ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Optical Switching and Networking ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Optical Switching and Networking

www.elsevier.com/locate/osn

Optimizing offset times in Optical Burst Switching networks with variable Burst Control Packets sojourn times

A.E. Martínez*, J. Aracil, J.E. López de Vergara

Networking Research Group, Escuela Politécnica Superior, Universidad Autónoma de Madrid. Ciudad Universitaria de Cantoblanco, 28049 Madrid, Spain

Received 19 May 2006; accepted 25 April 2007

Abstract

In this paper, we consider the case of an Optical Burst Switching (OBS) network where the Switch Control Units (SCU) *do not work at the peak rate*. As a consequence, some Burst Control Packets (BCPs) will have to wait in queue to be processed, and then the BCP sojourn time will be *variable*. On the contrary, the optical burst does not leave the optical domain and the delay suffered is close to the propagation delay. Hence, chances are that the BCP arrives *late* to a given switch and, in that case, the optical burst will be dropped. We propose a Load-adaptive Offset Time algorithm (LOT) that takes into account the BCP variable sojourn time for the offset time calculation. The algorithm performs on-line calculation of the Discrete Fourier Transform (DFT) of the BCPs waiting time pdf. Our findings show that this procedure is very efficient both in terms of bandwidth usage and processing load. For example, considering a Gaussian service time for the BCPs, it turns out than less than 45 coefficients are necessary to calculate the offset time for a SCU utilization factor larger than 0.1.

Keywords: Optical networks; Optical Burst Switching (OBS); Burst Control Packet (BCP); Switch Control Unit (SCU); Variable sojourn time; Offset time calculation; Load-adaptive Offset Time (LOT)

1. Introduction and problem statement

Optical Burst Switching (OBS) [1] is a transfer mode that is halfway between circuit switching and packet switching, thus providing intermediate switching granularity. The basic transmission unit is an optical burst, which is composed of several IP packets. A Burst Control Packet (BCP) is transmitted out of band an offset time prior to the transmission of the optical burst. Such BCP follows the same route as the burst, and serves for the switching matrix to be properly configured by the time the burst arrives. Note that the BCP suffers optoelectronic conversion, since the Switch Control Unit (SCU) is electronic. The same does not apply to the optical burst, that is swiftly transmitted without leaving the optical domain. Central to the concept of OBS is the offset time between optical burst and BCP, which will be denoted by δ. Such time separation must be long enough to allow configuration of all switches along the path from source to destination. It must be noted that the burst scheduling algorithms may require non-negligible execution time,

* Corresponding address: Computer and Telecommunications Engineering, Escuela Politicnica Superior, Universidad Autonoma de Madrid, Ciudad Universitaria de Cantoblanco, Calle Francisco Tomas y Valiente, 11 28049 Madrid, Spain. Tel.: +34 914972261; fax: +34 914972235.

E-mail address: antonio.martinez@uam.es (A.E. Martínez).

1573-4277/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.osn.2007.04.002

especially for large cross-connects. On the other hand, the optical burst is stored at the network edges for a time equal to δ. Thus, there is a trade-off between delay at the network edges and processing time that is left to the SCUs to process the BCP. Large values of δ are advantageous for the SCUs since they have more time to schedule the incoming burst. However, large values of δ also imply higher end-to-end delay and larger buffers at the ingress nodes. The value of δ is usually made equal to the number of switches along the route multiplied by the processing time at each switch. Thus, knowledge of the route is required in advance (source routing). If source routing is not adopted, and the route is not known in advance, then the maximum number of switches per route (network diameter) is considered in the calculation. While the latter approach ensures enough time separation between burst and BCP there is a penalty in end-to-end delay. Let us assume that the number of switches in the route is to K, which can be either equal to the exact number of hops (source routing) or to the network diameter in hops. Let Δ be the BCP sojourn time in a single switch. The usual approach is to consider that Δ is constant. As a result, $\delta = K\Delta$, i.e. the end-to-end BCP sojourn time is also assumed to be *constant*.

In order to quantify the value of Δ, one needs to consider that as bandwidth increases more processing speed is required at the SCU. A numerical example is proposed in [2]. If the minimum burst length is 1 ms and the OBS switching fabric has 64 input fibers, with 100 wavelengths each, the maximal number of BCPs that need to be processed is 6.4 million/s. In that case, the required processing time to avoid BCP queuing is equal to $\Delta = 156.25$ ns. However, in [3] the processing time is estimated in 2 μs. Such processing time includes the execution of the scheduling algorithm (Horizon [4], LAUC-VF [5], Min-SV [6]). Burst segmentation and preemption [7], if adopted, further complicate matters, since extra offset time is required. Needless to say, the lesser the value of Δ the more sophisticated the SCU becomes. Moreover, the foreseen technological evolution is toward providing more wavelengths per fiber and more fibers per switch. In order to keep the switch design cost-effective, the SCUs are not expected to operate at the peak rate. Thus, a BCP queue will build up and the processing time Δ will be *variable*. Note that Δ comprises queuing delay and service time. This paper is focused on the analysis of OBS performance under variable BCP sojourn time. This analysis is relevant for manifold reasons. First, dimensioning SCUs at the BCP peak rate is not cost-effective. Actually, optical networks are limited by the so-called electronic bottleneck. As the network bandwidth increases so does the processing speed that is required at the SCU, which is precisely the bottleneck. Secondly, in order to keep the SCU sojourn time constant one needs to upgrade the SCU processing power if the number of ports increases. Nevertheless, if the processing delay is allowed to be variable the switch scalability is greatly enhanced because the same SCU can serve a larger switch. In that case, the offset time should grow accordingly in order to adapt to longer sojourn times. However, the SCU processing speed remains unchanged. Third, a variable BCP sojourn time directly implies that the offset time has to be variable and adaptive to network load. Actually, if offset time is granted regardless of the actual network load some bursts will encounter a lightly loaded network and will thus have advantages over those that find worse load conditions. The former will experience better sojourn times at the SCUs. As a result, they will have a longer offset time in comparison with a burst that traverses a path with busier SCUs. Since longer offset times imply higher priority [1] this leads to unfairness for bursts that traverse loaded paths. One may argue that the processing delay contribution to the end-to-end delay is negligible, especially for long roundtrip time networks, but this is up to a certain extent only, since the processing capacity of the SCU is limited and the network bandwidth is growing steadily. From an engineering standpoint, the fact is that the SCU design is simpler and more cost-effective if it does not operate at the BCP peak rate. The network scalability is enhanced since bandwidth growth does not translate directly into electronic processing power growth. Furthermore, the relaxed processing capacity requirement gives scope for the implementation of new scheduling, preemption and segmentation algorithms. Such advantages motivate the analysis presented in this paper, that aims at providing insight on how to operate an OBS network with variable processing times. The overall objective of the paper is to analyze and devise techniques for providing an offset time that is long enough to minimize burst blocking, yet small enough not to jeopardize end-to-end delay. To do so, the offset time is calculated adaptively to the SCU load along the path from source to destination.

