Adaptive Optical Burst Switching

Thomas Bonald, Raluca-Maria Indre, Sara Oueslati

Adaptive Optical Burst Switching

Thomas Bonald
Telecom ParisTech
Paris, France
thomas.bonald@telecom-paristech.fr

Raluca-Maria Indre, Sara Oueslati
Orange Labs
Issy-les-Moulineaux, France
{ralucamaria.indre,sara.oueslati}@orange-ftgroup.com

***Abstract*—We propose a modified version of Optical Burst Switching (OBS) that adapts the size of switched data units to the network load. Specifically, we propose a two-way reservation OBS scheme in which every active source-destination pair attempts to reserve a lightpath and for every successful reservation, transmits an optical burst whose size is proportional to the number of active data flows. We refer to this technique as Adaptive Optical Burst Switching. We prove that the proposed scheme is optimal in the sense that the network is stable for all traffic intensities in the capacity region. We also evaluate the throughput and delay performance of adaptive OBS through both analysis and simulation in order to assess the practical load ranges at which the network may operate.**

***Index Terms*—Optical burst switching, random access, flow-level dynamics, stability, performance.**

I. Introduction

Ever increasing Internet traffic demand challenges the use of electronic switching in today's networks. The routing bottleneck can be alleviated by means of optical switching, which enables payload to be carried exclusively in the optical domain. In Wavelength Division Multiplexing (WDM) networks, a simple way of performing optical switching is by assigning to each source-destination (SD) pair a specific wavelength, a technique referred to as Optical Circuit Switching (OCS). While used in current IP over WDM networks to establish quasi-static, virtual point-to-point links, this technique is not scalable, since $O(N^2)$ wavelengths are required for N nodes, and does not adapt to the variations of the traffic matrix [1]. Future optical technologies must provide some form of dynamic *time sharing* of wavelength capacity so as to meet the demand of bursty traffic with a limited number of available wavelengths.

Proposed dynamic switching techniques differ with respect to the granularity of the switched data units, i.e., packets, bursts or flows. While conceptually ideal, Optical Packet Switching (OPS) is facing important technological challenges, such as the lack of optical random access memory and ultra-fast switching requirements, that question its viability in the near future [2]. By reserving optical resources on a much longer time scale, Optical Flow Switching (OFS) alleviates this issue but requires some form of traffic aggregation so as to improve wavelength utilization; the exact manner in which flows stemming from different users, starting and completing at different times, can be multiplexed and carried together remains an open issue [3], [4]. For these reasons, Optical Burst Switching (OBS) is generally considered as the most promising technology, as a feasible alternative to OPS without the flow-level traffic aggregation constraints of OFS.

In OBS networks, incoming IP packets are aggregated into optical bursts at the network edge before transmission. Most OBS architectures proposed to date rely on one-way reservation schemes [5], [6], [7]. Specifically, each optical burst is preceded by a control packet which is sent over a separate wavelength and processed electronically at each node in order to reserve the optical resources. The optical burst follows its control packet after an appropriate offset time without waiting for the confirmation of reservation. The major drawback of this technology is the high probability of burst collision it incurs, even when wavelength conversion is allowed at each node [8], [9]. In order to significantly reduce the burst collision probability, all OBS nodes need to provide full range conversion over a spectrum of around 100 wavelengths, a solution which is hardly feasible today and in the near future. Buffering the optical payload can only partially resolve contention due to the limited storage capacity of optical buffers implemented through fiber delay lines (FDL) [10], [11]. Techniques based on burst segmentation [12], [13] or deflection routing [14], [15] have also been proposed to alleviate contention but none is able to significantly improve performance.

Alternatively, burst collisions can be avoided by relying on a simple two-way reservation scheme, as explored in Wavelength-Routed OBS (WR-OBS) [16], [17]. Each optical burst must then wait for the confirmation of the reservation before entering the network. In such a network, the utilization of wavelength channels greatly depends on the ratio between the optical burst duration and the idle time, i.e. the time needed for the connection setup. For instance, if the two-way reservation scheme is combined with the so-called Just-In-Time (JIT) policy, the intermediate nodes are configured for the incoming burst immediately after the reception of the control packet. Since reservations have unspecified durations, JIT is easy to implement. However, by reserving resources for an unnecessarily long period of time, JIT is likely to incur very low utilization.

In an attempt to improve efficiency, the authors of [17] propose the Just-Enough-Time (JET) reservation policy, which aims at reducing the idle time. Under JET, reservations are delayed until the actual burst arrival and the resources are reserved only for the duration of the burst. JET suffers from several drawbacks. First, the size of each burst must be known

at the start of the reservation process and cannot be modified thereafter. As a consequence, packets that arrive during the connection setup cannot be appended to the optical burst. Next, JET requires each node to maintain complex reservation schedules; ensuring the accuracy of these schedules supposes network-wide synchronization, which is hardly feasible in large mesh networks. Finally, JET allows a single lightpath to be configured for each SD pair. In practice, it may be useful to configure multiple lightpaths per SD both to reduce the blocking probability and to quickly restore network connectivity in case of link failure.

In this paper, we propose a modified version of WR-OBS that is able to maximize the utilization of WDM channels while using simple JIT-based reservation. More precisely, we show that bandwidth utilization can be improved by simply adapting the size of the switched data units to the network load. While previous WR-OBS proposals create the optical burst based on deterministic parameters such as timers or size thresholds, the proposed scheme allows data units to be dynamically adapted to the traffic conditions. Specifically, each active SD pair attempts to reserve a lightpath and, once the reservation is successful, transmits a burst whose size is proportional to the *number of active flows*. We refer to this scheme as Adaptive Optical Burst Switching.

