On the early release of Burst-Control Packets in Optical Burst-Switched networks*

José Alberto Hernández and Javier Aracil

Universidad Autónoma de Madrid
Ctra. Colmenar Viejo, km. 15, 28049 Madrid, Spain
[Jose.Hernandez, Javier.Aracil]@uam.es,
WWW home page: http://www.ii.uam.es/~networking

Abstract. In Optical Burst-Switched networks, the so-called Burst-Control Packet is sent a given offset-time ahead of the optical data burst to advertise the imminent burst arrival, and reserve a time-slot at each intermediate node to allocate the incoming optical burst. This work proposes a methodology to estimate the number of packets to arrive in a given amount of time, in order to make it possible to send the BCP packet straightafter the first packet arrival and reduce the latency experienced during the burst-assembly process.
The following studies the impact of a wrong guess in terms of over-reservation of resources and waiting-time at the assembler, providing a detailed characterisation of their probability density functions. Additionally, a case example in a scenario with non-homogeneous Poisson arrivals is analysed and it is shown how to choose the appropriate burst-assembly algorithm values to never exceed a given over-reservation amount.

1 Introduction

Dense Wavelength Division Multiplexing [1] has been proposed as a promising physical layer technology for the forthcoming next-generation Internet, due to the huge amount of raw bandwidth provided, in the order of gigabits per wavelength, with more than one hundred wavelengths per optical fibre [2, 3]. In this light, the Optical Burst Switching (or just OBS) paradigm over DWDM physical layers arises as a cost-effective solution for the high utilisation and multiplexing of such tremendous amount of raw bandwith with relatively low switching complexity involved [4, 5].

In a typical OBS network, ingress nodes aggregate incoming packets into larger-size data bursts, which are transmitted all-optically through the network core. Such optical bursts do not suffer from optical/electrical/optical conversion at the intermediate nodes, leading to a fast and efficient transmission of large volumes of data.

Each optical burst has an associated Burst-Control Packet (or BCP), this is, a small-size packet which carries the control information to get its associated

* This work was funded by the European Union *e-Photon/ONe+* project and by the Spanish Ministry of Science and Education under the DIOR project

data burst delivered at the other end of the optical network. To do so, the BCP is sent a given offset time in advance of its associated data burst, and is processed electronically (it suffers O/E/O conversion) at each intermediate node along the path. Its main role is to advertise each intermediate node of the size and expected arriving time of its associated data burst [4, 6]. With this information, the core node can find and reserve a time-slot at the appropriate output wavelength, and consequently, can immediately switch the data burst in the optical domain as soon as it arrives. This way, the need for temporal buffering of optical data is removed, in contrast to electrical switches.

Typically, the BCP is generated and transmitted straightafter the data burst is assembled at the border node, since it must know the exact burst size and release time to inform the intermediate nodes' scheduler, under *Just-Enough-Time* (JET) scheduling [4, 6]. Hence, in addition to the delay suffered by the data packets during the burst assembly process, the packets suffer an extra delay given by the offset-time between the BCP and the data burst. In certain situations, such delay may be excessive.

To alleviate such long delay, this work proposes a mechanism to overlap the burst-assembly delay and the offset delay suffered by the the data packets. Essentially, after the first packet has arrived at the burst assembler, our algorithm generates and sends off the BCP to the next hop in the path. Such early BCP carries out a given burst-release time (which is equal to the offset time) and a rough estimation of the final size of the optical burst. The following studies how to make such estimation, and analyses its impact on the global network performance.

The remainder of this work is organised as follows: Section 2 studies the statistics of the burst generation time and hints how to estimate the final burst size. Section 3 validates the equations derived in the analysis section and further proposes a scenario to evaluate the benefits of the early BCP release mechanism. Finally, section 4 brings the main findings, conclusions and merits of this work.