Furthermore, the ingress node design will also benefit from an accurate calculation of the offset time. Note that the optical burst must be stored in a buffer for the duration of this offset time. Consequently, the longer the offset time, the larger the required buffer. The paper is structured as follows: Section 1.1 presents the state of the art. Section 2 is devoted to the analysis of burst loss probability. Since the processing time is now variable, the offset time can be dimensioned for a certain

Please cite this article in press as: A.E. Martínez, et al., Optimizing offset times in Optical Burst Switching networks with variable Burst Control Packets sojourn times, Optical Switching and Networking (2007), doi:10.1016/j.osn.2007.04.002

blocking probability objective, assuming that a burst that overtakes the corresponding BCP will be lost. In Section 3 we present an adaptive technique called LOT (Load-adaptive Offset Time) that dynamically adjusts the offset time to the current network state. Section 4 presents the conclusions and future work.

1.1. State of the art

There are related issues that support the research presented in this paper. However, to the best of our knowledge, none of them provides a technique to calculate the offset time for a case with BCP variable processing delay. This paper is motivated by the fact that the offset time calculation is a relevant issue. For example, if the switching speed is considered to be finite, offset times are very important. In [8], it is concluded that if the offset time approaches the maximum allowable delay for premium traffic, the burst assembly timeout must be very short. Then, the burst size becomes too small and the switching overhead due to the finite optical switching speed is comparable to the burst transmission time. The impact on the switch throughput is thus very significant. Nevertheless, if the offset time is decreased then the timeout can be increased. The burst size also increases and so does the throughput. The offset time calculation also becomes a very relevant issue when dealing with the scalability of SCUs. In [9], it is shown that both the SCU and the signaling channels have to be sized to process the amount of OBS traffic. It is stated that the loss of control packets (and associated bursts) due to saturation should be characterized, to be kept to a minimum. This implies that the loss of control packets actually happens, i.e. the SCU is not expected to operate at the peak rate and the BCP sojourn times are actually variable. Finally, an analytical expression for the burst end-to-end delay is given in [10], but the offset time value is not specified in the calculations. The algorithm described later in this paper is based on two pillars: (i) the sojourn time of the BCP on each SCU is variable, and (ii) the core switches should send state information to the edge nodes to calculate the offset time. The state of the art reveals a number of studies that support these assumptions.

1.1.1. Variable BCP sojourn time due to queuing delay

A case of variable BCP sojourn time due to network load is presented in [5]. It is stated that the maximum attainable throughput is upper bounded by $1/\xi$, being ξ the processing time of a BCP. If the offered load grows beyond such threshold the BCP will wait in queue, thus leading to variable delay in the SCU. To reduce the loss rate due to early burst arrivals, an Offset Time Compensation Algorithm is proposed in [11]. This algorithm is based on processing first those BCPs which are more delayed with respect to their initially scheduled offset times. The paper also presents the idea of dynamic offset times, which are periodically calculated based on the network load, using BCP delay estimation end-to-end. Burst losses can then be reduced by varying the offset time with this feedback information. However, an algorithm to obtain such dynamic offset times is not well formalized. A similar idea to process BCPs is proposed in [10], but with different traffic classes. In this case, a Fair Packet Queuing algorithm is used to provide fair bandwidth allocation. Moreover, the fact that SCUs may incorporate queues (thus making variable the sojourn time) has also been reported in [12], where bursts are stored in separate queues to perform scheduling in a DiffServ scenario. The experienced delay of low priority bursts depends on the high priority bursts workload. The above mentioned papers, however, do not provide a calculation of the offset time value based on the SCU load.

1.1.2. Variable processing time due to variable service time

The sojourn time can be variable due to waiting time in queue, but also due to the processing time, which is directly related to the scheduling algorithm used to allocate the incoming bursts. Such scheduling algorithms have different complexity order, which produces different processing times for each allocation. The complexity usually depends on the number of available voids, which is in turn dependent of the network load. In [13,14], a complexity analysis of scheduling algorithms is provided. It is shown that common algorithms such as LAUC-VF [5] have a complexity of $O(m)$, with m being the number of voids. Then, a set of geometry-based algorithms is proposed (*MinSV, MaxSV, MinEV, MaxEV*) that have a lower complexity ($O(\log(m))$), thus reducing the processing time due to scheduling. Such complexity orders serve to characterize the asymptotic behavior of the scheduler, that has a variable service time. In [15], prior scheduling algorithms are studied and it is stated that since OBS networks are open loop systems, they may often exhibit a worst-case performance. The same work also shows that the burst processing time is not constant, but a function of the scheduling algorithm and the traffic load. The use of processing time upper bounds (i.e. worst-case upper bounds) instead of average values is thus recommended. A new algorithm (VFO) is proposed

Please cite this article in press as: A.E. Martínez, et al., Optimizing offset times in Optical Burst Switching networks with variable Burst Control Packets sojourn times, Optical Switching and Networking (2007), doi:10.1016/j.osn.2007.04.002

which provides the best processing time upper bound. Several other papers are also focused on the scheduling algorithm's complexity order, thus supporting the fact that service time is variable. In [16] a slotted algorithm with $O(m/k)$ complexity is derived, where m is the number of elements in heads and ends arrays (heads and ends being the lists of the start and end times of the bursts whose reservation period falls within a slot), and k the number of slots per burst. This algorithm is particularly suitable for parallel implementations. In [17], rescheduling algorithms are proposed to achieve low computational complexity and high performance in order to avoid burst dropping, with an $O(w)$ improvement over LAUC-VF, being w the total number of wavelengths.

1.1.3. Transmission of the SCU load estimate

Our proposal is also based on the availability of a SCU load estimate to dynamically adapt the offset time. This implies that signaling messages will be exchanged between OBS switches, which will finally reach the edge burstifiers. In [18], some feedback schemes to send information back to the sources are mentioned, that serve to report a network load estimate to the source nodes. In [3], a *Link Scheduling State based Offset Selection* (LSOS) has been proposed that also requires the link states to be exchanged between routers and up to the network edges. The offset time is dynamically adjusted in order to ensure fairness (in the burst drop probability sense) regardless of the number of hops in the route. The results in [3] show the feasibility of such link state interchange if the network conditions remain stationary. However, the sojourn time in the SCUs is assumed constant. The variable delay assumption is thus a distinguishing feature of our work.