Adaptive OBS is sensitive to the traffic conditions: at low network load, it behaves like WR-OBS, bursts having some predefined, minimum size. As load increases, flows start to accumulate and the burst size increases proportionally, amortizing the reservation overhead and improving network utilization. Apart from proposing an adaptive variant of WR-OBS, the contributions of this paper are twofold. Firstly, we show that this simple adaptive scheme is in fact able to fully utilize the optical resources in the sense that it stabilizes the network for all traffic intensities in the capacity region. Secondly, we evaluate the throughput and delay performance of adaptive OBS through both analysis and simulation and derive the corresponding practical operational load ranges.

The rest of the paper is organized as follows. In the next section, we describe the proposed switching scheme. Sections III and IV present the stability analysis and the performance results, respectively. Section V concludes the paper.

II. Adaptive OBS

A. Network architecture

We consider a network of Wavelength Division Multiplex (WDM) links. Like in OBS networks, a specific wavelength is dedicated to the control plane, the corresponding traffic being processed electronically at each node. All other wavelengths are dedicated to the user plane; the corresponding traffic is optically switched, without any OEO conversion, from source to destination. The wavelength capacity is the scarce resource, not only because each fiber can carry a limited number of wavelengths, but because the complexity and cost of core and edge nodes grow with the number of wavelengths they support.

Edge nodes communicate with each other via optical bursts that may be routed through intermediate nodes and span multiple links. Several optical bursts can be simultaneously transmitted over the same link as long as they use different wavelengths. In the absence of wavelength converters, the optical bursts must use the same wavelength on all links on their path from the source to the destination. This wavelength continuity constraint can be relaxed if the optical switches are equipped with wavelength converters, that is devices that allow data to be switched from an incoming wavelength to a different outgoing wavelength.

In the described network, edge nodes must be equipped with one or several tunable transmitter(s), to be able to send one or several burst(s) on the appropriate wavelength(s), and with an array of fixed-tuned receivers, to be able to receive data on several wavelengths. Core network nodes are dynamic optical switches able to switch bursts over millisecond timescales. The optical switching fabric is reconfigured by an electronic control unit upon reception of a reservation request, as explained in the following.

B. Reservation scheme

Optical bursts are created at the edge of the network by assembling data packets as explained in detail in §II-C. When a source-destination (SD) pair has one burst ready for transmission, it becomes *active* and attempts to reserve an optical connection, we refer to as lightpath. Specifically, each SD pair has some predefined set of eligible paths in the network. At each reservation attempt, the source selects a subset of these paths and sends a *request* control packet on each of these paths. The request control packets collect the state of wavelengths on their way to the destination. Based on the data contained in the request control packets, the destination selects one of the available paths, if any. It then sends back a *reserve* control packet on the chosen path which is destined to reserve the optical resources at intermediate nodes. When the source receives the reserve control packet, it can immediately transmit data on the specified lightpath.

If no lightpath is available, the destination sends a *failure* control packet to inform the source of the occupancy of the optical resources; the source then reattempts a reservation after some random backoff time, imitating the Carrier Sense Multiple Access (CSMA) algorithm. Similarly, any SD pair that is still active after the transmission of an optical burst restarts the reservation process after some random backoff time. No time window is specified in the reservation process; the reserved resources are automatically released when the transmission of the optical burst is terminated. Each SD pair may also run several reservation processes in parallel so as to better exploit the optical resources.

Note that, due to the concurrent reservation processes of the SD pairs, the state of a link may change between the arrival of the request control packet of some SD pair and the reception of the associated reserve control packet, possibly causing the failure of the reservation. A failure control packet must then be transmitted by the corresponding optical node to both the source (to notify it) and the destination (to release the wavelengths reserved on downstream links, from that node

to the destination). We shall neglect this phenomenon of *backward blocking* in the following, most reservation failures being due to forward blocking, when request control packets find no available lightpath to the destination.

The described reservation process can be implemented via a signalling protocol such as RSVP-TE (Resource Reservation Protocol - Traffic Engineering), which has been standardized for the GMPLS (Generalized Multi-Protocol Label Switching) control plane. Configuring multiple paths per SD pair allows sources to quickly restore connectivity in case of link or node failure and to significantly reduce the blocking probability, as shown in Section IV.

C. Assembly mechanism

At each source, incoming data packets are electronically buffered according to their destination. These packets are then assembled into bursts that are characterized by some minimum size compatible with the switching capability of core optical nodes. Unlike conventional OBS, in which the size of the burst is insensitive to the traffic conditions, adaptive OBS allows the source to dynamically adjust the size of the burst to the network load. We use the number of active data flows as a measure of network congestion, as proposed in [18]. Specifically, the size of the burst sent by any SD pair is equal to the minimum burst size, say B, multiplied by the number of active data flows on this SD pair at the reception of the reserve control packet. A data flow here refers to any instance of application and is typically identified through the usual 5-uple of the IP header: source and destination IP addresses, source and destination ports, and protocol.

In principle, a data flow using some SD pair of the optical network may be considered as active as soon as it has at least one packet waiting for transmission in the corresponding buffer. This simple scheme would count all active flows, including voice-over-IP flows, http transfers and very short flows that do not contribute to the actual network load, as argued in [19] for instance. In practice, a minimum threshold on the number of buffered packets must be set to consider a data flow as active. For delay sensitive traffic, setting an appropriate threshold value is essential in order to limit queueing delays. In the following, we consider elastic data traffic only and consider a flow to be *active* as soon it has at least B bits in the buffer. We do not address the issue of QoS differentiation that may be enforced at the burst assembly, as proposed in [20] for instance.