2 Statistical analysis of the burst-release time

2.1 Problem statement

As previously stated, ingress OBS nodes aggregate packets together into the so-called bursts, which are converted to the optical plane. Throughout this work, packet arrivals shall be assumed to follow a Poissonian basis with rate λ packets/seg. This assumption is gaining in importance among the network research community, especially after the recent measurement-based studies in core Internet links [7, 8].

Let λ refer to the average rate of incoming packets per unit of time at the burst assembler, and let n refer to the number of packets in a burst. Without loss of generality, incoming packets are assume to have constant size. As shown in figure 1, packet interarrivals x_i, $i = 1, \ldots$ are exponentially distributed with parameter $\lambda = \frac{1}{EX}$.

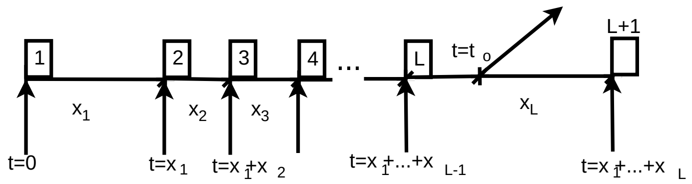

Fig. 1. Notation

Under these assumptions, the burst-assembler proceeds as follows: When the first packet arrives at the border node (packet no. 1 in the figure), the BCP is generated. Essentially, each BCP contains the information of (1) the time at which the burst shall be released, namely t_o, and (2) the number of packets in the optical burst, namely $\hat{L}$. Typically, t_o must not be smaller than the offset time, which is set by the network topology, and represents the amount of time that a BCP needs to configure all the intermediate nodes along the path. Hence, since t_o is fixed, the role of the burst-assembler is to guess the appropriate value of $\hat{L}$, taking into account that:

– It may well happen that the actual number of packets arriving within time t_o, say L, is smaller than the estimated burst-size $\hat{L}$. In this case, the optical burst must be released anyway at time t_o, and cannot wait for the $\hat{L} - L$ packets remaining to fulfill the optical burst. Hence, the BCP has reserved at intermediate nodes for $\hat{L}$ packets, whereas only $L < \hat{L}$ will actually occupy such scheduled time. Thus, the amount of over-reservation is $\hat{L} - L$ packets.
– On the other hand, it may well happen that the optical burst-size reaches the total of L packets before t_o, say at time $t < t_o$. However, the optical burst cannot be released before time t_o, because it is only guaranteed that the BCP has allocated space at time t_o. Therefore, the data burst must wait in queue at the intermediate node during $t_o - t$ units of time, and no resource over-reservation occurs.

The above clearly brings a trade-off when guessing/estimating the value of $\hat{L}$ packet arrivals before time t_o. A conservative estimation ($\hat{L}$ small enough) would lead to the over-reservation of resources at the intermediate nodes, whereas a tight estimation ($\hat{L}$ large) may produce buffer overloading at the burst-assembler. The following sections analysis the impact of choosing $\hat{L}$ small or large by means of, firstly the over-dimensioning probability distribution (when L small), secondly the distribution of waiting time in queue (when L large).

2.2 Probability distribution of the burst-release time

Under the assumption of Poissonian packet arrivals, the assembly time t for a L-sized burst follows a Gamma distribution with $L - 1$ degrees of freedom and parameter λ, as noted in [9, 10]. The Probability Density Function (pdf) for such assembly time is given by

$$\Gamma_t(L-1,\lambda) = \frac{\lambda^{L-1} t^{L-2}}{(L-2)!} e^{-\lambda t}, \quad t \geq 0 \tag{1}$$

with mean $E[t] = \frac{L-1}{\lambda}$ and standard deviation $Std[t] = \sqrt{\frac{L-1}{\lambda^2}}$.

In this light, since the BCP is released after the first packet arrival with information t_o and L, the probability to actually have $\hat{L}-1$ additional packet arrivals before release time t_o is given by:

$$P(t < t_o) = \int_0^{t_o} \frac{\lambda^{L-1} t^{L-2}}{(L-2)!} e^{-\lambda t} dt = \frac{\gamma_{\text{inc}}(L-1, \lambda t_o)}{(L-2)!} \tag{2}$$

where γ_{inc} refers to the incomplete gamma function[1].

