_{form\_ip}^i = \max(t_{ip}^m, t_{grid}^n, T_{max} + \min(t_{ip}^1, t_{grid}^1)) - t_{ip}^i \quad (6)$$

Similarly, the burst formation latency that a grid job  $p_{grid}^j$  experiences is

$$T_{form\_grid}^j = \max(t_{ip}^m, t_{grid}^n, T_{max} + \min(t_{ip}^1, t_{grid}^1)) - t_{grid}^j \quad (7)$$

Once a burst is launched, a packet (either IP or grid) has to wait for all packets (both IP and grid) in the burst that arrive earlier to be transmitted on the channel.

For an IP packet  $p_{ip}^i$  that satisfies the following condition

$$t_{grid}^{k-1} < t_{ip}^i < t_{grid}^k, \quad (8)$$

the transmission latency  $T_{tran\_ip}^i$  is

$$T_{tran\_ip}^i = \sum_{s=0}^{i-1} l_{ip}^s + \sum_{t=0}^{k-1} l_{grid}^t \quad (9)$$

Similarly, for a grid job  $p_{grid}^j$  that satisfies the following condition:

$$t_{ip}^{l-1} < t_{grid}^j < t_{ip}^l, \quad (10)$$

the transmission latency  $T_{tran\_grid}^j$  is

$$T_{tran\_grid}^j = \sum_{s=0}^{l-1} l_{ip}^s + \sum_{t=0}^{j-1} l_{grid}^t \quad (11)$$

The latency  $T_{latency\_ip}^i$  that an IP packet  $p_{ip}^i$  experiences is

$$T_{latency\_ip}^i = T_{form\_ip}^i + T_{tran\_ip}^i, \quad (12)$$

and the latency  $T_{latency\_grid}^j$  that a grid job  $p_{grid}^j$  experiences is

$$T_{latency\_grid}^j = T_{form\_grid}^j + T_{tran\_grid}^j \tag{13}$$

The average latency  $T_{avg\_latency\_ip}$  of the IP packets in the burst is

$$T_{avg\_latency\_ip} = \frac{1}{m} \sum_{i=1}^m T_{latency\_ip}^i \tag{14}$$

The average latency  $T_{avg\_latency\_grid}$  of the grid jobs in the burst is

$$T_{avg\_latency\_grid} = \frac{1}{n} \sum_{j=1}^n T_{latency\_grid}^j \tag{15}$$

Based on the above model, we have conducted simulations to study the effect of grid jobs on the latency in grid over optical burst switching under heterogeneous traffic conditions.

Figure 7 illustrates the variation in latency as a function of the grid job load. The grid job load is expressed as a fraction of occupancy of the trunk on which the burst is sent out. The IP packet load is similarly expressed as a fraction of the trunk occupancy. IP packets are assumed to be exponentially distributed with a mean length of 1.5 kBytes. The mean grid job size is 1.5 Mbytes with exponential distribution. As noted earlier, bursts can be formed either due to the maximum timeout interval or due to the maximum allowable burst length. The reduction in latency as the load increases is somewhat anti-intuitive. It is, however, explained by the fact that as the load increase, the maximum allowable burst is quicker to form, thus expediting its exit. The correlation in the latency reduction for both the IP packet and grid job is similarly explained. The average latency is the weighted latency of the two kinds of traffic considered, namely, the packet traffic and the grid job traffic.