2. Burst blocking probability

In a variable delay scenario, the offset time value should be large enough for the BCP not to overtake the burst. In our analysis, we wish to isolate burst drop events caused by contention from burst drop events produced by insufficient offset. Thus, in what follows, it will be assumed that the burst drop probability is given by the probability that the burst actually overtakes the BCP. This implies that resources for accommodating the burst are always available. If this is not the case, then the burst drop probability derived in this section is a lower bound of the overall burst drop probability. Fig. 1 shows an example of a BCP overtaking its associated burst due to insufficient offset time.

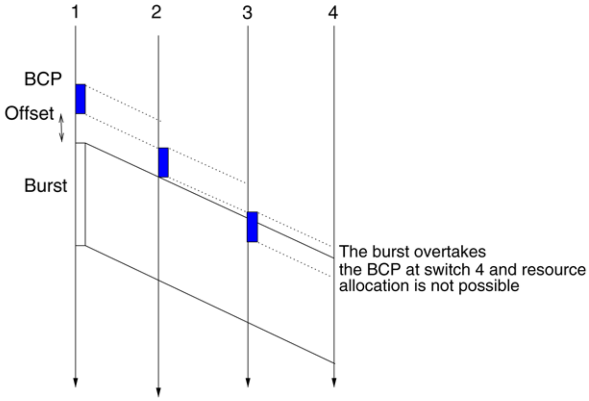
1
2
3
4
BCP
Offset
Burst
The burst overtakes the BCP at switch 4 and resource allocation is not possible

Fig. 1. A BCP overtakes the associated burst.

In Section 1.1, the state of the art was analyzed to show that the SCU sojourn time can be variable either due to queuing delay or variable service time delay (or both). In this section we analyze a case with constant service time and variable waiting time in queue. The constant service time assumption has also been suggested in [15], as a worst case upper bound. It will be assumed that the SCU incorporates a buffer, i.e., the SCU rate is not the peak BCP rate. Thus, a queue will build up that makes the BCP sojourn time in the SCU variable. For simplicity, it will be assumed that the average BCP service time equals unity (for example 1 μs, depending on the SCU processing speed). The BCP arrival at core OBS switches is the result of multiplexing traffic from many independent sources. As it has been shown in [19,20], under this condition this multiplexed traffic will approach to a Poisson process. On the other hand, the switch load will be assumed to be the same for all switches. The aim of this analysis is to study the strengths and drawbacks of a constant offset time versus a variable one. To this end, we consider a simple scenario with equally loaded switches. Actually, for identical deterministic BCP service times an expression for the convolution of waiting times of M/G/1 queues has been found in [21]. More specifically, the waiting time distribution $F_X(x) = P(X \le x)$ for K identical M/D/1 systems in tandem, with load ρ and service time unity is given by

$$F_X(x) = (1-\rho)^K \sum_{i=0}^{\lfloor x \rfloor} \sum_{j=0}^{K-1} \frac{(-1)^j}{i!j!} \times \binom{K+i-1}{i+j} (\rho(i-x))^{i+j} \, \mathrm{e}^{-\rho(i-x)}. \quad (1)$$

Let t_p refer to the offset time that guarantees a drop probability equal to p. Let us assume that the number of SCUs from source to destination is equal to K, with load

ρ and deterministic service time equal to unity, without loss of generality. Then,

$$t_p = K + F_X^{-1}(1-p) \tag{2}$$

where F_X is given by (1). In order to compare variable versus constant offset times, let $a \geq 1$ be a constant. For a constant offset time technique, it will be assumed that the offset time is given by

$$t_{\text{cons}} = aK. \tag{3}$$

Namely, the offset time value is set to a constant a multiplied by the number of hops in a given path, K. The K parameter can also be made equal to the network diameter (in the number of hops sense) [1]. It is important to remark that the SCU service time is made equal to unity. Thus, a can be interpreted as an amplification factor to account for queuing at the SCU. The constant offset time approach is common in the literature [1,5,22]. Fig. 2 shows the offset time values with variable and constant offset time policies in a light load scenario ($\rho = 0.1$), for a burst loss probability objective of 0.01 and number of hops from 1 to 10 in the x-axis. The same time units are used for both offset and SCU service time. The variable offset time curve shows the offset time to be allocated for a burst loss probability equal to 0.01. On the other hand, the constant offset time policy is evaluated with $a = \{1, 2\}$. If $a = 1$ then the burst loss probability is equal to $1 - (1-\rho)^K$ since the burst will not overtake its corresponding BCP if and only if the SCU queues are empty. For $\rho = 0.1$ and $K = 5$ this probability equals 0.409, i.e. more than 40% of the bursts will be lost. Furthermore, there is no way to decrease the blocking probability below 40%. For a case with $a = 2$, note that the burst loss probability is less than 0.01 because a longer offset time is provided in comparison to the variable offset time curve. However, this is at the expense of a substantially higher delay. Interestingly, the variable offset time curve shows sublinear increase. Indeed, if we let $X_1, \ldots, X_K$ the sojourn times at switches $1, \ldots, K$ then the tail probability of the end-to-end sojourn time

$$\begin{aligned} P(X > a) &= P(X_1 + \cdots + X_K > a) \\ &\leq \frac{E\{X_1 + \cdots + X_K\}}{a} = \frac{KE\{X_1\}}{a} \end{aligned} \tag{4}$$

for all $a > 0$ by Markov's inequality. It has been assumed that the average sojourn time at each switch is the same ($E\{X_1\}$). Fig. 3 shows the same curves for high load $\rho = 0.7$. Note that in this case a constant offset policy provides a burst loss probability which is higher than the objective value 0.01, even if $a = 2$, i.e. the SCU service time is doubled in the offset time calculation. Actually, in order to match the variable offset time performance, a should be set to seven, approximately. As the number of hops increases, the end-to-end delay becomes very large for the constant offset time scheme.

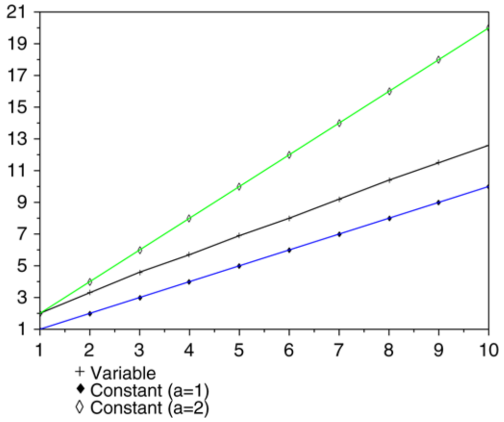

Fig. 2. Offset time value for 0.01 loss probability with light load ($\rho = 0.1$).