It is worth noticing here that such probability depends not only on the choice of t_o, but also on the value of L. Clearly, it is easier to complete L_1 packets within time $[0, t_o]$ than $L_2 > L_1$ within the same amount of time. This effect is shown in fig. 2.

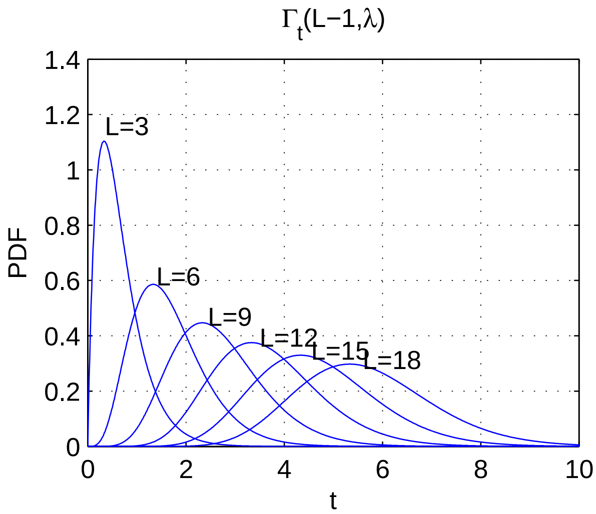

Fig. 2. Burst-release time distribution for various values of L

In this figure, the value of $\lambda = 3$ is fixed, but L takes values in the range $L \in \{3, 6, 9, 12, 15, 18\}$. As shown, given $t_o = 4$ fixed, the probability to complete L packets before t_o decreases the larger the value of L (see table 1).

Hence, the choice of L relatively small is a conservative estimation, that is, high probability to complete L packets before time t_o. Moreover, the choice of

[1] Note that $\gamma_{inc}(n, x) = \int_0^x t^{n-1} e^{-t} dt$

L	$P(t<4)$
3	0.9999
6	0.9924
9	0.9105
12	0.6528
15	0.3185
18	0.1013

Table 1. A few values of $\gamma_{\mathrm{inc}}(L-1, \lambda t_o)$.

small L would probably lead to a situation at which data bursts are completed in a time t much earlier than t_o, thus requiring to allocate them in memory for time $t_o - t$. However, the opposite (relatively large L values) leads to high probability of transmitting data bursts with less packets than predicted, thus over-loading the network.

The following analyses in detail the amount of over-reserved resources due to large values of L, and the burst waiting-time in queue due to early completion when small L.

2.3 Case 1: Over-reservation of resources

This section studies the first situation described above: The case at which the BCP reserves for a $\hat{L}$-sized optical burst, whereas the actual optical burst is of size $n < \hat{L}$ size. Let Y refer to the random variable that represents the excess reservation at intermediate nodes, that is, $Y = \hat{L} - n$. The probability mass function of Y, conditioned to $\hat{L}$, is given by:

$$\begin{aligned} \mathbb{P}(Y = m) &= \mathbb{P}(n = \hat{L} - m \text{ Poisson arrivals in } [0, t_o)) = \\ &= \frac{(\lambda t_o)^{\hat{L}-m}}{(\hat{L}-m)!} e^{-\lambda t_o} \end{aligned} \quad (3)$$

where $0 \leq n \leq \hat{L} - 1$. As shown, the random variable Y is distributed as a shifted Poisson distribution

Finally, the average over-reservation (in packets) is given by:

$$\mathbb{E}[Y] = \sum_{n=1}^{\hat{L}-1} (\hat{L} - n) \frac{(\lambda t_o)^{n-1}}{(n-1)!} e^{-\lambda t_o} \quad (4)$$