III. Stability Analysis

Before analysing stability, we present the network model and the resource allocation achieved by adaptive OBS.

A. Network model

Let L be the number of links and W_l the number of data wavelengths of link l (excluding the control wavelength). There are K source-destination (SD) pairs in the network. Each SD pair k is characterized by some set of eligible paths in the network. Path j of SD pair k is defined by some subset of links, $p_{kj} \subset \{1, \dots, L\}$.

Any burst transmission requires the prior reservation of some path from the source to the destination. Each reservation takes one round-trip time, denoted by δ_{kj} for SD pair k on path j. We consider the general case where SD pair k runs N_k reservation processes in parallel and can thus transmit up to N_k bursts simultaneously, (possibly on the same path, using different wavelengths). The source must then be equipped with at least N_k tunable transmitters.

We consider two types of networks, depending on the technology of the underlying optical switches:

- Wavelength conversion: A lightpath can use any available wavelength on each link. In particular, there is no need to specify the allocated wavelengths. The network state at time t is then described by some vector $y(t)$ whose component kj corresponds to the number of lightpaths reserved for SD pair k on path j at time t. The capacity contraints are given by:

$$\forall l = 1, \dots, L, \quad \sum_{k,j:l \in p_{kj}} y_{kj}(t) \leq W_l. \tag{1}$$

- No wavelength conversion: A lightpath must use the same wavelength from the source to the destination. To ensure connectivity, we then assume that all links have the same number of wavelengths, denoted by W. The network state at time t is described by some vector $y(t)$ whose kjw component is equal to 1 if some lightpath is reserved for SD pair k on path j and wavelength w at time t, and is equal to 0 otherwise. We still denote by $y_{kj}(t) = \sum_{w=1}^{W} y_{kjw}(t)$ the number of lightpaths reserved for SD pair k on path j at time t. Since a wavelength cannot be allocated to more than one SD pair, the capacity constraints become:

$$\forall l = 1, \dots, L, \ \forall w, \quad \sum_{k,j:l \in p_{kj}} y_{kjw}(t) \leq 1. \tag{2}$$

In both cases, the total number of reserved lightpaths of SD pair k, say $y_k(t) = \sum_j y_{kj}(t)$, cannot exceed N_k. We denote by $\mathcal{Y}$ the set of feasible states, that satisfy this constraint and either (1) or (2), depending on the considered network.

Let R be the optical line rate of each wavelength, in bit/s. The average throughput of SD pair k when state y is selected with probability $\pi(y)$ is given by:

$$\phi_k = R \sum_{y \in \mathcal{Y}} \pi(y) y_k.$$

We denote by ϕ the corresponding vector and refer to the *capacity region* as the set of vectors ϕ generated by all probability measures π on the set $\mathcal{Y}$. This defines the set of all throughput vectors that can be allocated to the SD pairs, using some centralized scheme for instance. In the rest of the section, we prove that adaptive OBS is able to fully exploit this capacity region, despite its distributed nature.

B. Resource allocation

Let x_k be the number of active flows on SD pair k; the pair becomes *active* as soon as $x_k > 0$. Whenever active, source k runs N_k reservation processes in parallel. Each process attempts to reserve a lightpath after some exponential backoff time of parameter ν. For simplicity, we assume[1] that a single path is attempted at random. Specifically, path j is attempted with probability $\alpha_{kj} > 0$, with $\sum_j \alpha_{kj} = 1$. In the absence of wavelength conversion, we assume that a single wavelength is attempted at random. If the reservation is successful, source k sends a burst of length $x_k B$, where B denotes the minimum burst size (in bits); otherwise, it reattempts a reservation after a new exponential backoff time of parameter ν.

As mentioned above, we neglect the phenomenon of backward blocking. Specifically, we assume that the reservation of source k starting at time t is successful if and only if the vector $y(t)+e_{kj}$ satisfies the capacity constraints (1) in case of wavelength conversion, or the vector $y(t) + e_{kjw}$ satisfies the capacity constraints (2) in the absence of wavelength conversion, where w denotes the attempted wavelength and e_{kj}, e_{kjw} are the corresponding unit vectors of $\mathcal{Y}$. The network state then changes instantaneously at time t, the actual transmission starting at time $t + \delta_{kj}$ for $x_k\tau$ time units, where $\tau = B/R$ denotes the transmission time of a burst of minimum size.

Under the above assumption, the reservation processes behave as a multiclass loss network of Engset type with class-k customers representing the N_k reservation processes of SD pair k. The associate stationary measure in state x is given by [21]:

$$u(x,y) = \prod_{k=1}^{K} \frac{N_k!}{(N_k - y_k)!} \prod_j \frac{(\alpha_{kj}\nu(\delta_{kj} + x_k\tau))^{y_{kj}}}{y_{kj}!},\ y \in \mathcal{Y}.$$

We obtain the stationary distribution of the resource allocation y in state x by normalization:

$$\pi(x,y) = \frac{u(x,y)}{\sum_{z\in\mathcal{Y}} u(x,z)}. \tag{3}$$

By the insensitivity property [21], this stationary distribution is independent of the distribution of the backoff times beyond the mean, provided the latter has a continuous, infinite support.