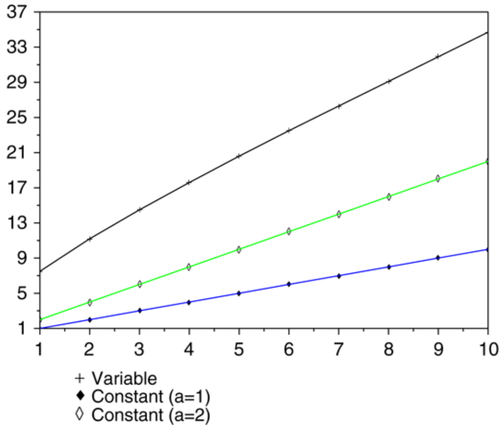

Fig. 3. Offset time value for 0.01 loss probability with high load ($\rho = 0.7$).

3. The Load-adaptive Offset Time (LOT) algorithm

In the previous section, we showed that the use of a variable offset time, adaptive to SCU load, provides considerable savings in end-to-end delay, for a given blocking probability objective. In this section we present the Load-adaptive Offset Time (LOT) algorithm, which provides an implementation of an adaptive offset time scheme. The LOT algorithm is based on the calculation of a given percentile of the end-to-end delay distribution. If the offset time is set to that particular value then the burst dropping probability

Please cite this article in press as: A.E. Martínez, et al., Optimizing offset times in Optical Burst Switching networks with variable Burst Control Packets sojourn times, Optical Switching and Networking (2007), doi:10.1016/j.osn.2007.04.002

is given by the percentile probability through (2). The proposed LOT algorithm is adaptive to the SCU load (which, in turn, relates to network load). In what follows we will compare the LOT algorithm against constant offset times policies, i.e. when the offset time is calculated as a constant multiplied by the network diameter. We will also provide a comparison with the analytical expressions presented in the previous section.

3.1. LOT rationale and justification

The BCP end-to-end delay is a random variable that will be denoted by T. Let us consider that K SCUs are present in the path from source to destination. The BCP sojourn time at SCU i is a random variable T_i, which can be characterized by its probability density function (pdf) $f_{T_i}(t), t \geq 0, i = 1, \dots, K$. The total BCP delay is then equal to the sum of random variables $T = \sum_{i=1}^{K} T_i$. Thus, the pdf of T results from the convolution

$$f_T = f_{T_1} * \cdots * f_{T_K}. \tag{5}$$

From this pdf the appropriate percentile can be derived in order to calculate the blocking probability, as in (2). However, note that an analytical expression for the pdfs $f_{T_i}, i = 1, \dots, K$ is very hard to obtain in practice. Assuming LAUC-VF scheduling is performed, for each BCP, the SCU will look for the smallest available void in the future such that the burst can be transmitted. This means that the SCU will perform a search operation per BCP that depends on the number of available voids in the output port. The BCP service time is related to the search algorithm that is employed and also to the output port load, since the number of available voids depends on the traffic load. As a conclusion, in order to derive the end-to-end sojourn time pdf an estimation of the individual pdfs is required. Then, the convolution (5) can be calculated. Estimation of the pdf and subsequent calculation of the convolution (5) can be performed using Fourier transform techniques, as explained in [23]. The proposed procedure comprises the following steps:

(1) *Estimation:* The individual sojourn time pdf $f_{T_i}, i = 1, \dots, K$ is estimated through on-line measurements at each SCU.
(2) *Sampling*: The individual pdf f_{T_i} is sampled at a sufficient rate to obtain a good approximation. If the pdf is bandwidth limited the Nyquist criterion can be applied to obtain the sampling rate.
(3) *Discrete Fourier Transform (DFT) calculation*: The DFT of the sampled pdf is calculated, yielding N coefficients.

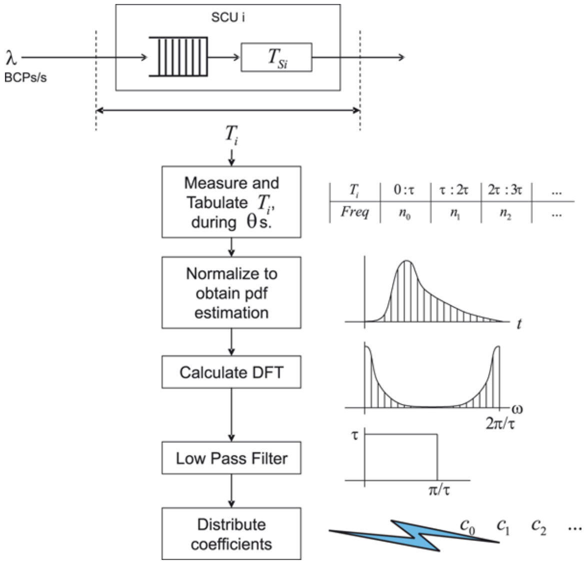
SCU i
λ
BCPs/s
T_{Si}
T_i
Measure and Tabulate T_i, during θs.
T_i | 0:τ | τ:2τ | 2τ:3τ | ...
Freq | n_0 | n_1 | n_2 | ...
Normalize to obtain pdf estimation
t
Calculate DFT
ω
2π/τ
τ
Low Pass Filter
π/τ
Distribute coefficients
c_0 c_1 c_2 ...

Fig. 4. LOT algorithm at the SCU.

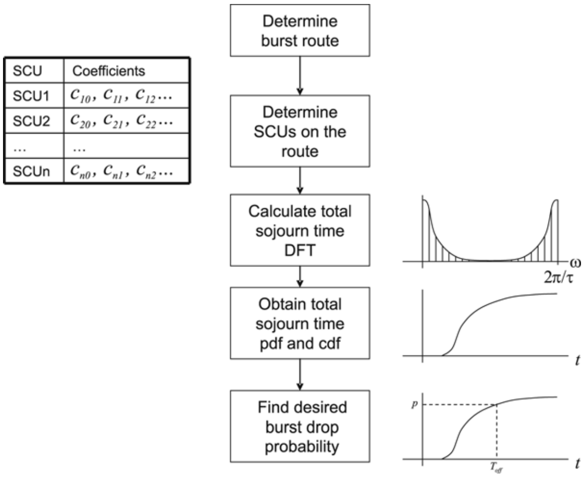
SCU | Coefficients
SCU1 | $c_{10}, c_{11}, c_{12}\dots$
SCU2 | $c_{20}, c_{21}, c_{22}\dots$
... | ...
SCUn | $c_{n0}, c_{n1}, c_{n2}\dots$
Determine burst route
Determine SCUs on the route
Calculate total sojourn time DFT
ω
2π/τ
Obtain total sojourn time pdf and cdf
t
Find desired burst drop probability
p
T_{off}
t

Fig. 5. LOT algorithm at the edge burstifier.