2.4 Case 2: Waiting time distribution

This section examines the second situation described above: The case at which the $\hat{L}$-th packet arrives at time $t < t_o$, thus fulfilling the data burst, and forcing the completed burst to be buffered for time $t_o - t$. Let Z refer to the random variable that represents the waiting-time in buffer, that is, $Z = t_o - t$. Then, it is clear that the probability density function of Z is the shifted gamma distribution:

$$
\begin{aligned}
f_Z(t) &= \Gamma_{t_o - t}(\hat{L}-1, \lambda) = \\
&= \frac{\lambda^{\hat{L}-1}(t_o - t)^{\hat{L}-2}}{(\hat{L}-2)!} e^{-\lambda(t_o - t)}, \quad 0 \le t \le t_0
\end{aligned} \tag{5}
$$

The average waiting time can be easily obtained by:

$$
\begin{aligned}
E[t_o - t] &= \int_0^{t_o} (t_o - t) \frac{\lambda^{\hat{L}-1} t^{\hat{L}-2}}{(\hat{L}-2)!} e^{-\lambda t} dt = \\
&= t_o \int_0^{t_o} \frac{\lambda^{\hat{L}-1} t^{\hat{L}-2}}{(\hat{L}-2)!} e^{-\lambda t} dt - \int_0^{t_o} \frac{\lambda^{\hat{L}-1} t^{\hat{L}-1}}{(\hat{L}-2)!} e^{-\lambda t} dt = \\
&= t_o \frac{\gamma_{\mathrm{inc}}(\hat{L}-1, \lambda t_o)}{(\hat{L}-2)!} - \frac{1}{\lambda} \frac{\gamma_{\mathrm{inc}}(\hat{L}, \lambda t_o)}{(\hat{L}-2)!}
\end{aligned} \tag{6}
$$

3 Experiments

The following experiments are focused on first, demonstrating the validity of eq. 3 and 5 above; and secondly, propose an algorithm to obtain the adequate value of $\hat{L}$ to meet a set of requirements, in an environment with non-homogeneous Poisson arrivals.

3.1 Validation

This experiment aims to show the validity of eq. 3 and 5 above. To do so, we have simulated a burst-assembler receiving $N = 10^6$ incoming packets on a Poissonian basis with parameter $\lambda = 100000$ packets/sec. The estimated size of outgoing optical bursts has been chosen as $\hat{L} = 50$ packets, and the burst-release time is $t_o = 53.52$ms. With these values, an amount of 25% of cases are not able to complete a 50-sized data burst before time t_o, as given by:

$$
P(t > t_o) = \frac{\gamma_{\mathrm{inc}}(\hat{L}-1, \lambda t_o)}{(\hat{L}-2)!} = 0.25
$$

whereas the other 75% do manage to fulfill the data bursts within time.

Figure 3 (top) shows the histogram of the waiting-time in queue of the 75% of cases (around $0.75 \times 10^6/50 \approx 15000$ simulated bursts) that achieve burst-completion within time together with the theoretical values given by equation 5. Similarly, figure 3 (bottom) shows the amount of over-reserved packets at intermediate nodes for the 25% of cases (around $0.25 \times 10^6/50 \approx 5000$ simulated bursts) that do not reach $\hat{L}$ within time. Again, the theoretical values, given by equation 3, are plotted together with the simulated results.