C. Flow-level dynamics

We now assume that data flows arrive according to a Poisson process of intensity $\lambda_k > 0$ at SD pair k and have exponential[2] flow sizes of mean σ_k bits. We denote by $\rho_k = \lambda_k\sigma_k$ the traffic intensity of pair k in bit/s and by ρ the corresponding vector.

Let $x(t)$ be the network state (in terms of the number of flows on each SD pair) at time t. Assuming that the flow time-scale is much slower than the burst time-scale, the throughput of SD pair k in state x is given by:

$$\phi_k(x) = R \sum_{y\in\mathcal{Y}} \pi(x,y) \sum_j \frac{x_k\tau}{\delta_{kj} + x_k\tau} y_{kj}. \tag{4}$$

The network state $x(t)$ then corresponds to that of a system of K coupled queues with arrival rates λ_k and service rates $\phi_k(x)/\sigma_k$. We say that the network is *stable* if the underlying Markov process is ergodic, meaning that the number of active flows on each SD pair achieves a stationary regime. We have the following key result, showing the optimality of adaptive OBS in terms of resource allocation:

Theorem 1: The network is stable whenever the vector ρ of traffic intensities lies in the interior of the capacity region.

Proof: If the vector of traffic intensities lies in the interior of the capacity region, there exist some $\epsilon > 0$, and some probability measure π on $\mathcal{Y}$ such that:

$$\forall k = 1,\dots,K, \quad \rho_k = R(1-2\epsilon)\sum_{y\in\mathcal{Y}} \pi(y) y_k. \tag{5}$$

Note that we can choose $\pi(y) > 0$ for all $y \in \mathcal{Y}$.

Define:

$$F(x) = \sum_{k:x_k>0} x_k\sigma_k \log(x_k\nu\tau).$$

By Foster's criterion [22], the network is stable if there exists some $\alpha > 0$ such that the corresponding drift, given by:

$$\begin{aligned}\Delta F(x) &= \sum_{k=1}^{K} \lambda_k (F(x+e_k) - F(x)) \\ &+ \sum_{k:x_k>0} \frac{\phi_k(x)}{\sigma_k}(F(x-e_k) - F(x)),\end{aligned}$$

satisfies $\Delta F(x) \le -\alpha$ in all states x but some finite number.

Using the convention $0\log(0) \equiv 0$, we have:

$$\Delta F(x) = G(x) + \sum_{k:x_k>0} \rho_k(x_k+1)\log(1+\frac{1}{x_k})$$

$$+ \sum_{k:x_k>0} \phi_k(x)(x_k-1)\log(1-\frac{1}{x_k}) + \sum_{k:x_k=0} \rho_k \log(\nu\tau),$$

with:

$$G(x) = \sum_{k:x_k>0} (\rho_k - \phi_k(x))\log(x_k\nu\tau).$$

Using (4) and (5), we obtain:

$$\begin{aligned}G(x) = R \sum_{k:x_k>0} \sum_{y\in\mathcal{Y}} &((1-2\varepsilon)\pi(y)y_k \\ &- \pi(x,y)\sum_j \frac{x_k\tau}{\delta_{kj}+x_k\tau} y_{kj})\log(x_k\nu\tau).\end{aligned}$$

Let:

$$v(x,y) = \prod_{k:x_k>0} (x_k\nu\tau)^{y_k}, \quad y \in \mathcal{Y}.$$

Using the fact that:

$$\log(v(x,y)) = \sum_{k:x_k>0} y_k \log(x_k\nu\tau),$$

[1] It turns out that this simple access scheme is sufficient for optimality. More complex schemes attempting several paths simultaneously are expected to improve performance and, in particular, to be also optimal.

[2] This assumption makes the network state Markovian but is not essential for the subsequent stability analysis.

we get:

$$G(x) = H(x) + R \sum_{k:x_k>0} \sum_{y\in\mathcal{Y}} \pi(x,y) \sum_j \frac{\delta_{kj}}{\delta_{kj}+x_k\tau} y_{kj} \log(x_k \nu\tau),$$

with:

$$H(x) = R \sum_{y\in\mathcal{Y}} ((1-2\varepsilon)\pi(y) - \pi(x,y)) \log(v(x,y)).$$

We then need the following lemma.

Lemma 1: Let:

$$v(x) = \max_{y\in\mathcal{Y}} v(x,y).$$

Then, for all states x but some finite number,

$$\sum_{y\in\mathcal{Y}} \pi(x,y) \log(v(x,y)) \geq (1-\epsilon)\log(v(x)).$$

Proof: Since $y_k \leq N_k$ for all k and $\alpha_{kj} > 0$ for all k, j, there exist positive constants β and β' such that, for all states x and all $y \in \mathcal{Y}$, $\beta \leq u(x,y)/v(x,y) \leq \beta'$. Let:

$$\mathcal{Y}(x) = \left\{ y \in \mathcal{Y} : \log(v(x,y)) \geq (1-\frac{\epsilon}{2}) \log(v(x)) \right\}.$$

We have:

$$\sum_{y\in\mathcal{Y}} \pi(x,y) \log(v(x,y)) \geq (1-\frac{\epsilon}{2}) \log(v(x)) \sum_{y\in\mathcal{Y}(x)} \pi(x,y).$$