(4) *Low-Pass filtering*: Not all the DFT coefficients have the same information to rebuild the original pdf. LOT selects those that are enough to rebuild the original pdf properly.
(5) *Sending coefficients to the edge nodes*: The DFT coefficients are transmitted to the edge burstifier. Note that only a limited number of coefficients are required.
(6) *Convolution*: The DFT of the convolution is obtained at the edge burstifier by multiplying the individual DFT coefficients. Then, the DFT is inverted to obtain the total sojourn time pdf.

LOT is based on an estimation of the sojourn time pdf and is thus adaptive to network load. Estimation

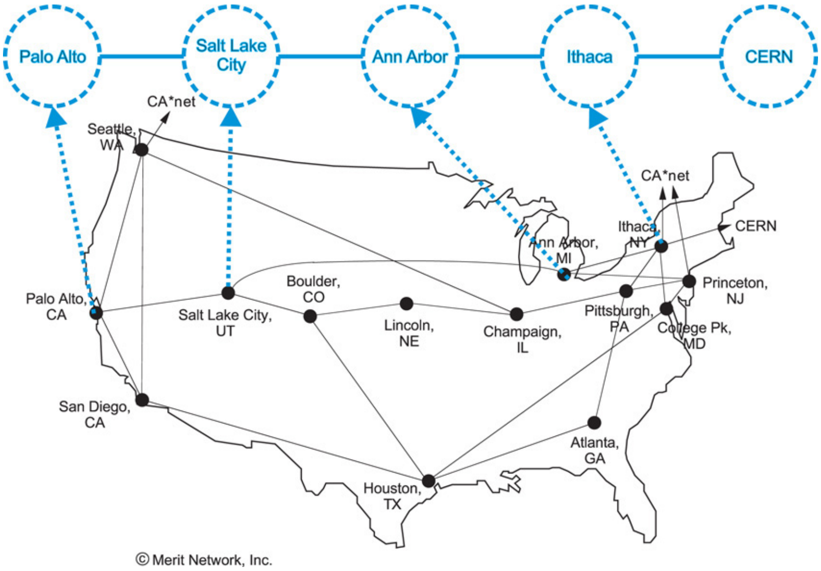

Fig. 6. NFSNet T1 Backbone and the path selected for the simulation.[1]

and sampling can be performed in a single step since the sojourn time samples can be grouped into bins beforehand. Namely, whenever a sojourn time sample is received the corresponding bin counter is increased, thus providing a discrete version of the pdf amenable for the calculation of the DFT coefficients. In fact, the DFT coefficients can be calculated on-line as the sojourn time samples are received. If the bin width is equal to τ s then the maximum DFT cutoff frequency is equal to π/τ. Fig. 4 shows the LOT algorithm block diagram at the SCU.

At the edge burstifier, the Fourier coefficients of the corresponding pdfs F_{T_i} are used to perform the convolution (5). Then, the resulting Fourier transform is inverted using an inverse DFT algorithm and the corresponding percentile for the end-to-end sojourn time is obtained. Fig. 5 shows the LOT algorithm block diagram at the edge burstifier.

3.2. *LOT performance evaluation*

To assess LOT performance, a simulation model has been implemented that incorporates the LOT algorithm at each SCU, which is depicted in Fig. 7. This system comprises five SCUs in order to mimic the maximum number of hops that a packet suffers from any source node in the NSFNET T1 Backbone [24] to the CERN node, under the shortest path criterion, as it can be seen on Fig. 6. As for the SCU processing time, i.e. the processing time per burst, not the sojourn time, we consider both a deterministic case and a Gaussian case. The former corresponds to a SCU which has the same service time per burst, as it was analyzed before. Such service time can be set, for example, to a processing time upper bound. The second case considers non-deterministic service times, assuming that their distribution is Gaussian. Each SCU receives traffic not only from the downstream SCU in the path under analysis but also from other adjacent SCUs (cross traffic), and from the edge node that assembles the local traffic. This is consistent with the fact the core OBS switches will multiplex traffic from a large number of sources. For this reason, the traffic at each OBS switch can be considered to follow a Poisson process.

3.2.1. *Deterministic SCU service time*

In this section we provide an evaluation of LOT assuming that the SCU service time is deterministic and the BCP arrivals follow a Poisson distribution, i.e. the SCU can be modeled as an M/D/1 system. From the LOT perspective, this is a *worst-case analysis, since the pdf is not bandwidth-limited*. On the other hand, goodness-of-fit to the exact analytical distribution will be assessed (refer to (1)). Two load scenarios have been evaluated: (i) *light load* with SCU utilization factor $\rho = 0.1$ and (ii) *high load* with SCU utilization factor $\rho = 0.7$. For each scenario we obtain the offset time for a given blocking probability objective with both LOT and the exact analytical expression (2). Fig. 8 shows the burst offset value (y-axis) normalized to the

[1] Map source: NSFNET postscript maps from ftp://ftp.uu.net/inet/maps/nsfnet/.

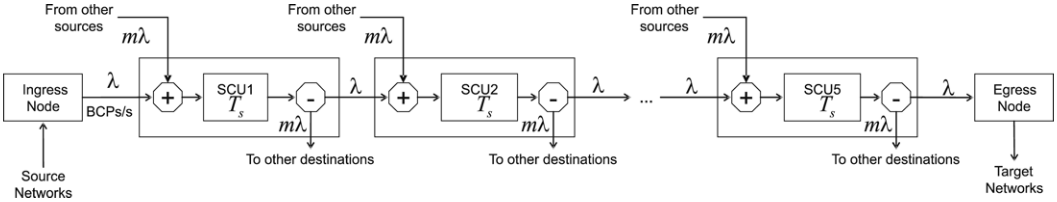

Fig. 7. Simulated system to test the LOT algorithm.