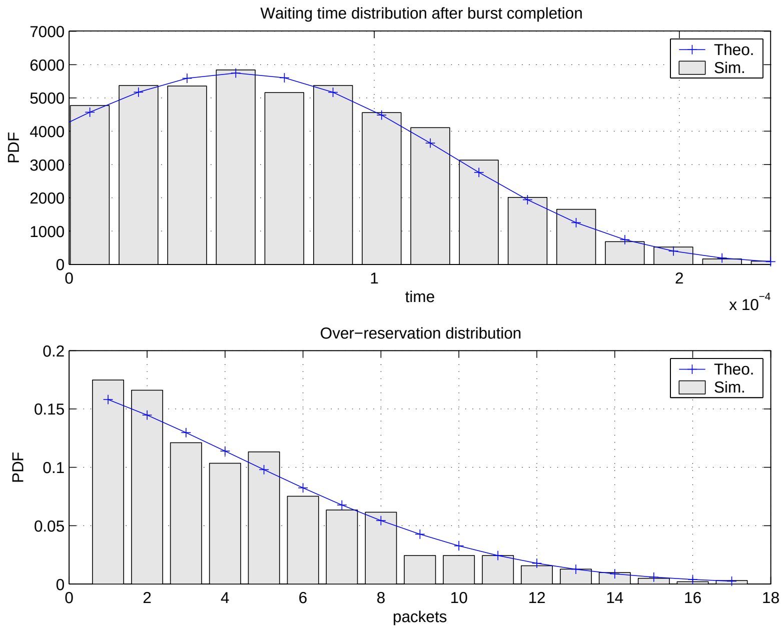

Fig. 3. Waiting-time (top) and over-reservation (bottom) distributions for $\hat{L} = 50$ packets

Obviously, the bottom figure is less accurate than the top figure, since in the bottom figure only 25% of the total cases constitute the histogram, whereas the top figure has been computed with 75% of the cases. Overall, it turns out that the analytical expressions match the simulation results very well.

3.2 Numerical example

This experiment presents a mechanism to obtain the appropriate value of $\hat{L}$ in a sample scenario with non-homogeneous Poisson arrivals. Indeed, incoming packets have been assumed to follow a Poissonian distribution with changing λ. For simplicity, we have assumed:

$$\lambda(n) = 10^5 + 2 \cdot 10^5 \cos^2(\frac{2\pi}{1000}n) \quad \text{packets/sec}$$

This is a value of λ ranging from 100000 to 300000 packets/sec with period 500 samples.

Obviously, the burstifier does not know the real value of λ at each instant and has to estimate it. In the experiment we have considered the well-known Exponential-Weighted Moving Average algorithm to estimate λ. Essentially, such algorithm proceeds as follows: For every new packet arrival with interarrival time x_n from the previous one, $n \geq 1$, we estimate the average interarrival time as:

$$\hat{\hat{x}}_n = \frac{W}{W+1}\hat{\hat{x}}_{n-1} + \frac{1}{W+1}x_n$$

for some value of W. With this value, we compute the estimated $\hat{\lambda}$ as $\hat{\lambda}_n = \left(\hat{\hat{x}}_n\right)^{-1}$ since $\lambda = 1/EX$. The choice of parameter W is a measure of the memory of the estimation. That is, W small gives more weight to new samples than W large. However, W provides a smoother estimate of λ. For highly changing environments, a small value of W is preferred. In our case, figure 4 shows the evolution of the estimated $\hat{\lambda}$ for several values of W. Clearly, $W = 25$ shows the best behaviour in terms of high accuracy in the estimation with fast tracking of the changes in λ.