Moreover,

$$\begin{aligned}
\sum_{y\notin\mathcal{Y}(x)} \pi(x,y) &= \frac{\sum_{y\notin\mathcal{Y}(x)} u(x,y)}{\sum_{y\in\mathcal{Y}} u(x,y)}, \\
&\leq \frac{\beta'}{\beta} \frac{\sum_{y\notin\mathcal{Y}(x)} v(x,y)}{\sum_{y\in\mathcal{Y}} v(x,y)}, \\
&\leq \frac{\beta'}{\beta} \frac{(|\mathcal{Y}| - |\mathcal{Y}(x)|) v(x)^{1-\frac{\epsilon}{2}}}{\max_{y\in\mathcal{Y}} v(x,y)}, \\
&= \frac{\beta'}{\beta} \frac{(|\mathcal{Y}| - |\mathcal{Y}(x)|)}{v(x)^{\frac{\epsilon}{2}}}.
\end{aligned}$$

Since $v(x)$ tends to $+\infty$ when $|x| = \sum_{k=1}^K x_k$ tends to $+\infty$, this quantity is less than $\epsilon/2$ for all states x but some finite number. We deduce that in all states x but some finite number:

$$\begin{aligned}
\sum_{y\in\mathcal{Y}} \pi(x,y) \log(v(x,y)) &\geq (1-\frac{\epsilon}{2})^2 \log(v(x)), \\
&\geq (1-\epsilon)\log(v(x)).
\end{aligned}$$

■

In view of Lemma 1, we have for all states x but some finite number:

$$H(x) \leq -\epsilon R \sum_{y\in\mathcal{Y}} \pi(y) \log(v(x,y)) + (1-\epsilon) R \left(\sum_{y\in\mathcal{Y}} \pi(y) \log(v(x,y)) - \log(v(x)) \right).$$

Since $v(x,y) \leq v(x)$ for all states x, the second term is non-positive and we deduce that for all states x but some finite number:

$$H(x) \leq -\epsilon R \sum_{y\in\mathcal{Y}} \pi(y) \log(v(x,y)).$$

Since $\pi(y) > 0$ for all $y \in \mathcal{Y}$, this expression tends to $-\infty$ when $|x| = \sum_k x_k$ tends to $+\infty$. The differences $\Delta F(x) - G(x)$ and $G(x) - H(x)$ being upper bounded, we deduce that there exists $\alpha > 0$ such that $\Delta F(x) \leq -\alpha$ for all states x but some finite number. ■

IV. Performance results

This section is devoted to the performance analysis of adaptive OBS. All links carry the same number of wavelengths, W. We assume full wavelength conversion in the practically interesting case $W = 8$, so as to limit the conversion range; the issue of wavelength continuity is addressed in §IV-E. The optical line rate is equal to $R = 10$ Gbit/s, yielding a total capacity of $WR = 80$ Gbit/s per link. All SD pairs have the same traffic intensity and a single reservation process. The mean flow size is set to $\sigma = 2.5$ MB. Unless otherwise specified, the minimum burst size is set to $B = 10$ Mbit, corresponding to a burst duration of $\tau = B/R = 1$ ms. The attempt rate is $\nu = 1$ ms^{-1}. The results are derived from the simulation of 10^7 jumps of the underlying Markov process, after a warm-up period of 10^6 jumps.

A. Throughput and delay metrics

We define the flow throughput as the ratio of the mean flow size to the mean flow duration. According to Little's formula [22], the flow throughput on SD pair k is given by:

$$\gamma_k = \frac{\rho_k}{\mathrm{E}[x_k]}. \tag{6}$$

Since SD pair k transmits bursts of mean size $\mathrm{E}[x_k]B$ bits, the mean delay between two bursts of SD pair k follows again from Little's formula:

$$\theta_k = \frac{\mathrm{E}[x_k]B}{\rho_k} = \frac{B}{\gamma_k}. \tag{7}$$

Thus, the mean delay between two successive bursts is equal to the ratio of the minimum burst size to the flow throughput.

B. A single link

We first compare the behaviour of adaptive OBS to that of WR-OBS on a single link shared by K SD pairs, as illustrated by Figure 1. The link load is defined as:

$$\varrho = \frac{\sum_{k=1}^K \rho_k}{WR}.$$

We assume that all SD pairs have the same round-trip time, denoted by δ and taken equal to 1 ms. In the simple case $W = 1$, we deduce from (3) the probability that the link is reserved for SD pair k (with $x_k > 0$) under adaptive OBS:

$$\frac{\nu(\delta + x_k\tau)}{1 + \sum_{i:x_i>0} \nu(\delta + x_i\tau)}.$$

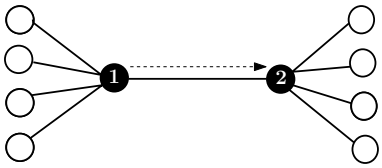


Fig. 1. Single link shared by $K = 16$ SD pairs.

Using (4), we obtain the throughput of SD pair k in state x:

$$\phi_k(x) = R\frac{x_k\nu\tau}{1+\sum\limits_{i:x_i>0}\nu(\delta+x_i\tau)}.$$

Note that the total throughput $\sum_k \phi_k(x)$ tends to R when $\sum_k x_k \to \infty$; thus adaptive OBS is able to entirely utilize the available capacity. The associate Markov process is ergodic provided $\varrho < 1$, cf. Theorem 1.

Under WR-OBS, a single burst is transmitted at each reservation completion; the throughput of SD pair k becomes:

$$\tilde{\phi}_k(x) = R\frac{\nu\tau}{1+\sum\limits_{i:x_i>0}\nu(\delta+\tau)}.$$

The total throughput is less than $R\tau/(\delta+\tau)$ when all routes are active. We deduce that the associate Markov process is transient as soon as $K\rho > R\tau/(\delta+\tau)$, corresponding to a maximum link load of $\tau/(\delta+\tau) = 0.5$.