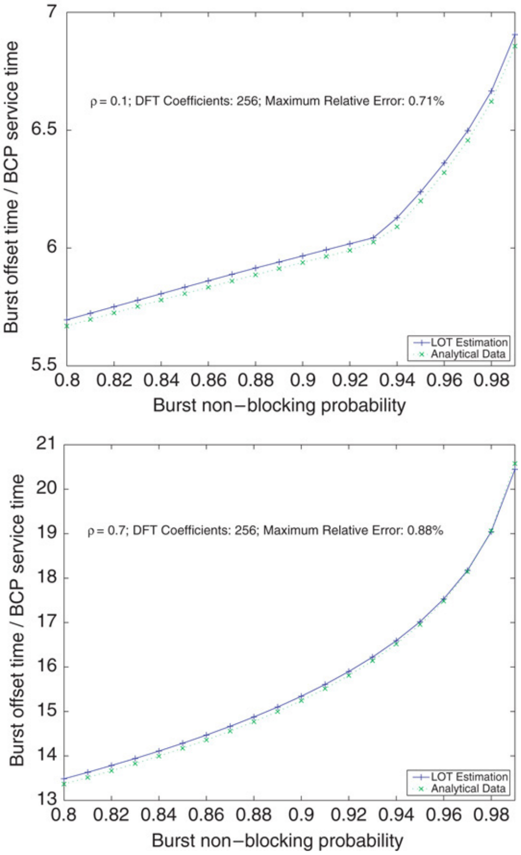

Fig. 8. Offset value versus blocking probability (LOT and analytical (2)) for light load -top- and high load -bottom- for a deterministic SCU.

BCP service time for a given burst blocking probability objective (x-axis) for light load (top) and heavy load (bottom). The number of pdf bins and DFT coefficients is equal to 512. The number of sojourn time samples is equal to 10,000. Note that the LOT algorithm matches the analytical results very closely. The number of DFT coefficients used to obtain Fig. 8 is equal to 512, that is equal to the number of bins of the original SCU sojourn time pdf. As the DFT of the pdf is even, only 256 coefficients have to be sent from each SCU to the edge burstifier. However, as each coefficient is a complex number, the total amount of information to be sent is the same as sending the original pdf samples. As not all the DFT coefficients provide the same information to reconstruct the original pdf, it is possible to truncate the DFT. Then, only those coefficients that are needed to obtain an accurate reconstruction of the original pdf are sent. To do so, we must take into account (i) a criterion to determine the number of coefficients to send, and (ii) a criterion to determine the goodness of fit to the pdf obtained with them. In order to determine the number of coefficients to send, we choose the *power spectral density criterion*. Considering the pdf as an electrical signal, the square of its DFT represents the power spectral density function. Then, the number of DFT coefficients to be sent is given by the power percentage of the original pdf that they cover. In our simulation experiments, the number of coefficients covers from the 10% to the 100% of the total power in steps of 10%, for each utilization factor. Thus, ten approximations to the total sojourn time pdf are obtained. To evaluate the quality of the approximations, we used a goodness of fit measure to the analytical distribution, given by (1). In fact, if the reconstructed end-to-end delay distribution at the edge matches the analytical expression (1) then the offset time will be closed to the theoretical optimum which is given by (2), for a given blocking probability objective. We choose the chi-square test as a goodness of fit measure, and a pdf is considered good if it passes the test with a significance level of 1%. Among the ten different pdfs that were obtained for each utilization factor, with different number of coefficients, we chose the one that passes the test with the minimum number of DFT coefficients. Fig. 9 shows the average, among the five SCUs simulated, of the minimum number of coefficients to match this condition vs. the SCUs service factor. As it was explained previously, the deterministic SCU is a worst case for the LOT algorithm due to the fact that *the pdf is not bandwidth-limited*. As a graphical example, Fig. 10 shows the discretized sojourn time pdf, with 512 bins and utilization factor equal to 0.7. The figure shows the abrupt changes in the pdf, which translate into high frequency components. For this reason, for low utilization factors, all the DFT coefficients are required to satisfy the chi-square

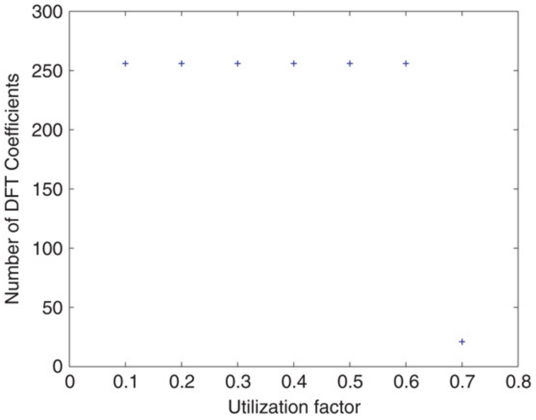

Fig. 9. Variation of DFT coefficients required to pass the chi-square test with a significance level of 1% vs. SCU utilization factor in a deterministic SCU.

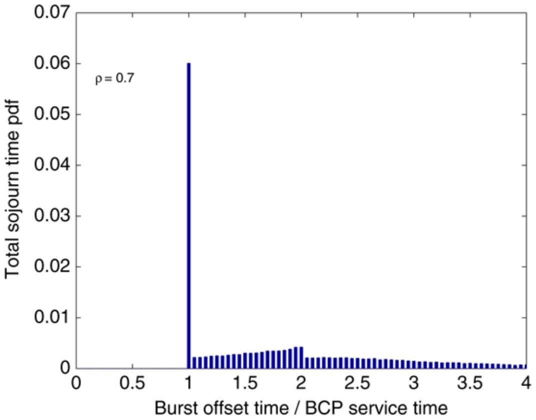

Fig. 10. Fragment of the sojourn time discretized pdf at the first SCU (512 bins, deterministic service time, $\rho = 0.7$).

criterion. However, as the utilization factor increases, the pdf tends to "enlarge" from the original delta impulse and becomes less abrupt. Then, the number of necessary DFT coefficients to represent the original pdf decreases, as the spectral power is more concentrated toward lower frequencies.

3.2.2. Gaussian SCU service time

When the BCP sojourn time is not assumed to be constant, the sojourn time pdf is smoother, and the aliasing effect is reduced with an appropriate value of the sampling parameter τ. For example, the sojourn time discretized pdf for Gaussian-distributed BCP services times, for an utilization factor of 0.7, is depicted in Fig. 11. The BCP service time mean is equal to 1 and the standard deviation is equal to 0.1. We note that now the pdf can be considered *bandwidth limited*. It turns out that, under all the utilization factor values considered, the number of coefficients that cover the 90% of the total pdf power is less than 45. In this case, no closed analytical expression can be found for the end-to-end sojourn time pdf, because numerical inversion of the end-to-end delay moment generating function is required. Thus, we use the total sojourn time pdf obtained from the simulation. Applying the same process as in the deterministic case, we obtain the number of coefficients needed to reconstruct the total sojourn time pdf. The average number of coefficients vs. the utilization factor is shown on Fig. 12. To obtain this figure we discarded those cases where the invert transform, due to the small amount of coefficients used to calculate them, presented negative values that were relevant. As it can be seen on the figure, the number of coefficients is reduced in comparison to the deterministic case. A maximum of only 44 coefficients are required for all utilization factors. Fig. 13 presents