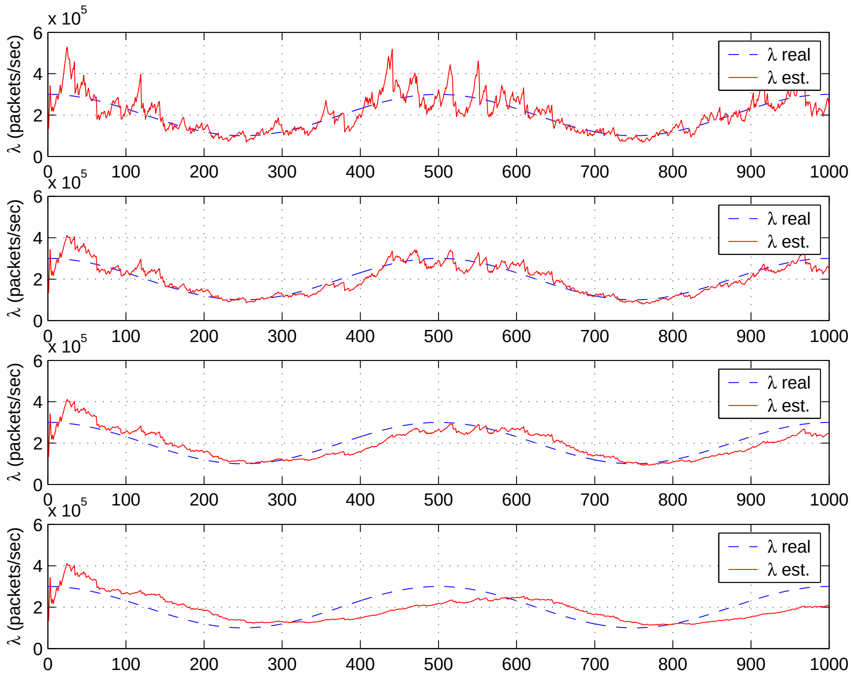

Fig. 4. Example of EMWA for $W = 10$ (top), $W = 25$ (middle-top), $W = 50$ (middle-bottom) and $W = 100$ (bottom) for the case $\lambda(n) = 10^5 + 2 \cdot 10^5 \cos^2(\frac{2\pi}{1000}n)$ packets/sec

In our case, we have chosen the value $W = 25$. Figure 5 (top) shows the real value of λ along with the estimated $\hat{\lambda}$ using the EWMA algorithm with $W = 25$.

Concerning the remaining experiment parameters, we have chosen the value $t_o = 10.19$ ms, which is fixed by the network topology, and a strategy of designing $\hat{L}$ on attempts to have bursts than exceed this value no more than 10% of the times. That is:

$$\text{Find } \hat{L} \text{ such that } \frac{\gamma_{\text{inc}}(\hat{L}-1, \hat{\lambda}t_o)}{(\hat{L}-2)!} = 0.9$$

as pointed out in equation 2.

With this parameter set, figure 5 shows the evolution of $\hat{lambda}_n$ (fig. 5 top), the predicted size (fig. 5 middle) and waiting-time in queue (fig. 5 bottom) with both real λ and estimated $\hat{\lambda}$.

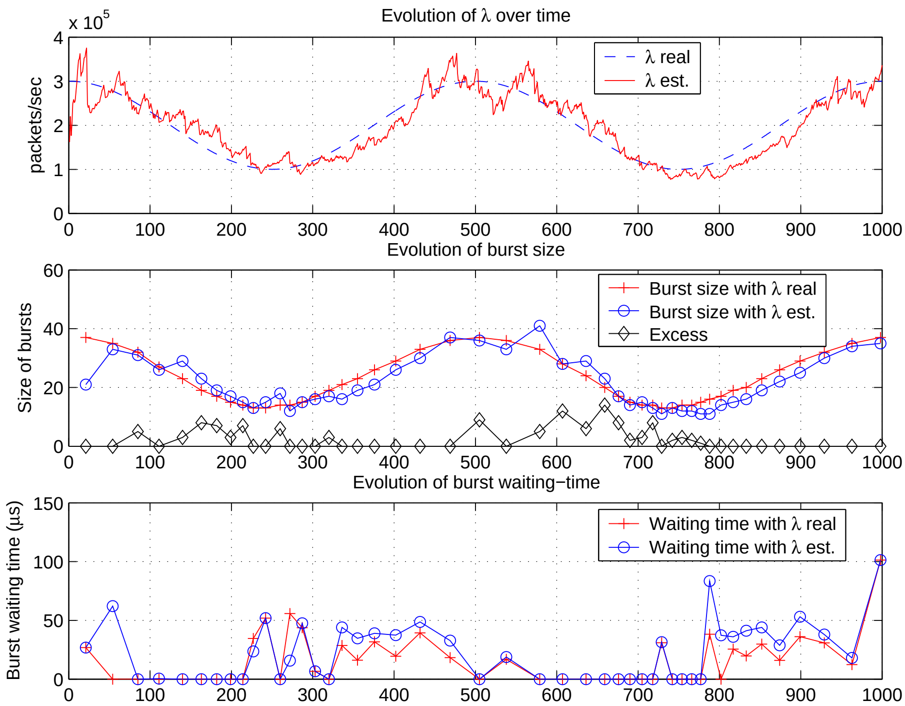

Fig. 5. Numerical example: evolution of $\hat{lambda}_n$ (top); estimated optical burst size along with their optimal values (middle); and, burst-waiting time in queue along with the optimal values (bottom)