Figure 2 shows the corresponding flow throughput (6) with respect to the link load ϱ for $W = 8$ wavelengths and $K = 16$ SD pairs. While both schemes have the same flow throughputs when $\varrho \to 0$, namely

$$R\frac{\nu\tau}{1+\nu(\delta+\tau)} \approx 3.3 \text{ Gbit/s}, \tag{8}$$

their performance differs significantly as load grows. Under WR-OBS, the throughput drops to 0 when $\varrho \to 0.5$. Under adaptive OBS, on the other hand, the throughput decreases gradually as ϱ grows from 0 to 1, showing the interest of burst size adaptation.

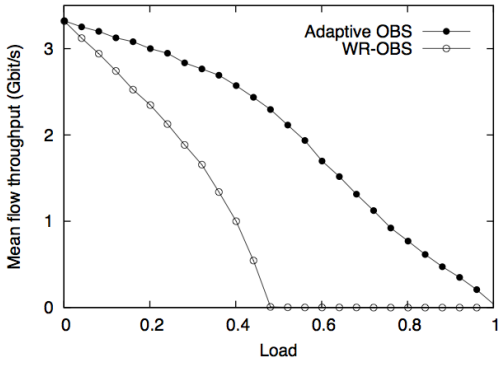


Fig. 2. Throughput performance of adaptive OBS and OBS for a single link with $W = 8$ wavelength channels shared by $K = 16$ SD pairs.

C. Setting the minimum burst size

We now investigate the impact of the minimum burst size, B. We still consider the case of a single link. Figure 3 gives the flow throughput and the mean delay of each SD pair for $B = 1, 10$ and 100 Mbit, corresponding to respective burst durations τ of 0.1, 1 and 10 ms. In view of (8), the flow throughput when $\varrho \to 0$ is respectively equal to 0.5, 3.3 and 8.3 Gbit/s. As expected, the flow throughput increases with the minimum burst size. However, setting large values of B (e.g., 100 Mbit) also increases the delays between subsequent bursts. The minimum burst size should typically be set so that the corresponding transmission time is of the same order as the reservation delay (backoff and round-trip time), namely $B = 10$ Mbit for the considered parameters.

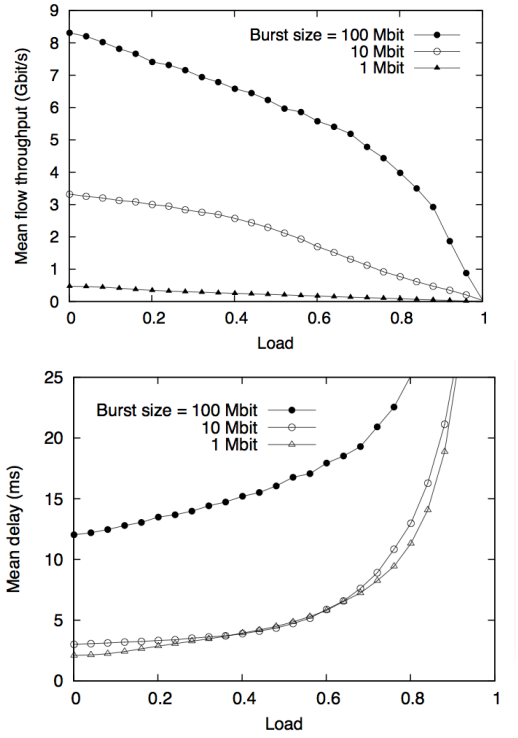


Fig. 3. Impact of the minimum burst size on the throughput (top) and delay (bottom) performance for a single link with $W = 8$ wavelength channels shared by $K = 16$ SD pairs.

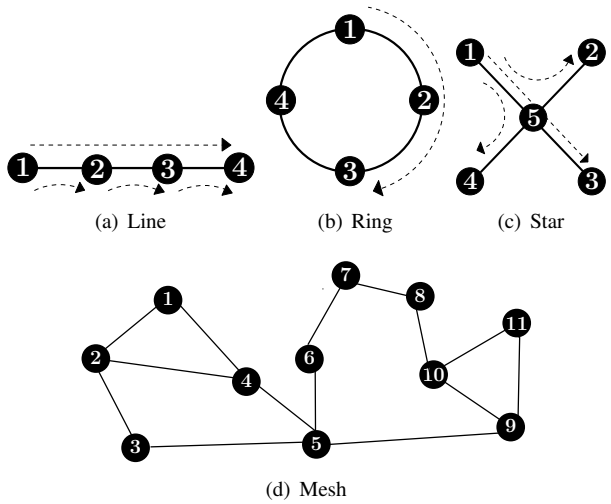
D. Networks

We now analyse the performance of adaptive OBS in various types of networks. Specifically, we consider the four network topologies of Figure 4. The ring, the star and the mesh topology are simple models for a metropolitan area network, a peering node, and a backbone network, respectively. Figure 4 depicts the considered routes for the bus, ring and star networks. The mesh network has one route per pair of nodes (i, j) such that $i < j$. Each route is shared by M distinct edge SD pairs, the corresponding edge nodes being omitted for simplicity. Table I summarizes the network parameters.

Topology	L	W	M	K	Route length
Bus	3	8	6	24	1 hop and 3 hops
Ring	4	8	6	24	2 hops
Star	4	8	4	48	2 hops
Mesh	14	8	1	55	from 1 to 5 hops

TABLE I
NETWORK PARAMETERS

There is a single eligible path per SD pair, taken as the shortest path in number of hops (see §IV-F for the impact of multipath reservation). We define the network load as the load of the most loaded link(s). For instance, the load of the mesh network is defined as that of link $5-9$, which is shared by 24 SD pairs. For simplicity, we assume that each link has a round-trip time of 1 ms.