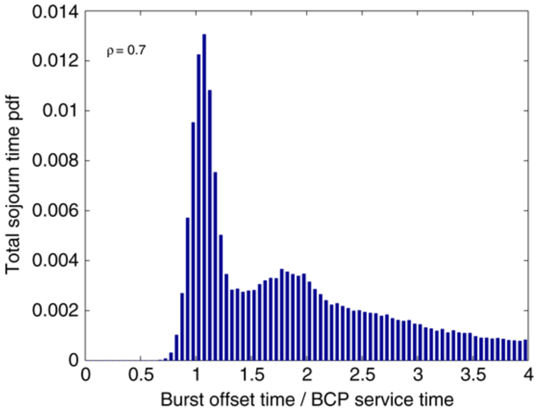

Fig. 11. Fragment of the sojourn time discretized pdf at the first SCU (512 bins, $N(1, 0.1)$ service time, $\rho = 0.7$).

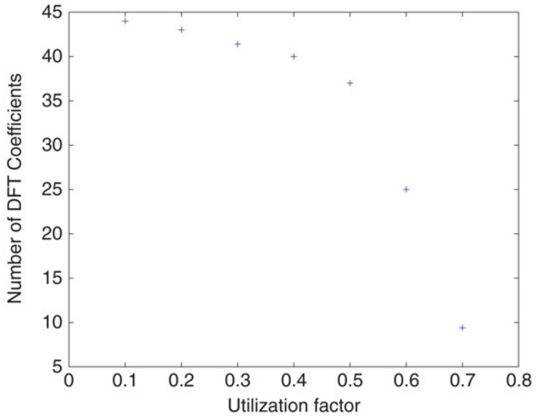

Fig. 12. Variation of DFT coefficients required to pass the chi-square test with a significance level of 1% vs. SCU utilization factor in a $N(1, 0.1)$ SCU.

Please cite this article in press as: A.E. Martínez, et al., Optimizing offset times in Optical Burst Switching networks with variable Burst Control Packets sojourn times, Optical Switching and Networking (2007), doi:10.1016/j.osn.2007.04.002

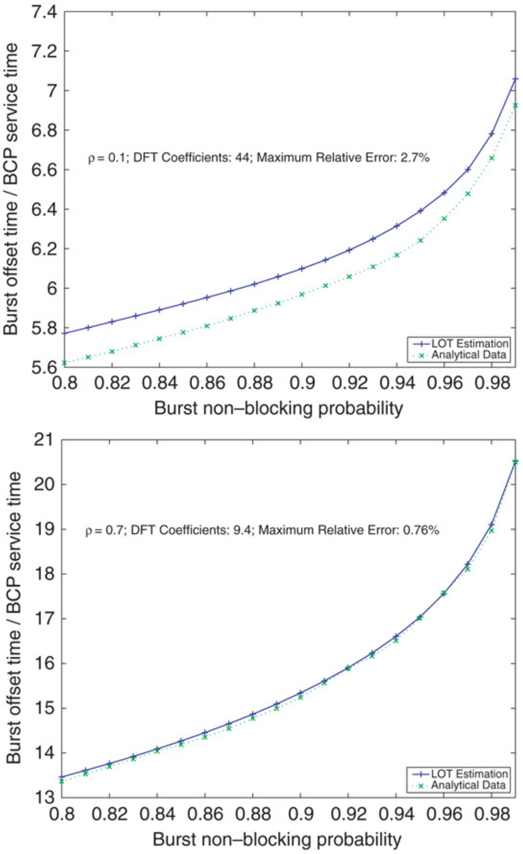

Fig. 13. Offset value versus non-blocking probability (LOT and simulation) for light load -top- and high load -bottom- for a N(1,0.1) SCU.

the estimated offset times values normalized to the SCU service time versus the non-blocking probability using the minimum number of DFT coefficients, compared with the simulation results. Even though the number of coefficients is very small, the offset estimation is very accurate, showing a maximum relative error (compared to the simulation results) of 2.7% in the worst case (light load).

3.3. Adaptability to load changes

It can be argued that the performance of LOT is highly dependent on the sojourn time stationarity. In this section we evaluate the dynamic behavior of LOT against load changes. The first step of LOT is the estimation of the sojourn time pdf. For this purpose, each SCU will record the sojourn time for each BCP, namely the time elapsed from the BCP arrival to the queue until the BCP leaves the SCU. This data collection is performed during θ s. On the one hand, θ has to be long enough to ensure that the number of sojourn time samples recorded is sufficient to have a good approximation to the pdf. On the other hand, θ has to be short enough to ensure that no significant load changes occur during the data collection, namely to ensure that the sojourn times come from a stationary distribution. Let us consider the event $A_k = \{\text{BCP sojourn time in the range}[k\tau, (k+1)\tau]\}$, with $k = 1, \ldots, M$, being M the number of bins and τ the bin width. Let $p_k = P(A_k), k = 1, \ldots, M$. Let n be the total number of samples collected during θ, and n_k the number of occurrences of A_k during θ, the ratio $\overline{p}_k = n_k/n$ is the point estimate of p_k. According to [23], for large n the sampling distribution of p_k is $N(p_k, \sqrt{p_k(1-p_k)/n})$. Under these conditions, the confidence interval of p_k with a confidence coefficient γ can be approximated by

$$\overline{p}_k \pm z_u \sqrt{\frac{\overline{p}_k(1-\overline{p}_k)}{n}} \tag{6}$$

where z_u is the u percentile of the standard normal density, and u is derived from the confidence coefficient by the expression $u = 1 - (1-\gamma)/2$. We can use (6) to determine the number of samples we have to collect to obtain an accurate estimation of p_k. Let us choose a confidence interval width which is smaller than a fraction α of the mean $\overline{p}_k$. Then, the following expression for n can be obtained from (6)

$$n \geq \frac{4z_u^2}{\alpha^2}\left(\frac{1}{\overline{p}_k} - 1\right). \tag{7}$$

We can calculate n to have the desired accuracy for the most representative bins on the pdf. As the number of bins is large (512 in our case), the probability p_k is very small in most of them. Thus, we restrict the condition given by (7) to bins with $p_k \geq 0.001$. Setting $\alpha = 1/3$ and $\gamma = 0.9$, we obtain as a result that $n \geq 97{,}302$ samples. Assuming, for example, that the SCU is dealing with 32 wavelengths, each one carrying a STM-64 signal (9,621,504 Kbps payload), being the average burst size equal to 160 Kbits, and the fiber utilization factor equal to 0.5, it turns out that the BCP arrival rate to the SCU is 962,150.4 BCPs/s. Thus, the time interval θ for our previous estimation is as low as 101.13 ms. We believe this time interval is short enough to ensure that the load does not change significantly. Additionally, when the load increases more sojourn time samples are collected in the time interval θ and the pdf estimation accuracy also increases. On the contrary, when the traffic decreases, larger values of θ are needed. Accordingly, LOT will increase the value of θ, but in the transient LOT will be tracking a pdf with higher network load. Thus, LOT provides an offset time that is always conservative with the target burst drop probability.