As shown, when the estimation $\hat{\lambda}$ and the real value of λ are close, both the predicted size and waiting-time in queue are close too. Additionally, it is worth remarking that, in those cases where there is an excess of packet reservation, the waiting-time in queue is null (see for instance the inverval $n \in [100, 200]$), and viceversa.

Finally, it is worth emphasising that the algorithm finds the appropriate $\hat{L}$ which produces over-reservation only 10% of the times. However, sometimes the algorithm finds such value based on a wrong estimate of λ, thus leading to situations with higher over-reservation of resources than the designed 10% of the cases.

4 Summary and conclusions

This work proposes to send the Burst Control Packet of a given optical burst as soon as the first packet comprising the burst has arrived at the burst-assembler. The burst-release time information is set by the network topology, and the final size of the optical burst is thus estimated as the amount of expected packets arriving until burst-release. Furthermore, such number must not be chosen as the expected number of arrivals, but it can be set to any particular value such that the probability to have a smaller number of incoming packets is small.

The impact of choosing a small or large value is further analysed and concluded that, the larger its value, the more likely to over-reserve resources in the network which, indeed, has a clear impact in the global network performance. On the other hand, the smaller the estimated value, the more likely to not exceed it, thus reducing the amount of unnecessary time-slot reservation at intermediate nodes, at the expense of having to temporarily allocate the completed burst until its release. Nevertheless, in a case or another, the burst-assembly delay and the offset-time delay are overlapped, thus reducing the former and providing a more efficient early burst release.

References

1. C. Brackett, "Dense Wavelength Division Multiplexing networks: Principles and applications," *IEEE JSAC*, vol. 8, no. 6, pp. 948–964, 1990.
2. J. Turner, "Terabit burst switching," *J. High Speed Networks*, vol. 8, pp. 3–16, 1999.
3. S. Verma, H. Chaskar, and R. Ravikanth, "Optical Burst Switching: A viable solution for Terabit IP backbone," *IEEE Network*, vol. Nov/Dec, pp. 48–53, 2000.
4. C. Qiao and M. Yoo, "Optical Burst Switching (OBS) – A new paradigm for an optical Internet," *J. High Speed Networks*, vol. 8, pp. 69–84, 1999.
5. Y. Chen, C. Qiao, and X. Yu, "Optical Burst Switching: A new area in optical networking research," *IEEE Network*, vol. May/June, pp. 16–23, 2004.
6. M. Yoo and C. Qiao, "Just-Enough Time (JET): A high-speed protocol for bursty traffic in optical networks," in *Proc. IEEE/LEOS Conf. Tech. Global Info. Infrastructure*, Aug. 1997, pp. 26–27.
7. T. Karagiannis, M. Molle, M. Faloutsos, and A. Broido, "A nonstationary Poisson view of Internet traffic," in *IEEE INFOCOM*, 2004.
8. P. Haga, K. Diriczi, G. Vattay, and I. Csabai, "Understanding packet pair separation beyond the fluid model: The key role of traffic," in *IEEE INFOCOM*, 2006.
9. X. Yu, J. Li, Y. Chen, and C. Qiao, "Traffic statistics and performance evaluation in Optical Burst Switching networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 22, no. 12, pp. 2722–2738, 2004.
10. J. A. Hernández, J. Aracil, V. López, and J. E. López de Vergara, "On the analysis of burst-assembly delay in OBS networks and applications in delay-based service differentiation," to appear in Photonic Network Communications.