(a) Line (b) Ring (c) Star

(d) Mesh

Fig. 4. Considered network topologies.

Figure 5 gives the flow throughput with respect to the network load ϱ for the considered networks. As for a single link, the flow throughput when $\varrho \to 0$ is given by (8) and thus depends on the round-trip time. It then decreases gradually to 0 as ϱ grows from 0 to 1, except for those routes of the mesh backbone network that do not go through the most loaded links and have positive throughput at load $\varrho = 1$. Longer routes experience lower flow throughput due to longer reservation delays. Equivalent simulation results (not reported here) show that under WR-OBS, the flow throughput drops to 0 at load less than 0.3.

The mean delay of an SD pair essentially depends on the route length and on the load of the corresponding links. At low load, it is simply given by the reservation delay plus the minimum burst duration, leading to delays of the order of a few milliseconds as for a single link (see Figure 3). As load grows, reservations are more likely to fail, increasing the mean delay accordingly. Assuming target mean delays ranging from 10 to 20 ms, we give in Table II the corresponding operational load ranges for the considered network topologies. We note that relatively high network loads can be sustained in all cases.

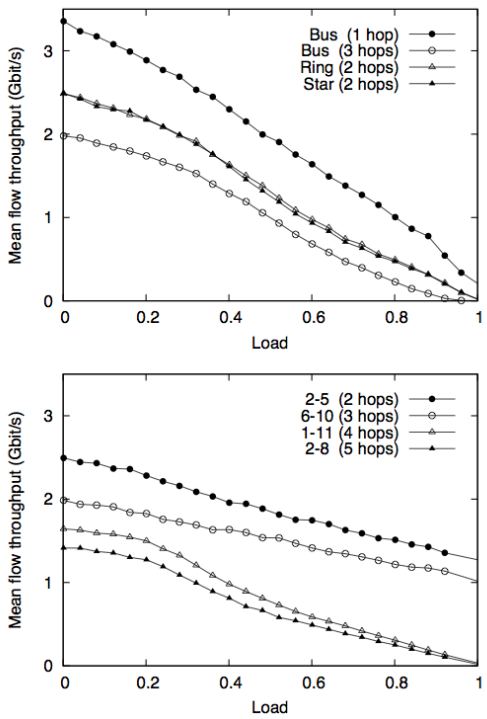


Fig. 5. Throughput performance of adaptive OBS in the line, ring, star topology (top) and in the mesh backbone network (bottom).

Network topology	Load range
Line	51% – 68%
Ring	60% – 81%
Star	60% – 78%
Mesh	36% – 64%

TABLE II
PRACTICAL OPERATIONAL LOAD RANGES

E. Wavelength continuity

To assess the performance of adaptive OBS in the absence of wavelength conversion, we consider the 3-link line of Figure 4. The long route must now utilize the same wavelength from source to destination. The impact of this wavelength continuity constraint is shown by Figure 6. It turns out that the short routes benefit from the higher contention suffered by the long route, slightly improving their performance. We note that adaptive OBS is still able to fully utilize network capacity, as predicted by Theorem 1.

F. Multi-path reservation

Finally, we evaluate the impact of multiple paths eligible for reservation on each SD pair, as described in §II-B, under full wavelength conversion. To this end, we consider the mesh backbone network in which 36 routes traverse the network from left to right by using either link $5-6$ or link $5-9$. We assume that each SD pair on each of these 36 routes has two eligible paths, one through link $5-6$ and another through link $5-9$. Figure 7 gives the flow throughput, averaged over the 36 routes, as a function of the overall load of these two links.

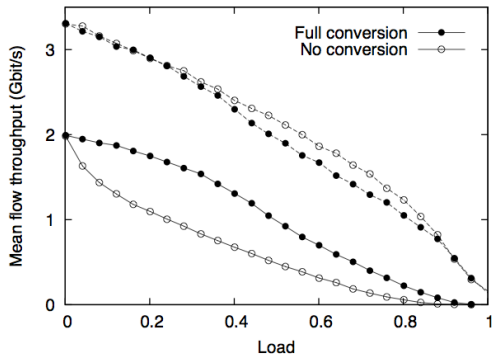


Fig. 6. Impact of wavelength continuity constraint on the throughput performance for the 3-link line; the dotted line represents the 3 hop route.

As expected, the multi-path reservation reduces the reservation failure probability and improves performance, yielding a gain of approximately 30% on network capacity.

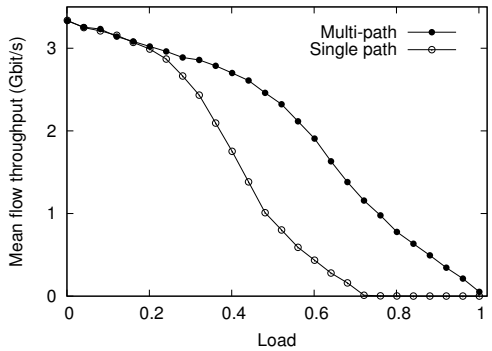


Fig. 7. Impact of multi-path reservation on throughput performance in the mesh backbone network.