4. Conclusions and future work

The case of OBS networks with variable sojourn time at the SCUs is considered in this paper. An adaptive offset time algorithm (LOT) is proposed that takes into account the SCU load and applies to all service time distributions. LOT allows us to attain a blocking probability objective, due to overtake of burst with respect to BCP, with variable BCP sojourn times at the SCUs. Two different BCP service time distributions are assumed: constant service time, that can be considered a worst case analysis, and Gaussian service time. In both cases, LOT produces offset estimations that are very close to the actual sojourn time experienced by the BCPs, with minimal computational cost and bandwidth consumption.

From the present results, our future work is related with two main areas. The first one is the analysis of the characteristics of the BCP service time at the SCU for the different scheduling algorithms that are commonly used. This study will provide information on the kind of pdfs that will be present on the SCUs and the number of coefficients that will be needed in a real case to accurately represent them. The second area deals with the determination of the dynamic LOT behavior when the network load changes. This work will include the determination of the moments when a SCU has to resend information about its sojourn time probability distribution.

References

[1] C. Qiao, M. Yoo, Optical burst switching (obs) — A new paradigm for an optical internet, Journal of High-Speed Networks 8 (1).

[2] X. Yu, J. Li, X. Cao, Y. Chen, C. Qiao, Traffic statistics and performance evaluation in optical burst switched networks, IEEE/OSA Journal of Lightwave Technology 22 (11) (2004) 2722–2738.

[3] S.K. Tan, G. Mohan, K.C. Chua, Link scheduling state information based offset management for fairness improvement in wdm optical burst switching networks, Computer Networks 45 (6) (2004) 819–834.

[4] J.S. Turner, Terabit burst switching, Journal of High Speed Networks 8 (1) (1999) 3–16.

[5] Y. Xiong, M. Vandenhoute, H.C. Cankaya, Control architecture in optical burst-switched WDM networks, IEEE Journal on Selected Areas in Communications 18 (10) (2000) 1838–1851.

[6] J. Xu, C. Qiao, J. Li, G. Xu, Efficient channel scheduling algorithms in optical burst switched networks, in: IEEE Infocom, 2003.

[7] V.M. Vokkarane, J.P. Jue, Prioritized burst segmentation and composite burst-assembly techniques for qos support in optical burst-switched networks, IEEE Journal on Selected Areas in Communications 21 (7) (2003) 1198–1209.

[8] N. Barakat, E. Sargent, On optimal ingress treatment of delay-sensitive traffic in multi-class obs systems, in: Proc. 3rd International Workshop on Optical Burst Switching, WOBS 2004, 2004.

[9] K. Sriram, D.W. Griffith, S. Lee, N.T. Golmie, Optical burst switching: Benefits and challenges, in: Proc. First International Workshop on Optical Burst Switching, WOBS 2003, 2003.

[10] A. Kaheel, H. Alnuweiri, Quantitative qos guarantees in labeled optical burst switching networks, in: Proc. Global Telecommunications Conference GLOBECOM '04, vol. 3, 2004, pp. 1747–1753.

[11] I.-Y. Hwang, J.-H. Ryou, H.-S. Park, Offset-time compensation algorithm — qos provisioning for the control channel of the optical burst switching network, Lecture Notes in Computer Science 3391 (2005) 362–369.

[12] K. Long, R.S. Tucker, C. Wang, A new framework and burst assembly for ip diffserv over optical burst switching networks, in: Proceedings of Globecom 2003, 2003, pp. 3159–3164.

[13] J. Xu, C. Qiao, J. Li, G. Xu, Efficient channel scheduling algorithms in optical burst switched networks, in: Proc. IEEE INFOCOM 2003, 2003.

[14] J. Xu, C. Qiao, J. Li, G. Xu, Efficient burst scheduling algorithms in optical burst-switched networks using geometric techniques, IEEE Journal on Selected Areas in Communications 22 (9) (2004).

[15] J. Li, C. Qiao, J. Xu, D. Xu, Maximizing throughput for optical burst switching networks, in: Proc. INFOCOM 2004, 2004.

[16] M. Phùng, K. Chua, G. Mohan, M. Motani, T. Wong, P. Kong, On ordered scheduling for optical burst switching, Computer Networks 48 (6) (2005) 891–909.

[17] S.K. Tan, G. Mohan, K.C. Chua, Algorithms for burst rescheduling in wdm optical burst switching networks, Computer Networks 41 (1) (2003) 41–55.

[18] G. Thodime, V.M. Vokkarane, J.P. Jue, Dynamic congestion-based load balanced routing in optical burst-switched networks, in: Proc. IEEE Globecom 2003, vol. 5, 2003, pp. 2694–2698.

[19] X. Yu, Y. Chen, C. Qiao, Study of traffic statistics of assembled burst traffic in optical burst switched networks, in: Proceedings of Opticomm 2002, 2002, pp. 149–159.

[20] M. Izal, J. Aracil, On the influence of self similarity on optical burst switching traffic, in: Proceedings of GLOBECOM 2002, 2002.

[21] O. Osterbo, Models for end-to-end delay in packet networks, Tech. Rep. 4, Telenor research and development, 2003.

[22] Y.C.S. Sheesia, C. Qiao, Performance comparison of OBS and SONET in metropolitan ring networks, IEEE Journal on Selected Areas in Communications 22 (8) (2004) 1474–1482.

[23] A. Papoulis, S. Pillai, Probability, Random Variables and Stochastic Processes, McGraw-Hill, 2002.

[24] K.C. Claffy, G.C. Polyzos, H.-W. Braun, Traffic characteristics of the T1 NSFNET backbone, in: INFOCOM (2), 1993, pp. 885–892.

Please cite this article in press as: A.E. Martínez, et al., Optimizing offset times in Optical Burst Switching networks with variable Burst Control Packets sojourn times, Optical Switching and Networking (2007), doi:10.1016/j.osn.2007.04.002