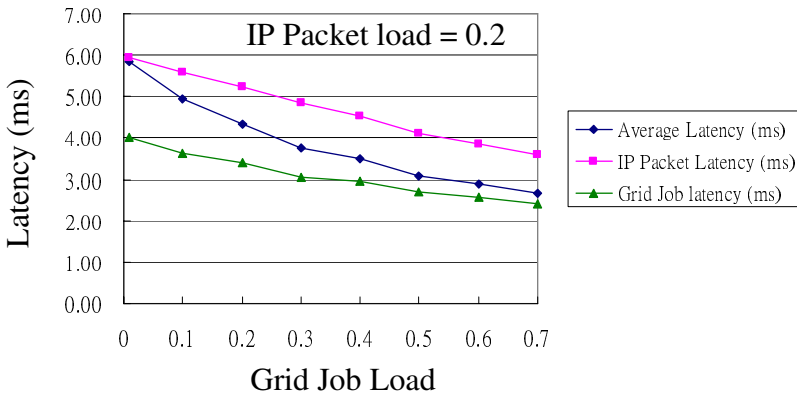


Fig. 7. Latency with fixed IP packet traffic load

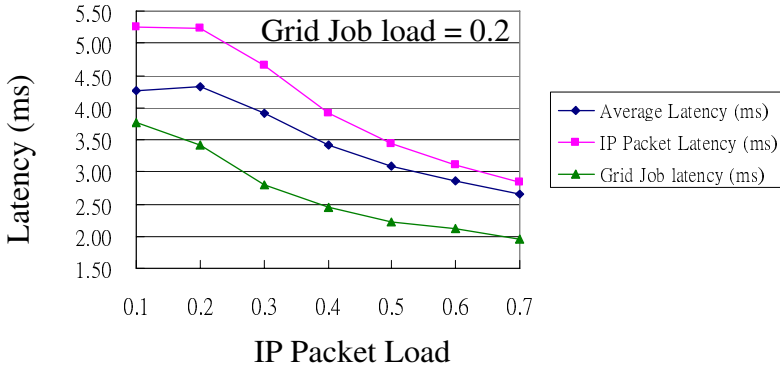


Fig. 8. Latency with fixed grid job traffic load

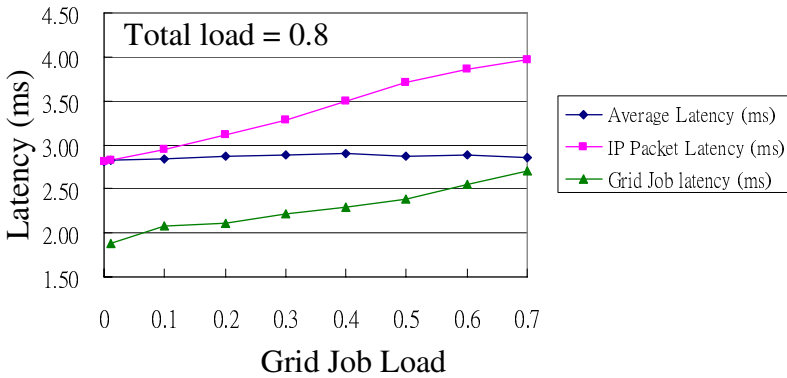
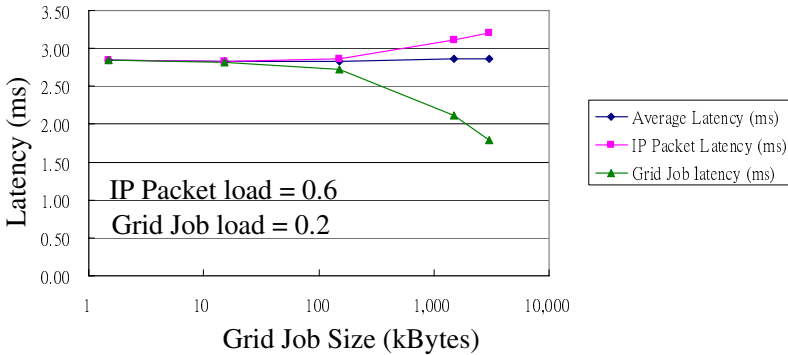


Fig. 9. Latency with fixed total traffic load

Figure 8 presents latency as a function of the grid job load under the condition that the packet load is varied to keep the overall load constant at 0.8. The scenario would imply that as the grid job load increases (as shown on the X-axis), the packet load would correspondingly reduce to keep the overall load constant. The figure clearly indicates the relatively low sensitivity of the average latency to the mixture of traffic as long as the total traffic is kept constant.

Figure 9 shows the variation of latency to the IP packet load, when the grid job load is kept constant at 20% of the trunk capacity. As in Fig. 7, the latency falls as the load increases because a higher level of load implies a quicker exit of the burst because of the maximum allowable burst size.

Figure 10 illustrates the impact of the grid job size on latency when both the packet load and the grid job load are kept constant at occupancies of 0.6 and 0.2 respectively. As shown in Fig. 8, latency is largely insensitive to the grid job size as long as the overall load is kept constant. However, the packet load latency does increase because



**Fig. 10.** Latency with varying grid job sizes

packets are crowded out by longer grid jobs. The grid jobs experience lower latency because larger grid jobs utilize the bandwidth more efficiently.

## 5 Conclusion

In this paper, we have envisioned that grid jobs co-exist with IP/10GE traffic in optical burst switching networks in order to achieve economy of scale. We have investigated the latency of grid jobs when transported over optical burst switching networks. We have also studied the impact of jumbo size grid jobs on the co-existing IP/10GE traffic. Simulation results have shown that in a Grid over OBS, grid jobs consistently have lower latency than co-existing IP/10GE traffic, with a slightly elevated latency for IP/10GE traffic when the size of grid jobs increases. We conclude that given the fact that OBS can efficiently handle an enormous amount of bandwidth made available by DWDM technology, Grid over Optical Burst Switching is a cost effective way to provide grid services, even for latency sensitive grid computing applications.

## References

1. Fox, G., Hey, A.J.G., Hey, T., Berman, F.: Grid Computing: Making the Global Infrastructure a Reality. John Wiley & Sons, Chichester (2003)
2. Joshy, J., Craig, F.: Grid Computing. Prentice Hall Ptr (2004)
3. Turner, J.S.: Terabit burst switching. *Journal of High Speed Networks* 8(1), 3–16 (1999)
4. Qiao, C., Yoo, M.: Optical Burst Switching (OBS) - A new paradigm for an optical internet. *Journal of High Speed Networks* 8, 69–84 (1999)
5. Leenheer, M.D., Thysebaert, P., Volckaert, B., Turck, F.D., Dhoedt, B., Demeester, P., Simeonidou, D., Nejabati, R., Zervas, G., Klouididis, D., O'Mahony, M.J.: A View On Enabling-Consumer Oriented Grids Through Optical Burst Switching. *IEEE Communications Magazine*, 124–131 (2006)

6. Simeonidou, D., Nejabati, R., Zervas, G., Klouidis, D., Tzanakaki, A., O'Mahony, M.J.: Dynamic Optical-Network Architectures and Technologies for Existing and Emerging Grid Services. *Journal of Lightwave Technology* 23(10), 3347–3357 (2005)
7. Simeonidou, D., Nejabati, R.: Grid Optical Burst Switched Networks (GOBS). Global Grid Forum Draft (2005)
8. Xiong, Y., Vandenhoute, M., Cankaya, H.C.: Control architecture in optical burst-switched WDM networks. *IEEE Journal on Selected Areas in Communication* 18, 1838–1851 (2000)