V. Conclusion

We have proposed a modified version of two-way reservation OBS that we refer to as Adaptive Optical Burst Switching. The burst size is adapted to the traffic conditions so as to fully utilize network resources. The proposed scheme is indeed provably optimal in this sense. We have analysed the throughput and delay performance of adaptive OBS and shown that relatively high loads can be sustained for reasonable performance targets.

On the theoretical side, the distributed resource allocation achieved by adaptive OBS resembles that obtained in wireless networks under adaptive CSMA algorithms, see e.g. [23], [24]. While the proofs of optimality have similar structures, some constraints like the signalling delays, the wavelength conversion and the multipath reservation are specific to optical networks.

Future work will be focused on the ability of adaptive OBS to react to node or link failures, thanks to multipath reservation. We also intend to relax the assumption of exponential flow sizes in the proof of optimality. Other practically interesting issues include the impact of backward blocking and the analysis of end-to-end delays, accounting for the burst assembly mechanism.

References

[1] M. Herzog, M. Maier, and M. Reisslein, "Metropolitan area packet-switched WDM networks: A survey on ring systems," *IEEE Communications Surveys and Tutorials*, vol. 6, no. 1-4, pp. 2–20, 2004.

[2] M. Maier, *Optical Switching Networks*. Cambridge Univ. Press, 2008.

[3] V. Chan, G. Weichenberg, and M. Medard, "Optical flow switching," in *Invited Paper, International Workshop on Optical Burst/Packet Switching*, 2006.

[4] G. Weichenberg, V. Chan, and M. Médard, "Performance analysis of optical flow switching," in *Proc. of IEEE International Conference on Communications*, 2009, pp. 1–6.

[5] C. Qiao and M. Yoo, "Optical burst switching (OBS) - A new paradigm for an optical Internet," *Journal of High Speed Networks*, vol. 8, pp. 69–84, 1999.

[6] J. Y. Wei and R. I. McFarland, "Just-In-Time signaling for WDM optical burst switching networks," *Journal of Lightwave Technology*, vol. 18, pp. 2019–2037, 2000.

[7] I. Baldine, H. G. Perros, G. Rouskas, and D. Stevenson, "JumpStart: A Just-in-Time signaling architecture for WDM burst-switched networks," *IEEE Communications Magazine*, vol. 40, pp. 82–89, 2002.

[8] K. Dolzer, C. Gauger, J. Spaeth, and S. Bodamer, "Evaluation of reservation mechanisms for optical burst switching," *International Journal of Electronics and Communications*, vol. 55, 2001.

[9] Z. Rosberg, H. L. Vu, M. Zukerman, and J. White, "Performance analyses of optical burst switching networks," *IEEE Journal of Selected Areas in Communications*, vol. 21, pp. 1187–1197, 2003.

[10] A. G. Fayoumi and A. P. Jayasumana, "Performance model of an optical switch using fiber delay lines for resolving contentions," in *Proc. of IEEE Local Computer Networks*, 2003, pp. 178–186.

[11] X. Lu and B. Mark, "Performance modeling of optical-burst switching with fiber delay lines," *IEEE Transactions on Communications*, vol. 12, pp. 2175–2183, 2004.

[12] C.-W. Tan, G. Mohan, and J. C. S. Lui, "Achieving multi-class service differentiation in WDM optical burst switching networks: A probabilistic preemptive burst segmentation scheme," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. S-12, pp. 106–119, 2006.

[13] M. Neuts, Z. Rosberg, H. L. Vu, J. White, and M. Zukerman, "Performance enhancement of optical burst switching using burst segmentation," in *Proc. of IEEE International Conference on Communications*, 2003, pp. 1828–1832.

[14] X. Wang, H. Morikawa, and T. Aoyama, "Deflection routing protocol for burst switching WDM mesh networks," in *Proc. of Terabit Optical Networking Conference*, 2000.

[15] C. Hsu, T. Liu, and N. Huang, "Performance analysis of deflection routing in optical burst-switched networks," in *IEEE INFOCOM*, 2002, pp. 66–73.

[16] M. Duser and P. Bayvel, "Analysis of a dynamically wavelength-routed optical burst switched network architecture," *Journal of Lightwave Technology*, vol. 20, pp. 574–585, 2002.

[17] Z. Lan, H. Guo, W. Jian, and J. Lin, "Performance of a distributed WR-OBS control architecture," *Chinese Optics Letters*, vol. 3, pp. 196–198, 2005.

[18] S. Ben Fredj, T. Bonald, A. Proutière, G. Régnié, and J. W. Roberts, "Statistical bandwidth sharing: a study of congestion at flow level," in *Proc. of ACM SIGCOMM*, 2001, pp. 111–122.

[19] P. Key, L. Massoulié, A. Bain, and F. Kelly, "A network flow model for mixtures of file transfers and streaming traffic," in *Proc. of International Teletraffic Conference 18*, 2003.

[20] T. Bonald, R.-M. Indre, S. Oueslati, and C. Rolland, "On virtual optical bursts for qos support in obs networks," in *Proc. of ONDM*, 2010.

[21] T. Bonald, "Insensitive traffic models for communication networks," *Discrete Event Dynamic Systems*, vol. 17, no. 3, pp. 405–421, 2007.

[22] P. Robert, *Stochastic networks and queues*, Springer, Ed., 2003.

[23] T. Bonald and M. Feuillet, "On the stability of flow-aware CSMA," in *IFIP Performance*, 2010.

[24] J. Ni, B. Tan, and R. Srikant, "Q-CSMA: Queue-length based CSMA/CA algorithms for achieving maximum throughput and low delay in wireless networks," in *IEEE INFOCOM*, 2010.